sustainability

MDPI

Article

A Novel Minimal-Cost Power Allocation Strategy for Fuel Cell Hybrid Buses Based on Deep Reinforcement Learning Algorithms

Kunang Li [1], Chunchun Jia [1], Xuefeng Han [2] and Hongwen He [1,*]

[1] National Engineering Research Center for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China; likunang1@163.com (K.L.); jiachunchun@163.com (C.J.)
[2] China North Vehicle Research Institute, Beijing 100072, China; xfhan85@gmail.com
* Correspondence: hwhebit@bit.edu.cn

Abstract: Energy management strategy (EMS) is critical for improving the economy of hybrid powertrains and the durability of energy sources. In this paper, a novel EMS based on a twin delayed deep deterministic policy gradient algorithm (TD3) is proposed for a fuel cell hybrid electric bus (FCHEB) to optimize the driving cost of the vehicle. First, a TD3-based energy management strategy is established to embed the limits of battery aging and fuel cell power variation into the strategic framework to fully exploit the economic potential of FCHEB. Second, the TD3-based EMS is compared and analyzed with the deep deterministic policy gradient algorithm (DDPG)-based EMS using real-world collected driving conditions as training data. The results show that the TD3-based EMS has 54.69% higher training efficiency, 36.82% higher learning ability, and 2.45% lower overall vehicle operating cost compared to the DDPG-based EMS, validating the effectiveness of the proposed strategy.

Keywords: fuel cell; energy management strategy; hybrid electric bus; TD3; battery degradation

Citation: Li, K.; Jia, C.; Han, X.; He, H. A Novel Minimal-Cost Power Allocation Strategy for Fuel Cell Hybrid Buses Based on Deep Reinforcement Learning Algorithms. *Sustainability* **2023**, *15*, 7967. https://doi.org/10.3390/su15107967

Academic Editor: Pablo García Triviño

Received: 22 March 2023
Revised: 9 April 2023
Accepted: 29 April 2023
Published: 12 May 2023

1. Introduction

In response to the energy scarcity crisis and climate warming, clean energy represented by hydrogen has received widespread attention [1,2]. The proton exchange membrane fuel cell (PEMFC) is a power generation system that converts the chemical energy in hydrogen into electrical energy [3,4]. Compared with internal combustion engines, proton exchange membrane fuel cells have the advantages of no pollution, high energy density and efficiency, and low noise [2,5]. PEMFCs and power batteries together constitute fuel cell hybrid electric vehicles (FCHEV) [6], which are able to overcome the problem of soft output voltage of PEMFC systems and have the characteristics of fast refueling and zero emissions [7,8]. Energy management strategy determines the power distribution between the fuel cell and the power battery [9], which has a significant impact on the economy and performance of the FCHEB. EMSs can be categorized into three groups: rule-based [10], optimization-based [11], and learning-based [12].

Rule-based EMSs distribute energy demand by fixed rules or fuzzy rules based on the existing engineering experience [13]. Examples include the logic threshold control strategy, thermostat strategy, and power follower strategy [14,15]. Zhang et al. [16] proposed a control strategy based on the state of charge (*SOC*) logic threshold. When the battery *SOC* is in a high state, the battery discharges, and the number of fuel cell stack operations decrease; when the battery *SOC* is in a low state, the number of fuel cell stack operations increases, and the battery is charged. Alexandre et al. [17] proposed a fuzzy logic controller for fuel cell hybrid electric vehicles, compared it with a dynamic programming (DP) algorithm-based control method, and improved the proposed fuzzy strategy using a genetic algorithm.

These methods require little computing power, but they have the problem of poor energy-saving performance and lack adaptability to different working conditions [18].

