International Journal of Precision Engineering and Manufacturing-Green Technology (2023) 10:245–267
https://doi.org/10.1007/s40684-022-00476-2

Online ISSN 2198-0810
Print ISSN 2288-6206

REVIEW

# Recent Progress in Learning Algorithms Applied in Energy Management of Hybrid Vehicles: A Comprehensive Review

Dezhou Xu[1,2] · Chunhua Zheng[1] · Yunduan Cui[1] · Shengxiang Fu[1,3] · Namwook Kim[4] · Suk Won Cha[5]

Received: 2 July 2022 / Revised: 22 August 2022 / Accepted: 27 September 2022 / Published online: 31 October 2022

## Abstract

Hybrid vehicles (HVs) that equip at least two different energy sources have been proven to be one of effective and promising solutions to mitigate the issues of energy crisis and environmental pollution. For HVs, one of the core supervisory control problems is the power distribution among multiple power sources, and for this problem, energy management strategies (EMSs) have been studied to save energy and extend the service life of HVs. In recent years, with the rapid development of artificial intelligence and computer technologies, learning algorithms have been gradually applied to the EMS field and shortly become a novel research hotspot. Although there are some brief reviews on the learning-based (LB) EMSs for HVs in recent years, a state-of-the-art and thorough review related to the applications of learning algorithms in HV EMSs still lacks. In this paper, learning algorithms applied in HV EMSs are categorized and reviewed in terms of the reinforcement learning algorithms and deep reinforcement learning algorithms. Apart from presenting the recent progress of learning algorithms applied in HV EMSs, advantages and disadvantages of different learning algorithms and LB EMSs are also discussed. Finally, a brief outlook related to the further applications of learning algorithms in HV EMSs, such as the integration towards autonomous driving and intelligent transportation system, is presented.

**Keywords** Hybrid vehicle · Energy management strategy · Reinforcement learning · Deep reinforcement learning · Recent progress

## Abbreviations

| | |
|---|---|
| AI | Artificial intelligence |
| AC | Actor-critic |
| A3C | Asynchronous advantage AC |
| AD | Autonomous driving |
| CV | Computer vision |
| DDPG | Deep deterministic policy gradient |
| DP | Dynamic programming |
| DRL | Deep reinforcement learning |
| DQN | Deep Q-network |
| DDQN | Double deep Q-network |
| ECMS | Equivalent consumption minimization strategy |
| EMS | Energy management strategy |
| EREV | Extended-range electric vehicle |
| FC | Fuel cell |
| FCHEV | Fuel cell hybrid electric vehicle |
| FCEV | Fuel cell electric vehicle |
| GA | Genetic algorithm |
| HESS | Hybrid energy storage system |
| HEV | Hybrid electric vehicle |
| HV | Hybrid vehicle |
| HETV | Hybrid electric tracked-vehicle |
| HTV | Hybrid tracked vehicle |
| HEB | Hybrid electric bus |
| HIL | Hardware-in-loop |
| ICE | Internal combustion engine |

Dezhou Xu and Chunhua Zheng contributed equally to this work.

This paper is an invited paper (Invited Review).

✉ Suk Won Cha
swcha@snu.ac.kr

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

[2] School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China

[3] School of Mechanical Engineering, Shenyang University of Technology, Shenyang 110870, China

[4] Department of Mechanical Engineering, Hanyang University, 55 Hanyangdeahak-ro, Sangnok-gu, Ansan, Gyeonggi-do 15588, South Korea

[5] School of Mechanical and Aerospace Engineering, Seoul National University, San 56-1, Daehak-dong, Gwanak-gu, Seoul 151742, South Korea

| | |
|---|---|
| ITS | Intelligent transportation system |
| LB | Learning-based |
| LSTM | Long short-term memory |
| MPC | Model predictive control |
| MDP | Markov decision process |
| MC | Markov chain |
| NN | Neural network |
| NDP | Neuro-dynamic programming |
| OB | Optimization-based |
| PHEV | Plug-in hybrid electric vehicle |
| PHEB | Plug-in hybrid electric bus |
| PMP | Pontryagin's minimum principle |
| PER | Prioritized experience replay |
| PPO | Proximal policy optimization |
| RB | Rule-based |
| RL | Reinforcement learning |
| SC | Supercapacitor |
| SDP | Stochastic dynamic programming |
| SOC | State of charge |
| SARSA | State-action-reward-state-action |
| SAC | Soft actor-critic |
| SPaT | Signal phase and timing |
| TD | Temporal difference |
| TD3 | Twin delayed DDPG |
| TPM | Transition probability matrix |
| V2V | Vehicle to vehicle |
| V2I | Vehicle to infrastructure |

## 1 Introduction

The energy crisis problem has become increasingly serious worldwide, and according to related statistics, the petroleum source will be used up within 50 years if there are no effective solutions [43, 141, 143, 145, 146]. Traditional vehicles driven by internal combustion engines (ICEs) consume more than 30% of the total energy, especially petroleum sources, and exhaust 25–30% of the total greenhouse gas every year in the world [97]. The governments around the world plan to reduce the number of traditional ICE vehicles gradually for the energy conservation and emission reduction. Under the background of calling for the sustainable development, hybrid vehicles (HVs) are regarded as one of the effective ways to cope with the energy crisis and environmental pollution problems, which are able to realize both higher fuel economy and lower pollutant emissions. In this paper, HVs refer to hybrid electric vehicles (HEVs), plug-in HEVs (PHEVs), extended-range EVs (EREVs), fuel cell HEVs (FCHEVs), and electric vehicles (EVs) with hybrid energy storage systems (HESSs). The common power sources of HVs include battery packs, ICEs, fuel cells (FCs), and supercapacitors (SCs) [21, 35, 97].

One of the key issues in HVs is to design optimal energy management strategies (EMSs) to efficiently allocate the demand power of vehicles among different power sources, which has a significant influence on the fuel economy and the lifetime of power sources and further on the popularization of HVs [110]. EMSs have been studied by many universities and research institutions for more than 20 years, in which the energy management problems can be regarded as nonlinear and time-varying optimal control problems. Figure 1 shows the schematic diagram of the energy management problem of HVs. In general, EMSs can be classified into rule-based (RB) EMSs, optimization-based (OB) EMSs, and learning-based (LB) EMSs. In the early years, RB and OB EMSs were commonly adopted, among which RB EMSs present the excellent real-time performance but require a lot of reliable human expertise and engineering experience for better control effects, while OB EMSs obtain outstanding optimization effects but demand the precise future

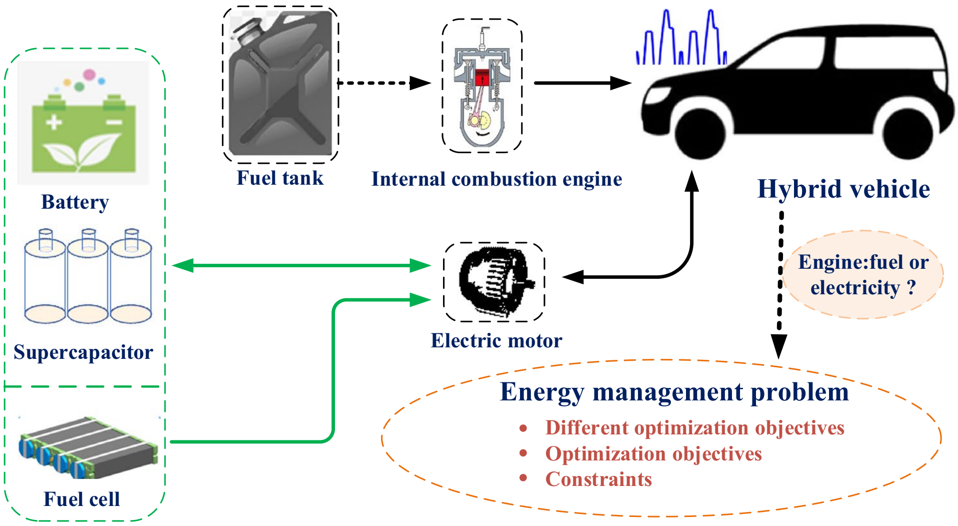

Fuel tank
Internal combustion engine
Hybrid vehicle
Battery
Supercapacitor
Fuel cell
Electric motor
Engine:fuel or electricity ?
Energy management problem
Different optimization objectives
Optimization objectives
Constraints

**Fig. 1** Schematic diagram of energy management problem of HVs

driving-related information in advance which hinders the real-time implementations [73, 74]. With the rapid development of computer technologies and artificial intelligence (AI), learning algorithms have been gradually applied to the research of HV EMSs in recent years. Besides acquiring satisfactory optimization effects, LB EMSs also have huge potential for real-time implementations, which can be taken as a trade-off between RB EMSs and OB EMSs. However, there are some intractable problems to be solved when developing LB EMSs, such as the requirement on plenty of driving-related training datasets, the tedious training time, the instability of training process, and the difficulty in objective-function settings.

In this paper, learning algorithms are categorized into reinforcement learning (RL) algorithms and deep reinforcement learning (DRL) algorithms according to whether deep neural networks (NNs) are employed to approximate some important factors. For RL algorithms, different types including the Q-learning, temporal-difference (TD) learning, and Dyna-style have been gradually applied in the EMS field since 2014. Especially, the Q-learning is the most commonly utilized algorithm among the RL algorithms. DRL algorithms can be further categorized into the deep Q-network (DQN), double DQN (DDQN), deep deterministic policy gradient (DDPG), twin delayed DDPG (TD3), and soft actor-critic (SAC), which are the mainstream algorithms in the current research on DRL-based EMSs. To the best of our knowledge, [59, 89] are the pioneers of the DRL algorithms' applications in HV EMSs [59, 89]. A detailed classification of learning algorithms applied in EMSs is described in Fig. 2.

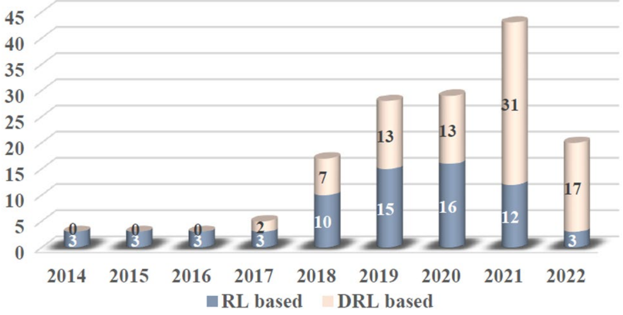


**Fig. 3** Statistics of the published papers related to LB EMSs from 2014 to 2022

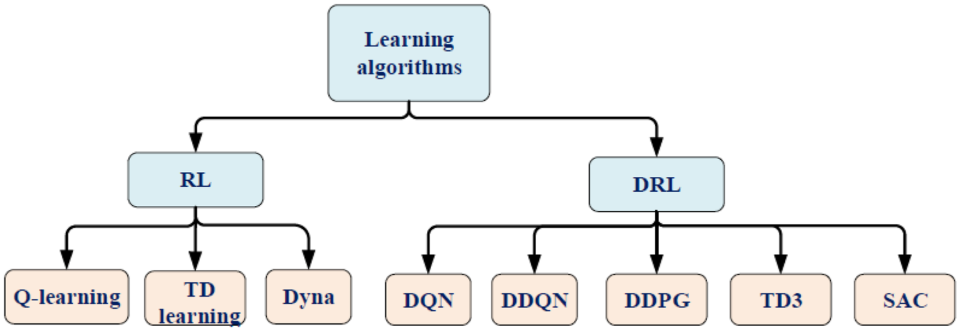
Figure 3 depicts the tendency of the number of published literature related to LB EMSs from 2014 to 2022 (until April 2022). It is evident that LB EMSs have become a research hotspot in the HV area with increasing attention in recent years. Although some articles already reviewed the RL & DRL-based EMSs [20, 41, 98], a more comprehensive and original review of the recent progress in learning algorithms applied in the HV EMS field still lacks. To fill in this gap, a thorough review is carried out meticulously in this paper targeting at the time since 2017. Additionally, the advantages and disadvantages of each LB EMS are discussed and an outlook on the future trends of LB EMSs is also presented, which includes integrating with emerging intelligent technologies, such as the autonomous driving (AD) and intelligent transportation system (ITS).

The remainder of this paper is organized as follows: applications of RL & DRL algorithms in HV EMSs are elucidated in Sects. 2 and 3, respectively; a summary of LB EMSs, a discussion on the advantages and disadvantages of LB EMSs, and an outlook for the future trends of LB EMSs are presented in Sect. 4; finally Sect. 5 concludes the whole paper.

## 2 RL in HV EMSs

RL algorithms solve optimal decision-making problems by self-learning without prior knowledge, which include three key factors: the agent, environment, and reward [102]. The goal of RL algorithms is to maximize the cumulative scalar reward through continuous interactions between the agent and the environment. The agent ultimately learns an optimal control strategy through a continuous trial and error search process [41, 70–74]. Additionally, the Markov property is a distinct characteristic of RL algorithms, where the changes of future system states are only relevant to the current system states. In that case, the decision process of RL algorithms

**Fig. 2** A classification of the learning algorithms applied in EMSs



is named the Markov decision process (MDP) [102]. The MDP can be described by a tuple$\{S, A, P, R\}$, where $S$ and $A$ represent the state space and the action space, respectively, $P : S \times A \times S \to [0, 1]$ represents the transition probability among all states, and $R : S \times A \to \mathbb{R}$ represents the reward. The primary objective of RL is to learn an optimal policy $\pi$ that maps state $S$ to optimal action$A$, with maximum accumulate reward$R = \sum_{t=0}^{T} \gamma^t \bullet r(t)$, where $\gamma \in$ [1] denotes the discount factor [102].

In RL algorithms, the Q-value is utilized to measure and evaluate the sum of long-term rewards under the executive action, i.e., the larger Q-value means the corresponding action is more likely to be adopted. Q-value updates based on the Bellman equation, which is shown as follows [102]:

The state value $V(s)$ is defined as follows and it can be regarded as the discount sum of numerical rewards:

$$V(s) = \mathbb{E}\left(\sum_{t=0}^{T} \tau^t \cdot r_{t+1}\right) \tag{1}$$

where $r_{t+1}$ represents the rewards at the time-step of $t + 1$. When the value of state $s$ is optimal, the value function $V(s)$ can be rewritten as Eq. (2):

$$V^*(s) = \min_{a \in A} \sum_{st+1 \in S} p^a_{s \to s_{t+1}} \left( r^a_{s \to s_{t+1}} + \gamma \cdot V^*\left(s_{t+1}\right) \right) \tag{2}$$

where $p^a_{s \to s_{t+1}}$ denotes the probability of shifting from state $s$ to state $s_{t+1}$ and $r^a_{s \to s_{t+1}}$ denotes the reward of making a transition from state $s$ to state $s_{t+1}$.

