Hierarchical Q-learning network for online simultaneous optimization of energy efficiency and battery life of the battery/ultracapacitor electric vehicle

Xu, Bin; Zhou, Quan; Shi, Junzhe ; Li, Sixu

DOI:
10.1016/j.est.2021.103925

License:
Other (please provide link to licence statement

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Xu, B, Zhou, Q, Shi, J & Li, S 2022, 'Hierarchical Q-learning network for online simultaneous optimization of energy efficiency and battery life of the battery/ultracapacitor electric vehicle', *Journal of Energy Storage*, vol. 46, 103925. https://doi.org/10.1016/j.est.2021.103925

Journal of Energy Storage 46 (2022) 103925

Contents lists available at ScienceDirect

Journal of Energy Storage

journal homepage: www.elsevier.com/locate/est

Research Papers

Hierarchical Q-learning network for online simultaneous optimization of energy efficiency and battery life of the battery/ultracapacitor electric vehicle

Bin Xu [a], Quan Zhou [b,*], Junzhe Shi [c], Sixu Li [d]

[a] The University of Oklahoma, School of Aerospace and Mechanical Engineering, 865 Asp Ave, Norman, OK 73019, USA
[b] University of Birmingham, Vehicle and Engine Research Centre, Birmingham, UK
[c] University of California, Berkeley, Department of Civil and Environmental Engineering, Berkeley, CA 94720, USA
[d] University of California, Berkeley, Department of Mechanical Engineering, 6141 Etcheverry Hall, Berkeley, CA 94720, USA

ARTICLE INFO

Keywords:
Energy management
Reinforcement learning
Q-learning
Electric vehicle
Ultracapacitor
Battery

ABSTRACT

Reinforcement learning has been gaining attention in energy management of hybrid power systems for its low computation cost and great energy saving performance. However, the potential of reinforcement learning (RL) has not been fully explored in electric vehicle (EV) applications because most studies on RL only focused on single design targets. This paper studied on online optimization of the supervisory control system of an EV (powered by battery and ultracapacitor) with two design targets, maximizing energy efficiency and battery life. Based on a widely used reinforcement learning method, Q-learning, a hierarchical learning network is proposed. Within the hierarchical Q-learning network, two independent Q tables, Q1 and Q2, are allocated in two control layers. In addition to the baseline power-split layer, which determines the power split ratio between battery and ultracapacitor based on the knowledge stored in Q1, an upper layer is developed to trigger the engagement of the ultracapacitor based on Q2. In the learning process, Q1 and Q2 are updated during the real driving using the measured signals of states, actions, and rewards. The hierarchical Q-learning network is developed and evaluated following a full propulsion system model. By introducing the single-layer Q-learning based method and the rule-based method as two baselines, performance of the EV with the three control methods (i.e., two baseline and one proposed) are simulated under different driving cycles. The results show that the addition of an ultracapacitor in the electric vehicle reduces the battery capacity loss by 12%. The proposed hierarchical Q-learning network is shown superior to the two baseline methods by reducing 8% battery capacity loss. The vehicle range is slightly extended along with the battery life extension. Moreover, the proposed strategy is validated by considering different driving cycle and measurement noise. The proposed hierarchical strategy can be adapted and applied to reinforcement learning based energy management in different hybrid power systems.

1. Introduction

Over the past a few years, vehicle electrification has gained momentum in the automotive field for the perspective of energy saving [1] and environmental protection [2]. Electric car companies like Tesla are leading the electrification revolution, which pressures giant Original Equipment Manufacturers (OEMs) like General Motors to shift from internal combustion engine cars to electric cars as General Motors announced new brand logo advocating for vehicle electrification.

As the main power source for the electric vehicle (EV), battery is facing many challenges in real-world operations, such as long charging time [3], high replacement cost [4], short range [5], and lack of charging infrastructures [6]. To reduce the battery replacement cost, an effective way is to regulate the battery usage profile to extend battery life. According to the mechanism of battery degradation, battery should reduce the operation at high current charging/discharging, high temperature, and high state-of-charge (SOC) [7]. Ultracapacitor is used as the second power source of electric vehicle to improve vehicle acceleration performance and reduce battery charging/discharging current [8]. Ultracapacitor is a type of energy storage device with high power density and low energy density [9]. The high-power density characteristic can compensate the low power density of lithium-ion battery at high

[*] Corresponding author: University of Birmingham, Vehicle and Engine Research Centre, Birmingham, UK.
E-mail addresses: binxu@ou.edu (B. Xu), q.zhou@bham.ac.uk (Q. Zhou), junzhe@berkeley.edu (J. Shi), sixu_li@berkeley.edu (S. Li).

https://doi.org/10.1016/j.est.2021.103925
Received 15 October 2021; Received in revised form 27 November 2021; Accepted 28 December 2021
Available online 5 January 2022

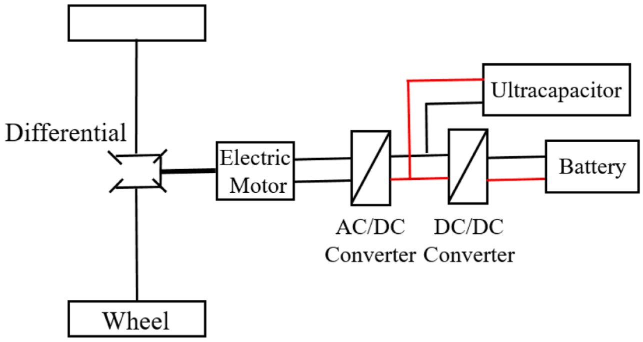

Fig. 1. Configuration of the vehicle powertrain.

power demand driving scenarios during vehicle acceleration and braking. The peak power reduction can effectively slows down lithium-ion battery degradation [10] and thus possibly avoids battery replacement over the vehicle life span. In addition, ultracapacitor has long cycle life up to 1 millions cycles [11] and its aging effect is usually not considered in vehicle application [8].

Energy management, which controls the onboard energy flows, is one of the key functionalities in EVs and a bad energy management system could lead to discounted performance [12]. For battery electric vehicles (BEV), all the traction power and regenerative braking power are supplied and absorbed by the lithium-ion battery. Different from BEV, the addition of the ultracapacitor increases flexibility in energy management of the battery/ultracapacitor electric vehicle (BUEV). Derived from expert knowledge or heuristic rules, rule-based methods are commonly used in vehicle energy management [13]. Rule-based methods are computationally efficient, but their performance are not optimal and highly dependent on the developers' experience [14]. They can only achieve acceptable performance under given conditions at early-stage proof-of-concept but hardly be optimal to various driving conditions that beyond the developers' experience [15].

Optimization-based methods are shown to be superior to the rule-based methods and can be obtained offline or online. Offline optimization (e.g., dynamic programming [16]) can determine the theoretical global optimum solution in given driving cycles but are computational costly and cannot be directly used for real-time control [17], whereas it requires prior knowledge of the driving cycle and is computationally intense. Zhou et al. developed parametric [32] and non-parametric models [33] to implement dynamic programming results into real-time control. Online optimization (e.g., Equivalent consumption minimization strategy [18], Q-learning [19][34]) requires less computational cost and is capable of achieving nearly global optimal solutions in real-time [20,21]. Nassef et al. implemented the equivalent consumption minimization strategy (ECMS) and the mine blast optimization in a battery/ ultracapacitor/ fuel cell system to minimize the energy consumption [12]. Sun et al. applied Q-learning in online optimization of the energy management strategy for a battery-ultracapacitor-fuel cell vehicle [22]. Compared to the same vehicle controlled by ECMS, 7–10% hydrogen can be saved via Q-learning under highway and city driving cycles. In summary, most of the existing literatures in battery-ultracapacitor electric vehicle area focused on saving of energy or fuel but have not paid sufficient attention on maximization of battery life which is also an important design target in EV.

A new hierarchical Q-learning network is proposed in this paper to enable online simultaneous optimization of energy efficiency and battery life of the BUEV. It allocates two independent Q tables, Q1 and Q2, in two control layers for control action execution and reinforcement learning. In addition to the power-split layer, which determines the power split ratio between battery and ultracapacitor based on the knowledge stored in Q1, an upper layer is developed to trigger the engagement of the ultracapacitor based on Q2. The engagement layer provides extra control over the ultracapacitor and adds more flexibility in the energy management system, which has the potential to improve the energy efficiency and battery life for the electric vehicle. In the learning process, Q1 and Q2 are updated during the real driving using the measured signals of vehicle states, actions, and rewards. The research is based on a full set of propulsion system model including the battery aging model that was identified with experimental data. The proposed hierarchical Q-learning framework is compared with a single-layer Q-learning and a rule-based strategy for benchmarking. The academic contributions of this paper are summarized as follows,

1) A new optimization-based energy management strategy is developed based on the hierarchical Q-learning network including two Q-learning agents, which can develop their optimal policies in a way like human learning but more computational accurate in action execution to ensure the control performance.
2) By interpreting the hierarchical Q-learning with optimal policy map extracted from Q values, the optimal vehicle operating settings, including vehicle speed, EM torque, battery power and SOC, ultracapacitor power and SOC, are obtained for different driving cycles.

The rest of the paper is organized as follows: the propulsion system model is presented in Section 2, followed by the introduction of the hierarchical Q-learning framework in Section 3. Section 4 details the hierarchical Q-learning interpretation, comparative study of three strategies and validation of the proposed strategy. Finally, the paper ends with conclusion in Section 5.

2. Vehicle and propulsion system model

The vehicle of interest in this study is a passenger electric vehicle powered by lithium-ion battery and ultracapacitor. Ultracapacitor adds flexibility in the power sources and shares the power demand with the battery. The topology of the battery/ ultracapacitor electric vehicle is shown in Fig. 1. Both the ultracapacitor and battery connect to the AC/ DC converter, which supplies power for the electric motor in propulsion mode and absorbs power from the electric motor in regenerating braking mode. The electric motor connects to the differential, via which the power is transmitted to/ from the wheels. In this section, the models of the components shown in Fig. 1 are presented, including ultracapacitor, battery, AC/DC converter, DC/DC converter, and electric motor. Additionally, vehicle dynamics model is presented for the calculation of vehicle speed and driver model is presented for the calculation of acceleration/ braking pedal positions.