The optimization-based strategies convert the energy allocation problem to a mathematical optimization problem and solve the optimal or suboptimal solution by means of optimization algorithms [19,20]. The optimization-based EMS can be subdivided according to whether or not it can be applied in real time. The global optimization strategy is able to obtain the optimal control sequence based on known global working conditions, such as dynamic programming and Pontryagin's minimum principle [21,22]. Xu et al. [23] proposed a power allocation strategy based on a dynamic programming algorithm to allocate power between fuel cell engine and lithium-ion battery system to reduce operating costs. However, global optimization strategies require either a priori knowledge or high computational costs, resulting in this type of strategy being implemented only offline and, therefore, generally used only as a benchmark for comparison [24,25]. The real-time optimization strategies approach the energy management effect of the optimal solution by meeting real-time requirements. The most typical of these strategies are the equivalent consumption minimization strategy (ECMS) and model predictive control (MPC). Zeng et al. [26] proposed an adaptive ECMS that periodically updates the equivalence factor based on the predicted power through a local optimization process to converge the battery *SOC* and ensure fuel economy. Chen et al. [27] proposed a model predictive control-based optimization strategy for fuel cell hybrid vehicle energy management considering the variation rate of fuel cell current to improve the durability and performance of PEMFC. However, the control effectiveness of EMS based on real-time optimization is limited by low adaptability to operating conditions or high model accuracy requirements.

With the emergence of artificial intelligence algorithms, learning-based EMS is gaining more and more attention and can avoid the shortcomings of rule-based and optimization-based energy management strategies. Learning-based strategies use trial-and-error mechanisms to gradually evaluate the strategy based on interaction information and reward feedback between the strategy and the vehicle system so that the agent can eventually learn the optimized control strategy. Li et al. [28] proposed an EMS for FCHEV based on speedy Q-learning, which pre-initializes the Q-table of the RL algorithm by power distribution-related rules to improve the convergence speed. Hsu et al. [29] proposed an energy management strategy based on a reinforcement learning approach for fuel cell and battery hybrid vehicles and compared the battery *SOC* maintenance and fuel consumption with a fuzzy logic-based approach. Yang et al. [30] proposed a power allocation method based on Q-learning, which considers system safety, economy, and fuel cell durability to design a real-time reference path for power allocation. These studies have explored the application of reinforcement learning for energy management problems. However, deep reinforcement learning methods are introduced because, although these Q-learning-based strategies achieve some results, they fall into the dimensionality catastrophe problem when the dimensionality increases. Zheng et al. [31] proposed a deep Q-network-based power allocation method for FCHEV which considers hydrogen consumption and fuel cell degradation. Guo et al. [32] proposed an advanced dueling-double-deep Q-network-based energy management strategy to achieve a reasonable balance between system degradation and hydrogen consumption at a low economic cost. The deep Q learning (DQL) algorithm effectively solves the dimensional explosion problem. However, the discrete nature of the actions of the DQL algorithm limits the effectiveness of the control due to the continuity of energy management. Huang et al. [33] proposed an EMS based on DDPG algorithm for a range extend fuel cell hybrid vehicle to achieve optimal power allocation between fuel cell and power battery in pure electric mode and the range extend mode. DDPG introduces an actor-critical framework based on deep Q-learning compared with other deep reinforcement learning methods [34]. Zheng et al. [35] proposed a DDPG-based energy management strategy, improved the efficiency of the algorithm by prioritizing experience replay techniques, and verified the effectiveness of the proposed strategy in comparison with DP. DDPG maintains continuous states and actions and is therefore better adapted

to the continuous control problem of energy management [36,37]. DDPG is now widely used in energy management problems and has achieved good performance, but DDPG overestimates Q -value, which makes it difficult to achieve optimal control [38].

Given these inherent problems, this paper proposes an EMS for FCHEBs on the basis of TD3 algorithm to reduce hydrogen costs and battery aging while considering the *SOC* maintenance of the lithium-ion battery and fuel cell durability. This paper contributes to related research in the following two aspects:

(1) The TD3 algorithm is used to solve the Q-value overestimation problem in the DDPG algorithm, and the proposed strategy uses real collected bus driving conditions as the training set, thus improving training efficiency and optimization.

(2) Battery aging and fuel cell power variation limits are embedded in the strategy framework to reduce hydrogen consumption and extend energy source life, thereby achieving coordination of overall vehicle operating costs and energy source durability.

