applied sciences

MDPI

Article

Energy Management for a Power-Split Plug-In Hybrid Electric Vehicle Based on Reinforcement Learning

Zheng Chen [1], Hengjie Hu [1], Yitao Wu [1], Renxin Xiao [1], Jiangwei Shen [1,*] and Yonggang Liu [2, 3,*]

[1] Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming 650500, China; chen@kmust.edu.cn (Z.C.); huhengjie1995@163.com (H.H.); yitaowumail@gmail.com (Y.W.); xrx1127@foxmail.com (R.X.)
[2] State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing 400044, China
[3] School of Automotive Engineering, Chongqing University, Chongqing 400044, China
* Correspondence: shenjiangwei6@163.com (J.S.); andyliuyg@cqu.edu.cn (Y.L.);

Received: 24 October 2018; Accepted: 30 November 2018; Published: 4 December 2018

Abstract: This paper proposes an energy management strategy for a power-split plug-in hybrid electric vehicle (PHEV) based on reinforcement learning (RL). Firstly, a control-oriented power-split PHEV model is built, and then the RL method is employed based on the Markov Decision Process (MDP) to find the optimal solution according to the built model. During the strategy search, several different standard driving schedules are chosen, and the transfer probability of the power demand is derived based on the Markov chain. Accordingly, the optimal control strategy is found by the Q-learning (QL) algorithm, which can decide suitable energy allocation between the gasoline engine and the battery pack. Simulation results indicate that the RL-based control strategy could not only lessen fuel consumption under different driving cycles, but also limit the maximum discharge power of battery, compared with the charging depletion/charging sustaining (CD/CS) method and the equivalent consumption minimization strategy (ECMS).

Keywords: energy management strategy; Markov decision process (MDP); plug-in hybrid electric vehicles (PHEVs); Q-learning (QL); reinforcement learning (RL)

1. Introduction

In recent years, as the greenhouse effect and air pollution have become increasingly severe, green energy attracts more attention in all walks of life. In automotive industry, exhaust emission from conventional fuel vehicles is an important factor that causes the environmental pollution. Developing new energy vehicles (NEVs) has shown its significance in reducing emission and lessening induced air pollution. Currently, NEVs can be mainly classified into three types, i.e., fuel cell vehicles, battery electric vehicles (BEVs) and hybrid electric vehicles (HEVs), and they are usually equipped with an energy storage system, such as a battery pack or a super-capacitor [1,2]. For BEVs, it can be powered purely by the battery pack or the super-capacitor. Plug-in hybrid electric vehicles (PHEVs) are considered to combine advantages of both BEVs and HEVs [3]. Compared with HEVs, the prominent advantage of PHEVs is that the battery pack can be recharged by the external charging plug, thereby supplying certain all electric range (AER). Compared with BEVs, the controller of PHEVs can start the engine to sustain the battery when a certain battery state of charge (SOC) threshold is reached and meanwhile supply the extended driving range. Consequently, it is critical to manage the power distribution between the battery and the engine properly in PHEVs.

Energy management strategy (EMS) of PHEVs is responsible for power and energy distribution among different energy storage systems, such as gasoline engine and electromotor. Different control

tradeoff of energy management target is mentioned in related literatures [4,5] including fuel economy improvement [6], and tailpipe emission reduction [7]. Rule based and optimization based methods are mostly considered, as discussed by the authors of [8]. Rule based methods are relatively easier to exploit and are widely employed in practice [9,10]. In [9], a classified rule based EMS is designed, which emphasizes on different operating modes of PHEVs, and simulation results yields satisfied emission reduction. However, these rule based strategies highly depend on design process and engineering experience, thus leading to longer design time [11]. On the contrary, modern real-time and global optimization based algorithms can be applied with provable optimal guarantee. In particular, dynamic programming (DP), adopted by many researchers, is generally treated as an emblematic algorithm among all the optimal methods [12–15]. In [12], the investigators proposed an intelligent EMS based on DP, by which numerical simulation results manifest the improved fuel economy dramatically. Quadratic programming (QP) is also a mature algorithm to search for the optimal result with affordable operational budgets [16], compared with DP. Pontryagin minimum principle (PMP) [17] and equivalent consumption minimization strategy (ECMS) [18] are also widely adopted in EMS of PHEVs. In addition, model predictive control (MPC) [19], is extensively investigated as a real-time optimization manner applying to EMS of PHEVs. Furthermore, intelligent algorithms such as simulated annealing (SA) optimization [17], neural network (NN) [20], genetic algorithm (GA) [21] are also employed for EMS of PHEVs in recent years.

Nowadays, with development of artificial intelligence (AI) technology, reinforcement learning (RL) is becoming more and more popular in various fields including robotic control, intelligent system, and energy management of power grids. In [22], a parallel control architecture based on the RL technology is applied for robotic manipulation, thereby enabling robots to easily adapt to the environment variation. RL is also introduced in the field of energy management of PHEVs in [23–30]. In [23], the investigators find that the RL based EMS cannot only guarantee the vehicle dynamic performance, but also improve the fuel economy, and as a result, can outperform stochastic dynamic program (SDP) in terms of adaptability and learning ability. In [24], the Kullback–Leibler (KL) divergence technique is applied to calculate the power transition probability matrices of the RL algorithm to find the optimal power distribution ratio between the battery and the super-capacitor. Simulation results show that this kind of control policy cannot only effectively decrease the battery charging frequency and control the maximum discharging current, but also maximize the energy efficiency to cut down the overall cost under diverse conditions. In [25], a novel RL based method is proposed combining with the remaining travel distance estimation, and the controller could continuously search for the optimal strategy and learn from the previous process. In [26], a RL method called TD (λ)-learning is employed for the HEV, and simulation results manifest that the RL based policy can improve the fuel economy by 42%. In [27], a blended real-time control strategy is proposed based on the Q-learning (QL) method to balance the overall performance and optimality. A bi-level control strategy is proposed in [28], in which the fuzzy encoding predictor and the KL divergence rate are employed to predict the driver's power demand in the higher level, and the lower level is mainly focused on employing the RL algorithm to find the optimal solution.

Based on the above discussion, it is imperative to further apply the RL technique for energy management of power-split PHEVs. Hence, the main motivation of the energy management strategy is to further refine the battery power based on the RL by selecting proper state and action variables. As a result, the objectives for both optimal fuel economy and battery power restriction can be met at the same time, thereby prolonging the battery life potentially. For the sake of achieving the target, the powertrain of a power-split PHEV is modeled and analyzed first. Subsequently, considering that the proposed method should be applicable in most driving conditions, the Markov chain is adopted to estimate the transition probability matrix regarding demanded power under different driving cycles. Finally, the QL algorithm is conducted to develop and finally form the EMS towards reaching the optimal target. Furthermore, the proposed EMS is compared with the CD/CS strategy to validate the optimality under different driving cycles by simulations. The rest of this article is structured as follows: Section 2 describes the simplified vehicle structure and the fuel consumption model. In Section 3, the RL based framework is proposed to realize the optimal EMS. In Section 4,

corresponding simulations prove the proposed method is superior to the CD/CS algorithm. Section 5 concludes the article.

2. PHEV Powertrain Model

In this paper, the model under study is a power-split PHEV derived from Autonomie. A typical power-split PHEV model is the Toyota Prius PHEV. The powertrain structure of the vehicle is shown in Figure 1, which consists of a 39 ampere-hour (Ah) traction battery pack, a gasoline engine, a final drive, a planetary transmission and two electric motors, i.e., Motor 1 and Motor 2. The engine, Motor 1 and Motor 2 connect with the planet carrier, the ring gear and the sun gear, respectively. As can be seen in Figure 1, motor 2 is employed to provide a significant portion of the electric power, and motor 1 is mainly used as a generator. The main parameters are listed in Table 1.

