Abstract—To improve the driving mobility and energy efficiency of connected autonomous electrified vehicles, this paper presents an integrated longitudinal speed decision-making and energy efficiency control strategy. The proposed approach is a hierarchical control architecture, which is assumed to consist of higher-level and lower-level controls. As the core of this study, model predictive control and reinforcement learning are combined to improve the powertrain mobility and fuel economy for a group of automated vehicles. The higher-level exploits the signal phase and timing and state information of connected autonomous vehicles via vehicle to infrastructure and vehicle to vehicle communication to reduce stopping at red lights. The higher-level outputs the optimal vehicle velocity using model predictive control technique and receives the power split control from the lower-level controller. These two levels communicate with each other via a controller area network in the real vehicle. The lower-level utilizes model-free reinforcement learning method to improve the fuel economy for each connected autonomous vehicle. Numerical tests illustrate that vehicle mobility can be noticeably improved (traveling time reduced by 30%) by reducing red light idling. The effectiveness and performance of the proposed method are validated via comparison analysis among different energy efficiency controls (fuel economy promoted by 13%).

Index Terms—Connected autonomous vehicles, Model predictive control, Energy efficiency control, Reinforcement learning, Intelligent transportation system

The work was supported in part by Supported by the State Key Laboratory of Mechanical System and Vibration (Grant No. MSV202016). (Corresponding author: Bo Wang)

T. Liu is with College of Automotive Engineering, Chongqing University, Chongqing 400044, China, and also with Department of Mechanical and Mechatronics Engineering, University of Waterloo, N2L 3G1, Canada. (email: tengliu17@gmail.com)

B. Wang is with School of Mathematics and Statistics, Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081, China. (email: wangbo89630@bit.edu.cn)

D. CAO is with Department of Mechanical and Mechatronics Engineering, University of Waterloo, Ontario N2L3G1, Canada. (dongpu.cao@uwaterloo.ca)

X. Tang and Y. Yang are with State Key Laboratory of Mechanical Transmissions, College of Automotive Engineering, Chongqing University, Chongqing, 400044, PR China. (email: tangxl0923@cqu.edu.cn, YYL@cqu.edu.cn)

I. INTRODUCTION

In recent years, connected autonomous vehicle technology has aroused interest in the automobile companies as well as in the academic community [1]. Researchers are exploring ways to achieve mobility, safety and environmental impact goals via communication between vehicle and vehicle or infrastructure (V2V or V2I). Vehicle information (e.g. position, velocity, and acceleration) and traffic information (congestion and light timing) are able to be shared among the connected vehicles to reduce or eliminate crashes and enhance current operational practices of drivers [2].

To promote powertrain mobility and energy efficiency, reducing the red light idling and minimizing the braking times are the promising solutions for connected autonomous vehicles (CAVs) [3]. For example, Asadi et al. proposed a strategy using model predictive control (MPC) to decrease idle time at red lights and fuel consumption by signal phase and timing (SPAT) information [4]. A predictive optimal velocity planning algorithm was proposed in Ref. [5] by using the probabilistic traffic SPAT information. The relevant simulation results indicate that the presented method could increase energy efficiency via probabilistic timing and real-time phase data. Furthermore, Jin et al. proposed a power-based longitudinal control algorithm which can be applied in a connected eco-driving system [6]. The roadway grade and vehicle’s brake specific fuel consumption are considered in this work when calculating an optimal speed profile with respect to energy savings and emissions reduction.

In current academic or industrial communities, CAVs are usually founded on or refitted from electrified vehicles, such as the hybrid electric vehicles (HEVs) and battery electric vehicles (BEVs) [7]. These vehicles have the potential to improve fuel economy and reduce pollutant emissions by embedding the electric storage systems into the powertrain [8]. Energy management is a significant technology for HEV and it means searching the optimal power distribution for a hybrid powertrain [9]. In general energy management problem, HEVs designers aim to optimize a pre-selected cost function (e.g. fuel consumption, harmful emissions, and running cost) by optimizing the power split controls of multiple energy storage sources while satisfying the driving power demand [10].

Many kinds of approaches have been utilized to formulate energy management strategies for HEVs, and the representative one is the optimization control theory-enabled method. For
example, as the specific driving cycle is known a priori, dynamic programming (DP) algorithm [11] and equivalent consumption minimization strategy (ECMS) [12] have been developed to obtain the globally optimal power split decision for HEVs. The authors in [13] treated DP-based results as a benchmark to estimate the learning-based method in an energy management problem for a hybrid tracked vehicle. By computing the feasible range of optimal equivalent factor in ECMS, Ref. [14] presented an adaptive EMCS-enabled power split policy for a particular HEV configuration. Also, convex programming (CP) is another approach to derive the globally optimal power split control for HEVs [15]. The carbon reductions arising from renewable energy integration and the CP framework are quantified in [16], and the authors stated the proposed method could efficiently improve battery lifetime and reduce carbon dioxide emissions.

Furthermore, when the full driving cycle is not given prior, stochastic dynamic programming (SDP) [17], model predictive control (MPC) [18] and reinforcement learning (RL) [19] are studied to achieve real-time optimal controls. For example, Stellato combined approximate DP and direct MPC to handle integer optimal control problems over long prediction horizons [20]. It is able to reduce the computational burden and enable sampling times below 25 μs. Ref. [21] discussed the control performance of different RL algorithms, Q-learning and Sarsa, in an online energy management problem. The pros and cons of these algorithms are analyzed and compared in the same hybrid powertrain. However, most literature did not involve the energy efficiency and fuel economic controls of autonomous electrified vehicles in the connected environments.