2.1. Ultracapacitor

Given the power demand $\mathrm{P_{cap}}$ [W], the terminal voltage $\mathrm{U_{cap,t}}$ [V] and current $\mathrm{I_{cap}}$ [A] of the ultracapacitor can be calculated using (1) and (2) [20]. $\mathrm{U_{cap,oc}}$ [V] is open circuit voltage, and $\mathrm{R_{cap}}$ [Ω] is the capacitor resistance.

$$\mathrm{U_{cap,t}} = \mathrm{U_{cap,oc}} - \mathrm{I_{cap}R_{cap}} \tag{1}$$

$$\mathrm{I_{cap}} = \frac{\mathrm{P_{cap}}}{\mathrm{U_{cap,t}}} \tag{2}$$

The current of the ultracapacitor can also be expressed as a function of capacitance $\mathrm{C_{cap}}$ [Ah], maximum voltage $\mathrm{U_{cap,max}}$ [V] and state-of-charge $\mathrm{SOC_{cap}}$ [%] as shown in (3), which derives the SOC in (4) using integration. $\mathrm{SOC_{cap}(0)}$ [%] is the initial charge of the ultracapacitor.

$$I_{cap} = C_{cap}U_{cap,\max}S\dot{O}C_{cap} \tag{3}$$

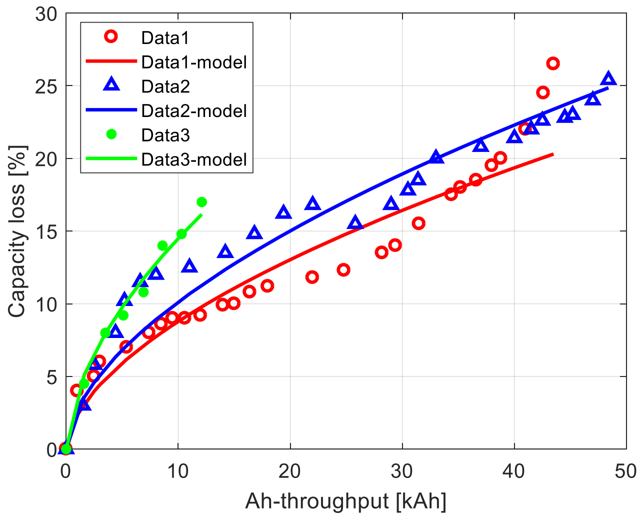

Fig. 2. Comparison of experimental data and the results from identified model.

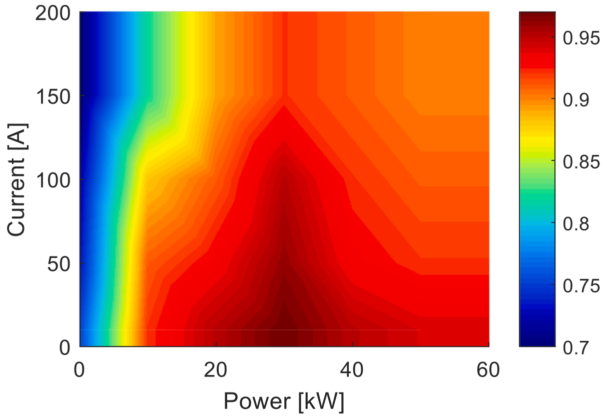

Fig. 3. DC/DC converter efficiency map.

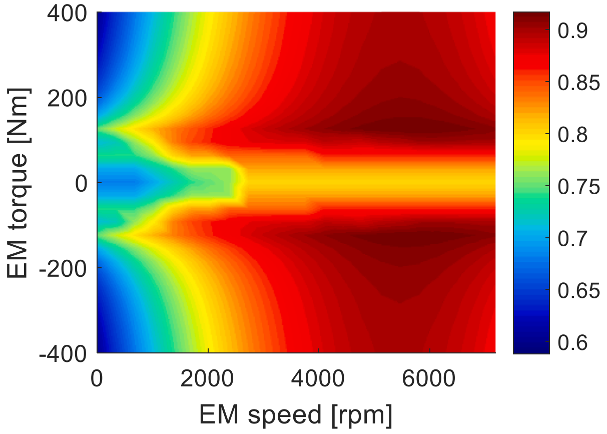

Fig. 4. Electric motor (EM) efficiency map.

$$\mathrm{SOC_{cap}}(t) = \mathrm{SOC_{cap}}(0) - \frac{\int_0^t \mathrm{I_{cap}}(\tau)d\tau}{\mathrm{C_{cap}U_{cap,max}}} \tag{4}$$

2.2. Battery

Given the power demand $\mathrm{P_{bat}}$ [W], the terminal voltage $\mathrm{U_{bat,t}}$ [V] and current $\mathrm{I_{bat}}$ [A] of the battery are calculated using (5) and (6). $\mathrm{R_{bat}}$ [Ω] is the internal resistance of the battery.

$$\mathrm{U_{bat,t}} = \mathrm{U_{bat,oc}} - \mathrm{I_{bat}R_{bat}} \tag{5}$$

$$\mathrm{I_{bat}} = \frac{\mathrm{P_{bat}}}{\mathrm{U_{bat,t}}} \tag{6}$$

The battery current can also be expressed as a function of battery nominal capacity $\mathrm{Q_{bat}}$ [Ah] and battery SOC as shown in (7) [21], using which the battery SOC can be derived in (8). $\mathrm{SOC_{bat}}(0)$ is the initial charge of the battery.

$$I_{bat} = Q_{bat}\dot{SOC}_{bat} \tag{7}$$

$$\mathrm{SOC_{bat}}(t) = \mathrm{SOC_{bat}}(0) - \frac{\int_0^t \mathrm{I_{bat}}(\tau)d\tau}{\mathrm{Q_{bat}}} \tag{8}$$

Battery degradation is discussed in this study, thus an empirical battery capacity loss model (severity factor model) is adopted [22]. The capacity loss $\mathrm{Q_{loss}}$ [Ah] is expressed as a function of severity factor σ and Ah-throughput in (9). z is a coefficient to be identified.

$$\mathrm{Q_{loss}}(\mathrm{SOC_{bat}}, \mathrm{I_{bat,c}}, \mathrm{T_{bat}}, \mathrm{Ah}) = \sigma(\mathrm{SOC_{bat}}, \mathrm{I_{bat,c}}, \mathrm{T_{bat}})\mathrm{Ah}^z \tag{9}$$

The expression of severity factor is shown in (10), which is a function of battery SOC, C-rate $\mathrm{I_{bat,c}}$ and temperature $\mathrm{T_{bat}}$ [K]. α, β, δ are the coefficients to be identified. E is the activation energy, which is 31,500 [$\mathrm{Jmol^{-1}}$]. R is the universal gas constant, which is 8.3145 [$\mathrm{JK^{-1}mol^{-1}}$].

$$\sigma(\mathrm{SOC_{bat}}, \mathrm{I_{bat,c}}, \mathrm{T_{bat}}) = (\alpha\mathrm{SOC_{bat}} + \beta)\exp\left(\frac{-E + \delta\mathrm{I_{bat,c}}}{\mathrm{RT_{bat}}}\right) \tag{10}$$

The C-rate is calculated using battery current and battery nominal capacity as follows:

$$\mathrm{I_{bat,c}} = \frac{|\mathrm{I_{bat}}|}{\mathrm{Q_{bat}}} \tag{11}$$

Using experimental data from [23] and [24], the coefficients of the severity factor model are identified. For the identification results, z is 0.5715, α is 2.0161, β is 4398.5, and δ is 112. The results from the identified model and the experimental data are shown in Fig. 2. The R squared values for the three curves are 0.9085, 0.9458 and 0.9871, respectively.

2.3. Converters

In this study, there are two converters: AC/DC converter and DC/DC converter. The efficiency map of the DC/DC converter is shown in Fig. 3 [20]. The efficiency of the AC/DC converter is assumed to be 92%.

2.4. Electric motor

The electric motor has two modes: traction and generation. The efficiency map of the electric motor is shown in Fig. 4. The electric motor is in traction mode when the torque is positive and in generation mode when the torque is negative.

2.5. Driver

The driver model determines the acceleration pedal position θ_{acc} [%] and braking pedal position θ_{brk} [%] as follows:

$$\theta_{\mathrm{acc}} = \begin{cases} \min(1.0, \mathrm{u_{driver}}) * 100\%, \text{ if } \mathrm{u_{driver}} > 0 \\ 0, \text{ if } \mathrm{u_{driver}} \le 0 \end{cases} \tag{12}$$

$$\theta_{\mathrm{brk}} = \begin{cases} \min(1.0, -\mathrm{u_{driver}}) * 100\%, \text{ if } \mathrm{u_{driver}} < 0 \\ 0, \text{ if } \mathrm{u_{driver}} \ge 0 \end{cases} \tag{13}$$

Table 1
Specification of the vehicle and energy storage units.