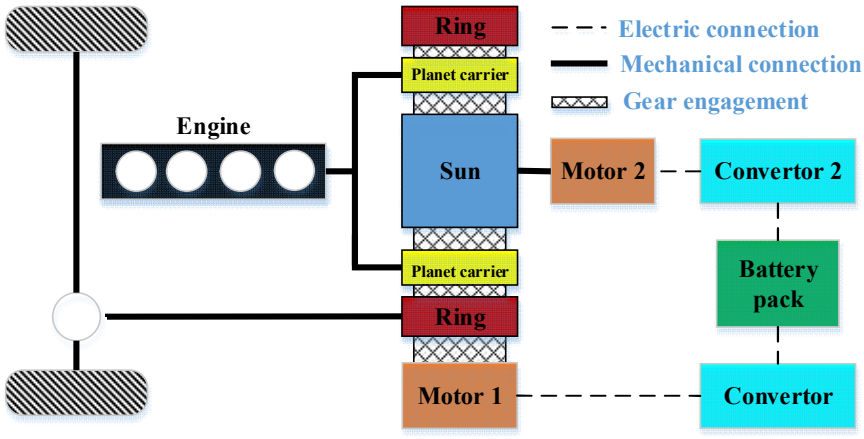


Figure 1. Power-Split plug-in hybrid electric vehicle (PHEV) powertrain structure.

Table 1. Main parameters of power-split PHEV.

Parts	Parameters	Value
Vehicle	Mass	1801 kg
Battery	Rated capacity	39 Ah
Motor 1	Peak power	50 kW
	Rated power	25 kW
Motor 2	Peak power	30 kW
	Rated power	15 kW
Engine	Rated power	57 kW
Planetary gear set	Sun gear	30
	Ring gear	78

2.1. Energy Management Problem

This paper focuses on minimizing the total fuel consumption. Hence, the fuel index β can be established as,

$$\beta = \min F_{total} = \min \int_0^T F_{rate} dt \quad (1)$$

where F_{total} is the total fuel consumption, F_{rate} donotes the fuel rate. T is the total driving time. For the sake of calculating the fuel rate by appropriate simplification, F_{rate} can be determined as,

$$F_{rate} = f(T_{eng}, \omega_{eng}) \quad (2)$$

where ω_{eng}, T_{eng} denote the speed and the torque of engine, respectively. To minimize the fuel consumption, the relationship between the vehicle power request and the fuel consumption needs to be analyzed in detail.

2.2. Power Request Model

Given a certain driving cycle, the power required to drive the vehicle powertrain can be calculated as,

$$P_{req} = (F_f + F_w + F_i)v \tag{3}$$

where P_{req} is the vehicle request power, F_f, F_w, and F_i represent the resistance derived from the road, air drag and vehicle inertial, respectively. v denotes the driving velocity. The resistances, that merely associated with vehicle and environment parameters, can be expressed as,

$$\begin{cases} F_f = mgf \\ F_w = C_d A v^2 / 21.15 \\ F_i = \delta mg \end{cases} \tag{4}$$

where m is the total mass, f denotes the road resistance coefficient, g is the gravity coefficient, A is the frontal area of the vehicle, C_d is the aerodynamic drag coefficient, and δ is the rotational mass coefficient. As shown in Figure 1, the power flow equations can be formulated to describe the corresponding power flow, as:

$$\begin{cases} P_{req} = P_{final} \cdot \eta_{final} \\ P_{final} = (P_{mot1} + P_{mot2} + P_{eng}) \cdot \eta_{gear} \\ P_{bat} = (P_{mot1}/\eta_{c1} + P_{mot2}/\eta_{c2}) + P_{acc} \\ P_{eng} = f_{eng}(T_{eng}, \omega_{eng}) \end{cases} \tag{5}$$

where P_{final} is the driveline power, P_{mot1}, P_{mot2}, and P_{eng} are the output power of motor 1, motor 2 and engine, respectively. P_{acc} denotes the power of electric accessories and is assumed to be a constant value, i.e., 220 W. η_{gear}, η_{final} and η_c are the transmission efficiency factor of gear, final drive and electric convertor, respectively. As seen in Figure 1, the planetary gear set works as the coupling device that connects the engine and the motors, and the corresponding dynamic equations are expressed as follows:

$$\begin{cases} \omega_{eng} = \dfrac{1}{1+i_{gear}}\omega_{mot2} + \dfrac{i_{gear}}{1+i_{gear}}\omega_{mot1} \\ T_{eng} = -(1+i_{gear})T_{mot2} = -\dfrac{1+i_{gear}}{i_{gear}}T_{mot1} \\ \omega_{ring} = \omega_{mot1} = \dfrac{v}{r_{whl}} r_{final} \end{cases} \tag{6}$$

where i_{gear} is the transmission ratio of the planetary gear, ω_{mot1}, ω_{mot2}, and ω_{ring} are the speed of motor 1, motor 2 and ring gear, respectively; T_{mot1} and T_{mot2} are the torque of two motors; r_{whl} denotes the radius of the wheel and r_{final} is the final driveline ratio. In this article, we choose to ignore the inertial of planet gear, sun gear and ring gear for ease of managing the energy distribution.

Based on the above descriptions, the instantaneous fuel consumption F_{rate} can be redefined as:

$$F_{rate} = f(T_{eng}, \omega_{eng}) = f(P_{bat}, P_{req}, v) \tag{7}$$

Now we can find that F_{rate} can be directly determined by P_{bat}, thus it is necessary to model the battery and analyze its power relationship.

2.3. Battery Model

To analyze the power relationship of the battery, a simplified battery model is presented here, which consists of an internal resistor and an open circuit voltage source, and the corresponding calculation equations of the battery model can be described as:

$$\begin{cases} P_{bat} = OCV \cdot i_{bat} - i_{bat}^{\ 2} R_{\text{int}} \\ i_{bat} = \dfrac{OCV - \sqrt{OCV^{\ 2} - 4R_{\text{int}} P_{bat}}}{2R_{\text{int}}} \\ SOC = SOC_{init} - \dfrac{1}{C_{bat}} \int_0^t i_{bat} dt \end{cases} \tag{8}$$

where OCV denotes the battery open circuit voltage, i_{bat} is the battery current, R_{int} is the battery internal resistance, C_{bat} is the battery capacity, SOC is the battery SOC and SOC_{init} is its initial value. Detailed battery parameters varying with SOC are shown in Figure 2. It can be found that R_{int} decreases from 0.1403 ohm to 0.09 ohm and OCV ranges from 165 V to 219.7 V.

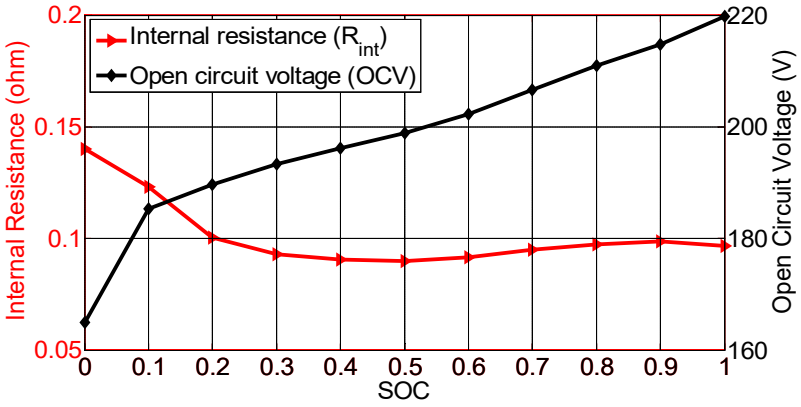


Figure 2. OCV and R_{int} variation with state of charge (SOC).