Overall, an integrated combination of CAV and HEV attracted more and more attention in recent years. How to improve powertrain mobility and fuel economy of this synthetical architecture is an important and comprehensive research hotspot [22]. However, few literature has considered and discussed the automation function of CAV and energy utilization of HEV together [23]. In order to fill that literature gap, this paper aims to optimize an integrated objective for a combination of CAV and HEV, which consists of the powertrain mobility and fuel economy. The proposed system is a hierarchical control architecture, which is assumed to consist of higher-level and lower-level controls, see Fig. 1 as an illustration. The entire goal of this work is promoting driving mobility and reducing fuel consumption for each automated electrified vehicle in the connected environment (formulated in (2) and (8)). The presented system aims to achieve this objective in two steps, the vehicle speed is optimized by the higher-level control first and then the energy management policy is derived in the lower-level.

Especially, MPC is utilized to solve the longitudinal speed decision-making problem in higher-level and RL is employed to research the energy management problem in lower-level. The higher-level exploits the SPAT and state information provided by V2I and V2V communication to reduce stopping at red lights. By doing this, the waiting time at the signal lights can be available shortened in order to promote efficiency and mobility. The higher-level outputs the optimal vehicle velocity for each vehicle and receives the power split control from the lower-level controller. The lower-level aims to improve fuel economy for each vehicle, which indicates it would manage the power distribution between different energy storage systems according to the obtained speed profile.

Three perspectives are contributed to this paper: (1) a novel hierarchical control architecture is presented to provide a holistic method for improving mobility and energy efficiency for connected autonomous electrified vehicles; (2) MPC approach is applied to make a decision that focuses on improving powertrain mobility and enhancing vehicle safety via decreasing red light idling; (3) model-free RL technique is applied to obtain optimal power split control between the engine and battery for each electrified vehicle. Numerical tests illustrate that vehicle mobility can be noticeably promoted by reducing red light idling. Comparison of different energy efficiency control strategies demonstrates the effectiveness and control performance of the presented method.

The rest sections of our paper are organized as follows. Section II describes the higher-level architecture of MPC framework for longitudinal speed decision-making in con-
gested road conditions; Section III illustrates the formulation of a power split control problem in electrified powertrain, and the structure of the lower-level model-free RL algorithm is also introduced; In Section IV, the vehicle and experiment parameters are first elaborated, and experiment results of higher-level and lower-level performance comparison are presented; Section V summarizes the key takeaways.

II. HIGHER-LEVEL PROBLEM FORMULATION AND MODEL PREDICTIVE CONTROLLER

In the higher-level, we focus on optimizing the traction or braking force to minimize the red light idling and energy consumption for traction of each connected autonomous vehicle. The position and velocity information of each vehicle can be obtained by V2V communication for their near neighborhood. The SPAT information of traffic lights is sent to each vehicle by roadside units through the wireless network. The proposed controller consists of two parts: First, a target velocity is calculated to avoid stopping at red lights, and second, MPC strategy is formulated to minimize the pre-defined cost function.

A. Higher-Level Problem Formulation

For longitudinal control, the dynamics of the vehicle $i$ is described as [24]:

$$
\begin{aligned}
\dot{x}_i &= f_i(x_i, u_i) \\
\dot{f}_i(x_i, u_i) &= \begin{bmatrix}
v_i \\
-\frac{1}{2M_i}C_d\rho_A v_i^2 - \mu g \sin \theta + u_i
\end{bmatrix}
\end{aligned}
$$

(1)

where $x_i = [s_i, v_i]$, $s_i$ and $v_i$ denote the position and velocity of vehicle $i$. The control action $u_i$ is the traction or braking force per unit mass. The symbols $M_i$, $C_d$, $\rho$, $A$, $\mu$, and $\theta$ denote vehicle mass, drag coefficient, air density, vehicle frontal area, rolling friction coefficient, and road gradient, respectively.

In the higher-level, the optimization objective can be computed by dividing the total distance from fuel consumption, which is denoted by $\sum \dot{m}_i'(t) \Delta t / s_i'$. Here $\dot{m}_i'(t)$ is the fuel consumption rate, $\Delta t$ is the time step, and $s_i'$ is the distance traveled by vehicle $i$ during the time interval $[0, t]$. The following expression describes the optimal control problem in higher-level [22]:

$$
\begin{aligned}
\arg\min_{u_i(t)} & \sum_{t=0}^{t_f} \frac{1}{s_i'} \sum_{t=0}^{t_f} \dot{m}_i'(t) \Delta t \\
\dot{m}_i'(t) &= \frac{1}{\eta_f'\eta_{LHV}} P_e(t) \\
P_e &= a_v v_i^3 + M_i g v_i (\mu + \theta) + \beta_1 [v_i M_i u_i] + \beta_2 [- \eta_f'\eta_i v_i M_i u_i] \\
s.t. \alpha_i &= \frac{1}{2} C_d \rho_A A, \beta_i = \begin{cases} 0 & \text{if } u_i \leq 0 \\ 1 & \text{otherwise} \end{cases}, \beta_i' = \begin{cases} 0 & \text{if } u_i > 0 \\ 1 & \text{otherwise} \end{cases} \\
v_{\min} \leq v_i \leq v_{\max}, u_{\min} \leq u_i \leq u_{\max}
\end{aligned}
$$

(2)

where $\eta_f'$ is the transmission efficiency connecting the fuel tank and powertrain, $H_{LHV}$ represents the fuel lower heating value for vehicle $i$, $P_e$ is the traction power depending on the traction force, and $\eta_f'\eta_i$ is the recuperation efficiency acquired from the lower-level controller [22]. $v_{\min}$, $v_{\max}$, $u_{\min}$, and $u_{\max}$ depict the upper and lower threshold values for the allowable speed and possible acceleration, respectively. Hence, the higher-level controller generates the optimal velocity profile for each connected vehicle and inputs them into the lower-level for energy efficiency control.