A step further, the optimal policy is determined according to Eq. (3):

$$\pi^* = \mathrm{argmin} \sum_{st+1 \in S} p^a_{s \to s_{t+1}} \left( r^a_{s \to s_{t+1}} + \gamma \cdot V^*\left(s_{t+1}\right) \right) \tag{3}$$

The optimal Q-value called the optimal state-action value can be deduced by Eqs. (2) and (3) as follows:

$$Q'\left(s_t, a_t\right) \leftarrow Q\left(s_t, a_t\right) + \alpha \cdot \left[ r\left(s_t, a_t\right) + \gamma \cdot \max Q\left(s_{t+1}, a_{t+1}\right) - Q\left(s_t, a_t\right) \right] \tag{4}$$

where $\alpha$ is the learning rate; $Q'\left(s_t, a_t\right)$ is the Q-value to be updated at the next time-step;$Q\left(s_t, a_t\right)$ denotes the calculated Q-value under the current state $s_t$ and action $a_t$;$r\left(s_t, a_t\right)$ denotes the current reward under the current state $s_t$ and action $a_t$; $Q\left(s_{t+1}, a_{t+1}\right)$ denotes the estimated Q-value for the next state $s_{t+1}$ and next action $a_{t+1}$.

RL algorithms applied in the HV EMS field include the Q-learning, TD learning, policy iteration, state-action-reward-state-action (SARSA), and Dyna-style, etc. The characteristics of the main RL algorithms are summarized in Table 1. On the one hand, according to whether the agent needs to learn the state transition probability model of the environment, RL algorithms can be classified into the model-based and the model-free algorithms. Both the Q-learning and TD learning algorithms belong to the model-free algorithm. On the other hand, according to whether the policy that the agent uses for a given state is the same as the policy it updates, RL algorithms can be categorized into the on-policy and the off-policy algorithms, where the former is for the same case. On-policy algorithms include the SARSA, and off-policy algorithms include the Q-learning and TD learning. The fundamental difference between the SARSA and the Q-learning is that the SARSA estimates the action value $Q_\pi(s, a)$, while the Q-learning estimates the state value $Q(s, a)$. In other words, the SARSA selects the action according to the current policy, while the Q-learning is more greedy and thus chooses the action that corresponds to the maximum Q-value. Specifically, the Dyna-style is a framework of RL algorithms, which combines the model-based and the model-free algorithms, including the Dyna-H, Dyna-Q, Dyna-1, and Dyna-2. Different from the Q-learning, the Dyna operates by iteratively interacting with the environment, thus, its training time is longer than the Q-learning. Furthermore, the Q-learning is a representative RL algorithm, it is a quasi-optimal decision-making approach while it needs to conduct state-action space discretization, and as the action-value function of the Q-learning is linked to long-term returns instead of immediate cost, the agent selects an action with minimum action value at every state throughout every episode [3, 102].

**Table 1** Characteristics of the main RL algorithms

| RL algorithm | On/Off-policy | State-action space $(S, A)$ | Estimate value | Action space exploration | References |
|---|---|---|---|---|---|
| Q-learning | Off-policy | Discrete | State value | $\varepsilon -$ greedy | [8, 12, 68, 102, 103] |
| TD learning | Off-policy | Discrete | State value | $\varepsilon -$ greedy | |
| Dyna-style | On-policy and Off- policy | Discrete | State value | $\varepsilon -$ greedy | |
| SARSA | On-policy | Discrete | Action value | $\varepsilon -$ greedy | |

The common agent-environment interaction process of the RL algorithm is depicted in Fig. 4. At every time-step, the agent chooses an action $a_t$ randomly based on the current state $s_t$, and then the environment feedbacks the corresponding scalar reward $r_t$ to the agent according to $s_t$ and $a_t$. Then the state changes to $s_{t+1}$ at next time-step. Such a process continues until the training is finished. When the RL is applied to the HV EMS field, the particular vehicle model and the driving conditions can be the environment; the vehicle status-related parameters can be the states, such as the battery state of charge (SOC), vehicle velocity, vehicle power demand, and torque demand; the power split-related parameters can be the actions, such as the battery output power, the torque of the ICE or the motor. More importantly, the reward function can be set based on different optimization objectives, such as the minimization of the energy consumption and the lifespan extension of the main energy source, etc. Through the continuous interaction with the environment and the proper hyperparameter tuning, the agent finally learns the optimal EMS.

With the rapid development of computer technologies and AI, the RL can be truly implemented into practical applications. The energy-management problem can be regarded as an MDP, where the vehicle states at the next moment are only relevant to the current states and independent of the historical state. The RL-based EMSs can be deemed as an optimum map from the states to the actions. Therefore, RL algorithms have become increasingly popular in the HV EMS field because they not only can obtain near-optimal effects, but also are suitable for online applications [41]. Nevertheless, there are also a few defects in RL-based EMSs, such as the discrete states and the difficulty in continuous control, which both lead to the problem of the "curse of dimensionality".

### 2.1 Q-learning in HV EMSs

Specific applications of the Q-learning algorithm in HV EMSs are depicted in Fig. 5. In general, the application process of RL-based EMSs experiences two parts, i.e. the offline training and the online application. The offline training is executed on the computers to derive ideal EMSs through continuous trial and error processes, and after the training, the ideal EMSs are put into real controllers. Specially, the LB EMSs in the controllers can still retain learning abilities to adapt to new driving conditions by collecting new driving data during online applications and retraining the EMSs.

In HV EMSs, the Q-learning is usually utilized to obtain the optimal control strategy between batteries and ICEs [3, 10, 49, 49, 50, 50, 76, 126–130, 134, 141, 143, 145, 146]. Additionally, to evaluate the optimality of the Q-learning-based EMSs, OB EMSs, such as the dynamic programming (DP) and Pontryagin's minimum principle (PMP), are often served as the benchmarks. For example, Lee et al. applied a model-based Q-learning to learn the characteristics of a current given driving environment and adaptively changed the control policy through learning, the simulation results presented the Q-learning-based EMS possessed quasi-optimal effects compared with the DP-based EMSs [49, 50]. In addition, the Q-learning is also utilized to optimize the power split between the FC and the battery. Hsu et.al employed the Q-learning to optimize the dynamic energy management between the FC and the battery, and compared with a fuzzy logic-based EMS, their proposed EMS reduced fuel consumption around 6% and maintained the stability of the battery SOC effectively [38]. In addition, considering the impact of temperature and current on the battery aging, Sarvaiya et al. presented a research to analyze the battery life optimization effect with different EMSs, including the thermostat, fuzzy logic, ECMS, and Q-learning, and the comparison results showed that the best fuel economy was achieved by the Q-learning-based EMS [92]. Q-learning uses the maximum action value as an approximation for the

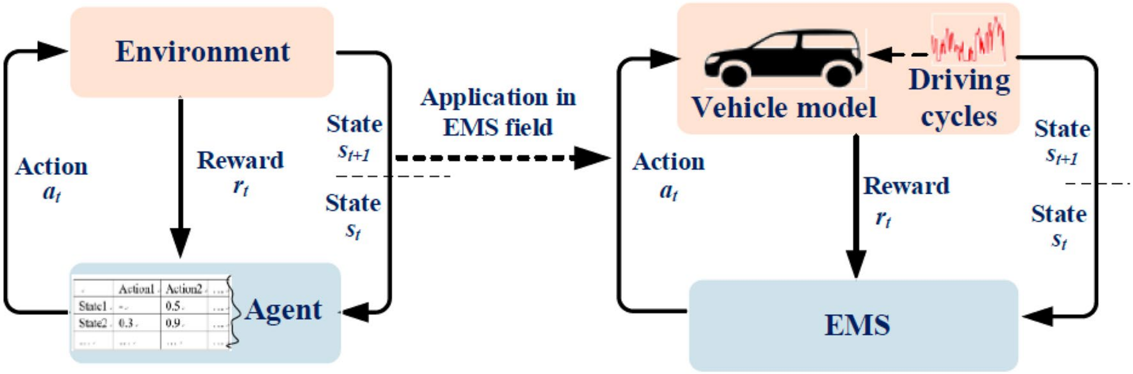


**Fig. 4** The principle and application of RL algorithms

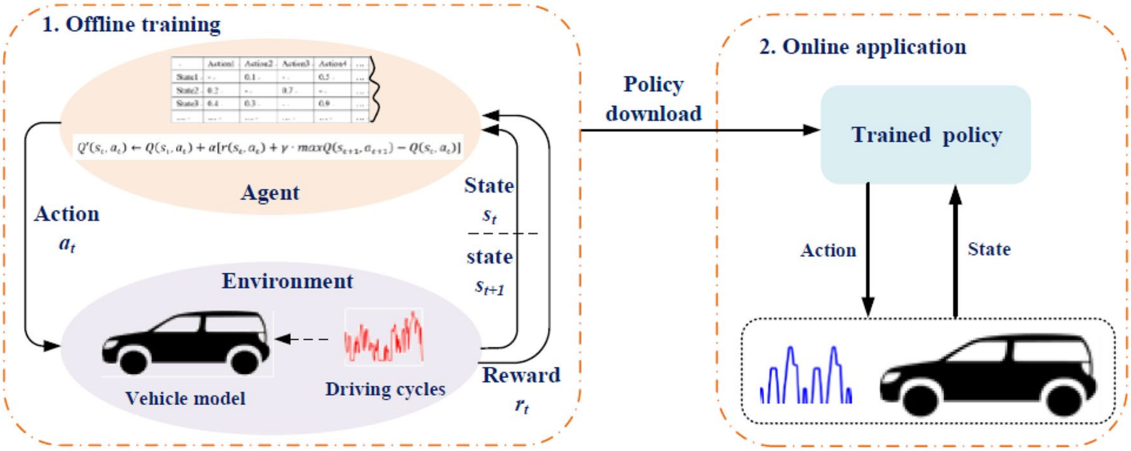


Fig. 5 Application process of Q-learning algorithm in HV EMSs

maximum expected action value. However, a positive bias would be introduced in the conventional Q-learning, in other words, the Q-value tends to be overestimated, which does harm to the optimality. To cope with this problem, the double Q-learning with two Q-functions was proposed by Hasselt in 2015 [32], and it has been applied to the HV EMS field to obtain more satisfied EMSs [9, 95, 117, 121]. For instance, aiming at reducing the overestimation of the merit-function values, Shuai et al. proposed two heuristic action execution policies based on the double Q-learning, i.e. the max-value-based policy and the random policy, to improve the vehicle energy efficiency and maintain the battery SOC. The simulation results demonstrated that the proposed EMS achieved at least 1.09% higher energy efficiency than the conventional double Q-learning [95].

In general, to improve the convergence rate, optimization effects, and online application/adaptability of the Q-learning-based EMSs further, lots of researchers integrated other optimization algorithms or some tricks to original Q-learning algorithms. The tricks for improving performances of the Q-learning -based EMSs are summarized in Table 2. The optimization effect is the primary focus, thus optimization algorithms and some tricks are integrated with the Q-learning, such as the RB [54, 56, 57], the model predictive control (MPC) [9, 24, 70–74, 82], the GA [107, 129, 130], the Neuro-DP (NDP) [76], and the ECMS [126, 128]. For instance, Li et al. proposed an EMS that combined the Q-learning with the RB for a lithium battery and ultracapacitor hybrid energy system. When the car was in a braking condition or the lithium battery and the ultracapacitor

Table 2 Tricks for improving performances of the Q-learning-based EMSs

| Performance | Tricks | Powertrain structure | References |
|---|---|---|---|
| Optimization effect | MPC | HEV, PHEV | [9, 24, 70–74, 82] |
| | GA | FCHEV,<br>EV with HESS | [107, 129, 130] |
| | RB | EV with HESS | [54, 56, 57] |
| | NDP | PHEV | [66] |
| | ECMS | HEV | [126, 128] |
| | LSTM | HTV | [70–74] |
| Convergence rate | MC &TPM | PHEB | [22, 23] |
| | Speedy Q-learning /fast Q-learning | HETV, FCHEV | [15, 52, 53] |
| | TL | HEV | [49, 50, 70–74] |
| Real-time performance/ adaptability | KL divergence | HEV,<br>EV with HESS | [5, 124, 124, 125] |
| | MC prediction | HEV | [11, 67, 75, 136, 137] |
| | NNs | Plug-in FCEV,<br>PHEV | [63, 64, 142, 144, 147] |
| | Cosine similarity and forgetting factor | PHEV | [63, 64] |

energy were not enough, the output power of the lithium battery and the ultracapacitor was directly allocated based on the rules. Otherwise, the rule and Q-learning strategy took over, and the lithium battery output power was obtained through the Q-table. Results showed that the energy loss of the hybrid energy system reduced from 9 to 7% by the proposed EMS compared with the RB EMS [54, 56, 57]. Liu et al. combined the MPC and Q-learning to improve the powertrain mobility and fuel economy for a group of automated vehicles, where the higher-level controller outputs the optimal vehicle velocity using the MPC technique while the lower-level controller decides the power split, and the simulation results showed that the vehicle traveling time reduced by 30% through reducing red-light idling and the proposed method increased fuel economy by 13% compared with different energy efficiency controls [70–74]. In addition, Bin Xu et al proposed an ensemble RL strategy that is the integration of the Q-learning, ECMS, and thermostatic methods to improve the fuel economy further, and the simulation result showed that the proposed EMS achieved a 3.2% fuel economy improvement compared with the Q-learning alone [126, 128]. Moreover, the long short-term memory (LSTM) was introduced to improve the Q-learning-based EMS for a hybrid tracked vehicle by Liu et al. It contained two levels, where the higher level included a parallel system that consists of a real powertrain system and an artificial system and a bidirectional LSTM network was used to train the synthesized data from the parallel system while the lower level utilized the trained data to determine the EMS with the Q-learning framework. The simulation results showed that the proposed EMS improved the energy efficiency significantly compared with the conventional RL and DRL approaches [70–74]. Particularly, Zhou et al. proposed a new model-free Q-learning-based EMS with the capability of 'multi-step' learning to enable the all-life-long online optimization, and the simulation results indicated that the proposed model-free EMS reduced the energy consumption by at least 7.8% for the same driving conditions compared with the model-based one [152]. Moreover, an EMS based on a hierarchical Q-learning network was proposed for the EV with the HESS by Xu et al. Two independent Q-tables, i.e. Q1 and Q2, were allocated in two control layers, where the lower layer was utilized to determine the power split ratio between the battery and the ultracapacitor based on the knowledge stored in Q1, while the upper layer was developed to trigger the engagement of the ultracapacitor based on Q2. The results indicated that the proposed EMS reduced the battery capacity loss by 8% and 20% compared with the single-layer Q-learning and the no ultracapacitor cases respectively, as well as extending the range [131–133].