From the above analysis, we can find that if the battery power is predetermined, the energy distribution strategy inside the vehicle can be achieved. By this manner, the control strategy distributions can be ascertained by the battery power. In order to ensure safety of all components and consider their power limitations and performance extension, some constraint conditions are imposed:

$$\begin{cases} P_{bat\_\min} \le P_{bat} \le P_{bat\_\max} \\ P_{mot1\_\min} \le P_{mot1} \le P_{mot1\_\max} \\ P_{mot2\_\min} \le P_{mot2} \le P_{mot2\_\max} \\ P_{eng\_\min} \le P_{eng} \le P_{eng\_\max} \\ P_{req\_\min} \le P_{req} \le P_{req\_\max} \\ SOC_{\min} \le SOC \le SOC_{\max} \end{cases} \tag{9}$$

where parameters with subscripts min and max mean their corresponding minimum and maximum values, respectively. In the next step, the RL based strategy is introduced to achieve the energy management of the PHEV.

3. Reinforcement Learning for Energy Management

To apply the RL for energy management of PHEVs, we need to build the vehicle power transition probability model first.

3.1. Transition Probablity Model

Markov chain model is a discrete time and state stochastic process with Markov property, of which the state is a sequence with multiple finite random variables. In this process, the selection of the next state is related to the current state and the current action, and does not show any relationship with the previous historical state. In addition, the change of state is independent of time, but is transferred by probability. According to the finite-state Markov chain driver model introduced in [31], the actual driving cycle can be considered as the stochastic Markov chain. The request power is treated as a stochastic variable and can be modeled by the Markov chain. To obtain the transition probability matrix, several standard driving cycles shown in Figure 3 are recorded and analyzed to estimate the transition probability matrix of the demanded power. The selected driving cycles not only include urban, suburban and highway driving conditions, but also involve some intense speed profiles, of which the velocity scale, the acceleration and deceleration frequency can cover most of the driving conditions. According to speed profiles of partially selected driving cycles depicted in Figure 3, the transition probability of the demand power can be calculated based on the maximum likelihood estimation, as:

$$\begin{cases} p_{s,s'} = \dfrac{n_{s,s'}}{n_s} \\ n_s = \sum_{k=1}^{K} p_{s,s'k} \end{cases} \tag{10}$$

where $n_{s,s'}$ represents the counted number transiting from s to s', and n_s is the total number for all transitions of s. $p_{s,s'}$ means the transition probability of the driver's power demand transferred from the current moment to the next moment at each velocity state.

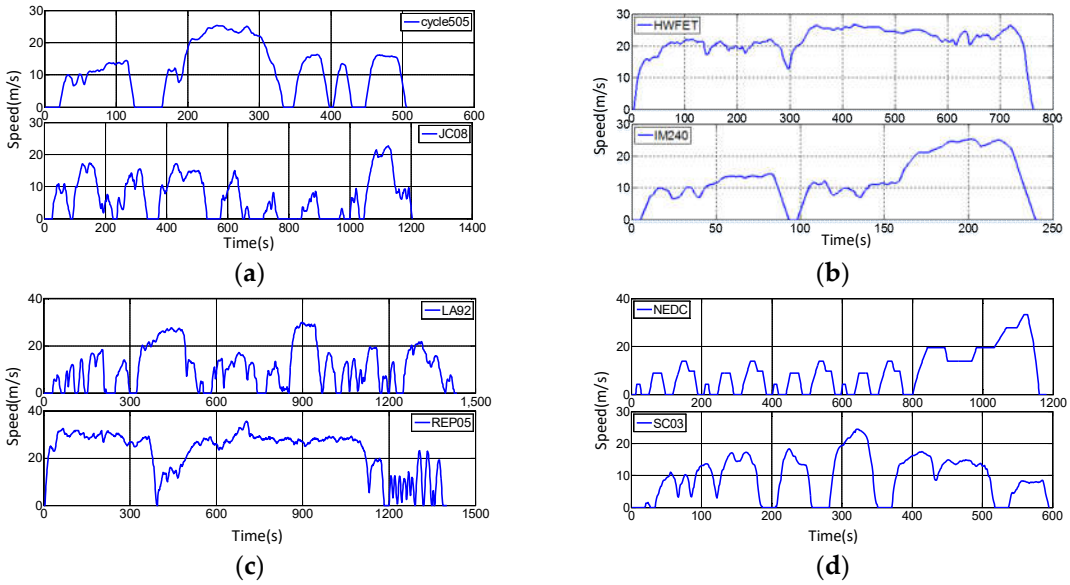
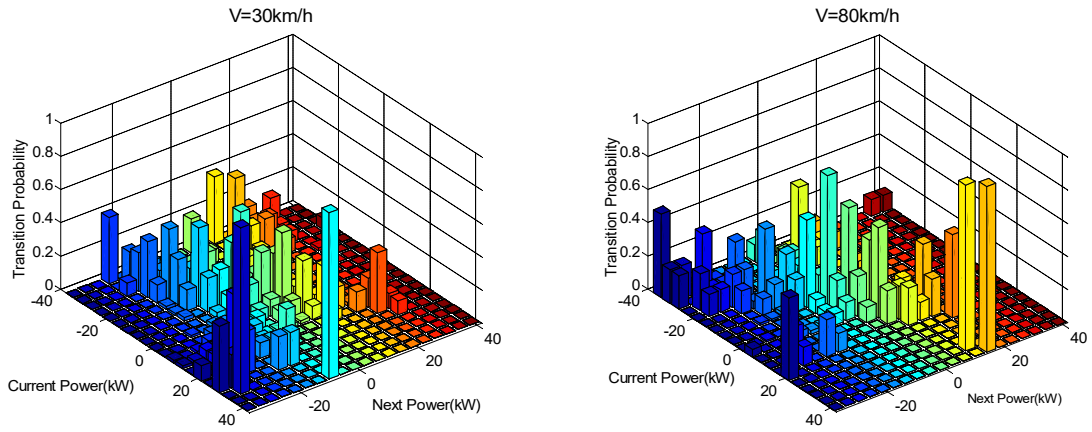


Figure 3. Drive cycle curves: (**a**) Cycle505 and JC08 cycles; (**b**) Highway Fuel Economy Test (HWFET) and IM240 cycles; (**c**) LA92 and REP05 cycles; and (**d**) New European Driving Cycle (NEDC) and SC03 cycles.

According to calculation based on the Markov chain, the transition probability matrix for vehicle speed of 30 km/h and 80 km/h are shown in Figure 4. It can be found that the request power scope is from −40 kW to 40 kW at speed of 30 km/h and the request power scope is from −80 kW to 80 kW at speed of 80 km/h. The transition probability is limited within 0.1 to 0.7, and most of the distribution is concentrated on a diagonal. In addition, it can be clearly seen from Figure 4 that the transiting probability of power request moving from the current state to the next state with different speed values is obviously different.



(**a**) (**b**)

Figure 4. The transition probability map. (**a**) The transition probability map at V = 30 km/h; (**b**) The transition probability map at V = 80 km/h.