B. Target Velocity Computation

The upcoming SPAT information is obtained a priori via the V2I communication for each connected vehicle. Based on this, the target velocity of each vehicle can be decided wisely to avoid idling at the red light. The target velocity at time instant $k$ for the vehicle $i$ is defined as following [23]:

$$
\begin{aligned}
\dot{v}_{\text{tar}}(k) &= \begin{cases} \frac{d_{\text{in}}(k)}{K_v t_{cy} - t_g - k} & \text{if Lig = red} \\
\frac{d_{\text{in}}(k)}{K_v t_{cy} + t_g - k} & \text{if Lig = green and } \frac{d_{\text{in}}(k)}{K_v t_{cy} - k} \leq v_{\max} \\
\frac{d_{\text{in}}(k)}{K_v t_{cy} + t_g - k} & \text{if Lig = green and otherwise}
\end{cases}
\end{aligned}
$$

(3)

where $d_{\text{in}}(k)$ is the distance between traffic signal and the vehicle location $(s_i(k), t_g)$, $t_g$, $t_f$, and $t_{cy}$ denote the red, green light and total cycle durations, respectively. $K_w$ represents the cycle number for the traffic light, and $\text{mod}$ denotes a function to reserve the remainder for division $k$ by $t_{cy}$.

Eq. (3) indicates that when the equation $d_{\text{in}}(k) / (K_w t_{cy} - k) \leq v_{\max}$ is satisfied and the status of the traffic light is green, the target velocity is selected as the maximum allowable speed. The speed conserving condition implies that the vehicle cannot pass through the current green light cycle although using the maximum speed. Thus, the vehicle needs to slow down and wait for the next green light window. Sometimes, the vehicle has to stop as there is no feasible velocity in the consecutive green light windows [24]. Overall, target velocity means the available speed to help each vehicle pass through the traffic signal at different green light windows.

Assuming the vehicle $i$ could go through the traffic signal at one of the green light windows, and thus there is a speed range for the target velocity because the green light window is a time interval. Assume $\dot{v}(k)$ denotes the lower threshold velocity and $\dot{v}(k)$ depicts the upper value, they are computed as [22]:

$$
\begin{aligned}
\text{arg min}_{v_i} \left\{ \sum_{t=0}^{t_f} \frac{1}{s_i'} \sum_{t=0}^{t_f} \dot{m}_i'(t) \Delta t \right\}
\end{aligned}
$$

(4)
target velocity based on the communication between each vehicle and traffic lights via V2I information exchange. Note that, the optimal velocity trajectory for energy efficiency control and the target velocity in the higher-level may be different because the latter does not consider energy efficiency. Hence, we compute the optimum velocity profile by considering the extra position information of the front vehicle $j$ over a finite future time horizon $T$ via V2V information exchange. The discussed driving scenario is depicted in Fig. 3, wherein the driving mobility of each vehicle can be interpreted as the traveling time for the same scenario and the energy efficiency could be quantified as the fuel consumption of each HEV [25, 26].

As the position information of the front vehicle $j$ and the target velocity of the current vehicle $i$ are available, the MPC approach is used to generate the best velocity profile. It can be formulated as a nonlinear optimization problem at time instant $k$ as follow [27]:

$$
J_i(k) = \arg \min_{u_i(t)} \sum_{t=k}^{k+T-1} \left[ \phi_1 \left( \hat{v}_i(t) - v_{\text{tar}}(k) \right)^2 \right] + \phi_2 X_{ij}^2 + \phi_3 (v_i(t) - v_{\text{tar}}(k))^2 + \phi_4 (u_i(t))^2
$$

where $S_0 + t_{\text{hw}} (v(t) - v_{\text{tar}}(k)) + (s_j - s_i)$

s.t. $v_i(t) \leq v_{\text{tar}}(k), \hat{u}_{\text{min}} \leq u_i(t) \leq \hat{u}_{\text{max}}, \forall t$

where $S_0$ and $t_{\text{hw}}$ represent the predefined critical distance and headway time, and the two inequality constraints are derived from Eqs. (4) and (6) respectively. In the cost function, the first term depicts the division of total fuel consumption by travel distance, the second items denote the desirable distance deviation between the vehicles $i$ and $j$, the third term restrains the velocity of vehicle $i$ close to the computed target velocity, and the last term aims to optimize control action [22].

The optimal energy efficiency velocity profile highly depends on the selection of the weights $\phi_1$ to $\phi_4$. The weights $\phi_1$ and $\phi_3$ are defined as a function of $(v_i(k) - v_{\text{tar}}(k))$ to ensure when the deviation is large, more emphasis is focused on energy efficiency rather than velocity tracking ($\phi_1$ high and $\phi_3$ low), and as the deviation is small, more attention is paid to the target velocity tracking ($\phi_1$ low and $\phi_3$ high). The weight $\phi_2$ is relevant with the distance deviation ($s_j - s_i$), and it increases as the deviation is small and it decreases as the deviation is large, finally the weight $\phi_4$ is a constant [23]. In conclusion, the speed range $[v_{\text{tar}}(k), v_{\text{tar}}(k)]$ serves two goals, first it helps choose some weights in the cost function, and second, it forces the constraints on the velocity deviation to avoid red light idling.

Three key processes are involved in the MPC algorithm: 1) search the optimization controls to minimize the pre-defined cost function during a prediction horizon; 2) obtain the first portion of the optimization controls and apply it into the physical plant; 3) moves forward the entire prediction horizon and repeat the first step [28]. Here we apply sequential quadratic programming to solve the nonlinear optimization problem (see Eq. (7)). Especially, the optimal control actions are computed using the command $\text{fmincon}$ in Matlab/Simulink and the optimal solution for time instant $k$ is applied to formulate the initialization.
of the decision variables at time instant $k+1$. Overall, the MPC-based optimal velocity trajectory calculation may take about one minute in higher-level.