To improve the convergence rate of the original Q-learning-based EMSs, the Markov Chain (MC) and transition probability matrix (TPM) are usually utilized to accelerate the training process [22, 23]. Additionally, the speedy Q-learning /fast Q-learning is also employed to accelerate the convergence rate [15, 52, 53]. Guo et al. proposed a Q-learning-PMP-based EMS for the plug-in hybrid electric bus (PHEB), where the control action was only updated at fixed time steps, and the reward was evaluated in the next 60 s, and the simulation results indicated that the training process of the RL-PMP was greatly accelerated [22, 23]. Liu et al. proposed an RL-based EMS that utilized a speedy Q-learning algorithm in the MC-based control policy computation [67, 75]. Interestingly, the transfer learning (TL) is also utilized to improve the convergence performance of Q-learning-based EMSs, [49, 50, 70–74]. For example, Liu et al. proposed a bi-level transfer RL-based adaptive EMS for an HEV, where how to transform the Q-tables in the RL framework via driving cycle transformation was solved in the upper level, while how to establish the corresponding EMS with the transferred Q-tables using the Q-learning was decided in the lower level, and comparison results indicated that the transferred RL-based EMS converged faster than the original RL-based EMS [70–74].

As for the better real-time performance/adaptability, some tricks such as the Kullback–Leibler (KL) divergence [5, 124, 125, 159], the MC prediction [11, 67, 75, 136, 137], and the NN [63, 64, 142, 144, 147] are integrated with the Q-learning algorithm. For instance, Cao et.al. utilized the KL divergence to determine when the TPM to be updated to make the proposed Q-learning-based EMS adapt to new driving conditions better, and the simulation results indicated that the proposed EMS can be employed in real-time [5]. A step further, they carried out a hardware-in-loop (HIL) simulation test to confirm the real-time control of the Q-learning-based EMS [125]. Chen et.al constructed a multi-step Markov velocity prediction model and applied it to the stochastic model predictive control after the accuracy validation, and the simulation results indicated that a single step calculation time of the Q-learning-based EMS controller was less than 57.15 microseconds, which proved it was real-time implementable [11]. Lin et.al proposed an online recursive EMS for FCHEVs based on the Q-learning, where the cosine similarity was combined with the forgetting factor to update both the TPM and the EMS, and the results showed the proposed strategy improved the fuel economy by 17.10%, 29.50%, and 38.20% at the trip distance of 150 km, 200 km, and 300 km respectively compared with the RB EMS, which indicated the strong adaptability of the proposed Q-learning-based EMS [63, 64]. Moreover, Liu et.al applied the fuzzy encoding and the nearest neighbor methods to realize the velocity prediction for the Q-learning-based predictive EMS, and the results in a HIL test indicated that the predictive controller could be put into the real-time application [69]. Zhang et al. put forward a bi-level EMS for PHEVs based on the Q-learning and MPC, in which the Q-learning was utilized to generate the SOC reference in the upper layer, while the MPC controller was designed to allocate the system power flows online with the short-period velocity predicted by the RBF NN in the lower layer. The results presented the ability of great robustness to tackle the inconsistent driving conditions, which meant the bi-level EMS could be implemented to the real-time application [142, 144, 147]. Besides integrating the Q-learning with a fuzzy logic controller to form an EMS, Wu et al. also employed the multi-time-scale prediction method to realize the future short-period driving cycle prediction and to achieve the real-time control [118, 122]. Sun et al. proposed an RL-based EMS for FCHEVs by utilizing the ECMS to tackle the high-dimensional state-action spaces and found a trade-off between the global learning and the real-time implementation [101]. Du et al. proposed a Q-learning-PMP algorithm to cope with the "curse of dimensionality" and realized the energy control of a PHEB effectively [22, 23]. Li et al. combined the Q-learning with deterministic rules for the real-time energy management between FCs and SCs [55].

### 2.2 Other RL Algorithms in HV EMSs

Other RL algorithms here refer to the Dyna-style [16, 68, 76, 82, 136, 137, 141, 143, 145, 146], the TD learning [8, 12, 18, 65, 90, 139], the SARSA [47], and the policy iteration [138], which are less implemented to the EMS field but achieve satisfactory optimization effects. For instance, the Dyna algorithm reached approximately the same fuel consumption as the DP-based global optimal solution, but the computational cost of which was substantially lower than that of the stochastic DP (SDP) [76]. Liu et al. also applied the Dyna agent to establish a heuristic planning energy management controller for the real-time fuel-saving optimization [68]. In addition, although the Dyna-H algorithm outperformed the Dyna in the convergence rate, it encountered terrible adaptability owing to the deficient training of the state-action pairs [16]. Furthermore, the TD learning was also applied to the EMS field. Chen et al. introduced a TD algorithm with historical driving cycle data to minimize the fuel consumption of a PHEB, which could achieve the real-time running without sacrificing the accuracy of the optimization compared with the RB EMS [8, 12]. The SARSA algorithm was also applied to design an EMS for an FCHEV [47]. Yin et al. applied a policy iteration algorithm to obtain the optimal control policy for a super-mild HEV [138]. Moreover, a model-based RL was utilized to optimize the fuel consumption for an FCEV by LEE et al., and the simulation results showed that the proposed EMS realized more fuel reduction than the RB EMS by 5.7% [48].

## 3 DRL in HV EMSs

Although RL algorithms have been widely applied to the HV EMS field and can obtain near-optimal control effects, there are still some deficiencies: on the one hand, RL algorithms have difficulty in dealing with high dimensional state-action spaces, named the "curse of dimensionality"; on the other hand, RL algorithms must be learned from a scalar reward signal that is frequently sparse, noisy, and delayed. Specially, the delay between actions and corresponding rewards can be thousands of time-steps long. To cope with the above problems, Mnih et al. tried to integrate the deep learning with the RL and then proposed the DQN algorithm [81]. A deep NN was utilized to approximate the Q-function in DRL algorithms instead of a tabular Q-table of Q-learning algorithms, hence the states are fed into the NN which outputs the corresponding actions subsequently. With the help of deep NNs, DRL algorithms can cope with the higher dimensional state and action spaces in the actual decision-making processes [108, 109]. In recent years, DRL algorithms have been successfully applied to a great many complex problems and proven that they even outperform human beings in some humanoid tasks (i.e., Playing chess and computer games) [88]. Owing to the conspicuous advantages that DRL algorithms have, the DRL-based EMSs have shortly become an active research hotspot after 2017.

DRL algorithms applied in the HV EMS field include the DQN and AC-based algorithms (i.e. DDPG, TD3, SAC), which have been widely applied to the EMS field so far. However, some deficiencies still exist in the DRL algorithms. The characteristics of the main DRL algorithms are summarized in Table 3. In the DQN, since the estimated Q-value and the target Q-value are calculated using the same NN, it is easy to overestimate the Q-value and thereby affect the optimality. Moreover, it cannot achieve the continuous control, which may cause control errors. Therefore, the DQN is not the optimal solution to handle energy-management problems. Given that, actor-critic (AC)-based DRL algorithms are proposed to figure out the above problems, but they still suffer from some demerits, such as the instability of training, the poor convergence, the sampling inefficiency, and the hyperparameter sensitivity [17, 27, 28, 36].

### 3.1 DQN in HV EMSs

DQN algorithms here refer to the DQN algorithm, the DDQN algorithm, and the dueling DQN algorithm, which are extensively applied to the HV EMS field. The DQN algorithm is one of the representative algorithms in the

**Table 3** Characteristics of the main DRL algorithms

| DRL algorithm | On/Off-policy | State-action space ($S$, $A$) | Neural network | Action space exploration | References |
|---|---|---|---|---|---|
| DQN | Off-policy | Continuous $\boldsymbol{S}$<br>Discrete $\boldsymbol{A}$ | Q-network | $\varepsilon$ − greedy | [17, 31, 84, 102] |
| DDPG | Off-policy | Continuous **(*S*, *A*)** | AC with single Q-network | Random noise $N$ | |
| TD3 | Off-policy | Continuous **(*S*, *A*)** | AC with two Q-networks | Random noise $N$ | |
| SAC | Off-policy | Continuous **(*S*, *A*)** | AC with two Q-networks | Maximum entropy | |

DRL, the update of which is based on the TD, i.e. the Q-network updates on every single step instead of every single episode, which improves the learning efficiency significantly. In the DQN, the trade-off between the exploration and exploitation needs to be properly addressed, and the $\varepsilon$ − greedy algorithm is widely applied to avoid the over exploration and over exploitation, where $\varepsilon$ denotes the degree of the exploration while $1-\varepsilon$ represents the degree of the exploitation. Similar to the Q-learning algorithm, the DQN also pursues the maximum cumulative reward, and the Q-value is calculated based on the Bellman equation as well, which is shown in Eq. (5). Specially, in the DQN, the main Q-network parameterized by $\theta$ is utilized to output actions, and the target Q-network parameterized by $\theta'$ is introduced to increase the training stability.

$$Q(s_t, a_t) = \mathrm{E}\left[r_{t+1} + \gamma \cdot \max_{a_{t+1}} Q(s_{t+1}, a_{t+1};\theta')\right] \tag{5}$$

In addition, the parameter $\theta$ is updated through the backpropagation and the gradient descent method. The network loss function is shown in Eq. (6) and the gradient loss function is calculated by Eq. (7), which is fed into the main Q-network during the training process.

$$L(\theta) = E\left[\left(r_t + \gamma \cdot \max_{a_{t+1}} Q(s_{t+1}, a_{t+1};\theta') - Q(s_t, a_t;\theta)\right)\right]^2 \tag{6}$$

$$\begin{aligned}&\nabla_\theta L(\theta)\\&= E\left[\left(r_t + \gamma \cdot \max_{a_{t+1}} Q'(s_{t+1}, a_{t+1};\theta') - Q(s_t, a_t;\theta)\right)\nabla_\theta Q(s_t, a_t;\theta)\right]\end{aligned} \tag{7}$$

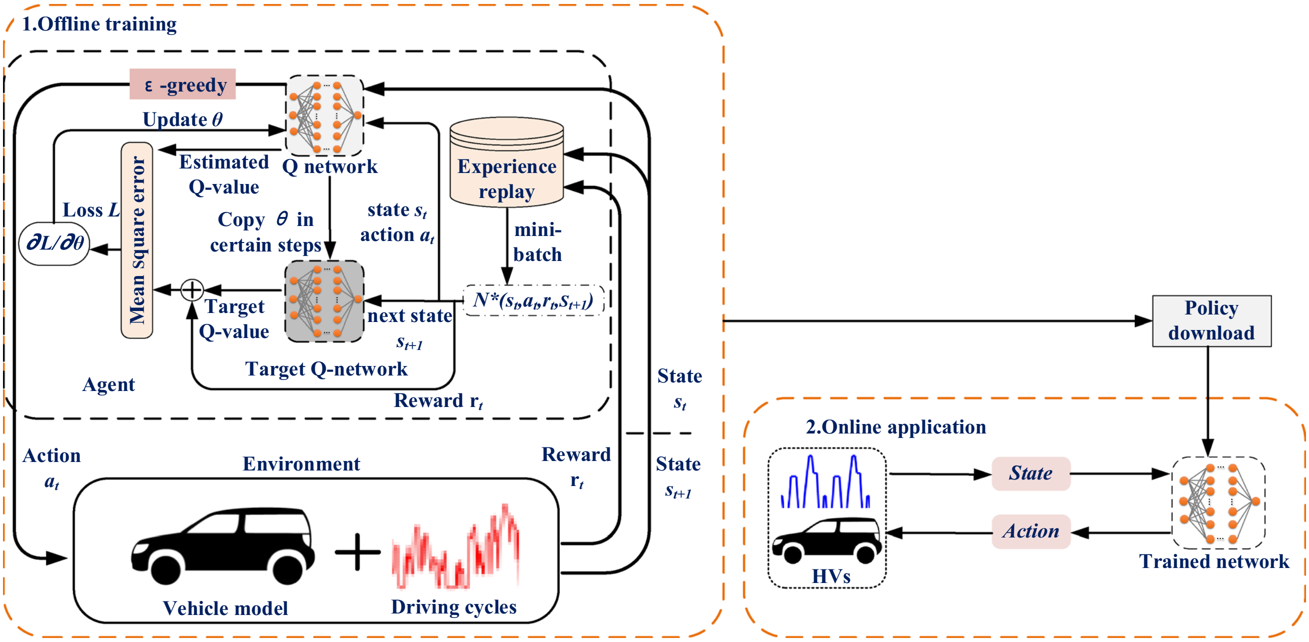


**Fig. 6** Applications of DQN algorithm in HV EMSs

Figure 6 describes the principle of the DQN algorithm and its application process. For the offline training part, at every time step, the agent outputs an action to the environment to execute based on the current policy; the environment turns to the state $s_{t+1}$ and sends the corresponding reward to the agent; the learning sample $(s_t, a_t, r_t, s_{t+1})$ is stored in the experience pool. The experience replay is introduced to the off-policy DQN algorithm to break the correlation of the training samples, and during the learning process, the agent selects a mini-batch experience samples from the experience pool to improve the learning efficiency. The state $s_t$ and the action $a_t$ feed into the main Q-network to calculate an estimated Q-value, while the state $s_{t+1}$ feeds into the target Q-network to create the target Q-value. The network loss is calculated by the reward $r_t$, the target Q-value, and the estimated Q-value according to Eq. (6), and the gradient loss $\partial L/\partial \theta$ is fed into the main Q-network to update the network parameter $\theta$. The target Q-network parameter θ' is updated every certain steps by copying parameters from the main Q-network, which increases the training stability. When the DQN algorithm is applied to the EMS field, it is also under the offline training first until the optimal polity is formulated. As for the policy download, it is similar to the Q-learning-based EMSs. For the online application part, the trained model simply maps the actual states to the actions and thus the calculation time is reduced greatly.

In recent years, the DQN has been widely applied to deal with the HV energy-management problems [2, 7, 8, 12, 59, 87–89, 98, 99]. For example, Chaoui et al. proposed a DQN-based EMS to derive optimal strategies to balance the SOC of all batteries, extend the battery lifespan, and reduce the battery frequent maintenance [7]. Some researchers also adopted the DQN to optimize the fuel economy of HEVs. Song et al. proposed a DQN-based EMS for PHEVs, and the results showed that the DQN-based EMS narrowed the fuel consumption gap to the DP benchmark to 6% [99]. Moreover, aiming at minimizing the summation of the hydrogen consumption, the FC degradation, and the battery degradation, Li et al. applied the DQN to develop an EMS for an FCHEV, and the simulation results presented that the optimization performance of the DQN-based EMS outperformed both the Q-learning-based EMS and the RB EMS [51].