The remainder of this paper is organized as follows: Section 2 presents the power system model of FCHEB. In Section 3, the TD3-based FCHEB energy management strategy is established. Section 4 compares and analyzes the proposed strategy with other benchmark strategies. Section 5 is the conclusion of this paper.

2. Configuration and Modeling

This section shows the configuration of FCHEB and describes the powertrain model, PEMFC model, battery model, and motor model of the FCHEB.

2.1. FCHEB Configuration

Figure 1 illustrates the FCHEB structure of this study. The fuel cell system transmits electricity to the DC/DC converter, and the electricity is transmitted to the DC bus after being boosted while the battery is connected directly to the DC bus. The PEMFC system consumes hydrogen to generate energy, while the battery receives electrical power from the fuel cell or brake energy recovery and discharges it when the system needs it. The PEMFC system and the power battery work together to transfer energy to the drive motor to meet the vehicle's operating requirements when the vehicle is in motion. The main parameters of the FCHEB are listed in Table 1.

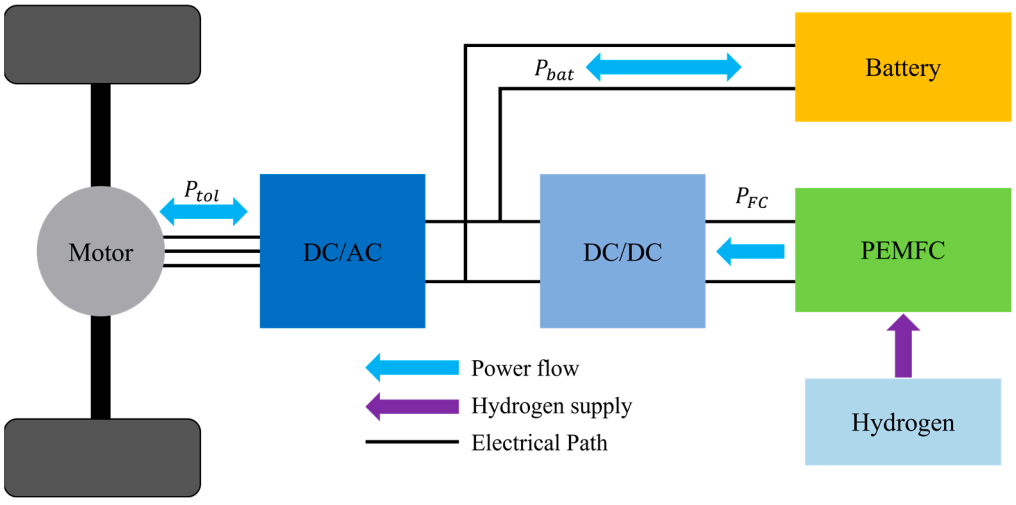

Figure 1. Configuration of the FCHEB.

Table 1. Main parameters of the FCHEB.

Components	Parameters	Values
Vehicle	Weight	12,000 kg
	Front area	8.16 m^2
	Rolling radius	0.466 m
	Rolling resistance coefficient	0.0085
PEMFC System	Rated power	60 kW
Battery	Nominal energy	47.3 kWh
	Battery capacity	90 Ah
Motor	Maximum torque	1800 Nm
	Maximum speed	5000 rpm
Transmission System	Main reduction ratio	6.2
DC/DC converter	Efficiency	0.9
DC/AC inverter	Efficiency	0.95

2.2. Powertrain Model

The vehicle is subject to rolling resistance, air resistance, acceleration resistance, and gradient resistance during the driving process. The total demand power of the vehicle can be obtained by overcoming these resistances, which can be calculated by Equations (1) and (2) [39]. The vehicle interior is powered by the power battery and fuel cell working together, so the vehicle power balance equation is formulated as Equation (3).