Parameters	Values
Vehicle curb weight	1722 kg
Number of batteries in series connection	98
Number of batteries in parallel connection	60
Rated capacity of a single battery cell	2.4 Ah
Number of ultracapacitor cells in series	50
Number of ultracapacitor cells in parallel	1
Ultracapacitor capacitance (single unit)	1200 F

where u_{driver} is the output of the driver control. The aim of the driver control is to follow the vehicle of the driving cycle. The control is composed of a feedforward term u_{ff} and a feedback term u_{fb}. The feedback term is shown in (14) and it is a PI control using vehicle tracking error e_v [m/s] as the input. P and I gains k_p,k_i are 0.25 and 0.03. The feedforward term is shown in (15), which is mainly a function of vehicle speed. $T_{EM,max}$/ $T_{EM,min}$ [Nm] are boundaries of electric motor torque. $\mathscr{E}$ is the vehicle static weight distribution, which is 49%/51%. H_{CG} is the height of vehicle central gravity, which is 0.5 [m]. m is the vehicle curb weight, which is 1620 [kg]. g is the gravity constant. r_{whl} is the wheel radius, which is 0.25 [m], B_w is the wheelbase, which is 2.55 [m], J_v is the vehicle inertia, which is 150 [kgm^2], c_0, c_1, c_2 are the road law coefficients, which are 105.95, 0.01 and 0.434, respectively.

$$u_{fb} = k_p e_v(t) + k_i \int_0^t e_v(\tau) d\tau \tag{14}$$

$$u_{ff} = \begin{cases} \dfrac{T_1}{T_{EM,max}}, \ T_1 \geq 0, \\ \left[(1-\mathscr{E}) - \dfrac{H_{CG}}{mgr_{whl}B_w}\right]\dfrac{T_1}{-T_{EM,min}}, \ T_1 < 0 \end{cases} \tag{15}$$

$$T_1 = \frac{\dot{v}J_v}{r_{whl}} + r_{whl}\left(c_0 + c_1 v + c_2 v^2\right) \tag{16}$$

The pedal positions from the driver model are used to calculate the vehicle torque and power demand as follows:

$$T_{dmd} = \theta_{acc}T_{EM,max} + \ \theta_{brake}T_{EM,min} \tag{17}$$

$$P_{dmd} = \omega_{EM}T_{dmd} \tag{18}$$

2.6. Vehicle dynamics

The force provided for vehicle acceleration ma [N] is the combination of four forces: aerodynamic force $\frac{1}{2}\rho C_d A v_{veh}^2$, rolling resistance force $\cos(\beta)f_{roll}mg$, gravitational force $\sin(\beta)mg$ and traction force F_{trc}, which is shown in (19). In the equation, ρ [kg/m^3] is the air density. C_d is the aerodynamic drag coefficient. A [m^2] is the frontal projection area of the vehicle. β is the slop of the road. f_{roll} is the rolling resistance coefficient.

$$ma = \frac{1}{2}\rho C_d A v^2 + \cos(\beta)f_{roll}mg + \sin(\beta)mg + F_{trc} \tag{19}$$

The traction force is connected to the electric motor torque T_{EM} [Nm] in (20). T_{whl} [Nm] is the torque applied to the wheels. r_{EM} is the gear reduction ratio of the electric motor.

$$F_{trc} = \frac{T_{whl}}{r_{whl}} = \frac{T_{EM}r_{EM}}{r_{whl}} \tag{20}$$

The electric motor speed is calculated using vehicle speed as follows:

$$\omega_{EM} = \omega_{whl}r_{EM} = \frac{v}{r_{whl}}r_{EM} \tag{21}$$

The electric motor power is calculated in (22) based on the electric motor torque, which equals to the torque demand calculated in (17). η_{EM} is the electric motor efficiency shown in Fig. 4.

$$P_{EM} = \begin{cases} \dfrac{\omega_{EM}T_{EM}}{\eta_{EM}}, \ discharge \\ \omega_{EM}T_{EM}\eta_{EM}, \ charge \end{cases} \tag{22}$$

The power of ultracapacitor and battery are calculated using following two equations:

$$P_{cap} = \begin{cases} r_{cap}P_{EM}, \ q_{cap} = 1 \\ 0, \ q_{cap} = 0 \end{cases} \tag{23}$$

$$P_{bat} = P_{EM} - P_{cap} \tag{24}$$

where q_{cap} is the ultracapacitor engage signal, r_{cap} is the power split ratio for ultracapacitor. The specification of the energy storage systems are listed in Table 1.

3. Supervisory control strategies

In this section, three supervisory control strategies are introduced and the diagram of the three strategies are shown in Fig. 5. Hierarchical Q-learning strategy has two layers, which are centered around Q_1 table and Q_2 table respectively, as illustrated in Fig. 5(a). Layer one determines the engagement of ultracapacitor and layer two determines the power split between battery and ultracapacitor. On the bottom of Fig. 5 (a), it shows the update of two Q tables utilize state, reward, and action information. In comparison, a single-layer Q-learning strategy that only considers power split ratio but with the engagement of ultracapacitor is developed in Fig. 5(b). The third strategy is a baseline strategy that does not consider ultracapacitor in the vehicle and all the power is supplied by battery. The comparison between Q-learning and rule-based in BUEV has been well studied and some details can be found in [25, 26].

3.1. Hierarchical Q-learning

The proposed hierarchical Q-learning strategy is a model-free strategy, which does not require a control-oriented model. It learns to control the usage of battery and ultracapacitor by interacting with the plant model. Due to the only reward, state and action information exchange between the strategy and the plant model and no physics principles are needed by the strategy, the plant model is regarded as a black box. Thus, once it is developed for one type of vehicle propulsion architecture, it can be easily adapted for different type of architectures. The diagram of the black box vehicle model and the hierarchical Q-learning is shown in Fig. 5. The black box model takes the two actions (i.e., ultracapacitor engage signal and ultracapacitor power split ratio) and outputs state and reward information to the hierarchical Q-learning strategy. In the middle of vehicle simulation, the Q-learning strategy only takes the state information form the vehicle model and output action. After the vehicle simulation, the state and reward information from the vehicle model and the corresponding action are utilized to update two Q tables. The updated Q tables are then applied to the next round of vehicle simulation.

In this study, the ultracapacitor engagement signal (q_{cap}) and ultracapacitor power split ratio (r_{cap}) are selected as the action space. The engagement signal is chosen between two values 0 and 1, whereas power split is chosen in the range of 0 and 1. Discretization detail is given later. Vehicle power demand (P_{dmd}) and ultracapacitor SOC (SOC_{cap}) are selected to form a two-dimensional state space. The lower and upper boundaries of the vehicle power demand are set to be −30 kW and 50 kW in this paper. The lower and upper boundaries of the ultracapacitor SOC are set to be 0%, and 100%, respectively.

The reward function in the reinforcement learning has the similar effect as the cost function in conventional optimal controls. In this vehicle propulsion system supervisory control problem, the energy

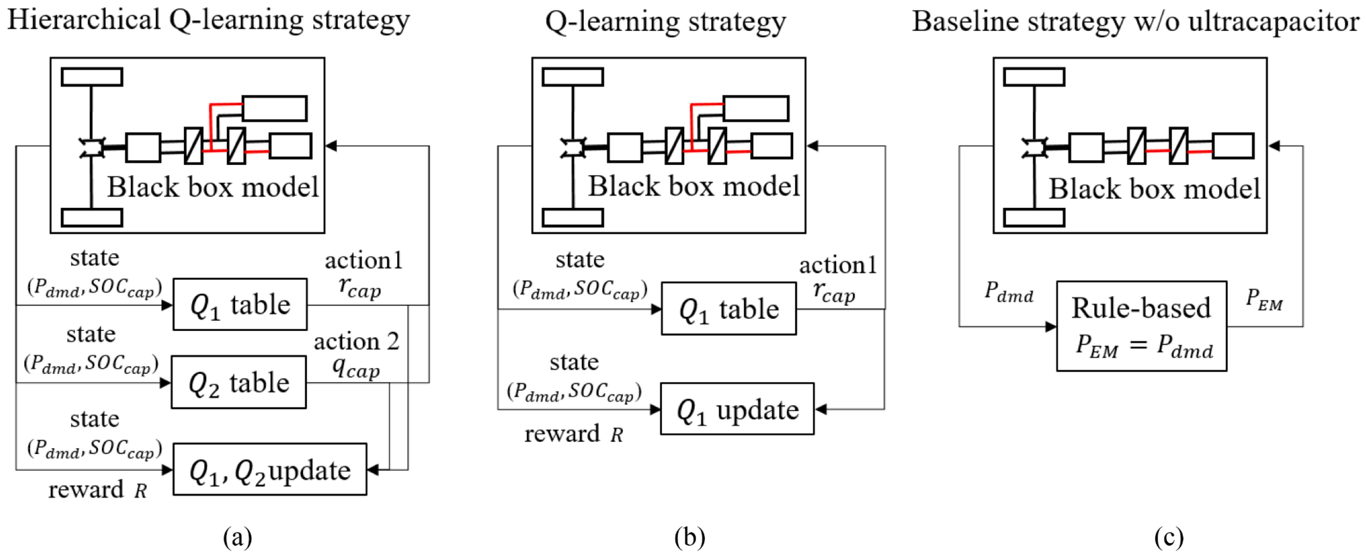

Fig. 5. Diagram of the three strategies used in this study: a) hierarchical Q-learning strategy; b) Q-learning strategy; and c) baseline strategy without ultracapacitor.

Table 2
Pseudocode of the hierarchical Q-learning algorithm.

```
Hierarchical Q-learning Algorithm

Initialize $Q_1(s, a_1)$, $Q_2(s, a_2)$ with zeros, for all $s \in S$, $(a_1, a_2) \in A(s)$.
Initialize $R_{tot}$ with zero.
for $i \in (1, \dots, N)$ do (for each episode):
    Experience exploration:
    Initialize s
    for $j \in (1, \dots, M)$ do (for each time step of episode):
        Choose action $(a_{1,j}, a_{2,j})$ at state s using policy derived from $Q_1, Q_2$
        ($\varepsilon$-greedy action selection method and $\varepsilon$ is constant at 0.05)
        Take action $(a_{1,j}, a_{2,j})$, observe $R_j$, $s_{j+1}$ from environment
    end for
    Experience evaluation:
    if $\sum_1^M R_j > R_{tot}$ do
        Q value function update:
        for $j \in (1, \dots, M)$ do (for each time step of episode):
            $Q_1(s_j, a_{1,j}) \leftarrow (1 - \mu_1) Q_1(s_j, a_{1,j}) + \mu_1 \left[ R_j + \gamma_1 \max_{a_1} Q_1(s_{j+1}, a_1) \right]$
            $Q_2(s_j, a_{2,j}) \leftarrow (1 - \mu_2) Q_2(s_j, a_{2,j}) + \mu_2 \left[ R_j + \gamma_2 \max_{a_2} Q_2(s_{j+1}, a_2) \right]$
        end for
        Experience evaluation criteria update:
        Initialize s
        for $j \in (1, \dots, M)$ do (for each time step of episode):
            Choose action $(a_{1,j}, a_{2,j})$ at state s using policy derived from
            $Q_1, Q_2$ ($\varepsilon$-greedy action selection method, $\varepsilon = 0$)
            Take action $(a_{1,j}, a_{2,j})$, observe $R_j$, $s_{j+1}$ from environment
        end for
        $R_{tot} \leftarrow \sum_1^M R_j$
    end if
end for
```

consumption of the battery and the ultracapacitor and the battery aging are considered in the reward function as follows:

$$R = -w_E \frac{\left(P_{bat} + P_{cap}\right) \Delta t}{E_{bat,norm} + E_{cap,norm}} - (1 - w_E) \frac{\sigma}{\sigma_{norm}} + b \tag{25}$$

where w_E represents the weighting factor of energy consumption, $E_{bat,norm}$, $E_{cap,norm}$ are used to normalize the energy consumption by the battery and ultracapacitor at Δt duration, σ_{norm} is used to normalize the severity factor, b is a constant positive bias to ensure a positive reward.