III. LOWER-LEVEL PROBLEM FORMULATION AND REINFORCEMENT LEARNING

In the lower-level, we aim to improve the energy efficiency of each connected autonomous electrified vehicle by deriving the optimal power split control between multiple energy sources. The sole state variable is the state of charge (SOC) in the battery and control action is the engine throttle percentage $th(t)$. After the higher-level gave the optimal speed profile, the relevant power demand can be calculated via the powertrain modeling. Then, the transition probability matrix (TPM) of the power demand can be calculated by the maximum likelihood estimate. Finally, the Q-learning algorithm in the RL framework is employed to evaluate the optimal control problem in the lower-level.

A. Lower-level problem

In the minimization problem of the energy efficiency for a parallel electrified vehicle, the optimization objective is the sum of the charge sustenance and fuel consumption. The problem can be mathematically established as follow [20, 21]:

$$ J = \arg \min_{\theta(t)} \sum_{t=0}^{T} \left[ \dot{m}^i_\text{dem} (th(t), \text{SOC}(t)) + \chi (\text{SOC}(t) - \text{SOC}(0))^2 \right] \Delta t $$

subject to:

$$ \text{SOC}_{\text{min}} \leq \text{SOC}(t) \leq \text{SOC}_{\text{max}}, \quad \text{th}_{\text{min}} \leq th(t) \leq \text{th}_{\text{max}} $$

$$ T_{\text{en}, \text{min}} \leq T_{\text{en}} (t) \leq T_{\text{en}, \text{max}}, \quad n_{\text{en}, \text{min}} \leq n_{\text{en}} (t) \leq n_{\text{en}, \text{max}} $$

$$ P_{b, \text{min}} \leq P_b (t) \leq P_{b, \text{max}}, \quad I_{b, \text{min}} \leq I_b (t) \leq I_{b, \text{max}} $$

where $\chi$ is a positive weighting factor (equal to 1000 in this work) to guarantee the initial and final SOC values stay the same. $\text{SOC}_{\text{min}}, \text{SOC}_{\text{max}}, \text{th}_{\text{min}},$ and $\text{th}_{\text{max}}$ are the boundary values of the SOC and throttle percentage. $T_{\text{en}}$ and $n_{\text{en}}$ are the engine torque and speed, $n_{\text{en}, \text{min}}, n_{\text{en}, \text{max}}, T_{\text{en}, \text{min}},$ and $T_{\text{en}, \text{max}}$ are the permitted lower and upper bounds. $P_b$ and $I_b$ are the battery power and current, $P_{b, \text{min}}, P_{b, \text{max}}, I_{b, \text{min}}, I_{b, \text{max}}$ are the threshold of the admissible sets for them.

The output torque and speed of engine are decided by the throttle percentage at each time step. The fuel consumption rate in the engine and the engine output power are defined as:

$$ P_{\text{en}} (t) = T_{\text{en}} (t) \times n_{\text{en}} (t) $$

$$ \dot{m}^i_\text{dem} (th(t)) = f_{\text{en}} (T_{\text{en}} (t), n_{\text{en}} (t)) $$

Modeling the battery with the equivalent circuit model [29], the state equation of the SOC is given by:

$$ \frac{d \text{SOC}}{dt} = -\frac{I_b (t)}{Q_b} = \frac{(V_{oc} - \sqrt{V_{oc}^2 - 4r_m^2 P^i_\text{en} (t)})}{2Q_b r_m} $$

where $Q_b$ denotes the battery rated capacity. $V_{oc}$ represents the open circuit voltage and $r_m$ describes the battery internal resistance. $V_{oc}$ and $r_m$ are relevant to the SOC.

B. Transition Probability Matrix

As the velocity profile is known in advance from the higher-level controller, the power demand to impel each connected autonomous electrified vehicle is computed as follow:

$$ P_{\text{dem}}^i = \delta M_{\text{en}} a_i V_i + \frac{C_d A_i}{21.15} V_i^3 + M_{\text{fr}} g \mu V_i $$

where $\delta$ is the mass factor, $a_i$ is the $i$-th vehicle acceleration.

For the purpose of keeping power balance in the powertrain, the power demand is satisfied by both of the engine and battery:

$$ P_{\text{dem}}^i = (P_{\text{en}}^i + P_{\text{fr}}^i) \eta^i $$

where $\eta^i$ is the traction motor efficiency for vehicle $i$. As the engine output power can be evaluated by Eq. (9), and then the battery power can be estimated from Eq. (12). The transition probability of the power demand is depicted via maximum likelihood estimator as:

$$ p_{nk,m} = P(P_{\text{dem},n} | P_{\text{dem},m}, V_k^i) = \frac{N_{nk,m}^i}{N_{nk}^i} $$

$$ N_{nk}^i = \sum_{m=1}^{M} N_{nk,m}^i, n, m = 1, 2, ..., M $$

where $N_{nk,m}^i$ denotes the number of times for the transition from
Based on the previous TPMs of power demand and vehicle modeling, RL technique is employed for energy efficiency control derivation. Especially, the Q-learning algorithm is implemented in Matlab with the Markov decision process (MDP) toolbox presented in [30], and the computer has an Intel quad-core CPU of 2.70 GHz a memory of 3.8 GB.

In the RL framework, a learning agent interacts with a stochastic environment. Five key variables are exploited to model the interaction, wherein S is the state variables set, A is the control actions set, P denotes the TPM for power request, r(s, a) ∈ R represents the reward function and τ ∈ (0, 1) means a discount factor.