$$P_{tol} = \frac{1}{\eta_t}(mgf\cos\alpha + \frac{1}{2}C_D A\rho v^2 + mg\sin\alpha + \delta ma)\cdot v \quad (1)$$

$$\eta_t = \eta_{DC/AC}\cdot\eta_{EM}\cdot\eta_{tra} \quad (2)$$

$$P_{tol} = P_{FC}\cdot\eta_{DC/DC} + P_{bat} \quad (3)$$

where the P_{tol} represents the total power; η_t represents the efficiency of FCHEB; m represents the weight of FCHEB; g represents the gravitational constant; f represents the rolling resistance coefficient; C_D represents the aerodynamic coefficient; A represents the front area; ρ represents air density; δ represents the correlation coefficient of rotating mass; a is the acceleration; v represents the vehicle velocity; α represents the road grade; $\eta_{DC/AC}$, η_{EM}, η_{tra}, $\eta_{DC/DC}$ is the efficiency of DC-AC converter, electric machine, driveline, and DC-DC converter. P_{FC} and P_{bat} are the power of PEMFC system and power battery, respectively.

2.3. PEMFC Model

This study uses a 60 kW fuel cell system as the power unit. The fuel cell hydrogen consumption and fuel cell efficiency at the corresponding fuel cell power were measured experimentally, and the fuel cell efficiency-power data were fitted by quintic curve fitting. The characteristic curve of the fuel cell is shown in Figure 2. Hydrogen mass consumption m_{H_2} of the system can be calculated as the following:

$$m_{H_2} = \int_0^t \frac{P_{FC}}{\eta_{FC}\cdot\rho_{H_2}}dt \quad (4)$$

$$\eta_{FC} = \frac{P_{FC}}{P_{H_2}} \quad (5)$$

where ρ_{H_2} represents the chemical energy density of H_2; η_{FC} represents the efficiency of the PEMFC system; and P_{H_2} is the lower heating value of H_2.

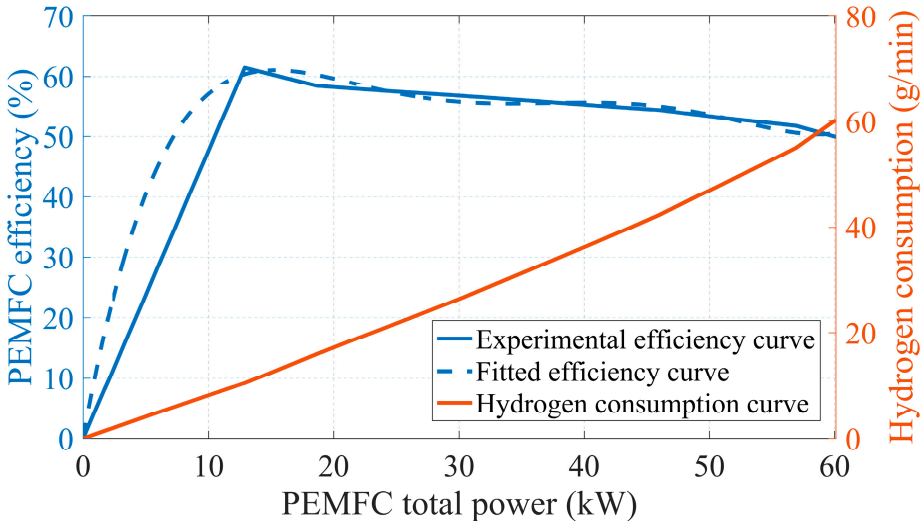

Figure 2. The efficiency and hydrogen consumption curve of the PEMFC system.

2.4. Battery Model

This research adopts an equivalent circuit consisting of a voltage source and a resistor connected in series with the voltage source to simulate a lithium-ion battery pack, which is formulated as Equations (6)–(8) [40]. The battery charging/discharging resistance and open circuit voltage data were measured experimentally for the corresponding battery *SOC*, and the characteristic curves of the battery are shown in Figure 3.

$$P_{bat} = V_{ocv} I_{bat} - I_{bat}^2 R_0 \tag{6}$$

$$I_{bat} = \frac{V_{ocv} - \sqrt{V_{ocv}^2 - 4R_0 P_{bat}}}{2R_0} \tag{7}$$

$$SoC(t+1) = SoC(t) - \frac{I_{bat}(t)\Delta t}{Q_{bat}} \tag{8}$$

where V_{ocv} is the open-circuit voltage; I_{bat} is the current; R_0 is the internal resistance; SoC is the state of charge; Q_{bat} is the nominal battery capacity.