There are two types of Q-learning methods based on the way Q value information storage: approximate method and tabular method. Approximate Q-learning method stores the Q value information in linear

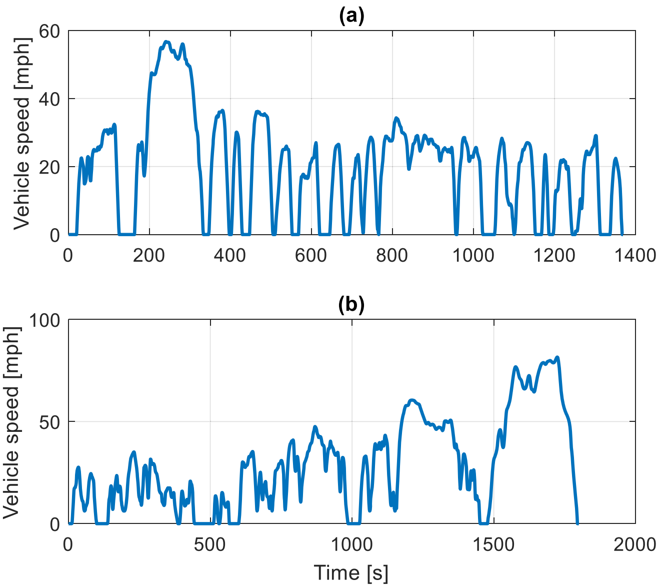

Fig. 6. Driving cycle speed profiles: (a) UDDS, and (b) WLTP.

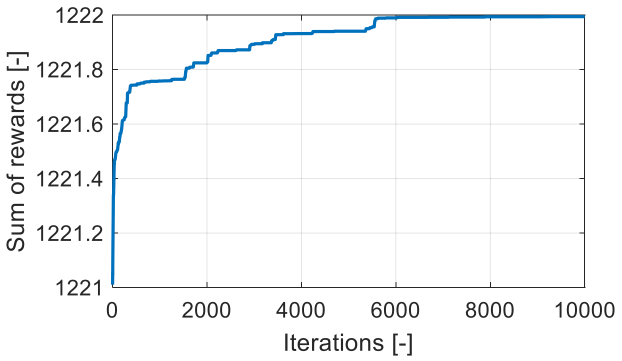

Fig. 7. The highest accumulated rewards along the iterations. (Each iteration is an entire UDDS driving cycle and the sum of rewards are calculated from the entire driving cycle. The curve is the average of 5 run).

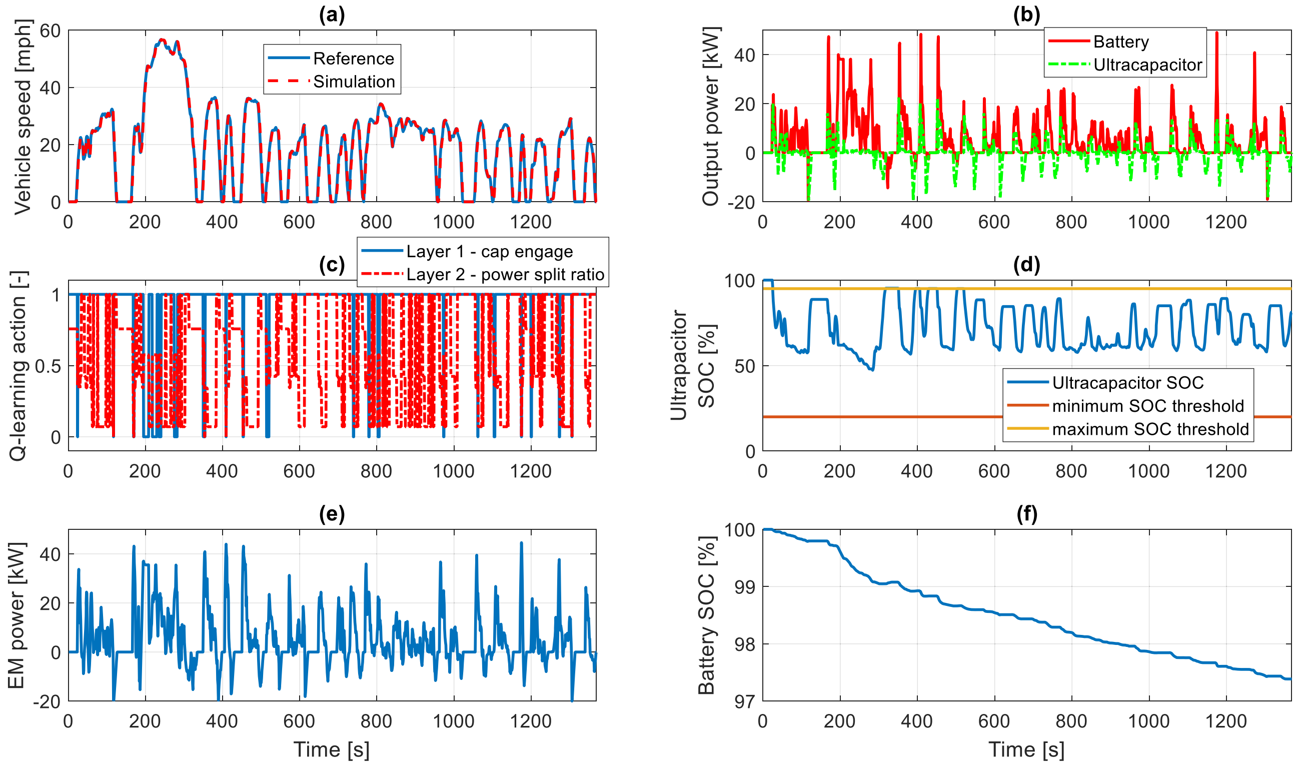

Fig. 8. Results from UDDS simulation using hierarchical Q-learning strategy: (a) Vehicle speed, (b) Output power of battery and ultracapacitor, (c) EM power, (d) ultracapacitor SOC, (e) Q-learning action, and (f) battery SOC. (cap engage: ultracapacitor engagement. 1 - engage, 0 – disengage.).

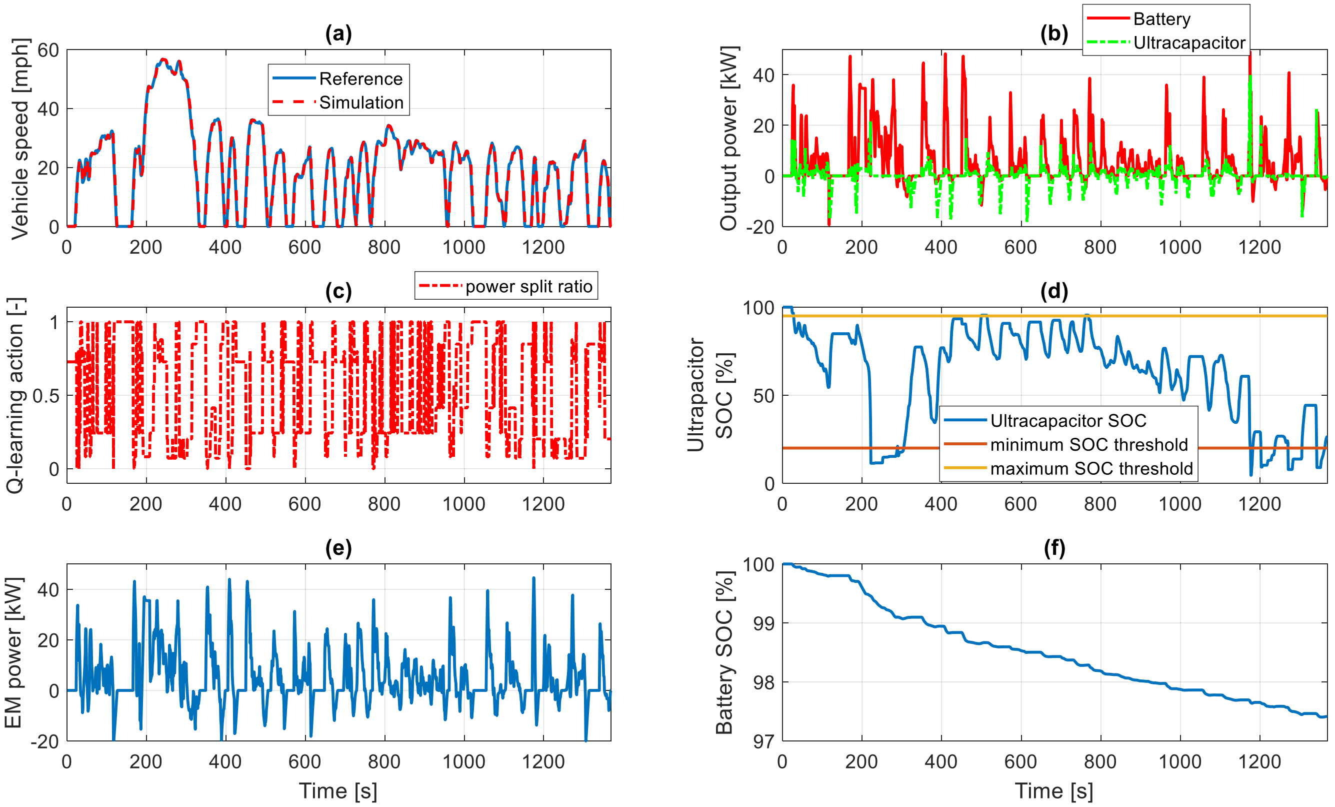

Fig. 9. Results from UDDS simulation using Q-learning strategy: (a) Vehicle speed, (b) Output power of battery and ultracapacitor, (c) EM power, (d) ultracapacitor SOC, (e) Q-learning action, and (f) battery SOC.