The sequence of control actions constitutes the control policy π. The optimal value function can be computed as follows [31]:

\[
V^*(s) = \min_{\pi} E \left( \sum_{t=0}^{\infty} \gamma^t r(s, a) \right) \\
= \min_{a} (r(s, a) + \tau \sum_{s' \in S} P(s, a, s') V^*(s')) \quad \forall s \in S
\]  

(14)

where \( p_{sa,s} \) denotes the transition probability. As the optimal value function is decided at each state variable, the relevant optimal control policy is determined as:

\[
\pi^*(s) = \arg \min_{a} (r(s, a) + \tau \sum_{s' \in S} P(s, a, s') V^*(s'))
\]  

(15)

The computation process of optimal control action can be interpreted as the matching design of state variable and control action. Hence, the \((s, a)\) is called state-action pair and the related action-value function \(Q(s, a)\) is given as follows:

\[
Q(s, a) = r(s, a) + \tau \sum_{s' \in S} P(s, a, s') Q(s', a')
\]

(16)

\[
Q^*(s, a) = r(s, a) + \tau \sum_{s' \in S} \min_{a'} Q(s', a')
\]

(17)

Note that, variable \( V^*(s) \) means the optimal value function that taking an optimal action \( a \) at state \( s \). Thus, \( V^*(s) = Q^*(s, a) \) and \( \pi^*(s) = \arg \min_a Q(s, a) \). Finally, the updated criterion for the optimal action-value function of the Q-learning algorithm is indicated by:

\[
Q(s, a) \leftarrow Q(s, a) + \gamma (r(s, a) + \tau \min_{a'} Q(s', a') - Q(s, a))
\]

(17)

where \( \gamma \in [0, 1] \) is a decaying factor of the Q-learning algorithm.

**TABLE I**

| Iteration Process of RL-Based Control |
|--------------------------------------|
| Method: Q-learning Algorithm |
| 1. Give a value for \( N \), \( Q(s, a) \) and \( s \) |
| 2. Repeat step by step \( k=1, 2, 3... \) |
| 3. Using \( \varepsilon \)-greedy policy to select control action \( a \) |
| 4. Compute \( r, s' \) based on \( s \) and \( a \) |
| 5. Compare and decide \( a^* = \arg \min_a Q(s', a) \) |
| 6. \( Q(s, a) \leftarrow Q(s, a) + \gamma (r(s, a) + \tau \min_{a'} Q(s', a') - Q(s, a)) \) |
| 7. Move to next step, \( s \rightarrow s' \) |
| 8. finish when \( s \) is the terminal |

To solve the optimal energy efficiency control problem, the state variables and control actions are discretized as follow: \( S = \{(SOC(t))[0.4 \leq SOC(t) \leq 0.8]\} \) and \( A = \{th(t) | 0 \leq th(t) \leq 1\} \). The set of reward function is \( R = \{m_1(s, a) + \chi (SOC(t)-SOC(0))^2\} \). The initial and final values of SOC are selected to avoid overcharge and over-discharge in the battery. Table I shows the calculation process of optimal control actions in Q-learning algorithm. First, the matrix of action-value function \( Q(s, a) \) and the state variable \( s \) are initialized. Then, in each episode, the control action is decided by \( \varepsilon \)-greedy policy at each step, and the action value function is updated by (17). Finally, one episode ends when the state variable is the terminal, and the whole process ends when the number of episodes reaches the setting.
value. The obtained matrix $Q(s, a)$ can be used to derive the control action in particular driving situation.

In this paper, the discount factor $\tau$ is taken as 0.96 to balance the importance of current and future rewards, the decaying factor $\gamma$ is equal to $1/\sqrt{k + 2}$ to accelerate the convergence rate and the value of $\varepsilon$ is $0.2 \times 0.99^k$ ($\varepsilon$ decreases with time steps) to select the control action at each step. The iterative times $N_k$ is 10000, and the sample time $\Delta t$ is 0.5 second. The effectiveness of the MPC and RL approaches for the higher-level, and lower-level problems are validated in the next section.

D. Parameters for SPAT and Powertrain

A signal lane road with eight vehicles constructs the simulation scenario, wherein the traffic lights are settled at every 500 m. Each parallel electrified vehicle has the same configuration and powertrain parameters, see Fig. 5 as an illustration. The primary parameters of them are listed in Table 2. The rated power for the diesel engine is 100 kW, and the maximum torque of diesel engine is 200 Nm as the rotational speed change from 110 rpm to 390 rpm. The maximum power, maximum torque and maximum rotational speed of the traction motor are 78 kW, 130 Nm, and 600 rpm, respectively. The battery pack is 6 Ah capacity with a nominal voltage of 250 V.

| Variables | Names                        | Numbers     |
|----------|------------------------------|-------------|
| $M_i$    | Vehicle weight               | 1000 kg     |
| $A_i$    | Fronted area                 | 2.25 m$^2$  |
| $C_d$    | Aerodynamic coefficient      | 0.3         |
| $\eta_{\text{tr}}$ | Efficiency of Transmission axle | 0.9       |
| $\eta_{\text{m}}$ | Efficiency of Traction motor | 0.95        |
| $\mu$    | Coefficient of Rolling resistance | 0.008    |
| $\theta$ | Road gradient                | 0           |
| $\rho_a$ | Air density                  | 1.205kg/m$^3$ |

For the SPAT parameters, the time intervals for the red and green light are 30 and 15 seconds ($t_r=30$ and $t_g=15$), respectively. The minimum and maximum allowable speeds for each connected vehicle are unified as 0 m/s and 20 m/s. Every vehicle’s initial position is given by random, and the initial relative distance between any two vehicles is also from 10 m to 20 m by random. The initial velocity of each connected autonomous electrified vehicle is also stochastic and selected from 10 m/s to 15 m/s.

IV. RESULTS AND DISCUSSION

The proposed MPC-enabled and RL-based longitudinal speed decision-making and energy efficiency control strategy for connected autonomous electrified vehicles is evaluated in this section. First, the performance of the MPC-enabled higher-level controller is examined by comparing with a baseline method. Then, the presented hierarchical system is compared with a benchmark system to assess its availability. The benchmark system employes MPC method in both higher and lower levels. Simulation results demonstrate that the proposed longitudinal speed decision-making and energy efficiency control strategy can highly reduce stopping at red lights and improve fuel efficiency.