3.2. Reinforcement Learning Algorithm

RL, as a significant machine learning method, can conduct repeated explorations in which the agent takes a series of actions in its environment to maximize its designated benefits. The agent-environment interaction for RL is illustrated in Figure 5.

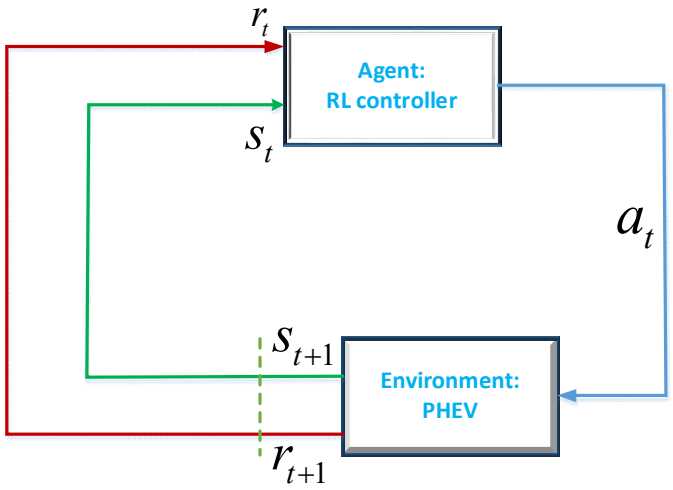


Figure 5. The agent-environment interaction.

The agent-environment interaction can be regarded as a Markov decision process, and the RL mainly focuses on solving the Markov decision process based on a series of iteration. In this paper, the state variable $s \in S$ includes the power request, SOC and the vehicle speed and the action variable $a \in A$ is the battery power. The reward function *r*, which evaluates the current action, is defined as the immediate fuel consumption of the engine.

The object function could be written as the total reward for the finite future at each state, which can be described as:

$$V^*(s) = E\left(\sum_{t=0}^{\infty} \gamma^t r_t\right) \tag{11}$$

where $\gamma \in [0,1]$ is the discount factor to guarantee convergence of the agent during the learning process. Since any state is different and each state is unique, the object function can be reformulated as:

$$V^*(s) = \min_{a \in A}(r(s,a) + \gamma \sum_{s' \in S} p_{sa,s'} V^*(s')) \tag{12}$$

where $p_{sa,s'}$ indicates the transition probability of state variables that change from s to s' based on action a, and $r(s,a)$ indicates the reward of applying action a to transfer from s to s'.

The optimal control strategy is determined by Bellman's principle:

$$\pi^*(s) = \arg\min_{a}(r(s,a) + \gamma \sum_{s' \in S} p_{sa,s'} V^*(s')) \tag{13}$$

As a popular candidate of RL algorithms, the QL algorithm is simple and easy to implement [32], and has been widely employed to solve the optimal value function of MDP. The QL algorithm can obtain a strategy to maximize the sum of expected discounted rewards by directly optimizing an iterated value function Q. According to the updated Q value, the agent needs to examine every action in each iteration to make sure that the learning process can converge. In terms of these merits, we employ the QL algorithm as the kernel algorithm to train, learn and finally achieve the energy management of PHEVs. In the QL algorithm, the Q value, i.e., the state-action value, can be written as:

$$Q^*(s,a) = r(s,a) + \gamma \sum_{s' \in S} p_{sa,s'} \min_{a} Q^*(s',a') \tag{14}$$

Furthermore, the updated rule of Q value can be described as:

$$Q(s,a) \leftarrow Q(s,a) + \eta(r + \gamma \min_{a'} Q(s',a') - Q(s,a)) \tag{15}$$

where $\eta \in [0,1]$ is a decaying factor.

According to the above discussion, the proposed method consists of a simplified vehicle model, a transition probability matrix, a reward matrix and the QL control strategy, where the reward matrix is computed via the simplified vehicle model and the control strategy is calculated according to the power transition matrix, the reward matrix and the QL algorithm feedback. Table 2 lists the pseudocode of the QL algorithm, and it can clearly illustrate the iterative process of QL algorithm. The optimal control strategy is derived through the iterative process shown in Table 2. Figure 6 summarized the detailed procedures of QL in Matlab [19]. First, the QL algorithm and the MDP as well as the related parameters are combined and discretized. Then, the power transition matrix is calculated based on the driver model. Based on the discrete variables and the simplified PHEV model, the reward matrix R is calculated. After iteration, the QL algorithm can be applied successfully to find the optimal energy management solution.

Table 2. The pseudocode of Q–Learning (QL) algorithm.

The QL Algorithm Framework
1. Arbitrarily initialize Q(s,a), S
2. Repeat each step
3. According to the Q(s,a) (ε-greedy policy), choose A
4. Take action A, observe R, S′
5. Update the Q(s,a), $S \leftarrow S'$
6. Until S is terminal

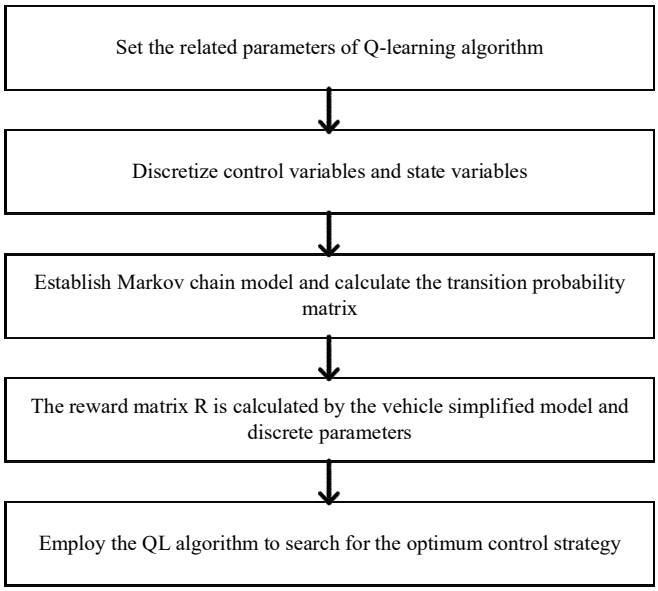


Figure 6. Procedures of the QL calculation.

The optimal control strategy based on the RL algorithm is shown in Figure 7. The battery power ranges from −12 kW to 12 kW, the required power range is limited within −45 kW to 45 kW, and the SOC ranges from 0.3 to 0.9. It can be found that the optimal battery power can be determined by state variables, i.e., the required power, SOC and the vehicle speed. Figure 8 shows the convergence process of the QL algorithm, where the mean discrepancy is applied to measure the difference of the Q values. We can find that with increase of the iterations, the mean discrepancy gradually decreases to 0. From this point, the effectiveness and convergence of the QL algorithm can be proved.