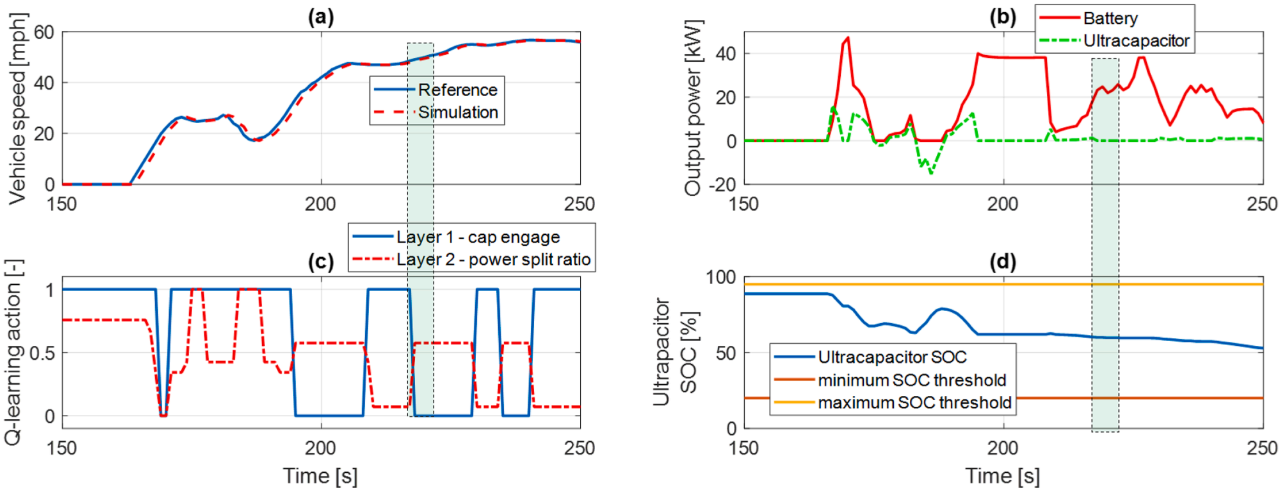

Fig. 10. Zoom-in of hierarchical Q-learning simulation results in Fig. 8:(a) Vehicle speed, (b) Output power of battery and ultracapacitor, (c) EM power, and (d) ultracapacitor SOC. (The highlighted green section is between 217 s and 222 s).

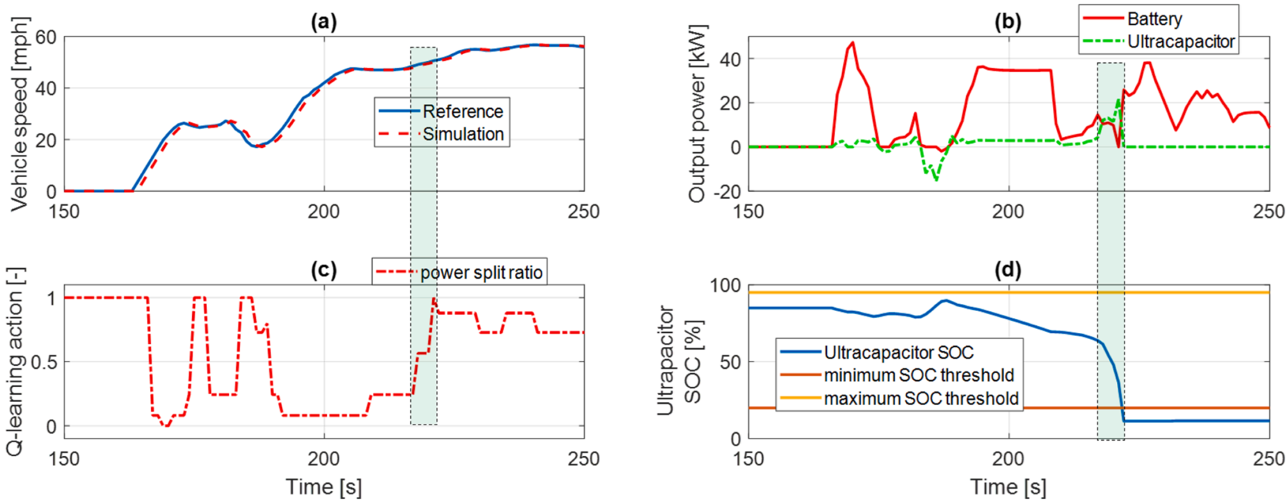

Fig. 11. Zoom-in of Q-learning simulation results in Fig. 9:(a) Vehicle speed, (b) Output power of battery and ultracapacitor, (c) EM power, and (d) ultracapacitor SOC. (The highlighted green section is between 217 s and 222 s).

or nonlinear correlations. During the learning process, the coefficients of the correlations are updated. Tabular Q-learning method directly stores the Q value information in lookup tables. Sutton et al. stated in a book that the approximate method only finds approximate value function [27]. This contrasts with the tabular method, which finds the exact optimal value function. Thus, this study focuses on the tabular Q-learning method. In this study, for each action, the Q values are stored as a vector, whose size is determined by the discretization dimension of state and action. The state discretization dimension is 5 for each state and action discretization dimension is 100. The detail of state and action discretization dimension study can be found in [28]. Therefore, for each action, the Q value vector size is 2500 (i.e., 5 × 5 × 100).

In this study, the pseudo-code of the hierarchical Q-learning algorithm is shown in Table 2. For the proposed hierarchical Q-learning strategy, the two Q vectors are firstly initialized with zeros. Then a full UDDS simulation is conducted, during which the two actions are taken by the Q-learning using ε-greedy policy method [29]. In the decision-making of each action, there is (1- ε) possibility that the action pointing to the largest Q value in the current state is chosen and there is ε possibility that the action is randomly selected within the upper and lower boundaries. ε is set as 0.05 in this study and the details of ε optimization can be found in [28]. After the two actions are decided and taken for one time step, an experience vector (s_i, a_i, R_i, s_{i+1}) is collected, where s_i is the state before the action taken, a_i is the action, R_i is the reward obtained in the state transition,s_{i+1} is the new state after action taken. Over the 1369s of UDDS driving cycle, 1369 sets of experience vector are collected. After the experience evaluation, the two Q tables are updated based on the following equations:

$$Q_1(s_j, a_{1,j}) \leftarrow (1-\mu_1)Q_1(s_j, a_{1,j}) + \mu_1\left[R_j + \gamma_1 \underset{a_1}{max}\, Q_1(s_{j+1}, a_1)\right] \tag{26}$$

$$Q_2(s_j, a_{2,j}) \leftarrow (1-\mu_2)Q_2(s_j, a_{2,j}) + \mu_2\left[R_j + \gamma_2 \underset{a_2}{max}\, Q_2(s_{j+1}, a_2)\right] \tag{27}$$

where μ_1, μ_2 are the learning rates, γ_1, γ_2 are the discount factors for the two actions of the hierarchical Q-learning. After the Q table update, the experience evaluation criteria is updated and another episode is conducted for a new round of action exploration. The learning process stopped when the number of predefined episodes is reached.

3.2. Q-learning and rule-based baseline without ultracapacitor

For the Q-learning strategy, the entire process is the same as the

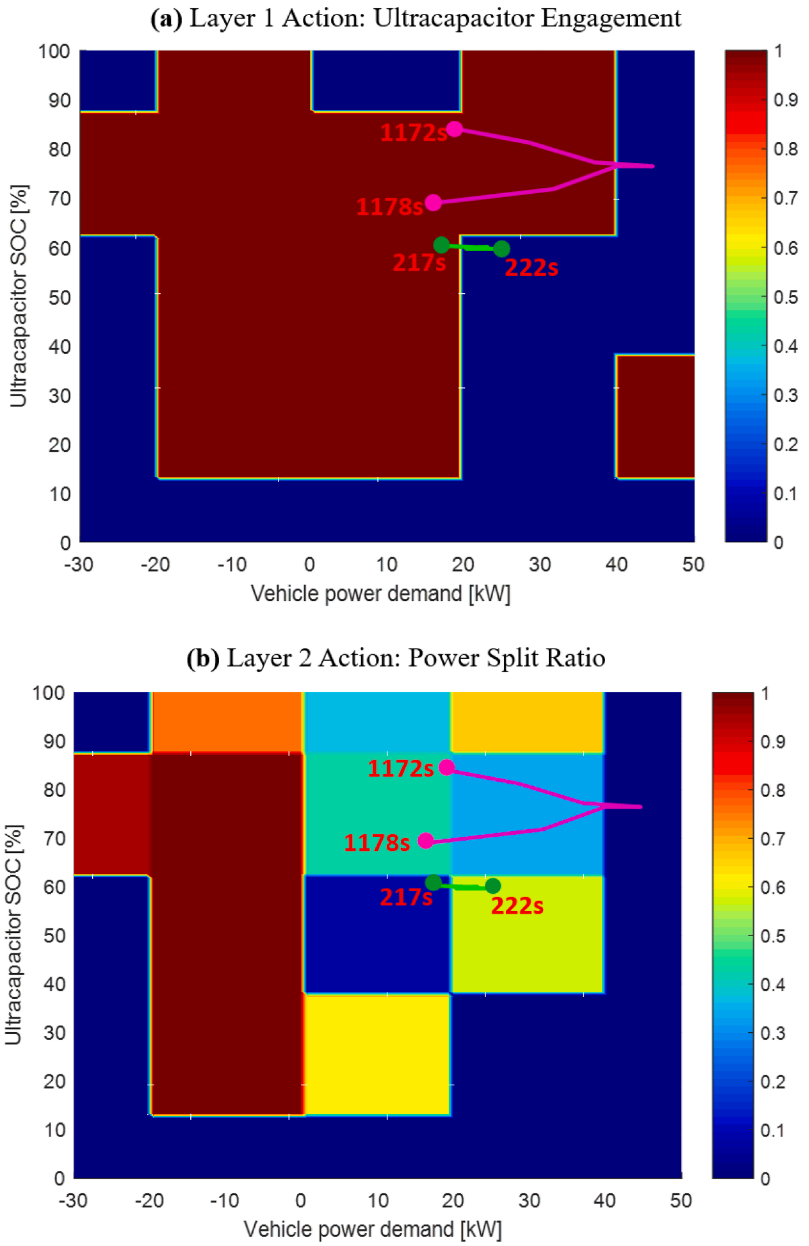

Fig. 12. Optimal policy map from the hierarchical Q-learning after the training: (a) ultracapacitor engage action from the layer 1, and (b) power split ratio between ultracapacitor and battery from the layer 2.

hierarchical Q-learning, except it only has one Q table and only makes one action decision (i.e., ultracapacitor power split ratio). The Q-learning strategy does not make decision on the high level ultracapacitor engage signal like the hierarchical Q-learning does. In this case, the ultracapacitor engage signal is always on.