Although the original DQN-based EMSs could attain satisfactory control effects, there are still some shortages, for example, being prone to overestimate the Q-value due to the simplex network and slow convergence rate. Hence, to calculate the Q-value and the estimated Q-value separately and speed up the convergence rate, the DDQN algorithm was proposed by Hasselt et.al [33], which has also been applied to the HV EMS field by researchers [1, 14, 30, 39, 42, 46, 116, 119, 120, 142, 144, 147, 156]. For instance, to improve the control performance of the DQL further, a DDQN-based EMS was proposed to optimize the fuel consumption performance of a hybrid electric tracked-vehicle (HETV) by Han et el., and the comparison results indicated that the fuel consumption of the DDQN-based EMS was less than that of the conventional DQN-based EMS by 7.1% while it reached 93.2% level of the DP benchmark [30]. Zhu et al. also applied the DDQN which can automatically develop the learning process to optimize the fuel economy of HEVs further [156]. In addition, to obtain a better policy evaluation and a faster convergence rate, a DQN with a dueling network architecture was proposed by Wang et al. [113], which has also been applied to obtain the optimal EMS [58, 88]. For instance, Qi et al. designed a DQN-based EMS with a dueling network structure, and comparison results showed that the DQN with the dueling network structure converged faster than the original DQN because the dueling network structure could learn from more useful driving records [88].

To obtain better performance on the basis of the original DQN-based EMSs, lots of efforts have been made. The tricks for improving the performance of the DQN-based EMSs are summarized in Table 4. To obtain better optimization effects, some tricks have been integrated to the DQN, such as the

**Table 4** Tricks for improving the performance of DQN-based EMSs

| Performance | Tricks | Powertrain structure | References |
|---|---|---|---|
| Optimization effect | BO | EV with HESS | [46] |
| | NAF & GA | PHEV | [157] |
| | Historical trips | EREV | [111] |
| Convergence rate | PER | Series HEV, PHEV, HEV, FCHEV | [14, 58, 104, 105, 142, 144, 147, 149, 158] |
| | Parallel computing | EV,<br>HEV | [34, 106] |
| | AMSGrad optimizer | HETV | [14, 25] |
| Real-time performance/ adaptability | DNN | HEV | [148] |
| | KL-divergence | HEV | [85, 86] |

Bayesian optimization (BO) [46], the normalized advantage function (NAF), and the genetic algorithm (GA) [157]. For instance, the hyperparameters of the RL have significant influences on the optimal performance [128], therefore Kong et al. proposed a DQL-based EMS for the HESS with the hyperparameters tuned by the BO, and the simulation results indicated that the proposed EMS acquired a better learning performance than the one using the random search and outperformed a near-optimal RB EMS on both the battery life-prolonging and the fuel economy [46]. Zou et al. proposed a self-adaptive DQL-based EMS for PHEVs, where the NAF and the GA were combined to obtain the optimal learning rate, and the simulation results presented that the evolved NAF-DQL reached the best fuel economy [157]. Moreover, Wang et al. utilized historical trips to train the agent of the DDQN-based EMS, and the simulation showed that an average of 19.5% in fuel economy improvement in miles per gallon gasoline equivalent was achieved on 44 test trips [111],

To accelerate the learning process of the DQN, the prioritized experience replay (PER) is one of the commonly used effective tricks [14, 58, 104, 105, 142, 144, 147, 149, 158], which can increase the replay frequency of valuable transition samples [93]. For example, Li et.al proposed a dueling DQN-based EMS, where the PER was introduced for more efficient data utilization during training, and the comparison results indicated that the proposed EMS achieved the faster convergence and the higher reward compared with the DQN-based EMS [58]. Qi et al. introduced a hierarchical structure into the DQL to formulate a DQL-H algorithm, where the high level was used to discretize the BSFV curve into sub-targets so that each sub-driving cycle can move towards the best direction of the fuel consumption, while the low level was used to solve the problem of the sparse reward, and the simulation showed that the DQL-H-based EMS realized better training efficiency than the DQL-based EMS [85, 86]. In addition, Li et.al. proposed a DQN-based EMS and designed a novel reward term to explore the optimal battery SOC range, and simulation results indicated that the training time and the computation time were reduced by 96.5% and 55.4% respectively compared with the Q-learning-based EMS [52, 53]. Moreover, parallel computing was applied to accelerate the training process of the DQN-based EMS as well [34, 106]. Besides, the AMSGrad optimizer was utilized to accelerate the training process of the DQN-based EMS by Guo et al. the simulation results indicated that the DQL with AMSGrad achieved a faster convergence rate than the traditional DQL with Adam optimizer [25],

To enhance the real-time performance/adaptability, Zhao et al. proposed an EMS composed of an offline deep NN construction phase and an online DQL phase, the offline DNN was applied to obtain the correlation between each state-action pair and its value function [148]. Moreover, typical features of the vehicle operation and KL-divergence were utilized to enhance the generalization ability and the real-time performance of the multi-agent DQN-based EMS by Qi [85, 86].

## 3.2 DDPG in HV EMSs

The DQN solved the "curse of dimensionality" problem of the RL successfully by utilizing NNs to approximate the Q-function. However, the control errors can be caused as the DQN algorithm must conduct the action space discretization, and the interval size of the discrete action spaces is another problem that needs to be well trade-off due to its great influence on the training and the optimization effects. Given that, an AC network was introduced to the DRL framework by Lillicrap et al. from Google DeepMind, thereby formulating the DDPG algorithm, and the DRL algorithm ultimately achieved continuous control [108, 109].

The basic principle of the AC is that an actor represents a currently followed policy with continuous action variable output, and then the actions are assessed and evaluated by the critic [108, 109]. On the one hand, the critic network updates its parameters through the TD approach, which is the same as the DQN algorithm, on the other hand, the actor network is updated with a deterministic policy gradient algorithm [96], and the output action is $a \leftarrow \mu(s|\theta^{\mu}) + N$. The critic network consists of a main network parameterized by $\theta^{Q}$ and a target network parameterized by $\theta^{Q\prime}$, the actor network comprises a main network parameterized by $\theta^{\mu}$ and a target network parameterized by $\theta^{\mu\prime}$. $\theta^{Q}$ and $\theta^{\mu}$ are updated according to the exponential smoothing [108, 109], as follows:

$$\begin{cases} \theta^{Q\prime} = \tau \cdot \theta^{Q} + (1-\tau) \cdot \theta^{Q\prime} \\ \theta^{\mu\prime} = \tau \cdot \theta^{\mu} + (1-\tau) v \theta^{\mu\prime} \end{cases} \tag{8}$$

where $\tau \ll 1$ represents the target smooth factor, it affects the update speed of the target networks and the learning stability of the agent.

The Q-value function of the DDPG algorithm is also learned by the Bellman equation, which is shown in Eq. (4), and the TD-error is calculated by $y(t) = r + \gamma \bullet \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$. Finally, the loss function is minimized by the gradient descent method, which is shown in Eq. (9).

$$L_k = \frac{1}{T} \sum_{t=1}^{T} \left( r + \gamma \cdot \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right)^2 \tag{9}$$

Based on the AC framework, the DDPG offers an effective approach for possible implementation in real-world vehicle hardware with low computational abilities, it is

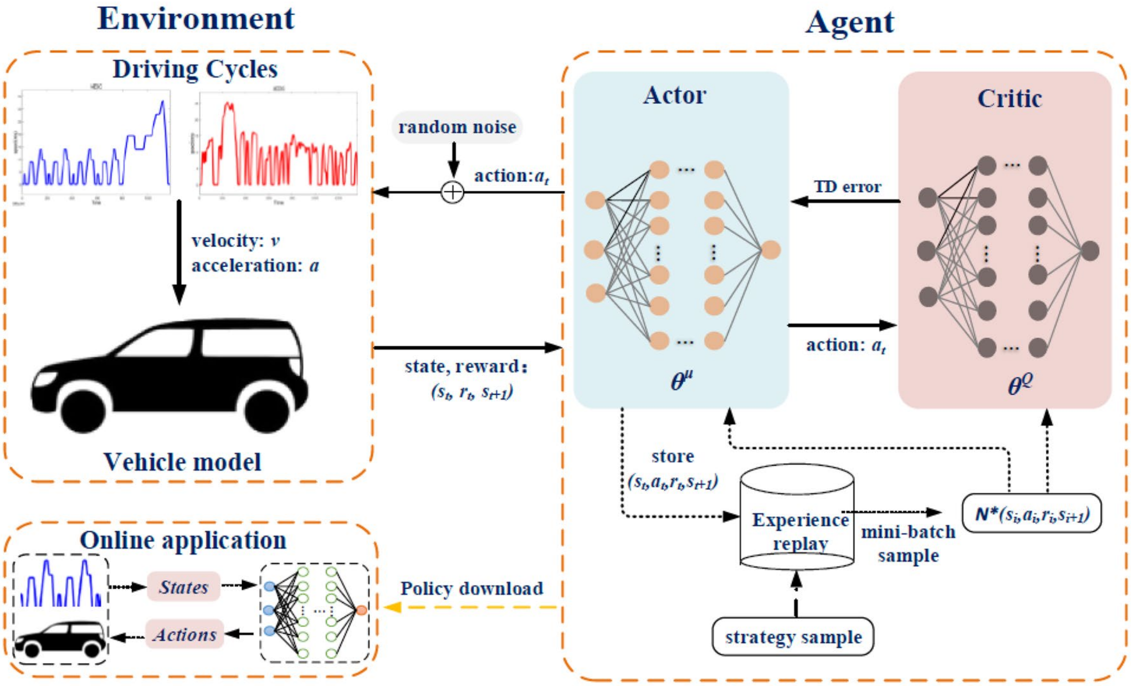


Fig. 7 Applications of DDPG algorithm in HV EMSs

regarded as one of the most promising algorithms to obtain optimal EMS. The framework of the DDPG and its applications in the HV EMS field is described in Fig. 7, differs from the above DQN algorithm, the DDPG agent comprises an actor network and a critic network, which are represented by $\theta^{\mu}$ and $\theta^{Q}$ respectively. At every time-step, the actor is guided to choose the action $a_t$ fed by the TD error from the critic, plus with random noises, such as Ornstein–Uhlenbeck (OU) noise and Gaussian noise, for the sake of improving the exploration ability. As for the online application process of the DDPG-based EMSs, it is similar to the DRL-based EMSs.

The DDPG has been widely applied to improve the fuel economy of HEVs [62, 91] and prolong the lifetime of the batteries and the FCs [154]. For instance, to derive efficient operating strategies for HEVs, Liessner et al. adopted the DDPG to realize fuel-efficient solutions without the route information [62]. Moreover, Zhou et al. proposed a long-term DDPG-based EMS to prolong the service time of the lithium battery and the proton electrolyte membrane FC, and the simulation results indicated that the proposed EMS reduced the attenuation of the FC and the lithium battery effectively [154].

A step further, to derive better performance, some tricks were integrated with the DDPG-based EMS framework by researchers, such as the computer vision (CV) [112], the terrain information [54, 56, 57], and the useful historical driving data, [40, 54, 56, 57, 72]. The tricks for improving the performance of the DDPG-based EMSs are summarized in Table 5. For instance, to obtain better optimization effects, Wang et al. developed an EMS that fused the DDPG with the CV, where a NN-based object detection algorithm was applied to extract the visual information for states input, and the simulation results indicated that the proposed EMS realized 96.5% fuel economy of the DP-based EMS in an urban driving cycle and outperformed the original DDPG agent by 8.8% [112]. In addition, Wu et al. integrated the expert knowledge with the DDPG-based EMS framework to achieve better "cold start" performance and power allocation control effect [117, 121].

To improve the convergence rate, the PER [114, 115, 118, 122], the expert knowledge [60, 70, 71]; [104, 105, 117, 121], and the TL [26, 60, 61, 131–133, 151, 153] were the common tricks that utilized by researchers. Lian et al. proposed an EMS that incorporated the expert knowledge including battery characteristics and the optimal brake specific fuel consumption curve with the DDPG to accelerate the learning process [60]. Moreover, Lian et al. proposed a DDPG-based EMS incorporated with the TL to achieve cross-type knowledge transfer, the simulation results showed that the proposed method obtained a 70% gap from baseline strategy in convergence efficiency and improved the generalization performance corresponding to the changes in vehicle parameters [60, 61].

To derive better real-time performance/adaptability, Zhang et al. utilized a Bayesian NN-based SOC shortage

**Table 5** Tricks of the DDPG-based EMSs

| Performance | Tricks | Powertrain structure | References |
|---|---|---|---|
| Optimization effect | CV | HEV | [112] |
| | Terrain information | HEB | [54, 56, 57] |
| | Historical trip information | HEV,<br>Series HEV, | [40, 54, 56, 57, 72] |
| | Expert knowledge | HEB | [117, 121] |
| Convergence rate | PER | EV,<br>PHEB | [114, 115, 118, 122] |
| | Expert knowledge | HEV | [60, 70, 71]; [104, 105, 117, 121] |
| | TL | HEV,<br>HTV,<br>PHEV | [26, 60, 61, 131–133, 151, 153] |
| Real-time performance/adaptability | DNN | PHEV | [141, 143, 145, 146] |
| | A time-varying weighting factor | HETV | [78] |
| | A large amount of driving cycles | PHEB | [118, 122] |

probability estimator to optimize an adaptive AC-based EMS parameterized by a DNN, and a novel advantage function was introduced to evaluate the energy-saving performance considering the long-term SOC dynamic, the simulation results indicated that the proposed EMS possessed both the adaptability and robust performance in complex and uncertain driving conditions by the average 4.7% energy saving rate [141, 143, 145, 146]. Ma et al. introduced a time-varying weighting factor into a DDPG-based EMS to update old network parameters with the experience learned from the most recent cycle segments, and the simulation results showed that the computational time of the proposed EMS with an online updating mechanism was greatly reduced [78]. Wu et al. proposed a DDPG-based EMS for a series–parallel PHEB, where the agent was trained with a large amount of driving cycles that generated from traffic simulation, and the experiments on the traffic simulation driving cycles showed the DDPG-based EMS presented a great generality to the different standard driving cycles [118, 122]. In addition, Han et al. proposed a multi-state DDPG-based EMS for a series HETV, where the lateral dynamics of the vehicle was systematically integrated into the DDPG-based EMS framework and a multidimensional matrix framework was applied to extract the parameters of the actor network from a trained DDPG-based EMS, and the HIL experiment results showed that the proposed DDPG-based EMS possessed the strong adaptability to different initial SOC values [29].