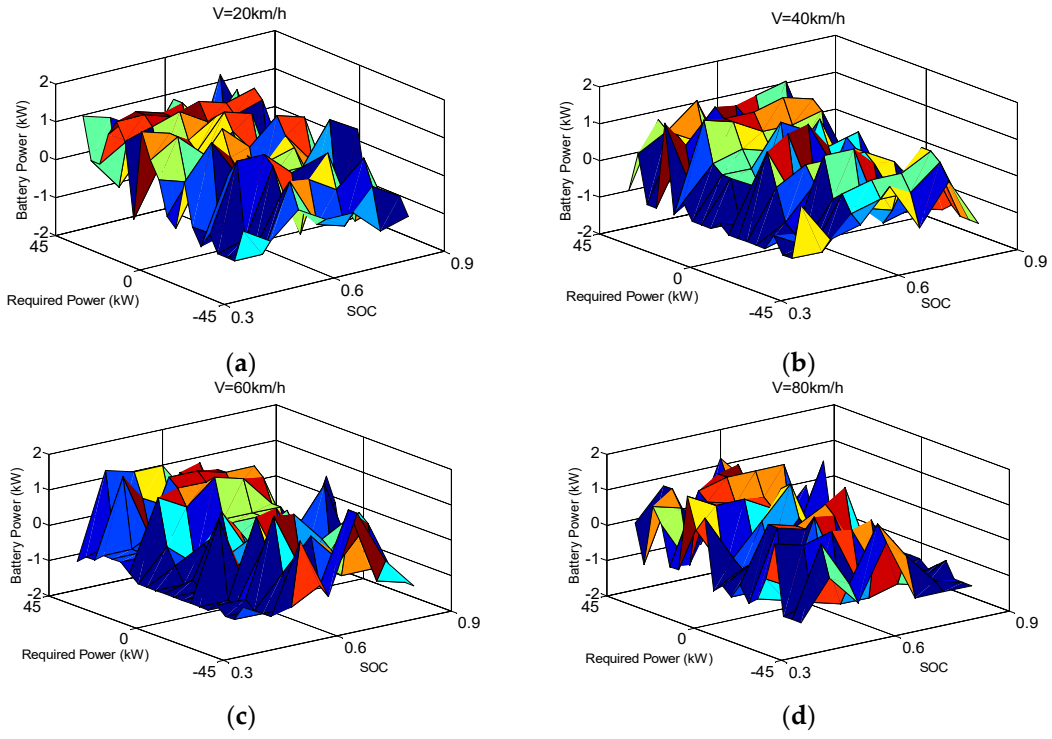


Figure 7. Optimal control strategy based on RL algorithm with different speeds. (**a**) The optimal control action variable at V = 20 km/h; (**b**) The optimal control action variable at V = 40 km/h; (**c**) The optimal control action variable at V = 60 km/h; and (**d**) The optimal control action variable at V = 80 km/h.

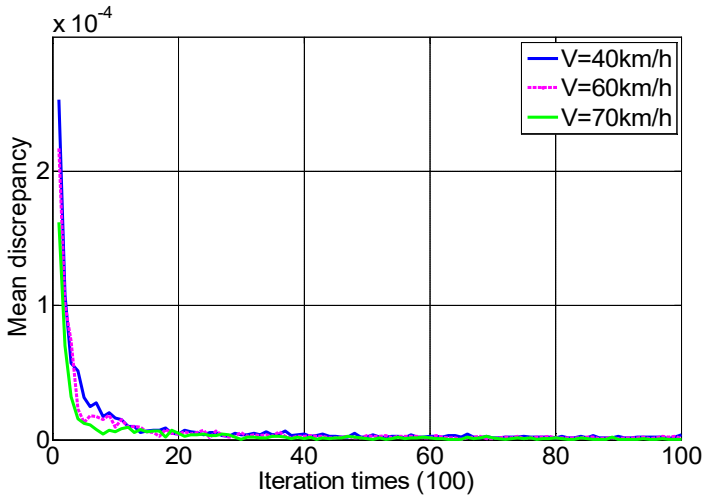


Figure 8. Mean discrepancy of the Q-values.

4. Simulation and Result

In this article, simulations are conducted based on the Autonomie and Matlab/Simulink. New European Driving Cycle (NEDC), Highway Fuel Economy Test (HWFET) and Urban Dynamometer Driving Schedule (UDDS), shown in Figure 9, are employed to verify the proposed strategy. The selected driving cycles can represent most of the driving pattern under different driving conditions.

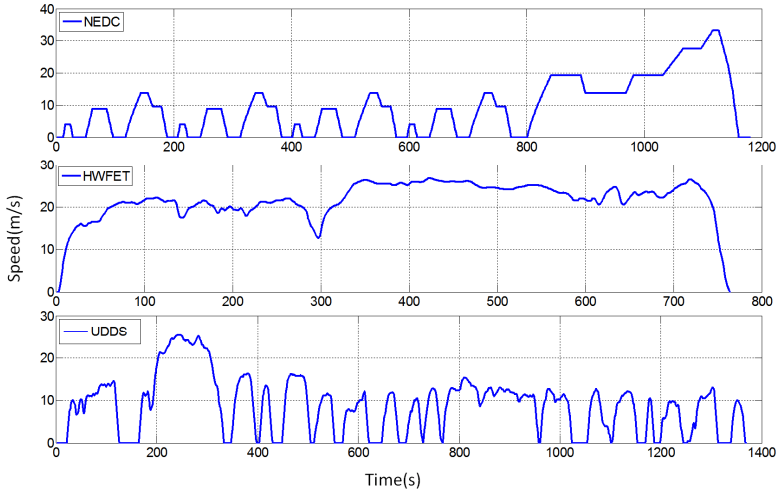


Figure 9. Profile of simulation driving cycles.

To compare the performance of the proposed method, the charge depletion/charge sustaining (CD/CS) algorithm is introduced as a benchmark, which is widely employed in actual applications. In addition, the ECMS is also employed to compare the performance of the proposed algorithm. For the CD/DS algorithm, the power distribution of the vehicle can be easily achieved by setting a series of control parameters without any pre-known information of driving conditions. During the CD stage, except for some specific situation, the engine generally remains shut down, and the tractive power is mainly provided by the battery until the SOC drops to a specified lower threshold (e.g., 30%). Then, the vehicle is powered by both the engine and the battery to remain SOC near the specified value under the CS stage. The detailed CD/CS control scheme can be described [12] as:

$$P_{bat}=\begin{cases}P_{req} & SOC>36\% \\ \min(27804.9,P_{req}) & 33\%\le SOC\le 36\% \\ \min(27804.9\cdot(SOC-0.3)/0.03,P_{req}) & 30\%\le SOC\le 33\% \\ \max(-28157.5\cdot(SOC-0.3)/0.03,P_{req}) & P_{req}<0,27\%\le SOC\le 30\% \\ \max(-28157.5\cdot(SOC-0.3)/0.03,P_{req}-P_{eng\_\max}) & P_{req}>0,27\%\le SOC\le 30\% \\ \max(-28157.5,P_{req}) & P_{req}<0,SOC<27\% \\ \max(-28157.5,P_{req}-P_{eng\_\max}) & P_{req}>0,SOC<27\% \end{cases} \quad (16)$$

where $P_{eng\_\max}$ represents the maximum power of engine.

The ECMS algorithm, as a classical real-time optimization algorithm, transfers the electric consumption of the battery to the equivalent fuel consumption and then tries to minimize the fuel consumption. During each time constant, the vehicle power request is distributed to the battery and the engine according to the minimum principle. By this way, the whole fuel consumption can be reduced and the fuel economy can be improved simultaneously. A typical solution of the ECMS can be formulated based on the Hamilton function, as:

$$H(x(t),u(t),\lambda(t),t)=F_{rate\_eng}(u(t),t)+\lambda\cdot f(x(t),u(t),t) \quad (17)$$

where λ is an equivalent factor that can be adjusted dynamically or can be fixed as a constant value. $x(t)$ and $u(t)$ are state variables and control variables, respectively. In this paper, $x(t)$ includes the battery SOC, the vehicle power demand, and the vehicle speed. Similar to before, $u(t)$ is the battery power. By solving (17), the optimal solution can be found and the final fuel consumption can be obtained.