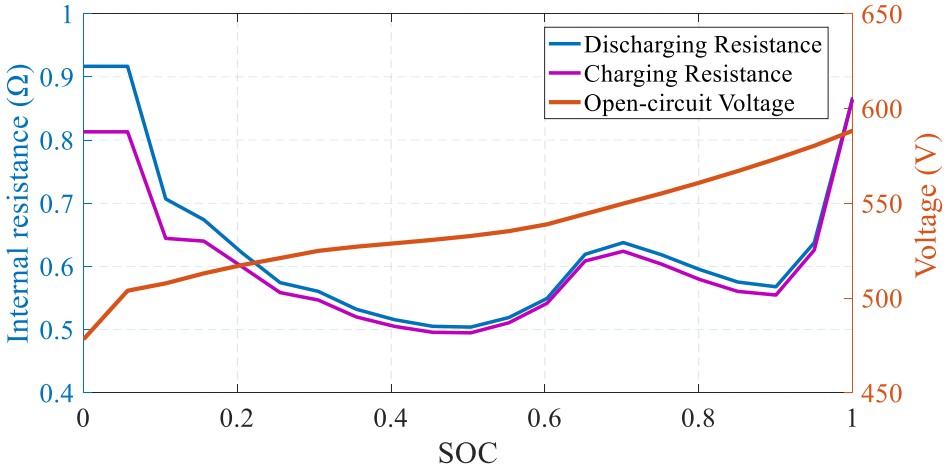

Figure 3. Characteristic curves of the battery.

In this study, a semi-empirical model is used to describe the capacity loss of the battery. The relationship between the capacity loss of the battery and Ah-throughput $A(c)$ at constant operating conditions can be formulated as follows:

$$Q_{loss} = B(c) \cdot \exp(\frac{-E_a(c)}{R \cdot T}) \cdot A(c)^z \tag{9}$$

where Q_{loss} is the capacity loss; it is generally believed that the automotive battery pack reaches the end of life at 20% capacity loss, so it takes 20; c is the C-rate; R is the ideal gas constant; T is the absolute temperature inside the battery, which is taken as 303 K; z represents the power law factor which is taken as 0.55; $B(c)$ is the pre-exponential factor; and E_a is the activation energy, which can be obtained by [41].

The Ah-throughput $A(c)$ can be derived from Equation (9), and the equivalent number of cycles $N(c)$ can be formulated as the following:

$$A(c) = \left[\frac{Q_{loss}}{B(c) \cdot \exp(\frac{-E_a(c)}{R \cdot T})}\right]^{\frac{1}{z}} \tag{10}$$

$$N(c) = \frac{3600A(c)}{Q_{bat}} \tag{11}$$

Hence, the governing equation of the state-of-health (SOH) can be expressed as:

$$\frac{dSOH(t)}{dt} = -\frac{|I_{bat}|}{2N(c)Q_{bat}} \tag{12}$$

2.5. Motor Model

The motor converts electrical energy into mechanical energy when the vehicle needs power and converts mechanical energy into electrical energy for energy recovery when the vehicle is braking. The efficiency map of the motor is illustrated in Figure 4. The efficiency of the motor can be formulated as follows:

$$\eta_{motor} = f(\omega_{motor}, T_{motor}) \tag{13}$$

where η_{motor} is the efficiency of the motor; ω_{motor} is the rotational speed of the motor; T_{motor} is the torque of the motor. After obtaining the speed and torque of the motor, we can determine the motor efficiency by means of a two-dimensional look-up table.