### 3.3 Other DRL algorithms in HV EMSs

Though the DDPG provides sampling-efficient learning and achieves continuous control, its applications are still notoriously challenging owing to its extreme brittleness and hyperparameter sensitivity according to [17] and Henderson [36]. In other words, the interplay between the deterministic actor networks and the Q-function typically makes it difficult to be stabilized and brittle to hyperparameter settings [17, 36]. Hence, motivated by the DDPG, some original DRL algorithms were proposed, such as the asynchronous advantage AC (A3C) [80], June), the proximal policy optimization (PPO) [94], the TD3 [19], and the SAC [27, 28].

These DRL algorithms have been applied to the HV EMS field by some researchers since 2019, as follows: the A3C [4, 150], the PPO [37, 44, 45, 155], the TD3 [13], [114, 115], [116, 119, 120, 151, 153], and the SAC [116, 119, 120, 123, 131–133]. For instance, Hofstetter et al. proposed a PPO-based EMS to optimize the fuel consumption between the ICE and the electric motor [37]. Moreover, The information of the vehicle to vehicle (V2V) and the vehicle to infrastructure (V2I) was employed as a part of state variables for the training of the PPO algorithm, and the local controller was utilized to improve the learning process by correcting the bad actions [44]. Furthermore, LSTM was also integrated with the PPO to optimize the fuel economy [155]. Additionally, Biswas et al. made a comparison among the A3C-based controller, the ECMS, and the RB controller, the results showed that the A3C-based controller possessed the better potential for real-time control [4]. In addition, the TD3 is an improved DDPG algorithm, it introduced two Q-networks to reduce the over estimation of the Q-value, and delayed the update of the actor network to increase the training stability. Zhou et.al integrated a heuristic RB local controller with the TD3 to design an EMS for HEVs, where a hybrid experience replay method including the offline computed optimal experience and the online learned experience was adopted to resolve the influence of the environmental disturbance, and the simulation results showed that the improved TD3-based EMS obtained the best fuel optimality, the fastest convergence speed, and the highest robustness compared with the DDQN, the Dueling DQN, and the DDPG-based

EMSs [151, 153]. Moreover, the SAC is an AC-based DRL algorithm based on the maximum entropy and can realize better sampling efficiency and learning stability with fewer samples through a smooth policy update [27, 28]. Wu et al. applied the SAC to develop an EMS for an HEB considering both the thermal safety and degradation of the onboard lithium-ion battery system, and the simulation results indicated that the training time was reduced by 87.5% and 96.34% respectively compared with the DQL-based EMS and the Q-learning-based EMS and the fuel economy was increased by 23.3% compared with the DQL-based EMS [116, 119, 120]. In addition, with the help of the parallel computing and the knowledge extracted from the DP-based EMS, Xu et al. developed an EMS based on the SAC for an EV with the HESS, which not only achieved a faster convergence rate by 205.66% compared with the DDPG-based EMS but also realized 8.75% and 6.09% improvements in reducing the energy loss compared with the DQN-based EMS and the DDPG-based EMS respectively and narrowed the gap with the DP-based EMS to 5.19% simultaneously [131–133].

# 4 Issues and Future Trends

## 4.1 Summary and Issues

Energy sources, optimization objectives, and benchmark strategies commonly used in the LB HV EMS field are summarized in the Table 6. The performance of LB EMSs can be improved from three aspects, i.e. the optimization effects, the convergence rate, and the real-time application/adaptability, by different tricks, details of which are listed in the Table 7.

In general, to obtain preferable control effects and improve the adaptability of the RL & DRL-based EMSs, especially the RL-based EMSs, the TPM of the demand power is widely calculated to figure out the demand power distribution [10, 11, 22, 23, 49, 50, 63, 64, 67, 70–75], [67, 75, 77, 118, 122, 124, 124, 125, 125, 138, 142, 144, 147, 152, 158, 159]. Specifically, the common calculation process of the TPM includes two stages: (1) modeling the driving cycles as stationary/finite MC firstly; (2) utilizing a recursive algorithm or nearest neighborhood method and maximum likelihood estimator to extract the TPM from the driving cycles [77].

In addition, the PER trick is also widely utilized to accelerate the learning processes of DRL-based EMSs [14, 104, 105, 114, 115, 118, 122, 142, 144, 147, 149, 151, 153, 158]. Its principle is as follows [93]: the core of the PER is to measure the importance of each transition $i$; a transition's TD error δ indicates how "surprising" or unexpected the transition is; the probability of being sampled is monotonic in a transition's priority while guaranteeing a non-zero probability even for the lowest-priority transition; then, the probability of sampling transition $i$ is defined as

$$P(i) = \frac{p_i^\lambda}{\sum p_i^\lambda} \tag{10}$$

where $p_i = |\delta i| + \epsilon > 0$ is the priority of transition $i$, $\epsilon$ is a small positive constant that prevents the edge-case of transitions not being revisited once their error is zero. The exponent $\lambda$ determines how much prioritization is used, $\lambda = 0$ corresponds to the original experience replay.

**Table 6** Summary of the scope of LB HV EMSs

| Energy source | Optimization objective | Benchmark |
|---|---|---|
| Battery + ICE | Fuel economy/Energy consumption | RB EMSs |
| Battery + FC + ICE | Lifetime extension | Charge-depleting and charge-sustaining, fuzzy logic, deterministic rule, state machine |
| FC + Battery | Efficiency improving | |
| Battery + SC | SOC maintaining | OB EMSs |
| FC + Battery + SC | | DP, ECMS, PMP |
| High-energy and high-power battery pack | | LB EMSs |
| | | Original Q-learning, DQN, and DDPG-based EMSs, etc |

**Table 7** Summary of means of LB EMS performance improvement

| Optimization effect | Convergence rate | Adaptability/Real-time performance |
|---|---|---|
| Combined with conventional algorithms;<br>- RB, ECMS, PMP;<br>Combined with other machine learning algorithms;<br>- NN-based;<br>Historical driving datasets of experienced drivers | Transfer learning;<br>trained with DP/ECMS/PMP results;<br>PER (DRL);<br>Parallel computing | MC + TPM + KL divergence, induced matrix norm/recursive algorithm, forgetting factor, cosine similarity;<br>MPC/NN for velocity prediction;<br>Integrate with ITS, cloud/edge computing |

Furthermore, to enhance the real-time application/adaptability of the LB EMSs, the KL divergence, the induced matrix norm, and the forgetting factor have been widely utilized to determine whether to update the TPM based on the threshold value of the difference between the old TPM and the new TPM during the online application process [5, 15, 63, 64, 67, 75, 124, 124, 125, 125, 159].

However, some issues still exist in the LB HV EMSs. For example, RL-based HV EMSs could derive quasi-optimal effects and converge well in specific driving conditions, but their adaptability turns poor rapidly when the driving environment and conditions change. Additionally, the discretization on the state and action spaces in conventional RL algorithms inevitably brings discretization errors, which make the RL-based EMSs fail to obtain accurate solutions. Moreover, the required calculation time and the storage space increase exponentially as the discrete accuracy increases [141, 143, 145, 146], which would lead to a bad convergence ability during the training process due to the "curse of dimensionality" [127, 134]. As for the DRL-based HV EMSs, on the one hand, though the DQN-related algorithms could derive excellent optimization effects, they cannot achieve continuous control and affect the control accuracy, which is similar to the RL-based EMSs. On the other hand, the AC-based DRL algorithms can realize continuous control and derive satisfactory optimization effects, but some key problems, such as the hyperparameter sensitivity, the instability of training, the brittle convergence performance, the sampling inefficiency, and the tedious training time, remain to be resolved [27]. Finally, the advantages and disadvantages of LB EMSs are summarized in Table 8.

## 4.2 Future Trends

The future trends of LB EMSs might be divided into two aspects: the development of learning algorithms and the on-board implementation of LB EMSs. Firstly, as the current learning algorithms still face the challenges of the tedious training time, the difficulty in objective-function settings, and the difficulty in hyperparameter tuning, more efficient and practical learning algorithms will be put forward. In addition, the inverse RL algorithms can be explored to find the optimal objective-function, and the auto machine learning can be employed to search the optimal hyperparameter set, and the transfer learning can also be integrated with the learning algorithms more deeply to enhance the generalization ability. As for the further implementation of LB EMSs, with the rapid development of the AD and the ITS, learning algorithms can be coordinated with the above technologies to form integrated EMSs (iEMSs) [110]. In addition, LB EMSs will become more intelligent and powerful with the help of the high performance chip, the cellular network, and the over-the-air update technology, which means LB EMSs can adapt to different driving conditions with high drivability and low energy cost. The AD technologies have been an active research hotspot since 2014, and a great many

**Table 8** Summary of advantages and disadvantages of the LB EMSs

| Algorithm | Advantages | Disadvantages | References |
|---|---|---|---|
| RL | | | |
| Q-learning | Quasi-optimal effect | Discrete state spaces and action spaces | [81] |
| TD learning | Better optimization performance than Q-learning | Tabular Q-table<br>Curse of dimensionality<br>Tedious training time | [8, 12, 139] |
| Dyna-style | Better optimization performance than Q-learning | Tedious training time | [68, 76] |
| DRL | | | |
| DQN | Fast convergence rate<br>Continuous state spaces<br>Approximate Q-function with a deep NN | Discrete action spaces<br>Discretization errors<br>Overestimated Q-value | [36, 102] |
| DDQN | More efficient and more stable training performance, faster convergence rate, and better optimization effects than DQN | Discrete action spaces<br>Discretization errors | [36, 102] |
| DDPG | Continuous state and action spaces<br>Can handle high dimensional state-action spaces | Tedious training time<br>Training instability<br>Hyperparameter sensitivity<br>Low sampling efficiency | [17, 19, 36] |
| TD3 | Excellent optimization effect<br>Easier hyperparameter tuning than DDPG | Tedious training time<br>Training instability | [19, 36] |
| SAC | Stable training<br>Soft update<br>Easy hyperparameter tuning<br>Random strategy<br>High sampling efficiency | Tedious training time | [28, 36] |

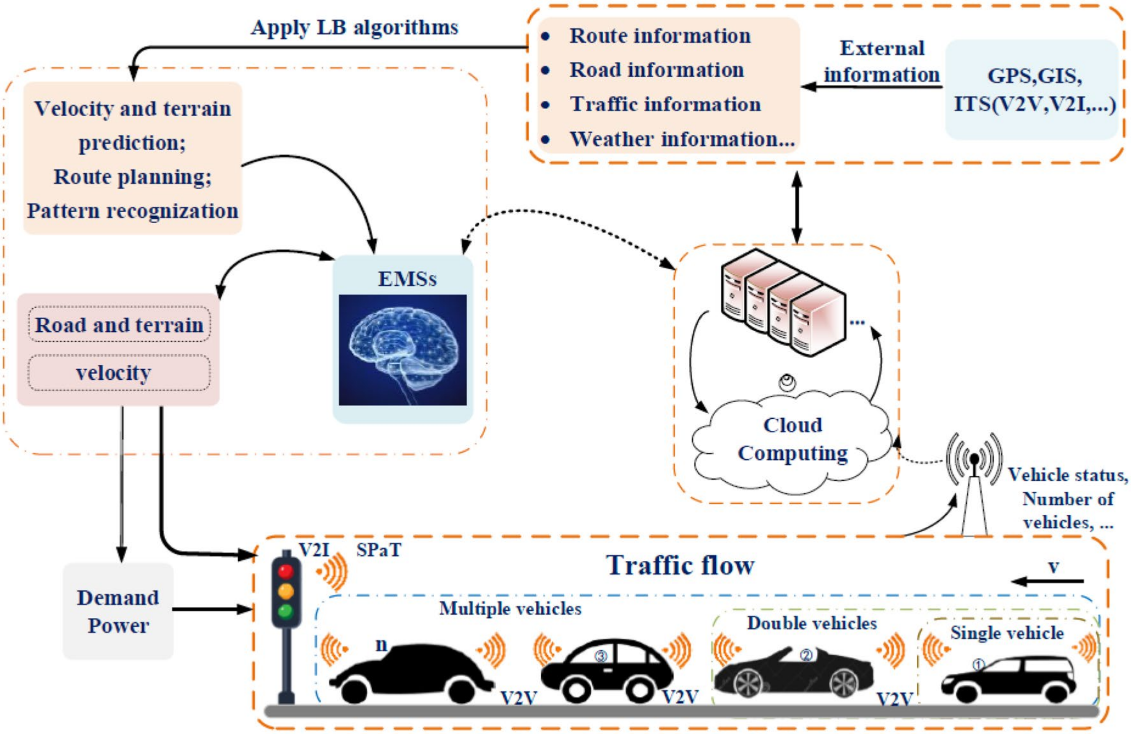


Fig. 8 Schematic diagram of LB EMSs integrated with AD and ITS

researchers and startups are devoting themselves to the level 3-level 4 autopilot at present. The ITS emerged with the pursuit of the smart city, and Yang et al. also reviewed the recent progress of HV EMSs based on the ITS [135]. Moreover, the V2X (vehicle to everything) technologies, including V2V, V2I, V2H (vehicle to house), and V2G (vehicle to grid), etc., are another hotspot under the trends of automobile intellectualization [79, 100]. By integrating with the above technologies, full-scale and instant information that affects the decision-making can be acquired by HV energy-management controllers, such as the velocity of the other vehicles, traffic conditions, and weather conditions. Moreover, cloud computing can be utilized to accelerate the learning processes. The global positioning system (GPS) and the geographic information system (GIS) can be also employed to access the road and terrain information, which is beneficial for forming a highly accurate and real-time LB EMS [6, 83]. The framework of HV LB EMSs with the AD and the ITS is demonstrated in Fig. 8 [140].

## 5 Conclusions

This paper presents a thorough review of the literature related to the LB EMSs of HVs. Detailed applications of RL & DRL algorithms in HV EMSs are described and the merits and the demerits of the LB EMSs are summarized and a preview for the future applications of the learning algorithms in HV EMSs is also carried out. On the one hand, this paper provides the developing trends of the applications of learning algorithms in HV EMSs for the researchers in the EMS field, as the LB EMSs have become an increasingly active research hotspot. On the other hand, this paper contributes to improving the current situation that there are fewer review papers targeting at the learning algorithms applied in the HV EMS field.