In simulation validation, three standard cycles are selected to splice multifarious and verifiable conditions. Cycle 1 is consisted of two NEDC cycles, one UDDS cycle and two HWFET cycles, Cycle 2 is comprised of two UDDS cycles, two NEDC cycles and two HWFET cycles, and Cycles 3 and 4 includes five and six HWFET cycles. Cycles 5 and 6 are consisted of six and seven UDDS cycles, respectively. The fuel consumption results with the SOC correction [3] are listed in Table 3. It can be found that compared with the CD/CS scheme, the RL based control strategy can effectively reduce the fuel consumption by 10.1%, 9.31%, 4.84%, 4.49%, 5.95% and 5.13% under different driving cycles. Compared with the ECMS, the RL algorithm can gain similar fuel consumption savings. Thus, the validity of RL based algorithm can be proved. More intuitively, Figure 10 shows the battery power comparison with respect to the proposed algorithm, the ECMS and the CD/CS scheme. The power range of the battery based on the RL algorithm is from −12 kW to 12 kW, while the battery power based on the CD/CS algorithm ranges from −30 kW to 5 kW. It can be recognized that the EMS based on the RL algorithm is capable of controlling the range of the battery power variation smaller than that of the CD/CS method, and the RL method can restrict the maximum battery discharge power. Here we can conclude that the EMS based on the RL control strategy can protect the battery and extend the battery life to some extent.

Table 3. Fuel economy comparison.

Driving Cycle	Strategy	Ending SOC (%)	Fuel Consumption (kg)	Saving (%)
Cycle 1	CD/CS	30.57	1.3205	-
	ECMS	30.21	1.2061	8.39
	RL method	30.45	1.1851	10.1
Cycle 2	CD/CS	30.57	1.7374	-
	ECMS	30.21	1.6067	7.15
	RL method	30.25	1.5702	9.31
Cycle 3	CD/CS	30.57	1.9951	-
	ECMS	30.21	1.8734	5.78
	RL method	30.25	1.8930	4.84

	CD/CS	30.57	2.6360	-
Cycle 4	ECMS	30.21	2.4819	5.60
	RL method	30.25	2.5121	4.49
	CD/CS	30.14	1.2574	-
Cycle 5	ECMS	29.31	1.1681	5.91
	RL method	29.33	1.1688	5.95
	CD/CS	30.14	1.6551	-
Cycle 6	ECMS	2931	1.5488	5.52
	RL method	29.32	1.5563	5.13

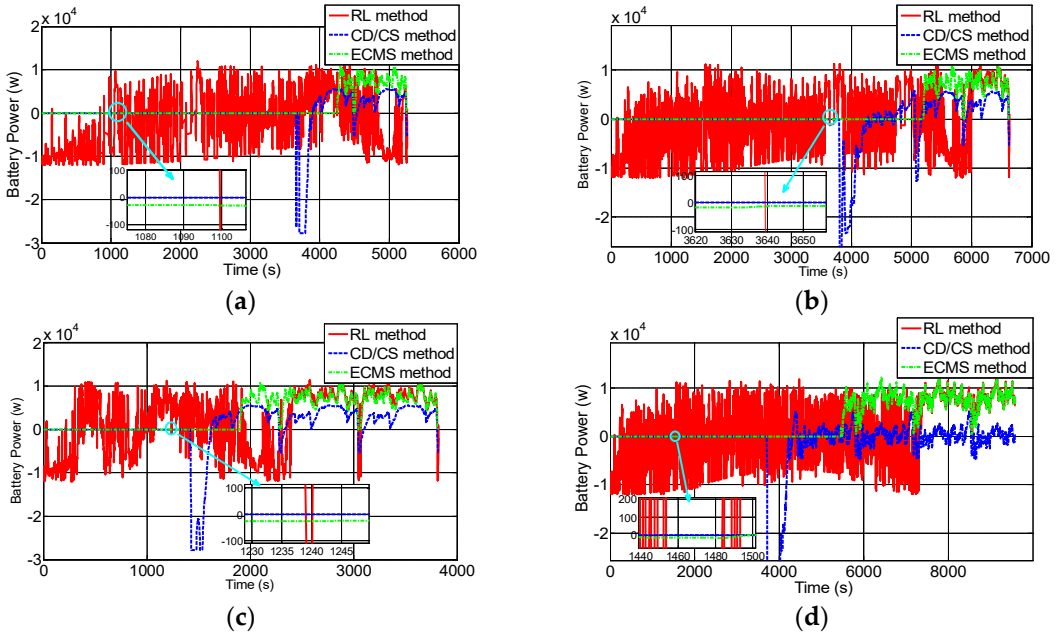


Figure 10. Battery power comparison under driving cycles. (**a**) Cycle 1; (**b**) Cycle 2; (**c**) Cycle 3; and (**d**) Cycle 6.

Figure 11 shows the SOC curve under different driving cycles. The initial SOC is supposed to be 90%, and the minimum SOC threshold is 30%. Compared with the results of the CD/CS scheme, the SOC downward trend based on the RL method is more smoothly. Figure 12 illustrates the fuel consumption under four driving cycles. According to Figures 11 and 12, we can find that the optimized control strategy does not take effect completely in the entire cycle, and works before the battery SOC drops to a certain value. Even so, the proposed algorithm can still effectively reduce the fuel consumption.

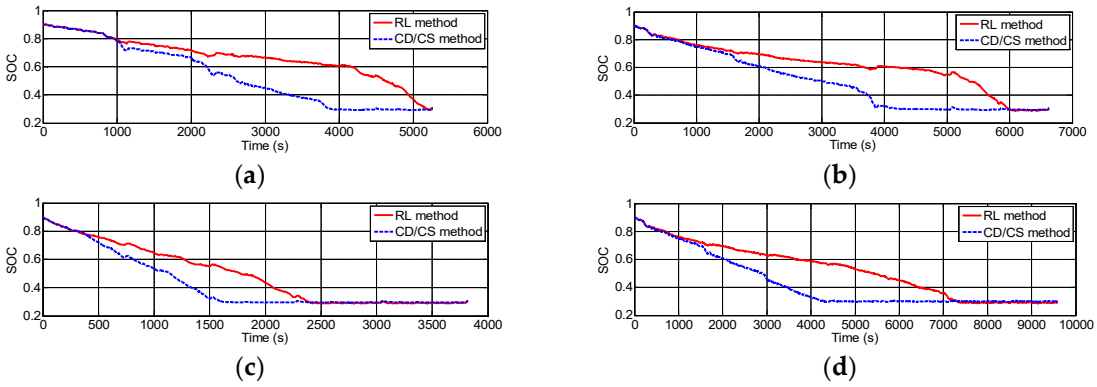


Figure 11. SOC comparison of driving cycles. (**a**) Cycle 1; (**b**) Cycle 2; (**c**) Cycle 3; and (**d**) Cycle 6.