A. Higher-Level Controller Evaluation

In this section, the MPC-based longitudinal speed decision-making control is compared with a baseline method to verify its effectiveness. Here, the baseline method means the connected vehicle apply V2I communication to calculate the target velocity in Eq. (3), and then executes the connected cruise control strategy [32].

Fig. 6 and Fig. 7 illustrate the vehicle trajectories (distance varies over time) in the MPC-based and baseline approaches for 10 minutes. The horizontal red bars depict the red lights duration time and the green bars point to the green lights’ intervals. The vehicles can only go through the traffic signal during the green bars period. The zoom-in plots in these figures indicate that connected autonomous vehicles move together until the traffic red lights split them. In the baseline method, the vehicles
split into two groups at 100 second, however, the vehicles in the MPC-based method split until 400 second. These results attribute to the limitation of road speed (the rest of vehicles cannot violate the limitation) and indicate that the MPC-based technique can significantly improve the system mobility.

To avoid chaotic and unclear exhibition, four random vehicles from eight vehicles are selected to display their running speeds in the defined scenario. Fig. 8 describes the velocity profiles for these vehicles (their order: 1, 3, 5, 6, a total of 8 vehicles). It is apparent that the vehicle velocities in the MPC-based method always stay positive which indicates that all the vehicles would not stop at every traffic signal light. Oppositely, the connected vehicles controlled by the baseline method have to stop at all traffic signals. Thus, the traveling time of each vehicle (pass through the same driving scenario) could be reduced by 30% (shown in Table 3) in the proposed method. It also suggests that longitudinal speed decision-making control strategy presented in the higher-level controller could effectively avoid stopping at red signal lights so that the system mobility can be highly promoted.

B. Evaluation of Integrated System

In this section, the proposed whole system is compared with a benchmark system to assess its effectiveness and merits. In this benchmark (baseline) system, the higher-level controller is the same as our proposed MPC-based longitudinal speed decision-making control strategy, while the lower-level still utilizes the MPC method [28] instead of the RL algorithm. Similarly, to prevent tedious and chaotic demonstrations of relevant results, two random vehicles are chosen to compare their energy management strategies. The order of these two vehicles is 2 and 7 (totally eight vehicles).

The velocity tracking in both vehicles using the RL algorithm in the lower-level controller are shown in Fig. 9. It illustrates that the lower-level controller is able to track the optimal velocity trajectory provided by the higher-level controller almost perfectly. Under this situation, the proposed method can avoid stopping at red lights and guarantee improving energy efficiency simultaneously.

Fig. 10 shows the state variable SOC evolution for these two vehicles in different approaches. For the same vehicle, the derived velocity trajectories are different in these two methods.
Thus, the related power demand and control actions are different, which results in the disparate variation of SOC curves. Due to the charge sustenance constraint in the lower-level controller, the final SOC value is close to the initial value. This implies that the battery power used by the electrified vehicle, leading to a decrease of SOC, which is compensated by charging with the power from the engine and regenerative braking power. Hence, we can evaluate the different methods’ control performance by comparing the power split and fuel consumption.

The power split controls in these two vehicles are described in Fig. 11. As Eqs. (9)-(12) stated, the battery and engine powers are finally decided by the throttle percentage. From Fig. 10, it is obvious that the power split controls in the proposed system are different from those of the benchmark system, which implies these two systems could achieve different control performance although the input speed profiles are the same. The working points of the engine in the different control strategies are given in Fig. 12. It is apparent that the engine working area under the proposed RL-based control locates in the lower fuel consumption region more frequently, compared to the MPC-based control. This implies that the RL method leads to higher energy efficiency than MPC technique. Hence, the proposed system could achieve lower fuel consumption compared with the baseline system.

To quantify the performance in driving mobility and energy efficiency [33, 34], Table 3 shows the corresponding fuel consumption after SOC-correction and traveling time in these two systems for two selected vehicles (Vehicle 2 and 7). Note that, the corrected fuel consumption in the proposed system is lower than that of the baseline system for both vehicles. These results illustrate that our proposed energy efficiency control can achieve higher control performance than MPC-based control in the lower-level. For the same vehicle, it is obvious that the traveling time in the proposed system is shorter than that of benchmark system. It means the vehicle could reduce red light idling and improve the average running speed, and thus promote the driving mobility by using the proposed system. Thus, the proposed MPC and RL system can improve both driving mobility and energy efficiency.

Furthermore, to evaluate the possible application in real-world, Table 4 displays the computation time of the proposed system and benchmark system for the chosen two vehicles. It can be discerned that the proposed system is shorter than the baseline one and it is feasible to be applied online for automated HEV. The future work includes extending the MPC and RL-enabled system into multiple driving scenarios, and thus it can help the vehicle improve the mobility and fuel economy in real-world environments.

### Table III

| Order | Algorithms | Fuel consumption (g) | Saving (%) | Traveling time (s) | Saving (%) |
|-------|------------|----------------------|------------|--------------------|------------|
| Vehicle 2 | Baseline | 137.1 | — | 685.9 | — |
| | RL | 118.8 | 13.3 | 479.3 | 30.1 |
| Vehicle 7 | Baseline | 145.5 | — | 707.4 | — |
| | RL | 126.4 | 13.1 | 496.5 | 29.8 |

* A 2.7 GHz microprocessor with 3.8 GB RAM was used.
* Traveling distance is 5000 m for each vehicle.
Improving Fuel Economy and Reducing Transactions on In-

The higher level controller and addressing complex decision-making tasks in the lower-level controller.

**REFERENCES**

[1] Kamal MAS, Taguchi S, Yoshimura T. Efficient Driving on Multilane Roads Under a Connected Vehicle Environment. IEEE Transactions on Intelligent Transportation Systems. 2016; 17: 2541-51.