With the upcoming big data and AI era, the future trends of the learning algorithms applied in the HV EMS field will be highly data-driven, and the edge computing and cloud computing can be brought to the LB EMS field to decrease the computational burden and enhance the real-time capability. In addition, other novel learning algorithms such as the multi-agent learning and distributed learning can also be applied to the HV EMS field. Furthermore, conventional RB and OB approaches can also be explored to be more deeply merged with the learning algorithms to formulate highly robust and efficient EMSs. A step further, the TL can be explored to be integrated with the LB EMSs as well, which is able to significantly reduce the cost of developing new EMSs among similar powertrain structure HVs.

**Acknowledgements** This research was supported by Shenzhen Science and Technology Innovation Commission (Grant No. KQJSCX20180330170047681, JCYJ20210324115800002), Department of Science and Technology of Guangdong Province (Grant

No. 2021A0505030056, 2021A0505050005), CAS PIFI program (2017VCA0032, 2021VEB0001, 2021VCC0001), Shenzhen Pengcheng Program, and the Hydrogen Energy Innovation Technology Development Program of the National Research Foundation of Korea (NRF) funded by the Korean government (Ministry of Science and ICT(MSIT)) (2019M3E6A1064695).

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding authors state that there is no conflict of interest.

## References

1. Aljohani, T. M., Ebrahim, A., & Mohammed, O. (2021). Real-Time metadata-driven routing optimization for electric vehicle energy consumption minimization using deep reinforcement learning and Markov chain model. *Electric Power Systems Research*. https://doi.org/10.1016/j.epsr.2020.106962
2. Alqahtani, M., & Hu, M. (2022). Dynamic energy scheduling and routing of multiple electric vehicles using deep reinforcement learning. *Energy*. https://doi.org/10.1016/j.energy.2021.122626
3. Biswas, A., Anselma, P. G., & Emadi, A. (2019). Real-time optimal energy management of electrified powertrains with reinforcement learning. In *2019 IEEE Transportation Electrification Conference and Expo (ITEC)* <Go to ISI>://WOS:000502391500035
4. Biswas, A., Anselma, P. G., Rathore, A., & Emadi, A. (2020). Comparison of Three Real-Time Implementable Energy Management Strategies for Multi-mode Electrified Powertrain *2020 IEEE Transportation Electrification Conference & Expo (ITEC)*, <Go to ISI>://WOS:000620344100091
5. Cao, J. Y., & Xiong, R. (2017). Reinforcement learning -based real-time energy management for plug-in hybrid electric vehicle with hybrid energy storage system proceedings of the 9th international conference on applied energy, <Go to ISI>://WOS:000452901602010
6. Chao, S., Moura, S. J., Xiaosong, H., Hedrick, J. K., & Fengchun, S. (2015). Dynamic traffic feedback data enabled energy management in plug-in hybrid electric vehicles. *IEEE Transactions on Control Systems Technology, 23*(3), 1075–1086. https://doi.org/10.1109/tcst.2014.2361294
7. Chaoui, H., Gualous, H., Boulon, L., & Kelouwani, S. (2018). Deep reinforcement learning energy management system for multiple battery based electric vehicles *2018 IEEE Vehicle Power and Propulsion Conference (VPPC)*.
8. Chen, I.-M., Zhao, C., & Chan, C.-Y. (2019). A Deep Reinforcement Learning-Based Approach to Intelligent Powertrain Control for Automated Vehicles *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*.
9. Chen, Z., Gu, H., Shen, S., & Shen, J. (2022). Energy management strategy for power-split plug-in hybrid electric vehicle based on MPC and double Q-learning. *Energy*. https://doi.org/10.1016/j.energy.2022.123182
10. Chen, Z., Hu, H. J., Wu, Y. T., Xiao, R. X., Shen, J. W., & Liu, Y. G. (2018). Energy management for a power-split plug-in hybrid electric vehicle based on reinforcement learning. *Applied Sciences-Basel*, ARTN 24940.3390/app8122494
11. Chen, Z., Hu, H. J., Wu, Y. T., Zhang, Y. J., Li, G., & Liu, Y. G. (2020). Stochastic model predictive control for energy management of power-split plug-in hybrid electric vehicles based on reinforcement learning. *Energy*. https://doi.org/10.1016/j.energy.2020.118931
12. Chen, Z., Li, L., Hu, X. S., Yan, B. J., & Yang, C. (2019). Temporal-difference learning-based stochastic energy management for plug-in hybrid electric buses. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/Tits.2018.2869731
13. Deng, K., Liu, Y., Hai, D., Peng, H., Löwenstein, L., Pischinger, S., & Hameyer, K. (2022). Deep reinforcement learning based energy management strategy of fuel cell hybrid railway vehicles considering fuel cell aging. *Energy Conversion and Management*. https://doi.org/10.1016/j.enconman.2021.115030
14. Du, G., Zou, Y., Zhang, X., Guo, L., & Guo, N. (2022). Energy management for a hybrid electric vehicle based on prioritized deep reinforcement learning framework. *Energy*. https://doi.org/10.1016/j.energy.2021.122523
15. Du, G. D., Zou, Y., Zhang, X. D., Kong, Z. H., Wu, J. L., & He, D. B. (2019). Intelligent energy management for hybrid electric tracked vehicles using online reinforcement learning. *Applied Energy, 251*, 1–16. https://doi.org/10.1016/j.apenergy.2019.113388
16. Du, G. D., Zou, Y., Zhang, X. D., Liu, T., Wu, J. L., & He, D. B. (2020). Deep reinforcement learning based energy management for a hybrid electric vehicle. *Energy*. https://doi.org/10.1016/j.energy.2020.117591
17. Duan, Y., Chen, X., Houthooft, R., Schulman, J., & Abbeel, P. (2016). *Benchmarking Deep Reinforcement Learning for Continuous Control* Proceedings of the 33rd International Conference on Machine Learning. PMLR, https://arxiv.org/abs/1604.06778
18. Fang, Y., Song, C., Xia, B., & Song, Q. (2015). An energy management strategy for hybrid electric bus based on reinforcement learning. 27th Chinese Control and Decision Conference Qingdao, China. https://doi.org/10.1109/CCDC.2015.7162814
19. Fujimoto, S., Hoof, H. v., & Meger, D. (2018). Addressing function approximation error in actor-critic methods proceedings of the *35th International Conference on Machine Learning (PMLR)*, Stockholm, Sweden.
20. Ganesh, A. H., & Xu, B. (2022). A review of reinforcement learning based energy management systems for electrified powertrains: Progress, challenge, and potential solution. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2021.111833
21. Geetha, A., & Subramani, C. (2017). A comprehensive review on energy management strategies of hybrid energy storage system for electric vehicles. *International Journal of Energy Research, 41*(13), 1817–1834. https://doi.org/10.1002/er.3730
22. Guo, H. Q., Du, S. Y., Zhao, F. R., Cui, Q. H., & Ren, W. L. (2019). Intelligent energy management for plug-in hybrid electric bus with limited state space. *Processes*. https://doi.org/10.3390/pr7100672
23. Guo, H. Q., Wei, G. L., Wang, F. B., Wang, C., & Du, S. Y. (2019). Self-learning enhanced energy management for plug-in hybrid electric bus with a target preview based SOC plan method. *IEEE Access, 7*, 103153–103166. https://doi.org/10.1109/Access.2019.2931509
24. Guo, L., Zhang, X., Zou, Y., Guo, N., Li, J., & Du, G. (2021). Cost-optimal energy management strategy for plug-in hybrid electric vehicles with variable horizon speed prediction and adaptive state-of-charge reference. *Energy*. https://doi.org/10.1016/j.energy.2021.120993
25. Guo, L., Zhang, X., Zou, Y., Han, L., Du, G., Guo, N., & Xiang, C. (2022). Co-optimization strategy of unmanned hybrid electric tracked vehicle combining eco-driving and simultaneous energy management. *Energy*. https://doi.org/10.1016/j.energy.2022.123309
26. Guo, X. W., Liu, T., Tang, B. B., Tang, X. L., Zhang, J. W., Tan, W. H., & Jin, S. F. (2020). Transfer deep reinforcement

learning-enabled energy management strategy for hybrid tracked vehicle. *IEEE Access, 8*, 165837–165848. https://doi.org/10.1109/Access.2020.3022944

27. Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor [arXiv:1801.01290v2]. International conference on machine learning. PMLR
28. Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., & Kumar, V. (2018). Soft Actor-Critic Algorithms and Applications [arXiv:1812.05905]. *arXiv preprint*
29. Han, R., Lian, R., He, H., & Han, X. (2021). Continuous reinforcement learning based energy management strategy for hybrid electric tracked vehicles. *IEEE Journal of Emerging and Selected Topics in Power Electronics*. https://doi.org/10.1109/jestpe.2021.3135059
30. Han, X. F., He, H. W., Wu, J. D., Peng, J. K., & Li, Y. C. (2019). Energy management based on reinforcement learning with double deep Q-learning for a hybrid electric tracked vehicle. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2019.113708
31. Hao Dong, Z. D., Shanghang Zhang. (2020). Deep reinforcement learning.
32. Hasselt, H. V. (2015). Double Q-learning. Advances in neural information processing systems, 23.
33. Hasselt, H. V., Guez, A., & Silver, D. (2016). Deep reinforcement learning with double Q-learning Proceedings of the AAAI Conference on Artificial Intelligence
34. He, H. W., Cao, J. F., & Cui, X. (2020). Energy optimization of electric vehicle's acceleration process based on reinforcement learning. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2019.119302
35. Hemmati, R., & Saboori, H. (2016). Emergence of hybrid energy storage systems in renewable energy and transport applications - A review. *Renewable & Sustainable Energy Reviews, 65*, 11–23. https://doi.org/10.1016/j.rser.2016.06.029
36. Henderson, P., Islam, R., Bachman, P., J. P., Precup, D., & Meger, D. (2017). Deep reinforcement learning that matters proceedings of the AAAI conference on artificial intelligence.
37. Hofstetter, J., Bauer, H., Li, W. B., & Waichtmester, G. (2019). Energy and emission management of hybrid electric vehicles using reinforcement learning. *IFAC-PapersOnLine, 52*(29), 19–24. https://doi.org/10.1016/j.ifacol.2019.12.615
38. Hsu, R. C., Chen, S.-M., Chen, W.-Y., & Liu, C.-T. (2016). A Reinforcement learning based dynamic power management for fuel cell hybrid electric vehicle *2016 Joint 8th International Conference on Soft Computing and Intelligent Systems*, <Go to ISI>://WOS:000392122900072
39. Hu, B., & Li, J. (2021). An edge computing framework for powertrain control system optimization of intelligent and connected vehicles based on curiosity-driven deep reinforcement learning. *IEEE Transactions on Industrial Electronics, 68*(8), 7652–7661. https://doi.org/10.1109/Tie.2020.3007100
40. Hu, D., & Zhang, Y. (2022). Deep reinforcement learning based on driver experience embedding for energy management strategies in hybrid electric vehicles. *Energy Technology*. https://doi.org/10.1002/ente.202200123
41. Hu, X. S., Liu, T., Qi, X. W., & Barth, M. (2019). Reinforcement learning for hybrid and plug-in hybrid electric vehicle energy management recent advances and prospects. *IEEE Industrial Electronics Magazine, 13*(3), 16–25. https://doi.org/10.1109/Mie.2019.2913015
42. Hu, Y., Li, W. M., Xu, K., Zahid, T., Qin, F. Y., & Li, C. M. (2018). Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning. *Applied Sciences-Basel*. https://doi.org/10.3390/app8020187
43. Huang, Y. J., Wang, H., Khajepour, A., He, H. W., & Ji, J. (2017). Model predictive control power management strategies for HEVs: A review. *Journal of Power Sources, 341*, 91–106. https://doi.org/10.1016/j.jpowsour.2016.11.106
44. Inuzuka, S., Xu, F. G., Zhang, B., & Shen, T. L. (2019). Reinforcement learning based on energy management strategy for HEVs *2019 IEEE Vehicle Power and Propulsion Conference (VPPC)*, <Go to ISI>://WOS:000532785000183
45. Inuzuka, S., Zhang, B., & Shen, T. (2021). Real-time HEV energy management strategy considering road congestion based on deep reinforcement learning. *Energies*. https://doi.org/10.3390/en14175270
46. Kong, H. F., Yan, J. P., Wang, H., & Fan, L. (2019). Energy management strategy for electric vehicles based on deep Q-learning using Bayesian optimization. *Neural Computing and Applications, 32*(18), 14431–14445. https://doi.org/10.1007/s00521-019-04556-4
47. Kouche-Biyouki, S. A., Naseri-Javareshk, S. M. A., Noori, A., & Javadi-Hassanehgheh, F. (2018). Power Management Strategy of Hybrid Vehicles Using Sarsa Method *26th Iranian Conference on Electrical Engineering (Icee 2018)*, <Go to ISI>://WOS:000482783300178
48. Lee, H., & Cha, S. W. (2021). Energy management strategy of fuel cell electric vehicles using model-based reinforcement learning with data-driven model update. *IEEE Access, 9*, 59244–59254. https://doi.org/10.1109/Access.2021.3072903
49. Lee, H., Kang, C., Park, Y. I., Kim, N., & Cha, S. W. (2020). Online data-driven energy management of a hybrid electric vehicle using model-based Q-learning. *IEEE Access, 8*, 84444–84454. https://doi.org/10.1109/Access.2020.2992062
50. Lee, H., Song, C., Kim, N., & Cha, S. W. (2020). Comparative analysis of energy management strategies for HEV: dynamic programming and reinforcement learning. *IEEE Access, 8*, 67112–67123. https://doi.org/10.1109/Access.2020.2986373
51. Li, J., Wang, H., He, H., Wei, Z., Yang, Q., & Igic, P. (2022). Battery optimal sizing under a synergistic framework with DQN-based power managements for the fuel cell hybrid powertrain. *IEEE Transactions on Transportation Electrification, 8*(1), 36–47. https://doi.org/10.1109/tte.2021.3074792
52. Li, W., Ye, J., Cui, Y., Kim, N., Cha, S. W., & Zheng, C. (2021). A speedy reinforcement learning-based energy management strategy for fuel cell hybrid vehicles considering fuel cell system lifetime. *International Journal of Precision Engineering and Manufacturing-Green Technology*. https://doi.org/10.1007/s40684-021-00379-8
53. Li, W. H., Cui, H., Nemeth, T., Jansen, J., Unlubayir, C., Wei, Z. B., Zhang, L., Wang, Z. P., Ruan, J. G., Dai, H. F., Wei, X. Z., & Sauer, D. U. (2021). Deep reinforcement learning-based energy management of hybrid battery systems in electric vehicles. *Journal of Energy Storage*. https://doi.org/10.1016/j.est.2021.102355
54. Li, Y., Tao, J., & Han, K. (2019). Rule and Q-learning based hybrid energy management for electric vehicle *2019 Chinese Automation Congress (CAC)*,
55. Li, Y., Tao, J. L., Xie, L., Zhang, R. D., Ma, L. H., & Qiao, Z. J. (2020). Enhanced Q-learning for real-time hybrid electric vehicle energy management with deterministic rule. *Measurement & Control, 53*(7–8), 1493–1503. https://doi.org/10.1177/0020294020944952
56. Li, Y. C., He, H. W., Khajepour, A., Wang, H., & Peng, J. K. (2019). Energy management for a power-split hybrid electric bus via deep reinforcement learning with terrain information. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2019.113762
57. Li, Y. C., He, H. W., Peng, J. K., & Wang, H. (2019). Deep reinforcement learning-based energy management for a series hybrid electric vehicle enabled by history cumulative trip information. *IEEE Transactions on Vehicular Technology, 68*(8), 7416–7430. https://doi.org/10.1109/Tvt.2019.2926472