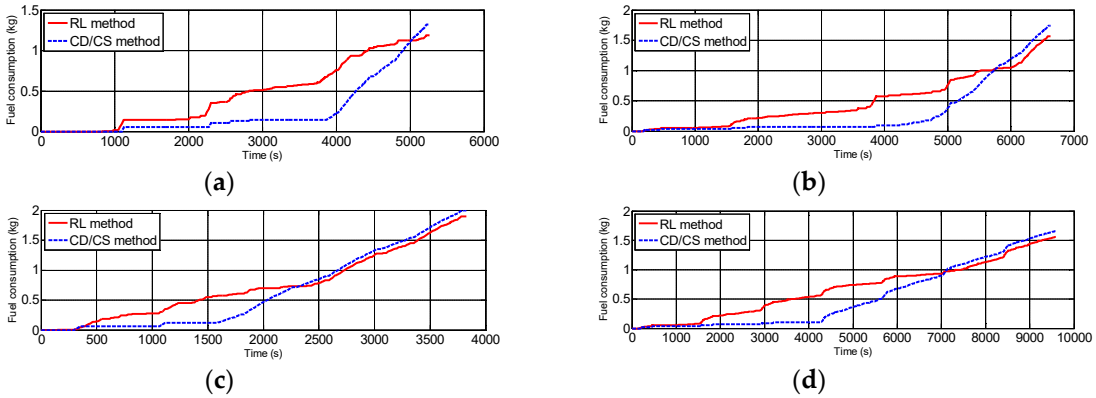


Figure 12. Fuel consumption results. (**a**) Cycle 1; (**b**) Cycle 2; (**c**) Cycle 3; and (**d**) Cycle 6.

To further discover improvements of the RL based strategy, the engine operating points for both RL based method and the CD/CS method under four driving cycles are depicted in Figure 13. It can be obviously found that by implementing the RL based algorithm, the engine working efficiency is higher than 30% in most cases. Compared with the CD/CS strategy, the proposed method can make the engine working points more densely in the high efficiency area. Moreover, it can be noticed that based on the RL based method, the majority of engine working points gather near the optimal operating line, not like that by the CD/CS algorithm. Therefore, it can explain that why the fuel consumption based on the proposed method is less than that based on the CD/CS method.

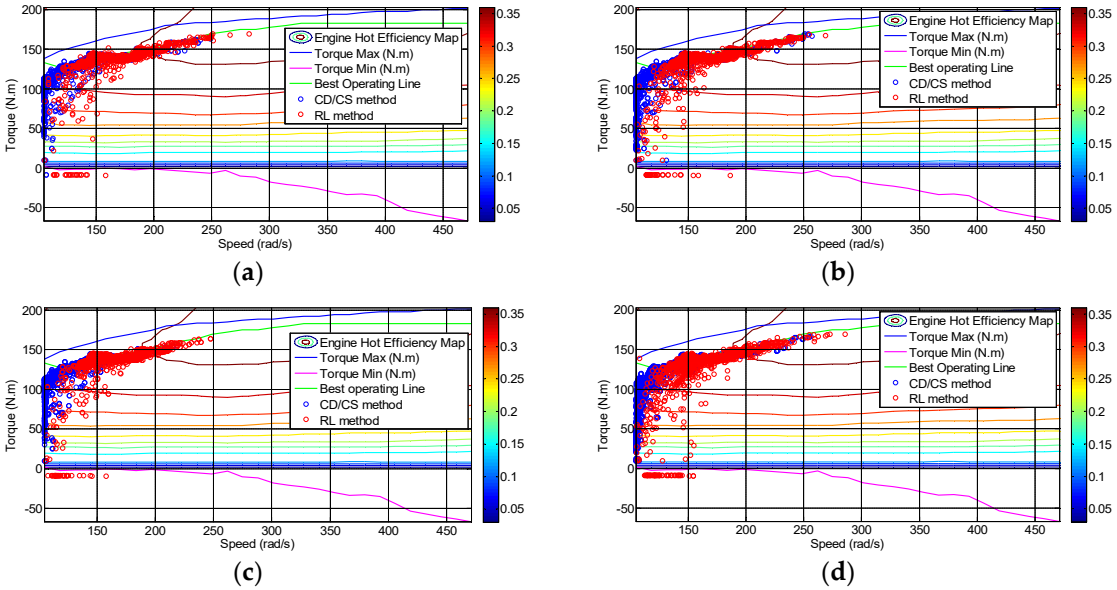


Figure 13. Engine hot efficiency results. (**a**) Cycle 1; (**b**) Cycle 2; (**c**) Cycle 3; and (**d**) Cycle 6.

5. Conclusions

In this paper, the Q-learning RL algorithm has been employed for the energy management of a power-split PHEV. The mathematical vehicle model is built after detailed powertrain analysis. By combining Q-learning method with MDP, the RL model of PHEV is constructed and the optimal result based on RL is obtained where the battery power is optimized. Three standard driving cycles are chosen for simulation verification. Simulation results manifest that the proposed RL algorithm can guarantee a preferable fuel consumption and show more effectiveness than the CD/CS algorithm. In addition, the proposed algorithm can restrict the battery current within a narrower range, thus extending the battery life cycle to some extent.

Our next step work will focus on exploring a more stable Markov chain model and more advanced optimization algorithm. In addition, the proposed algorithm will be further investigated

to update the transition probability matrix of the Markov driver chain in real time, and hardware-in-the-loop and actual vehicle validation will be conducted to verify the real control performance of the proposed method.

Author Contributions: Z.C. and H.H. drafted this paper, discussed combine reinforcement learning with Markov decision process. Y.W. and R.X. provided some energy management strategy suggestions. J.S. oversaw the research. Y.L. revised the paper and provide some technical help.

Funding: This work is supported by the National Science Foundation of China (Grant No. 61763021 and 51775063) in part and the National Key Research and Design Program of China (Grant No. 2018YFB0104000 and 2018YFB0104900) in part. Most importantly, the authors would also like to thank the anonymous reviewers for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Repp, S.; Harputlu, E.; Gurgen, S.; Castellano, M.; Kremer, N.; Pompe, N.; Woerner, J.; Hoffmann, A.; Thomann, R.; Emen, F.M.. Synergetic effects of Fe3+ doped spinel Li4Ti5O12 nanoparticles on reduced graphene oxide for high surface electrode hybrid supercapacitors. *Nanoscale* **2018**, *10*, 1877–1884, doi:10.1039/c7nr08190a.
2. Genc, R.; Alas, M.O.; Harputlu, E.; Repp, S.; Kremer, N.; Castellano, M.; Colak, S.G.; Ocakoglu, K.; Erdem, E.J.S.R. High-Capacitance Hybrid Supercapacitor Based on Multi-Colored Fluorescent Carbon-Dots. *Sci. Rep.* **2017**, *7*, 11222.
3. Martinez, C.M.; Hu, X.; Cao, D.; Velenis, E.; Gao, B.; Wellers, M. Energy Management in Plug-in Hybrid Electric Vehicles: Recent Progress and a Connected Vehicles Perspective. *IEEE Trans. Veh. Technol.* **2017**, *66*, 4534–4549, doi:10.1109/tvt.2016.2582721.
4. Trovao, J.P.F.; Santos, V.D.N.; Antunes, C.H.; Pereirinha, P.G.; Jorge, H.M. A Real-Time Energy Management Architecture for Multisource Electric Vehicles. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3223–3233, doi:10.1109/tie.2014.2376883.
5. Lu, X.; Sun, K.; Guerrero, J.M.; Vasquez, J.C.; Huang, L. State-of-Charge Balance Using Adaptive Droop Control for Distributed Energy Storage Systems in DC Microgrid Applications. *IEEE Trans. Ind. Electron.* **2014**, *61*, 2804–2815, doi:10.1109/tie.2013.2279374.
6. Sabri, M.F.M.; Danapalasingam, K.A.; Rahmat, M.F. A review on hybrid electric vehicles architecture and energy management strategies. *Renew. Sustain. Energy Rev.* **2016**, *53*, 1433–1442, doi:10.1016/j.rser.2015.09.036.
7. Hu, X.; Martinez, C.M.; Yang, Y. Charging, power management, and battery degradation mitigation in plug-in hybrid electric vehicles: A unified cost-optimal approach. *Mech. Syst. Signal Process.* **2017**, *87*, 4–16, doi:10.1016/j.ymssp.2016.03.004.
8. Feng, T.; Yang, L.; Gu, Q.; Hu, Y.; Yan, T.; Yan, B. A Supervisory Control Strategy for Plug-In Hybrid Electric Vehicles Based on Energy Demand Prediction and Route Preview. *IEEE Trans. Veh. Technol.* **2015**, *64*, 1691–1700, doi:10.1109/tvt.2014.2336378.
9. Wirasingha, S.G.; Emadi, A. Classification and Review of Control Strategies for Plug-In Hybrid Electric Vehicles. *IEEE Trans. Veh. Technol.* **2011**, *60*, 111–122, doi:10.1109/tvt.2010.2090178.
10. Peng, J.; He, H.; Xiong, R. Rule based energy management strategy for a series-parallel plug-in hybrid electric bus optimized by dynamic programming. *Appl. Energy* **2017**, *185*, 1633–1643, doi:10.1016/j.apenergy.2015.12.031.
11. Gao, Y.; Ehsani, M. Design and Control Methodology of Plug-in Hybrid Electric Vehicles. *IEEE Trans. Ind. Electron.* **2010**, *57*, 633–640, doi:10.1109/tie.2009.2027918.
12. Chen, Z.; Mi, C.C.; Xu, J.; Gong, X.; You, C. Energy Management for a Power-Split Plug-in Hybrid Electric Vehicle Based on Dynamic Programming and Neural Networks. *IEEE Trans. Veh. Technol.* **2014**, *63*, 1567–1580, doi:10.1109/tvt.2013.2287102.
13. Wu, J.; He, H.; Peng, J.; Li, Y.; Li, Z. Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus. *Appl. Energy* **2018**, *222*, 799–811, doi:10.1016/j.apenergy.2018.03.104.