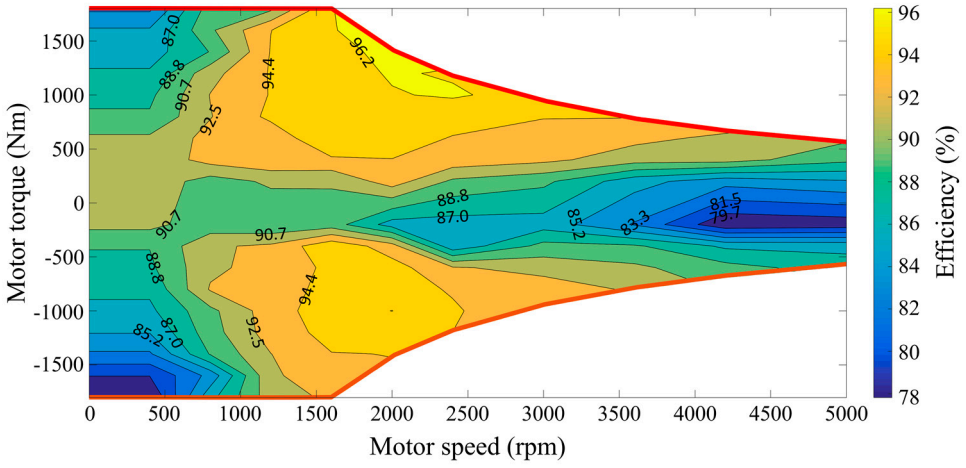

Figure 4. Map of motor efficiency.

3. TD3-Based Energy Management Strategy

In this section, the TD3 algorithm and its advantages are described in detail, and the TD3-based FCHEB energy management strategy is established.

3.1. TD3 Algorithm

The TD3 algorithm is an efficient algorithm that improves on the deep deterministic policy gradient algorithm. It combines the actor-critic framework with the addition of two types of deep neural networks: the critic network Q with parameters θ^Q, and the actor network μ with parameters θ^μ. The actor network is responsible for outputting actions a based on the input state s. The critic network obtains the current state s and the actions a output by the actor network and evaluates the actions accordingly to help the actor network to update. To improve the stability, the algorithm builds the target networks Q' with parameters $\theta^{Q'}$ and the target networks μ' with parameters $\theta^{\mu'}$ for the actor network and the critic network, respectively. Compared to DDPG, TD3 has two critic networks, $Q_1(s,a|\theta^{Q_1})$ and $Q_2(s,a|\theta^{Q_2})$, each of which corresponds to a target network, $Q'_1(s,a|\theta^{Q_1})$ and $Q'_2(s,a|\theta^{Q_2})$, respectively. In addition, the target network parameters and actor network parameters are updated relatively slowly and only when the critic network updates a certain number of steps, which can improve the stability of strategy learning.

In reinforcement learning, the Temporal Difference (td) update method is usually used to accelerate the learning process of Q-value estimation, where the td error can be calculated as Equations (14) and (15). In the calculation, the smaller action value of these two is chosen to calculate the Q-value:

$$y_{\text{target}}(t) = r(s_t, a_t) + \gamma \min_{i=1,2} Q'_i(s_{t+1}, \hat{a}_{t+1}|\theta^{Q'_i}) \tag{14}$$

$$\delta(t) = y_{\text{target}}(t) - Q_i(s_t, a_t|\theta^{Q_i}) \tag{15}$$

where $y_{\text{target}}(t)$ represents the target Q network under the corresponding state s_t and action a_t; r represents the instant reward; γ represents the discount factor to control future rewards; θ^{Q_i} and $\theta^{Q'_i}$ are the parameters of Q_i and target Q'_i network, respectively.