58. Li, Y. C., He, H. W., Peng, J. K., & Wu, J. D. (2018). Energy management strategy for a series hybrid electric vehicle using improved deep Q-network learning algorithm with prioritized replay *Joint International Conference on Energy, Ecology and Environment Iceee 2018 and Electric and Intelligent Vehicles Iceiv 2018*, <Go to ISI>://WOS:000468631900027
59. Li, Y. C., He, H. W., Peng, J. K., & Zhang, H. L. (2017). Power management for a plug-in hybrid electric vehicle based on reinforcement learning with continuous state and action spaces *Proceedings of the 9th International Conference on Applied Energy*, <Go to ISI>://WOS:000452901602066
60. Lian, R. Z., Peng, J. K., Wu, Y. K., Tan, H. C., & Zhang, H. L. (2020). Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle. *Energy*. https://doi.org/10.1016/j.energy.2020.117297
61. Lian, R. Z., Tan, H. C., Peng, J. K., Li, Q., & Wu, Y. K. (2020). Cross-type transfer for deep reinforcement learning based hybrid electric vehicle energy management. *IEEE Transactions on Vehicular Technology, 69*(8), 8367–8380. https://doi.org/10.1109/Tvt.2020.2999263
62. Liessner, R., Schroer, C., Dietermann, A. M., & Bäker, B. (2018). Deep Reinforcement Learning for Advanced Energy Management of Hybrid Electric Vehicles ICAART (2).
63. Lin, X., Zhou, B., & Xia, Y. (2021). Online recursive power management strategy based on the reinforcement learning algorithm with cosine similarity and a forgetting factor. *IEEE Transactions on Industrial Electronics, 68*(6), 5013–5023. https://doi.org/10.1109/tie.2020.2988189
64. Lin, X. Y., Zeng, S. R., & Li, X. F. (2021). Online correction predictive energy management strategy using the Q-learning based swarm optimization with fuzzy neural network. *Energy*. https://doi.org/10.1016/j.energy.2021.120071
65. Liu, C., & Murphey, Y. L. (2014). Power management for plug-in hybrid electric vehicles using reinforcement learning with trip information *2014 IEEE Transportation Electrification Conference and Expo (ITEC)*, Dearborn, MI, United states.
66. Liu, C., & Murphey, Y. L. (2017). Analytical greedy control and Q-learning for optimal power management of plug-in hybrid electric vehicles *2017 Ieee Symposium Series on Computational Intelligence (Ssci)*, <Go to ISI>://WOS:000428251402126
67. Liu, T., Hu, X., Zou, Y., & Cao, D. (2018). Fuel saving control for hybrid electric vehicle using driving cycles prediction and reinforcement learning
68. Liu, T., Hu, X. S., Hu, W. H., & Zou, Y. (2019). A heuristic planning reinforcement learning-based energy management for power-split plug-in hybrid electric vehicles. *IEEE Transactions on Industrial Informatics, 15*(12), 6436–6445. https://doi.org/10.1109/Tii.2019.2903098
69. Liu, T., Hu, X. S., Li, S. E., & Cao, D. P. (2017). Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle. *IEEE/ASME Transactions on Mechatronics, 22*(4), 1497–1507. https://doi.org/10.1109/Tmech.2017.2707338
70. Liu, T., Tan, W., Tang, X., Chen, J., & Cao, D. (2020). Adaptive energy management for real driving conditions via transfer reinforcement learning. *arXiv preprint* arXiv:2007.12560.
71. Liu, T., Tan, W., Tang, X., Zhang, J., Xing, Y., & Cao, D. (2020). Driving conditions-driven energy management for hybrid electric vehicles: a review. *arXiv preprint* arXiv:2007.10880.
72. Liu, T., Tang, X., Hu, X., Tan, W., & Zhang, J. (2020). Human-like energy management based on deep reinforcement learning and historical driving experiences. *arXiv preprint* arXiv:2007.10126.
73. Liu, T., Tian, B., Ai, Y. F., & Wang, F. Y. (2020). Parallel reinforcement learning-based energy efficiency improvement for a cyber-physical system. *IEEE-Caa Journal of Automatica Sinica, 7*(2), 617–626. https://doi.org/10.1109/Jas.2020.1003072
74. Liu, T., Wang, B., Cao, D., Tang, X., & Yan, Y. (2020). Integrated longitudinal speed decision making and energy efficiency control for connected electrified vehicles. *arXiv preprint* arXiv:2007.12565.
75. Liu, T., Wang, B., & Yang, C. L. (2018). Online Markov Chain-based energy management for a hybrid tracked vehicle with speedy Q-learning. *Energy, 160*, 544–555. https://doi.org/10.1016/j.energy.2018.07.022
76. Liu, T., Zou, Y., Liu, D. X., & Sun, F. C. (2015). Reinforcement learning-based energy management strategy for a hybrid electric tracked vehicle. *Energies, 8*(7), 7243–7260. https://doi.org/10.3390/en8077243
77. Liu, T., Zou, Y., Liu, D. X., & Sun, F. C. (2015). Reinforcement learning of adaptive energy management with transition probability for a hybrid electric tracked vehicle. *IEEE Transactions on Industrial Electronics, 62*(12), 7837–7846. https://doi.org/10.1109/Tie.2015.2475419
78. Ma, Z., Huo, Q., Zhang, T., Hao, J., & Wang, W. (2021). Deep deterministic policy gradient based energy management strategy for hybrid electric tracked vehicle with online updating mechanism. *IEEE Access, 9*, 7280–7292. https://doi.org/10.1109/access.2020.3048966
79. Martinez, C. M., Hu, X. S., Cao, D. P., Velenis, E., Gao, B., & Wellers, M. (2017). Energy management in plug-in hybrid electric vehicles: recent progress and a connected vehicles perspective. *IEEE Transactions on Vehicular Technology, 66*(6), 4534–4549. https://doi.org/10.1109/Tvt.2016.2582721
80. Mnih, V., Badia, A. P., Mirza, M., Graves, A., Harley, T., Lillicrap, T. P., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning international conference on machine learning. PMLR
81. Mnih, V., Kavukcuoglu, K., Silver, D., Ioannis, A. G., & Antonoglou. (2013). Playing Atari with Deep Reinforcement Learning. *arXiv preprint* arXiv:1312.5602.
82. Nyong-Bassey, B. E., Giaouris, D., Patsios, C., Papadopoulou, S., Papadopoulos, A. I., Walker, S., Voutetakis, S., Seferlis, P., & Gadoue, S. (2020). Reinforcement learning based adaptive power pinch analysis for energy management of stand-alone hybrid energy storage systems considering uncertainty. *Energy, 193*, 16–40. https://doi.org/10.1016/j.energy.2019.116622
83. Ozatay, E., Onori, S., Wollaeger, J., Ozguner, U., Rizzoni, G., Filev, D., Michelini, J., & Di Cairano, S. (2014). Cloud-based velocity profile optimization for everyday driving: a dynamic-programming-based solution. *IEEE Transactions on Intelligent Transportation Systems, 15*(6), 2491–2505. https://doi.org/10.1109/Tits.2014.2319812
84. Pineau, J., Bellemare, M. G., Islam, R., Henderson, P., & François-Lavet, V. (2018). An introduction to deep reinforcement learning. *arXiv preprint* arXiv:1811.12560*, 11*(3–4), 219–354. https://doi.org/10.1561/2200000071
85. Qi, C., Song, C., Xiao, F., & Song, S. (2022). Generalization ability of hybrid electric vehicle energy management strategy based on reinforcement learning method. *Energy*. https://doi.org/10.1016/j.energy.2022.123826
86. Qi, C., Zhu, Y., Song, C., Yan, G., Xiao, F., & Da, w., Zhang, X., Cao, J., & Song, S. (2022). Hierarchical reinforcement learning based energy management strategy for hybrid electric vehicle. *Energy*. https://doi.org/10.1016/j.energy.2021.121703
87. Qi, C. Y., Zhu, Y. W., Song, C. X., Cao, J. W., Xiao, F., Zhang, X., Xu, Z. H., & Song, S. X. (2021). Self-supervised reinforcement learning-based energy management for a hybrid electric vehicle. *Journal of Power Sources*. https://doi.org/10.1016/j.jpowsour.2021.230584

88. Qi, X. W., Luo, Y. D., Wu, G. Y., Boriboonsomsin, K., & Barth, M. (2019). Deep reinforcement learning enabled self-learning control for energy efficient driving. *Transportation Research Part C-Emerging Technologies, 99*, 67–81. https://doi.org/10.1016/j.trc.2018.12.018
89. Qi, X. W., Luo, Y. D., Wu, G. Y., Boriboonsomsin, K., & Barth, M. J. (2017). Deep reinforcement learning-based vehicle energy efficiency autonomous learning system *2017 IEEE Intelligent Vehicles Symposium* (IV), <Go to ISI>://WOS:000425212700191
90. Qi, X. W., Wu, G. Y., Boriboonsomsin, K., Barth, M. J., & Gonder, J. (2016). Data-driven reinforcement learning-based real-time energy management system for plug-in hybrid electric vehicles. *Transportation Research Record, 2572*(2572), 1–8. https://doi.org/10.3141/2572-01
91. Qu, X. B., Yu, Y., Zhou, M. F., Lin, C. T., & Wang, X. Y. (2020). Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2019.114030
92. Sarvaiya, S., Ganesh, S., & Xu, B. (2021). Comparative analysis of hybrid vehicle energy management strategies with optimization of fuel economy and battery life. *Energy*. https://doi.org/10.1016/j.energy.2021.120604
93. Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized experience replay. *arXiv preprint* arXiv:1511.05952.
94. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms arXiv preprint arXiv:1707.06347,
95. Shuai, B., Zhou, Q., Li, J., He, Y. L., Li, Z. Y., Williams, H., Xu, H. M., & Shuai, S. J. (2020). Heuristic action execution for energy efficient charge-sustaining control of connected hybrid vehicles with model-free double Q-learning. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2020.114900
96. Silver, D., Lever, G., Heess, N., Thomas Degris, Wierstra, D., & Riedmiller, M. (2014). Deterministic policy gradient algorithms *Proceedings of the 31st International Conference on Machine Learning (PMLR)*, Beijing, China.
97. Singh, K. V., Bansal, H. O., & Singh, D. (2019). A comprehensive review on hybrid electric vehicles: Architectures and components. *Journal of Modern Transportation, 27*(2), 77–107. https://doi.org/10.1007/s40534-019-0184-3
98. Song, C., Kim, K., Sung, D., Kim, K., Yang, H., Lee, H., Cho, G. Y., & Cha, S. W. (2021). A Review of optimal energy management strategies using machine learning techniques for hybrid electric vehicles. *International Journal of Automotive Technology, 22*(5), 1437–1452. https://doi.org/10.1007/s12239-021-0125-0
99. Song, C., Lee, H., Kim, K., & Cha, S. W. (2018). A Power Management Strategy for Parallel PHEV Using Deep Q-Networks *2018 IEEE Vehicle Power and Propulsion Conference (VPPC)*, <Go to ISI>://WOS:000459870300026
100. Song, Z., Song, K., & Zhang, T. (2019). State-of-the-art and development trends of energy management strategies for intelligent and connected new energy vehicles: a review *SAE Technical Paper Series*
101. Sun, H. C., Fu, Z. M., Tao, F. Z., Zhu, L. L., & Si, P. J. (2020). Data-driven reinforcement-learning-based hierarchical energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles. *Journal of Power Sources*. https://doi.org/10.1016/j.jpowsour.2020.227964
102. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.
103. Szepesvári, C. (2010). Algorithms for reinforcement learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 4*(1), 1–103. https://doi.org/10.2200/s00268ed1v01y201005aim009
104. Tang, X., Zhang, J., Pi, D., Lin, X., Grzesiak, L. M., & Hu, X. (2022). Battery health-aware and deep reinforcement learning-based energy management for naturalistic data-driven driving scenarios. *IEEE Transactions on Transportation Electrification, 8*(1), 948–964. https://doi.org/10.1109/tte.2021.3107143
105. Tang, X., Zhou, H., Wang, F., Wang, W., & Lin, X. (2022). Longevity-conscious energy management strategy of fuel cell hybrid electric vehicle based on deep reinforcement learning. *Energy*. https://doi.org/10.1016/j.energy.2021.121593
106. Tang, X. L., Chen, J. X., Liu, T., Qin, Y. C., & Cao, D. P. (2021). Distributed Deep reinforcement learning-based energy and emission management strategy for hybrid electric vehicles. *IEEE Transactions on Vehicular Technology, 70*(10), 9922–9934. https://doi.org/10.1109/Tvt.2021.3107734
107. Tao, J. L., Zhang, R. D., Qiao, Z. J., & Ma, L. H. (2022). Q-Learning-based fuzzy energy management for fuel cell/supercapacitor HEV. *Transactions of the Institute of Measurement and Control*. https://doi.org/10.1177/01423312211069488
108. Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint* arXiv:1509.02971*, 29*(11), 5174–5184. https://doi.org/10.1109/TNNLS.2018.2805379
109. Timothy P. Lillicrap, J. J. H., Alexander Pritzel ,Nicolas Heess,Tom Erez,Yuval Tassa,David Silver. (2015). Continuous control with deep reinforcement learning *IEEE Trans Neural Network Learning System*, https://www.ncbi.nlm.nih.gov/pubmed/29994078
110. Tran, D. D., Vafaeipour, M., El Baghdadi, M., Barrero, R., Van Mierlo, J., & Hegazy, O. (2020). Thorough state-of-the-art analysis of electric and hybrid vehicle powertrains: Topologies and integrated energy management strategies. *Renewable & Sustainable Energy Reviews, 119*, 1–29. https://doi.org/10.1016/j.rser.2019.109596
111. Wang, P. Y., Li, Y., Shekhar, S., & Northrop, W. F. (2019). A deep reinforcement learning framework for energy management of extended range electric delivery vehicles *2019 IEEE Intelligent Vehicles Symposium (IV), <Go to ISI>://WOS:000508184100244*
112. Wang, Y., Tan, H. C., Wu, Y. K., & Peng, J. K. (2021). Hybrid electric vehicle energy management with computer vision and deep reinforcement learning. *IEEE Transactions on Industrial Informatics, 17*(6), 3857–3868. https://doi.org/10.1109/Tii.2020.3015748
113. Wang, Z., Schaul, T., Hessel, M., Hasselt, H. v., Lanctot, M., & Freitas, N. d. (2016). Dueling Network Architectures for Deep Reinforcement Learning *Proceedings of the 33rd International Conference on Machine Learning (PMLR)*, New York, USA
114. Wei, H., Zhang, N., Liang, J., Ai, Q., Zhao, W., Huang, T., & Zhang, Y. (2022). Deep reinforcement learning based direct torque control strategy for distributed drive electric vehicles considering active safety and energy saving performance. *Energy*. https://doi.org/10.1016/j.energy.2021.121725
115. Wei, Z., Quan, Z., Wu, J., Li, Y., Pou, J., & Zhong, H. (2022). Deep deterministic policy gradient-DRL enabled multiphysics-constrained fast charging of lithium-ion battery. *IEEE Transactions on Industrial Electronics, 69*(3), 2588–2598. https://doi.org/10.1109/tie.2021.3070514
116. Wu, J., Wei, Z., Li, W., Wang, Y., Li, Y., & Sauer, D. U. (2021). Battery thermal- and health-constrained energy management for hybrid electric bus based on soft actor-critic DRL algorithm. *IEEE Transactions on Industrial Informatics, 17*(6), 3751–3761. https://doi.org/10.1109/tii.2020.3014599