14. Zhang, S.; Xiong, R. Adaptive energy management of a plug-in hybrid electric vehicle based on driving pattern recognition and dynamic programming. *Appl. Energy* **2015**, *155*, 68–78, doi:10.1016/j.apenergy.2015.06.003.
15. Xie, S.; Sun, F.; He, H.; Peng, J. Plug-In Hybrid Electric Bus Energy Management Based on Dynamic Programming. In *Clean Energy for Clean City: Cue 2016—Applied Energy Symposium and Forum: Low-Carbon Cities and Urban Energy Systems*; Energy Procedia, Yan, J., Wennersten, R., Chen, B., Yang, J., Lv, Y., Sun, Q., Eds.; Jinan, China, 2016; Volume 104, pp. 378–383.
16. Chen, Z.; Mi, C.C.; Xiong, R.; Xu, J.; You, C. Energy management of a power-split plug-in hybrid electric vehicle based on genetic algorithm and quadratic programming. *J. Power Sources* **2014**, *248*, 416–426, doi:10.1016/j.jpowsour.2013.09.085.
17. Chen, Z.; Mi, C.C.; Xia, B.; You, C. Energy management of power-split plug-in hybrid electric vehicles based on simulated annealing and Pontryagin's minimum principle. *J. Power Sources* **2014**, *272*, 160–168, doi:10.1016/j.jpowsour.2014.08.057.
18. Chen, Z.; Xia, B.; You, C.; Mi, C.C. A novel energy management method for series plug-in hybrid electric vehicles. *Appl. Energy* **2015**, *145*, 172–179, doi:10.1016/j.apenergy.2015.02.004.
19. Li, G.; Zhang, J.; He, H. Battery SOC constraint comparison for predictive energy management of plug-in hybrid electric bus. *Appl. Energy* **2017**, *194*, 578–587, doi:10.1016/j.apenergy.2016.09.071.
20. Murphey, Y.L.; Park, J.; Chen, Z.; Kuang, M.L.; Masrur, M.A.; Phillips, A.M. Intelligent Hybrid Vehicle Power Control-Part I: Machine Learning of Optimal Vehicle Power. *IEEE Trans. Veh. Technol.* **2012**, *61*, 3519–3530, doi:10.1109/tvt.2012.2206064.
21. Chen, Z.; Xiong, R.; Wang, C.; Cao, J. An on-line predictive energy management strategy for plug-in hybrid electric vehicles to counter the uncertain prediction of the driving cycle. *Appl. Energy* **2017**, *185*, 1663–1672, doi:10.1016/j.apenergy.2016.01.071.
22. Hester, T.; Quinlan, M.; Stone, P. RTMBA: A Real-Time Model-Based Reinforcement Learning Architecture for Robot Control. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 85–90.
23. Liu, T.; Zou, Y.; Liu, D.; Sun, F. Reinforcement Learning of Adaptive Energy Management With Transition Probability for a Hybrid Electric Tracked Vehicle. *IEEE Trans. Ind. Electron.* **2015**, *62*, 7837–7846, doi:10.1109/tie.2015.2475419.
24. Xiong, R.; Cao, J.; Yu, Q. Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle. *Appl. Energy* **2018**, *211*, 538–548, doi:10.1016/j.apenergy.2017.11.072.
25. Liu, C.; Murphey, Y.L. Power Management for Plug-in Hybrid Electric Vehicles using Reinforcement Learning with Trip Information. In *2014 IEEE Transportation Electrification Conference and Expo*, 2014.
26. Lin, X.; Wang, Y.; Bogdan, P.; Chang, N.; Pedram, M.; IEEE. Reinforcement Learning Based Power Management for Hybrid Electric Vehicles. In Proceedings of the 2014 IEEE/Acm International Conference on Computer-Aided Design, Dearborn, MI, USA, 15–18 June 2014; pp. 32–38.
27. Qi, X.; Wu, G.; Boriboonsomsin, K.; Barth, M.J. A Novel Blended Real-time Energy Management Strategy for Plugin Hybrid Electric Vehicle Commute Trips. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas, Spain, 15–18 September 2015; 10.1109/itsc.2015.167pp. 1002–1007.
28. Liu, T.; Hu, X. A Bi-Level Control for Energy Efficiency Improvement of a Hybrid Tracked Vehicle. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1616–1625, doi:10.1109/tii.2018.2797322.
29. Liu, T.; Hu, X.; Li, S.E.; Cao, D. Reinforcement Learning Optimized Look-Ahead Energy Management of a Parallel Hybrid Electric Vehicle. *IEEE-Asme Trans. Mechatron.* **2017**, *22*, 1497–1507, doi:10.1109/tmech.2017.2707338.
30. Zou, Y.; Liu, T.; Liu, D.; Sun, F. Reinforcement learning-based real-time energy management for a hybrid tracked vehicle. *Appl. Energy* **2016**, *171*, 372–382, doi:10.1016/j.apenergy.2016.03.082.
31. Hong, Y.Y.; Chang, W.C.; Chang, Y.R.; Lee, Y.D.; Ouyang, D.C. Optimal Sizing of Renewable Energy Generations in a Community Microgrid Using Markov Model. *Energy* **2017**, *135*, 68–74.
32. Holland, O.; Snaith, M. Extending the Adaptive Heuristic Critic and Q-Learning—From Facts to Implications. *Artif. Neural Netw.* **1992**, *2*, 599–602.

33. Hou, C.; Ouyang, M.; Xu, L.; Wang, H. Approximate Pontryagin's minimum principle applied to the energy management of plug-in hybrid electric vehicles. *Appl. Energy* **2014**, *115*, 174–189, doi:10.1016/j.apenergy.2013.11.002.

© 2018 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).