For the rule-based baseline strategy without ultracapacitor, no ultracapacitor is considered in the vehicle propulsion system. All the power to the electric motor is provided by the battery. This purpose of this strategy is to reveal the benefits of adding the ultracapacitor to the propulsion system.

4. Results analysis

Two driving cycles are considered in this study as shown in Fig. 6. Urban Dynamometer Driving Schedule (UDDS) driving cycle [30] is used in the proposed strategy training, while Worldwide Harmonized Light Vehicles Test Procedure (WLTP) driving cycle [31] is used in proposed strategy validation. The vehicle model and control strategies are built upon Matlab/Simulink environment.

4.1. Hierarchical Q-learning results

Over the 30,000 iterations, the highest sum of rewards is shown in Fig. 7, which presents the best result from the first iteration. 30,000 is selected based on the preliminary simulation results to ensure convergence. It takes around 5500 iterations for the sum of rewards to converge, and the sum of rewards does not increase significantly over the next 4500 iterations. In this study, the training process only occurs once and then the Q-learning parameters are fixed for aging and range simulation. To further improve the Q-learning performance, the training process can be re-activated after certain battery aging is detected, which makes Q-learning as an adaptive control. Details of adaptiveness analysis can be found in [15].

The key signals of the vehicle system obtained by the hierarchical Q-

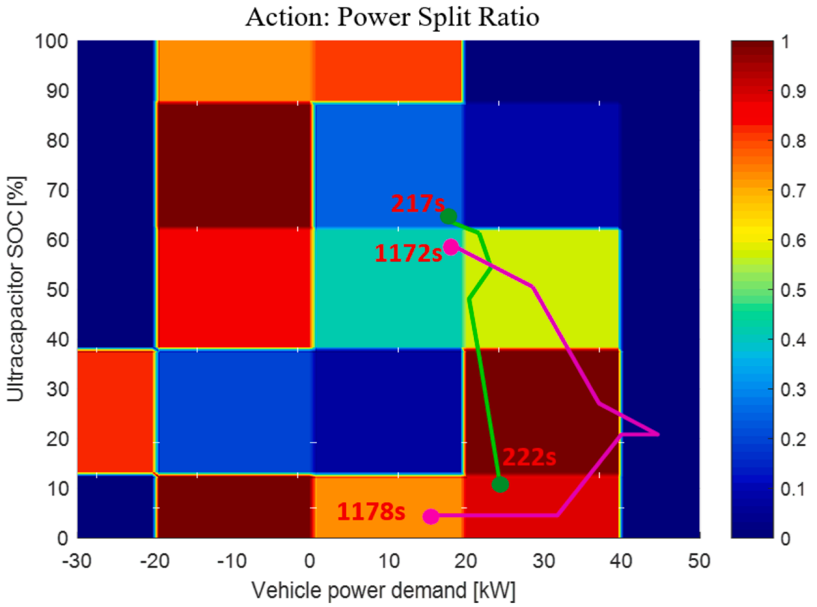

Fig. 13. Optimal policy map from the Q-learning after the training: power split ratio between ultracapacitor and battery.

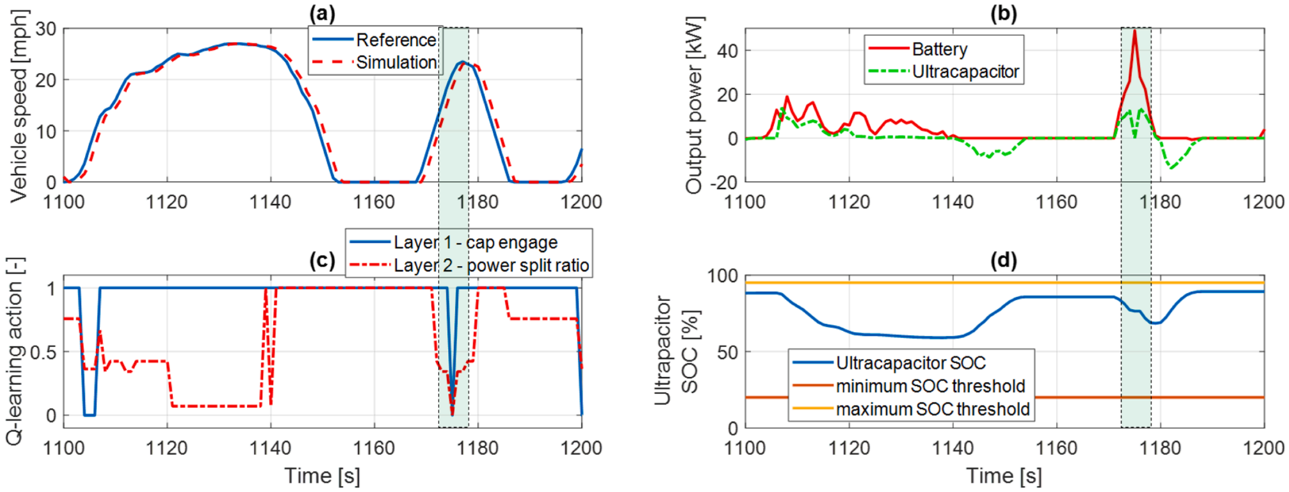

Fig. 14. Zoom-in of hierarchical Q-learning simulation results in Fig. 8:(a) Vehicle speed, (b) Output power of battery and ultracapacitor, (c) EM power, and (d) ultracapacitor SOC. (The highlighted green section is between 1172 s and 1178 s).

learning energy management strategy under the UDDS driving cycle are shown in Fig. 8, where Fig. 8(a) illustrates how the vehicle is controlled by the virtual driver to follow the velocity trajectory defined in UDDS cycle; Fig. 8(b) compares the power allocation between battery and ultracapacitor based on the control signals shown in Fig. 8(c); Fig. 8(d) shows the change of ultracapacitor SOC vs. time while providing the upper and lower SOC boundaries; and Fig. 8(e) and Fig. 8(f) show the power demand at electric power and the battery SOC, respectively. It indicated that the battery is mainly working in discharging conditions, while the ultracapacitor is working with similar charging and discharging power during the driving under UDDS cycle. That means the ultracapacitor takes charge of most regenerative braking and its SOC bounces back every time vehicle experiences a braking event. On the contrary, battery SOC keeps dropping due to lack of regenerative braking by the battery. Based on the control outputs shown in Fig. 8(c), the ultracapacitor is engaged at most of time and the power split ratio varies significantly across the entire driving cycle.

For the same vehicle driving under UDDS cycle, the vehicle performances obtained by single-layer Q-learning are shown in Fig. 9. One noticeable difference between the results obtained by hierarchical Q-learning and single-layer Q-learning can be found in the trajectories of ultracapacitor SOC. For the vehicle controlled by single-layer Q learning, the ultracapacitor SOC drops below the minimum threshold for several times, while it maintains above 50% in hierarchical Q-learning simulation. In addition, the power split ratio curve is not the same between the two strategies, which leads to the different ultracapacitor usage. The ultracapacitor power output in the positive region in the single-layer Q-learning is smaller than that in the hierarchical Q-learning. Zoom-in windows will be utilized later to analyze the SOC and power output difference between the single-layer and the hierarchical Q-learning.

For better visualization, the zoom-in plots for the vehicle performance from 150 s to 250 s are shown in Figs. 10 and 11, respectively. During this period, vehicle accelerates between 40 mph and 60 mph as shown in Fig. 10(a). A shorter time window is highlighted with green background between 217 s and 222 s. During this 5 s, ultracapacitor power is close to zero in Fig. 10(b) due to the engagement signal is zero as shown in Fig. 10(c). On the contrary, the ultracapacitor in single-layer

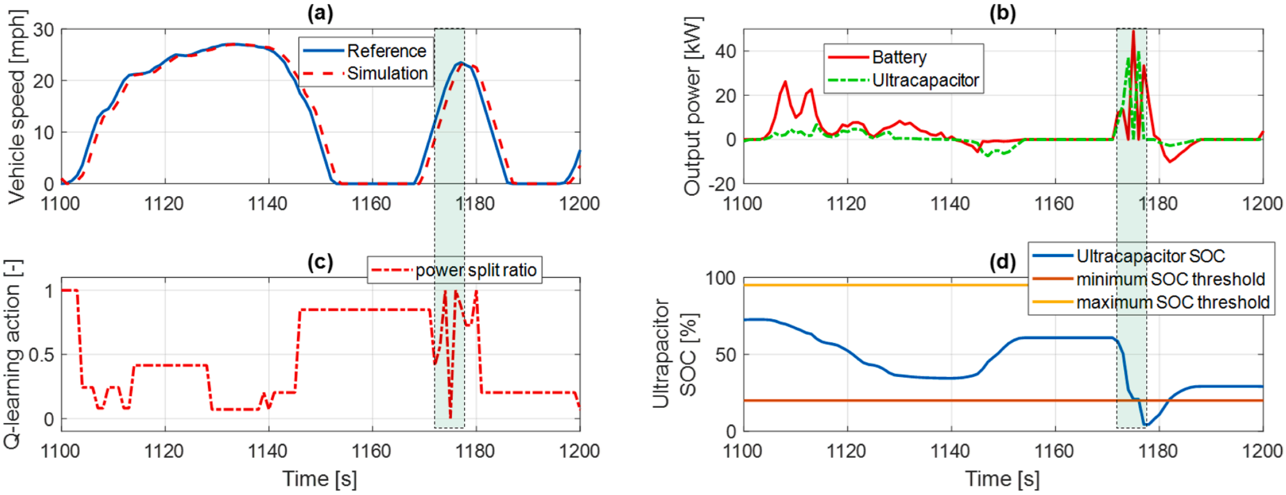

Fig. 15. Zoom-in of Q-learning simulation results in Fig. 9:(a) Vehicle speed, (b) Output power of battery and ultracapacitor, (c) EM power, and (d) ultracapacitor SOC. (The highlighted green section is between 1172 s and 1178 s).