117. Wu, J., Wei, Z., Liu, K., Quan, Z., & Li, Y. (2020). Battery-involved energy management for hybrid electric bus based on expert-assistance deep deterministic policy gradient algorithm. *IEEE Transactions on Vehicular Technology, 69*(11), 12786–12796. https://doi.org/10.1109/tvt.2020.3025627
118. Wu, J. L., Zou, Y., Zhang, X. D., Liu, T., Kong, Z. H., & He, D. B. (2019). An Online correction predictive EMS for a hybrid electric tracked vehicle based on dynamic programming and reinforcement learning. *IEEE Access, 7*, 98252–98266. https://doi.org/10.1109/Access.2019.2926203
119. Wu, P., Partridge, J., Anderlini, E., Liu, Y., & Bucknall, R. (2021). An intelligent energy management framework for hybrid-electric propulsion systems using deep reinforcement learning. *arXiv preprint* arXiv:2108.00256*, 2021*.
120. Wu, P., Partridge, J., Anderlini, E., Liu, Y. C., & Bucknall, R. (2021). Near-optimal energy management for plug-in hybrid fuel cell and battery propulsion using deep reinforcement learning. *International Journal of Hydrogen Energy, 46*(80), 40022–40040. https://doi.org/10.1016/j.ijhydene.2021.09.196
121. Wu, P., Partridge, J., & Bucknall, R. (2020). Cost-effective reinforcement learning energy management for plug-in hybrid fuel cell and battery ships. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2020.115258
122. Wu, Y. K., Tan, H. C., Peng, J. K., Zhang, H. L., & He, H. W. (2019). Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus. *Applied Energy, 247*, 454–466. https://doi.org/10.1016/j.apenergy.2019.04.021
123. Xiao, B., Yang, W., Wu, J., Walker, P. D., & Zhang, N. (2022). Energy management strategy via maximum entropy reinforcement learning for an extended range logistics vehicle. Available at SSRN: https://ssrn.com/abstract=4039638
124. Xiong, R., Cao, J. Y., & Yu, Q. Q. (2018). Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle. *Applied Energy, 211*, 538–548. https://doi.org/10.1016/j.apenergy.2017.11.072
125. Xiong, R., Duan, Y. Z., Cao, J. Y., & Yu, Q. Q. (2018). Battery and ultracapacitor in-the-loop approach to validate a real-time power management method for an all-climate electric vehicle. *Applied Energy, 217*, 153–165. https://doi.org/10.1016/j.apenergy.2018.02.128
126. Xu, B., Hu, X. S., Tang, X. L., Lin, X. K., Li, H. Y., Rathod, D., & Filipi, Z. (2020). Ensemble reinforcement learning-based supervisory control of hybrid electric vehicle for fuel economy improvement. *IEEE Transactions on Transportation Electrification, 6*(2), 717–727. https://doi.org/10.1109/Tte.2020.2991079
127. Xu, B., Malmir, F., Rathod, D., & Filipi, Z. (2019). Real-time reinforcement learning optimized energy management for a 48V mild hybrid electric vehicle SAE Technical Paper Series
128. Xu, B., Rathod, D., Zhang, D. R., Yebi, A., Zhang, X. Y., Li, X. Y., & Filipi, Z. (2020). Parametric study on reinforcement learning optimized energy management strategy for a hybrid electric vehicle. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2019.114200
129. Xu, B., Shi, J., Li, S., Li, H., & Wang, Z. (2021). Energy consumption and battery aging minimization using a q-learning strategy for a battery/ultracapacitor electric vehicle.pdf. *Energy, 229*:120705
130. Xu, B., Tang, X. L., Hu, X. S., Lin, X. K., Li, H. Y., Rathod, D., & Wang, Z. (2021). Q-learning-based supervisory control adaptability investigation for hybrid electric vehicles. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/Tits.2021.3062179
131. 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*. https://doi.org/10.1016/j.est.2021.103925
132. Xu, D., Cui, Y., Ye, J., Cha, S. W., Li, A., & Zheng, C. (2022). A soft actor-critic-based energy management strategy for electric vehicles with hybrid energy storage systems. *Journal of Power Sources, 524*, 231099.
133. Xu, J., Li, Z., Gao, L., Ma, J., Liu, Qi , & Zhao, Y. (2022). A comparative study of deep reinforcement learning-based transferable energy management strategies for hybrid electric vehicles. arXiv:2202.11514.
134. Xu, N., Kong, Y., Chu, L., Ju, H., Yang, Z. H., Xu, Z., & Xu, Z. Q. (2019). Towards a smarter energy management system for hybrid vehicles: a comprehensive review of control strategies. *Applied Sciences-Basel*. https://doi.org/10.3390/app9102026
135. Yang, C., Zha, M. J., Wang, W. D., Liu, K. J., & Xiang, C. L. (2020). Efficient energy management strategy for hybrid electric vehicles/plug-in hybrid electric vehicles: Review and recent advances under intelligent transportation system. *IET Intelligent Transport Systems, 14*(7), 702–711. https://doi.org/10.1049/iet-its.2019.0606
136. Yang, N., Han, L., Xiang, C., Liu, H., & Li, X. (2021). An indirect reinforcement learning based real-time energy management strategy via high-order Markov Chain model for a hybrid electric vehicle. *Energy*. https://doi.org/10.1016/j.energy.2021.121337
137. Yang, N. K., Han, L. J., Xiang, C. L., Liu, H., & Hou, X. Z. (2021). Energy management for a hybrid electric vehicle based on blended reinforcement learning with backward focusing and prioritized sweeping. *IEEE Transactions on Vehicular Technology, 70*(4), 3136–3148. https://doi.org/10.1109/Tvt.2021.3064407
138. Yin, Y., Ran, Y., Zhang, L., Pan, X., & Luo, Y. (2019). An energy management strategy for a super-mild hybrid electric vehicle based on a known model of reinforcement learning. *Journal of Control Science and Engineering*. https://doi.org/10.1155/2019/9259712
139. Yue, S. Y., Wang, Y. Z., Xie, Q., Zhu, D., Pedram, M., & Chang, N. (2014). Model-free learning-based online management of hybrid electrical energy storage systems in electric vehicles *IECON 2014–40th Annual Conference of the IEEE Industrial Electronics Society*, <Go to ISI>://WOS:000389471602155
140. Zhang, F. Q., Hu, X. S., Langari, R., & Cao, D. P. (2019). Energy management strategies of connected HEVs and PHEVs: Recent progress and outlook. *Progress in Energy and Combustion Science, 73*, 235–256. https://doi.org/10.1016/j.pecs.2019.04.002
141. Zhang, F. Q., Wang, L. H., Coskun, S., Pang, H., Cui, Y. H., & Xi, J. Q. (2020). Energy management strategies for hybrid electric vehicles: review, classification, comparison, and outlook. *Energies*. https://doi.org/10.3390/en13133352
142. Zhang, H., Fan, Q., Liu, S., Li, S. E., Huang, J., & Wang, Z. (2021). Hierarchical energy management strategy for plug-in hybrid electric powertrain integrated with dual-mode combustion engine. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2021.117869
143. Zhang, H. L., Peng, J. K., Tan, H. C., Dong, H. X., Ding, F., & Ran, B. (2020). Tackling SOC long-term dynamic for energy management of hybrid electric buses via adaptive policy optimization. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2020.115031
144. Zhang, J. Q., Jiao, X. H., & Yang, C. (2021). A double-deep Q-network-based energy management strategy for hybrid electric vehicles under variable driving cycles. *Energy Technology*. https://doi.org/10.1002/ente.202000770
145. Zhang, Q., Wu, K., & Shi, Y. (2020). Route planning and power management for PHEVs with reinforcement learning. *IEEE Transactions on Vehicular Technology, 69*(5), 4751–4762. https://doi.org/10.1109/Tvt.2020.2979623

146. Zhang, W., Wang, J. X., Liu, Y., Gao, G. Z., Liang, S. W., & Ma, H. F. (2020). Reinforcement learning-based intelligent energy management architecture for hybrid construction machinery. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2020.115401
147. Zhang, X. D., Guo, L. X., Guo, N. Y., Zou, Y., & Du, G. D. (2021). Bi-level Energy management of plug-in hybrid electric vehicles for fuel economy and battery lifetime with intelligent state-of-charge reference. *Journal of Power Sources*. https://doi.org/10.1016/j.jpowsour.2020.228798
148. Zhao, P., Wang, Y. Z., Chang, N., Zhu, Q., & Lin, X. (2018). A deep reinforcement learning framework for optimizing fuel economy of hybrid electric vehicles *2018 23rd Asia and South Pacific Design Automation Conference (Asp-Dac)*, <Go to ISI>://WOS:000426987100032
149. Zheng, C., Li, W., Li, W., Xu, K., Peng, L., & Cha, S. W. (2021). A Deep Reinforcement learning-based energy management strategy for fuel cell hybrid buses. *International Journal of Precision Engineering and Manufacturing-Green Technology*. https://doi.org/10.1007/s40684-021-00403-x
150. Zhou, J., Xue, Y., Xu, D., Li, C., & Zhao, W. (2022). Self-learning energy management strategy for hybrid electric vehicle via curiosity-inspired asynchronous deep reinforcement learning. *Energy*. https://doi.org/10.1016/j.energy.2021.122548
151. Zhou, J. H., Xue, S. W., Xue, Y., Liao, Y. H., Liu, J., & Zhao, W. Z. (2021). A novel energy management strategy of hybrid electric vehicle via an improved TD3 deep reinforcement learning. *Energy*. https://doi.org/10.1016/j.energy.2021.120118
152. Zhou, Q., Li, J., Shuai, B., Williams, H., He, Y. L., Li, Z. Y., Xu, H. M., & Yan, F. W. (2019). Multi-step reinforcement learning for model-free predictive energy management of an electrified off-highway vehicle. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2019.113755
153. Zhou, Q., Zhao, D., Shuai, B., Li, Y., Williams, H., & Xu, H. (2021). Knowledge implementation and transfer with an adaptive learning network for real-time power management of the plug-in hybrid vehicle. *IEEE Trans Neural Netw Learn Syst, 32*(12), 5298–5308. https://doi.org/10.1109/TNNLS.2021.3093429
154. Zhou, Y. F., Huang, L. J., Sun, X. X., Li, L. H., & Lian, J. (2020). A long-term energy management strategy for fuel cell electric vehicles using reinforcement learning. *Fuel Cells, 20*(6), 753–761. https://doi.org/10.1002/fuce.202000095
155. Zhu, Z., Gupta, S., Gupta, A., & Canova, M. (2021). A deep reinforcement learning framework for eco-driving in connected and automated hybrid electric vehicles. *arXiv preprint* arXiv:2101.05372.
156. Zhu, Z., Liu, Y., & Canova, M. (2020). Energy management of hybrid electric vehicles via deep Q-networks *2020 American Control Conference (ACC)*
157. Zou, R., Zou, Y., Dong, Y., & Fan, L. (2020). A self-adaptive energy management strategy for plug-in hybrid electric vehicle based on deep Q learning. *Journal of Physics: Conference Series*.
158. Zou, R. N., Fan, L. K., Dong, Y. R., Zheng, S. Y., & Hu, C. X. (2021). DQL energy management: An online-updated algorithm and its application in fix-line hybrid electric vehicle. *Energy*. https://doi.org/10.1016/j.energy.2021.120174
159. Zou, Y., Liu, T., Liu, D. X., & Sun, F. C. (2016). Reinforcement learning-based real-time energy management for a hybrid tracked vehicle. *Applied Energy, 171*, 372–382. https://doi.org/10.1016/j.apenergy.2016.03.082

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Dezhou Xu** received his M. S. from China University of Mining and Technology in 2022. He is currently working in United Automotive Electronic Systems.

**Chunhua Zheng** received her Ph.D. from Seoul National University in 2012. She is currently working as an associate professor in Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. Her research field includes energy management strategies of new energy vehicles and fuel cells.

**Yunduan Cui** is currently undertaking an associate professor in Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. He received Ph.D. from Nara Institute of Science and Technology, Japan in 2017, M.E from Doshisha University, Japan in 2014, and the B.E. from Xidian University, China in 2012. His research interests include machine learning, especially reinforcement learning in various robots and unmanned systems.

**Shengxiang Fu** received his B. S. from Shandong Institute of Petroleum and Chemical Technology in 2020. His research interest includes energy management strategies of hybrid vehicles and water and thermal management strategies of fuel cell systems.

**Namwook Kim** received his Ph.D. from Seoul National University in 2009. He is currently working as an associate professor in the Department of Mechanical Engineering, Hanyang University ERICA Campus at Ansan, South Korea.

**Suk Won Cha** received his Ph.D. from Stanford University in 2004. He is currently working as a professor in the Department of Mechanical Engineering, Seoul National University, South Korea.