[2] Liu T, Yu H, Guo H, Qin Y, Zou Y. Online Energy Management for Multimode Plug-in Hybrid Electric Vehicles. IEEE Transactions on Industrial Informatics. DOI: 10.1109/TII.2018.2880897.

[3] Mahler G, Vahidi A. Red light avoidance through probabilistic traffic signal timing prediction. IEEE Transactions on Intelligent Transportation Systems. 2014; 15: 2516-2532.

[4] Asadi B, Vahidi A. Predictive Cruise Control: Utilizing Upcoming Traffic Signal Information for Improving Fuel Economy and Reducing Trip Time. IEEE Transactions on Control Systems Technology. 2011; 19: 707-14.

[5] Mahler G, Vahidi A. An Optimal Velocity-Planning Scheme for Vehicle Energy Efficiency Through Probabilistic Prediction of Traffic-Signal Timing. IEEE Transactions on Intelligent Transportation Systems. 2014; 15: 2516-23.

[6] Jin Q, Wu G, Boriboonsomsin K, Barth MJ. Power-Based Optimal Longitudinal Control for a Connected Eco-Driving System. IEEE Transactions on Intelligent Transportation Systems. 2016; 17: 2900-10.

[7] Liu T, Zou Y, Liu D, Sun F. Reinforcement Learning of Adaptive Energy Management With Transition Probability for a Hybrid Electric Tracked Vehicle. IEEE Transactions on Industrial Electronics. 2015; 62: 7837-46.

[8] Qi X, Wu G, Boriboonsomsin K, Barth MJ. Development and Evaluation of an Evolutionary Algorithm-Based Online Energy Management System for Plug-In Hybrid Electric Vehicles. IEEE Transactions on Intelligent Transportation Systems. 2017; 18: 2181-91.

[9] Zou Y, Liu T, Liu D, Sun F. Reinforcement learning-based real-time energy management for a hybrid tracked vehicle. Applied Energy, 2016; 171: 372-82.

[10] Tang X, Zhang D, Liu T, KhajeiPour A, Yu H, Wang H. Research on the energy control of a dual-motor hybrid vehicle during engine start-stop process. Energy, 2019; 166: 1181-93.

[11] Wei Q, Liu D, Lin H. Value Iteration Adaptive Dynamic Programming for Optimal Control of Discrete-Time Nonlinear Systems. IEEE Trans Cybern. 2016; 46: 840-53.

[12] Han J, Kurn D, Park Y. Synthesis of Predictive Equivalent Consumption Minimization Strategy for Hybrid Electric Vehicles Based on Closed-Form Solution of Optimal Equivalence Factor. IEEE Transactions on Vehicular Technology. 2017; 66: 5604-16.

[13] Liu T, Wang B, Yang C. Online Markov Chain-based energy management for a hybrid tracked vehicle with speedy Q-learning. Energy, 2018; 160: 544-55.

[14] Rezaei A, Burl JB, Zhou B. Estimation of the ECMs Equivalent Factor Bounds for Hybrid Electric Vehicles. IEEE Transactions on Control Systems Technology. 2018; 26: 2198-205.

[15] Liu T, Hu X. A Bi-Level Control for Energy Efficiency Improvement of a Hybrid Tracked Vehicle. IEEE Transactions on Industrial Informatics. 2018; 14: 1616-25.
[16] Hu X, Zou Y, Yang Y. Greener plug-in hybrid electric vehicles incorporating renewable energy and rapid system optimization. Energy. 2016; 111: 971-80. DOI: 10.1016/j.energy.2016.03.077.

[17] Vagg C, Akehurst S, Bruce CJ, Ash L. Stochastic Dynamic Programming in the Real-World Control of Hybrid Electric Vehicles. IEEE Transactions on Control Systems Technology. 2016; 24: 853-66.

[18] Moussa HG, Hussenet GA, Abel-Jabar N, Ahmad SE. Use of Model Predictive Control and Artificial Neural Networks to Optimize the Ultrasonic Release of a Model Drug From Liposomes. IEEE Trans Nanobioscience. 2016; 15: 449-56.

[19] Liu T, Hu X, Li SE, Cao D. Reinforcement Learning Optimized Look-Ahead Energy Management of a Parallel Hybrid Electric Vehicle. IEEE/ASME Transactions on Mechatronics. 2017; 22: 1497-507.

[20] Stallato B, Geyer T, Goulart PJ. High-speed finite control set model predictive control for power electronics. IEEE Transactions on power electronics, 2017; 32(5): 4007-4020.

[21] Liu T, Zou Y, Liu D, Sun F. Reinforcement Learning-Based Energy Management Strategy for a Hybrid Electric Tracked Vehicle. Energies. 2015; 8: 7243-60.

[22] HomChaudhuri B, Lin R, Pisupati S. Hierarchical control strategies for energy management of connected hybrid electric vehicles in urban roads. Transportation Research Part C: Emerging Technologies. 2016; 62: 70-86.

[23] Loeb B, Kockelman KM. Fleet performance and cost evaluation of a shared autonomous electric vehicle (SAEV) fleet: A case study for Austin, Texas. Transportation Research Part A: Policy and Practice. 2019; 121: 374-385.

[24] Liu T, Hu X, Hu W, Y Zou. A Heuristic Planning Reinforcement Learning-Based Energy Management for Power-Split Plug-in Hybrid Electric Vehicles. IEEE Transactions on Industrial Informatics, 2019, DOI: 10.1109/TII.2019.2903098.

[25] Malakorn KJ, Park B. Assessment of mobility, energy, and environment impacts of IntelliDrive-based Cooperative Adaptive Cruise Control and Intelligent Traffic Signal control. 2010 IEEE International Symposium on Sustainable Systems and Technology, 2010; 1-6.

[26] Dallinger D, Krampe D, Wietzsch M. Vehicle-to-grid regulation services based on a dynamic simulation of mobility behavior. IEEE Transactions on Smart Grid. 2011; 2(2): 302-313.