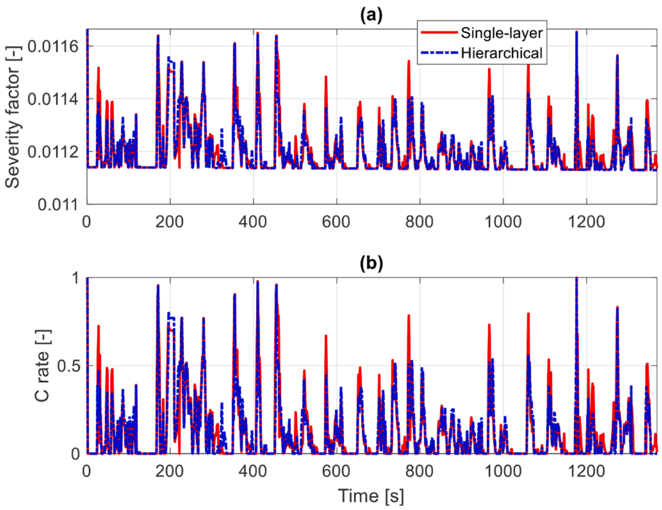

Fig. 16. Battery severity factor and C rate comparison of single-layer Q-learning and hierarchical Q-learning.

Table 3
Range, 500 driving cycle capacity loss and Ah-throughput comparison among three methods.

		Rule-based without ultracapacitor	Q-learning	Hierarchical Q-learning
UDDS	Range [miles]	251.78	253.30	255.35
	500 cycles capacity loss [%]	0.41	0.36	0.33
	Ah-throughput [Ah]	84.34	67.17	59.48

Q-learning keeps outputting power until ultracapacitor SOC drops to minimum threshold as shown in Fig. 11(b). The main difference is the ultracapacitor disengagement by the hierarchical Q-learning during the 5 s period, which avoids the drop SOC. Besides the SOC difference, the ultracapacitor output power in the hierarchical Q-learning is greater than that in the single-layer Q-learning in the scenarios of 175 s, 180 s and 190 s in Fig. 10(b) and Fig. 11(b). The greater ultracapacitor output power is the result of larger power split ratio at those three scenarios as shown in Figs. 10(c) and 11(c).

For the hierarchical Q-learning, the two-layer actions between 217 s and 222 s are shown in the optimal policy maps in Fig. 12. The optimal policy maps are generated by choosing the action corresponding to the maximum Q value in each state and the color of each region is mapped to the value in the color bar. As shown in Fig. 12(a), the ultracapacitor engage drops from 1 to 0 during the 217 s and 222 s. The Fig. 12(b) shows the trajectory of power split ratio from 0.2 to 0.6 in the same period. These two subplots explain the two actions shown in Fig. 10(c). The optimal policy map from Q-learning is shown in Fig. 13. It shows the detail trajectory of the power split ratio increases from 0.2 to 0.9 during 217 s and 222 s.

Another zoom-in window of Figs. 8 and 9 are shown in Figs. 14 and 15, respectively. The vehicle has three stops within the 100 s window. A 5 s section (1172 s - 1177 s) is highlighted with green background. The vehicle accelerates from 12 mph to 22 mph. The ultracapacitor SOC drops nearly 50% in the Q-learning strategy simulation (Fig. 15(d)), while it only drops 20% in the hierarchical Q-learning strategy simulation (Fig. 14(d)). The disengagement of ultracapacitor at 1175s help reduce the SOC drop in the hierarchical Q-learning. Again, the ultracapacitor power output in the hierarchical Q-learning at around 1150s is greater than that in the single Q-learning. Even though the ultracapacitor power output in the hierarchical Q-learning in the highlighted region (1175s) is smaller than that in the single-layer Q-learning, this occurs much less frequently as shown in Figs. 8(b) and 9(b).

The difference of the ultracapacitor output power leads to different battery aging effect of the two Q-learning, which can be explained by Fig. 16. The severity factor of the single-layer Q-learning is larger than that of the hierarchical Q-learning at most of the peaks. The larger severity factor leads to faster aging and larger battery capacity loss. The severity factor is defined in Eq. (10) and it mainly determined by battery C rate in this study as shown in Fig. 16(b). The trend of the two subplots in Fig. 16 is similar. Battery C rate is proportional to battery current, which is proportional to battery power output at a given battery voltage. Based on earlier analysis, ultracapacitor power output in the hierarchical Q-learning in most peaks are greater than that in the single-layer Q-learning, which means the battery power output of the two Q-learning have the reverse effect. Therefore, the battery C rate difference between two Q-learning is the result of the different ultracapacitor power output or battery power output.

4.2. *Comparison of three energy management strategies*

The vehicle single charge range and battery capacity loss generated from vehicle simulation using three strategies are summarized in

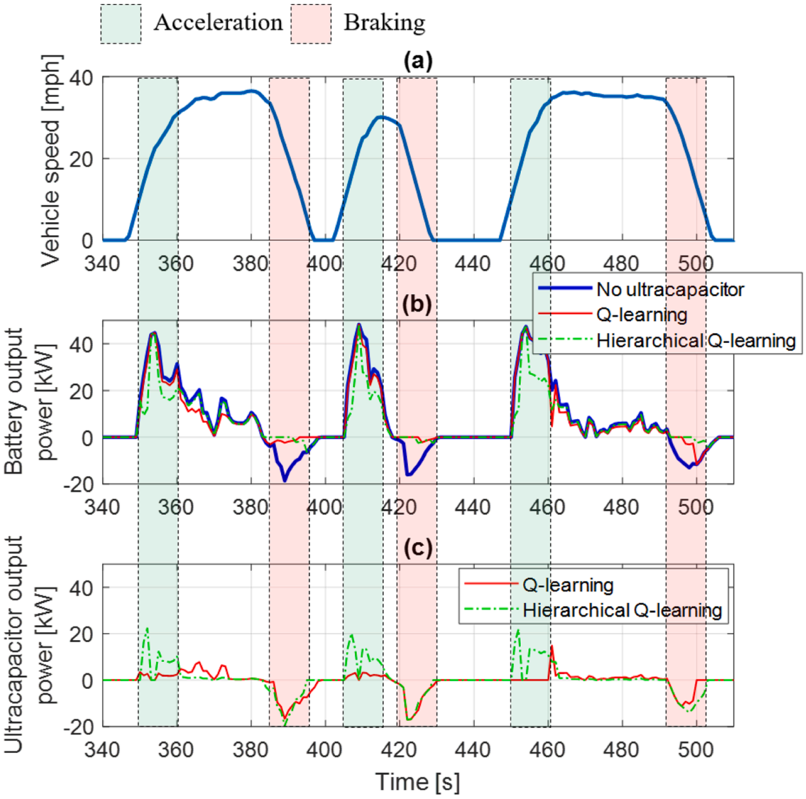

Fig. 17. Comparison of three energy management strategies using UDDS simulation results: (a) vehicle speed, (b) battery output power, and (c) ultracapacitor output power.

Table 4
Energy management strategies validation at WLTP driving cycle.

	No ultracapacitor	Q-learning	Hierarchical Q-learning
Range [miles]	216.53	216.92	218.51
500 cycles capacity loss [%]	0.61	0.56	0.53
1200 cycles capacity loss [%]	1.32	1.21	1.14

Table 5
Hierarchical Q-learning strategy validation considering velocity measurement noise.

	UDDS		WLTP	
	w/o noise	w/ noise	w/o noise	w/ noise
Range [miles]	255.35	255.30	218.51	217.61
500 cycles capacity loss [%]	0.33	0.34	0.53	0.55
1200 cycles capacity loss [%]	0.71	0.73	1.14	1.19

Table 3. From the range results, it is observed that the addition of ultracapacitor slightly increases the range. Using Q-learning and hierarchical Q-learning with the ultracapacitor, the range increases by 1.5 mile and 3.5 mile, respectively. More importantly, the addition of ultracapacitor substantially reduces the battery capacity loss. Ultracapacitor plus the single-layer Q-learning, the battery capacity loss drops from 0.41% to 0.36% over the 500 UDDS driving cycles, which is 12% reduction. The hierarchical Q-learning strategy reduces the capacity loss from 0.41% to only 0.33%, which is 20% reduction.

Compared with the vehicle equipped only battery, the vehicle equipped with battery and ultracapacitor reduces the overall Ah-throughput of the battery and reduces the peak power during acceleration and braking events (Fig. 17). The battery Ah-throughput from two reinforcement learning strategies is close to each other, but far below the battery Ah-throughput from the vehicle simulation without ultracapacitor. The difference in battery capacity loss among the three energy management strategies are similar to the difference in battery Ah-throughput. In addition to the Ah-throughput, the battery peak power also indicates the battery capacity loss difference. At a fixed voltage, the battery output power is proportional to the battery current. Large current leads to high severity factor and fast battery degradation. Thus, reduction of large battery output power slows down battery degradation. In Fig. 17(b), when the vehicle accelerates, the battery power curve from hierarchical Q-learning stays in the high-power areas shorter than the other two strategies. When the vehicle brakes, battery recovers the energy when ultracapacitor is not equipped, otherwise ultracapacitor recovers most of the energy.

4.3. Validation

In this subsection, the hierarchical Q-learning strategy is validated in two aspects: i) different driving cycle, ii) measurement noise. First, WLTP driving cycle is used in the simulation of each strategy. For the Q-learning strategy and hierarchical Q-learning strategy, the Q tables are trained in the UDDS driving cycle. Then, the optimal action are taken based on the trained Q tables in the validation simulation. The results of the validation are summarized in Table 4. Similar to the results in UDDS driving cycle, Hierarchical Q-learning shows substantial battery capacity loss reduction compared to the other two strategies in WLTP driving cycle. Additionally, Hierarchical Q-learning strategy slightly leads the range with the 218.51 miles. Therefore, the observation from WLTP driving cycle results is then consistent with the observation from UDDS driving cycle results.

Uncertainty usually exists in signal measurement, which is considered in the validation of the hierarchical Q-learning strategy. Vehicle velocity is one of the key parameters in the energy management. A noise with a normal distribution is added to the velocity signal, where the mean value μ is 0 and the standard deviation σ is 0.2. The validation results are summarized in Table 5. The results show that the consideration of measurement noise slightly affects the range and battery capacity loss. The range merely changes, while the 500 cycles capacity loss increases 0.01% and 0.02% for UDDS and WLTP driving cycles, respectively. Overall, the impacts of measurement noise are not significant.

5. Conclusion

A hierarchical Q-learning strategy is proposed for the supervisory control of a battery/ ultracapacitor electric vehicle. It considers a high layer ultracapacitor engage control and a low layer ultracapacitor power split ratio. The proposed strategy converges after 5000 iterations during the learning process. Compared to a baseline strategy without ultracapacitor, the proposed strategy reduces the battery capacity loss by 20% and increases the range by 1.5%. Compared to a single-layer Q-learning strategy, the proposed hierarchical Q-learning reduces the battery capacity loss by 12% and slightly increases the range. The proposed strategy is also validated in a different driving cycle. Compared with the baseline strategy without ultracapacitor and the single-layer Q-learning strategy, the proposed strategy reduces the battery capacity loss by 13% and 5%, respectively and maintains slightly longer range. Additionally, the proposed strategy is validated considering vehicle velocity measurement noise and the result is not significantly impacted by the addition of the measurement noise. Due to the model-free characteristics of the hierarchical Q-learning algorithm, the proposed strategy can be easily adapted for different hybrid power system applications.

In this study, some models are built with experimental data. However, the proposed strategy is not validated in experiments, which will be one of future research tasks. Additionally, the approximate Q-learning algorithm has the potential to improve the proposed strategy by considering more state variables and is worth exploration.

CRediT authorship contribution statement

Bin Xu: Conceptualization, Methodology, Software, Writing – original draft. **Quan Zhou:** Conceptualization, Methodology, Writing – review & editing. **Junzhe Shi:** Methodology, Investigation, Writing – review & editing. **Sixu Li:** Methodology, Software, Investigation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Qiu, G. Wang, New evaluation methodology of regenerative braking contribution to energy efficiency improvement of electric vehicles, Energy Convers. Manage. 119 (2016) 389–398.
[2] H. Dia, Rethinking urban mobility: unlocking the benefits of vehicle electrification. Decarbonising the Built Environment, Springer, 2019, pp. 83–98.
[3] P.M. Attia, A. Grover, N. Jin, K.A. Severson, T.M. Markov, Y.-H. Liao, M.H. Chen, B. Cheong, N. Perkins, Z. Yang, P.K. Herring, M. Aykol, S.J. Harris, R.D. Braatz, S. Ermon, W.C. Chueh, Closed-loop optimization of fast-charging protocols for batteries with machine learning, Nature 578 (7795) (2020) 397–402.
[4] P. Weldon, P. Morrissey, M. O`Mahony, Long-term cost of ownership comparative analysis between electric vehicles and internal combustion engine vehicles, Sustain. Cities Soc. 39 (2018) 578–591.
[5] C.J. Meinrenken, Z. Shou, X. Di, Using GPS-data to determine optimum electric vehicle ranges: a Michigan case study, Transp. Res. Part D 78 (2020), 102203.
[6] T. Chen, X.-P. Zhang, J. Wang, J. Li, C. Wu, M. Hu, H. Bian, A review on electric vehicle charging infrastructure development in the UK, J. Modern Power Syst. Clean Energy 8 (2) (2020) 193–205.
[7] A. Cordoba-Arenas, S. Onori, Y. Guezennec, G. Rizzoni, Capacity and power fade cycle-life model for plug-in hybrid electric vehicle lithium-ion battery cells containing blended spinel and layered-oxide positive electrodes, J. Power Sources 278 (2015) 473–483.
[8] J. Shen, S. Dusmez, A. Khaligh, Optimization of sizing and battery cycle life in battery/ultracapacitor hybrid energy storage systems for electric vehicle applications, IEEE Trans. Ind. Inf. 10 (4) (2014) 2112–2121.
[9] L. Zhang, X. Hu, Z. Wang, F. Sun, D.G. Dorrell, A review of supercapacitor modeling, estimation, and applications: a control/management perspective, Renew. Sustain. Energy Rev. 81 (2018) 1868–1878.
[10] Y. Wang, S.J. Moura, S.G. Advani, A.K. Prasad, Power management system for a fuel cell/battery hybrid vehicle incorporating fuel cell and battery degradation, Int. J. Hydrogen Energy 44 (16) (2019) 8479–8492.
[11] F. Odeim, J. Roes, A. Heinzel, Power management optimization of an experimental fuel cell/battery/supercapacitor hybrid system, Energies 8 (7) (2015) 6302–6327.
[12] A.M. Nassef, A. Fathy, H. Rezk, An effective energy management strategy based on mine-blast optimization technique applied to hybrid PEMFC/supercapacitor/batteries system, Energies 12 (19) (2019), 3796.
[13] Y. Wang, Z. Sun, Z. Chen, Energy management strategy for battery/supercapacitor/fuel cell hybrid source vehicles based on finite state machine, Appl. Energy 254 (2019), 113707.
[14] A. Castaings, W. Lhomme, R. Trigui, A. Bouscayrol, Comparison of energy management strategies of a battery/supercapacitors system for electric vehicle under real-time constraints, Appl. Energy 163 (2016) 190–200.
[15] B. Xu, X. Tang, X. Hu, X. Lin, H. Li, D. Rathod, Z. Wang, Q-learning-based supervisory control adaptability investigation for hybrid electric vehicles, IEEE Trans. Intell. Transp. Syst. (2021).
[16] R. Bellman, Dynamic programming, Science 153 (3731) (1966) 34–37.
[17] P. Pisu, G. Rizzoni, A comparative study of supervisory control strategies for hybrid electric vehicles, IEEE Trans. Control Syst. Technol. 15 (3) (2007). Art. no. 3.
[18] G. Paganelli, S. Delprat, T.M. Guerra, J. Rimaux, J.J. Santin, Equivalent consumption minimization strategy for parallel hybrid powertrains, in: Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No.02CH37367) 4, 2002, pp. 2076–2081. Mayvol. 4.
[19] C.J. Watkins, P. Dayan, Q-learning, Mach. Learn. 8 (3–4) (1992) 279–292.
[20] X. Han, H. He, J. Wu, J. Peng, Y. Li, Energy management based on reinforcement learning with double deep Q-learning for a hybrid electric tracked vehicle, Appl. Energy 254 (2019), 113708.
[21] B. Shuai, Q. Zhou, J. Li, Y. He, Z. Li, H. Williams, H. Xu, S. Shuai, Heuristic action execution for energy efficient charge-sustaining control of connected hybrid vehicles with model-free double Q-learning, Appl. Energy 267 (2020), 114900.
[22] H. Sun, Z. Fu, F. Tao, L. Zhu, P. Si, Data-driven reinforcement-learning-based hierarchical energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles, J Power Sources 455 (2020), 227964.
[23] J. Groot, “State-of-health estimation of li-ion batteries: cycle life test methods,” 2012.
[24] P. Spagnol, S. Onori, N. Madella, Y. Guezennec, J. Neal, Aging and characterization of li-ion batteries in a hev application for lifetime estimation, IFAC Proc. Vol. 43 (7) (2010) 186–191.
[25] H. Kong, J. Yan, H. Wang, L. Fan, Energy management strategy for electric vehicles based on deep Q-learning using Bayesian optimization, Neural. Comput. Appl. 32 (18) (2020) 14431–14445.
[26] Y. Li, J. Tao, K. Han, Rule and Q-learning based hybrid energy management for electric vehicle, in: 2019 Chinese Automation Congress (CAC), 2019, pp. 51–56.
[27] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, 2018.
[28] B. Xu, D. Rathod, D. Zhang, A. Yebi, X. Zhang, X. Li, Z. Filipi, Parametric study on reinforcement learning optimized energy management strategy for a hybrid electric vehicle, Appl. Energy 259 (2020), 114200.
[29] M. Wunder, M.L. Littman, M. Babes, Classes of multiagent q-learning dynamics with epsilon-greedy exploration, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 1167–1174.
[30] O. US EPA, Dynamometer Drive Schedules, US EPA, 2015. Sep. 16, https://www.epa.gov/vehicle-and-fuel-emissions-testing/dynamometer-drive-schedules (accessed Feb. 18, 2021).
[31] “What is WLTP: the Worldwide Harmonised Light Vehicle Test Procedure?,” WLTPfacts.eu. https://www.wltpfacts.eu/what-is-wltp-how-will-it-work/ (accessed Feb. 18, 2021).
[32] Quan Zhou, Dezong Zhao, Bin Shuai, Yanfei Li, Huw Williams, Hongming Xu, Knowledge implementation and transfer with an adaptive learning network for real-time power management of the plug-in hybrid vehicle, IEEE Transactions on Neural Networks and Learning Systems 32 (12) (2021) 5298–5308, https://doi.org/10.1109/TNNLS.2021.3093429. In this issue.
[33] Quan Zhou, Yanfei Li, Zhao Dezong, Ji Li, Huw Williams, Hongming Xu, Fuwu Yan, Transferable representation modelling for real-time energy management of the plug-in hybrid vehicle based on k-fold fuzzy learning and Gaussian process regression, Applied Energy 305 (2022), 117853, https://doi.org/10.1016/j.apenergy.2021.117853. In this issue.
[34] Quan Zhou, Ji Li, Bin Shuai, Huw Williams, Yinglong He, Ziyang Li, Hongming Xu, Fuwu Yan, Multi-step reinforcement learning for model-free predictive energy management of an electrified off-highway vehicle, Applied Energy 255 (2019), 113755, https://doi.org/10.1016/j.apenergy.2019.113755. In press.