[27] HomChaudhuri B, Vahidi A, Pisupati S. Fast Model Predictive Control-Based Fuel Efficient Control Strategy for a Group of Connected Vehicles in Urban Road Conditions. IEEE Transactions on Control Systems Technology. 2017; 25: 760-7.

[28] Yan F, Wang J, Huang K. Hybrid Electric Vehicle Model Predictive Control Torque-Split Strategy Incorporating Engine Transient Characteristics. IEEE Transactions on Vehicular Technology. 2012; 61: 2458-67.

[29] Fotouhi A, Auger DJ, Propp K, Longo S, Purkayastha R, O’Neill L, Waluś, S. Lithium-Ion Cell Equivalent Circuit Network Model Parameterization and Sensitivity Analysis. IEEE Transactions on Vehicular Technology. 2017; 66: 7711-21.

[30] Chadou I, Chopron G, Cors M J, Garcia F, Sabbadin R. MDPoolbox: a multi - platform toolbox to solve stochastic dynamic programming problems. Ecography. 2014; 37(9): 916-920.

[31] Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: A survey. Journal of artificial intelligence research, 1996; 4: 237-285.

[32] Jin I G, Orozco G. Dynamics of connected vehicle systems with delayed acceleration feedback. Transportation Research Part C: Emerging Technologies. 2014; 46: 46-64.

[33] Hu J, Shao Y, Sun Z, Wang M, Bared J, Huang Peter. Integrated optimal eco-driving on rolling terrain for hybrid electric vehicle with vehicle-to-vehicle communication. Transportation Research Part C: Emerging Technologies, 2016; 68: 228-244.

[34] Galus MD, Warach R A, Noembrini F, Steurs K, Georges G, Boulouchos K, Aehusen KW, Andersson G. Integrating power systems, transport systems and vehicle technology for electric mobility impact assessment and efficient control. IEEE Transactions on Smart Grid, 2012; 3(2): 934-949.

Teng Liu (M’2018) received the B.S. degree in mathematics from Beijing Institute of Technology, Beijing, China, in 2011. He received his Ph.D. degree in the automotive engineering from Beijing Institute of Technology (BIT), Beijing, in 2017. His Ph.D. dissertation, under the supervision of Pro. Fengchun Sun, was entitled “Reinforcement learning-based energy management for hybrid electric vehicles.” He worked as a research fellow in Vehicle Intelligence Pioneers Ltd for one year. Now, he is a member of IEEE VTS, IEEE ITS, IEEE SMC and IEEE/CAA. Dr. Liu is now a postdoctoral fellow at Department of Mechanical and Mechatronics Engineering, University of Waterloo, Ontario N2L3G1, Canada. Dr. Liu has more than 8 years’ research and working experience in renewable vehicle and connected autonomous vehicle. His current research focuses on reinforcement learning (RL)-based energy management in hybrid electric vehicles, RL-based decision making for autonomous vehicles, and CPSS-based parallel driving. He has published over 30 SCI papers and 10 conference papers in these areas. He received the Merit Student of Beijing in 2011, the Teli Xu Scholarship (Highest Honor) of Beijing Institute of Technology in 2015, “Top 10” in 2018 IEEE VTS Motor Vehicle Challenge and sole outstanding winner in 2018 ABB Intelligent Technology Competition. Dr. Liu is a workshop co-chair in 2018 IEEE Intelligent Vehicles Symposium (IV 2018) and has been reviewer in multiple SCI journals, selectively including IEEE Trans. Industrial Electronics, IEEE Trans. on Intelligent Vehicles, IEEE Trans. Intelligent Transportation Systems, IEEE Transactions on Systems, Man, and Cybernetics: Systems, IEEE Transactions on Industrial Informatics, Advances in Mechanical Engineering.

Bo Wang received B. S. degree in Mathematics from Beijing Institute of Technology, China in 2011. He received Ph. D. degree in Mathematics from Beijing Normal University, China, in 2016. He is currently an Assistant Professor at Beijing Institute of Technology, China. His research focuses on partial differential equations and their applications.

Dongpu Cao received the Ph.D. degree from Concordia University, Canada, in 2008. He is currently an Associate Professor and Director of Driver Cognition and Automated Driving (DC-Auto) Lab at University of Waterloo, Canada. His research focuses on vehicle dynamics and control, driver cognition, automated driving and parallel driving, where he has contributed more than 170 publications and 1 US patent. He received the ASME AVTT’2010 Best Paper Award and 2012 SAE Arch T. Colwell Merit Award. Dr. Cao serves as an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE/ASME TRANSACTIONS ON MECHATRONICS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS and ASME JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT AND CONTROL. He has been a Guest Editor for VEHICLE SYSTEM DYNAMICS, and IEEE TRANSACTIONS ON SMC. SYSTEMS. He has been serving on the SAE International Vehicle Dynamics Standards Committee and a few ASME, SAE, IEEE technical committees, and serves as Co-Chair of IEEE ITSS Technical Committee on Cooperative Driving.

Xiaolun Tang received B.S. degree in Mechanics engineering and M.S. Degree in Vehicle Engineering from Chongqing University, China, in 2006 and 2009, respectively. He received the Ph.D. degree in Mechanical Engineering from Shanghai Jiao Tong University, China, in 2015. He is currently an Assistant Professor at the State Key Laboratory of Mechanical Transmissions and the Department of Automotive Engineering, Chongqing University, Chongqing, China. His research focuses on Hybrid Electric Vehicles (HEVs), Vehicle Dynamics, Noise and Vibration, and Transmission Control.

Yalian Yang received the Ph.D. degree in mechanical engineering from Chongqing University, Chongqing, China, in 2002. He is currently a Professor with the State Key Laboratory of Mechanical Transmissions, Chongqing University. His research interests include alternative energy drivelines, virtual reality driving simulator, and vehicle dynamic simulation and control.