The control actions a are obtained from the target actor network μ' while adding clipped normal distribution noise based on the idea of smoothing. This process can be formulated as follows:

$$\hat{a}_{t+1} = \mu'(s_{t+1}|\theta^{\mu'}) + \xi,\ \xi \sim \text{clip}(\mathcal{N}(0, \sigma^2), -c, c) \tag{16}$$

The loss function is minimized by training the critic network to approximate the Q-value to the target Q-value. The loss function $L(\theta^Q)$ and its gradient can be expressed as follows:

$$L(\theta^{Q_i}) = E[\left(y_{\text{target}}(t) - Q_i(s_t, a_t|\theta^{Q_i})\right)^2] \tag{17}$$

$$\nabla_{\theta^{Q_i}} L(\theta^{Q_i}) = E\Big[(y_{\text{target}} - Q_i(s_t, a_t|\theta^{Q_i}))\nabla_{\theta^{Q_i}} Q_i(s_t, a_t|\theta^{Q_i})\Big] \tag{18}$$

where $E(\cdot)$ represents the mathematical expectation.

The Q-value can be maximized by training the actor network. The objective function $J(\theta^\mu)$ and its gradient can be expressed as the following:

$$J(\theta^\mu) = E[Q_1(s_t, \mu(s_t))] \tag{19}$$

$$\nabla_{\theta^\mu} J(\theta^\mu) = E\Big[\nabla_a Q_1(s_t, a_t|\theta^{Q_1})\nabla_{\theta^\mu}\mu(s_t|\theta^\mu)\Big] \tag{20}$$

The gradient descent update formula for the actor network and the critic network can be expressed as follows:

$$\theta^{Q_i} \leftarrow \theta^{Q_i} + \alpha \cdot \nabla_{\theta^{Q_i}} L(\theta^{Q_i}) \tag{21}$$

$$\theta^{\mu} \leftarrow \theta^{\mu} + \beta \cdot \nabla_{\theta^{\mu}} J(\theta^{\mu}) \tag{22}$$

where α is the learning rate of the critic-network; and β is the learning rate of the actor-network.

The target networks ${Q'}_i$ and μ' are updated through soft updating by the equation,

$$\theta^{Q'_i} \leftarrow \tau\theta^{Q_i} + (1-\tau)\theta^{Q'_i} \tag{23}$$

$$\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'} \tag{24}$$

where $\tau(\tau \ll 1)$ is the factor to control updating speed.

3.2. TD3-Based EMS

Considering the operation of the FCHEB and the control objectives of the system, the state space and action space are set as follows:

$$s = [v, acc, SOC, SOH] \tag{25}$$

$$a = [P_{FC} | P_{FC} \in [0, 60\text{kW}]] \tag{26}$$

The optimization objective of this study is to reduce the hydrogen consumption of FCHEB and the aging of the battery based on the consideration of *SOC* maintenance and fuel cell durability. Based on the above requirements, the reward function is set as follows:

$$r = -\left\{C_{H_2} + C_{bat} + L_{SOC} + L_{FC}\right\} \tag{27}$$

$$C_{H_2} = \rho_1 \cdot [m_{\text{H}_2}(t)] \tag{28}$$

$$C_{bat} = \rho_2 \cdot Q_{bat} \cdot \Delta SOH \tag{29}$$

$$L_{SOC} = \gamma \cdot [SOC(t) - SOC_{tar}]^2 \tag{30}$$

$$L_{FC} = \delta \cdot |\Delta P_{FC} / \Delta P_{FC\text{max}}| \tag{31}$$

where $m_{\text{H}_2}(t)$ is the mass of hydrogen consumption; SOC_{tar} is the target *SOC* value, taken as 0.6; ΔP_{FC} is the power variation of PEMFC; $\Delta P_{FC\text{max}}$ is the maximum power variation of PEMFC, which is taken as 3000 W; and ρ_1 and ρ_2 are the unit price of hydrogen and the battery replacement cost, respectively. The unit price of hydrogen is \$4/kg and the battery replacement price is \$178.41/kg in this paper [9]. The corresponding weights are γ and δ, taken as 1200 and 1, respectively [42].

The architecture of TD3-based strategy is shown as Figure 5, and its pseudocode is enclosed within Table 2. The TD3 system obtains the current state of the FCHEB system and the reward since the last action to output the corresponding action. The vehicle controller unit gets the action output by the TD3 system and controls the vehicle system. The vehicle system enters a new state based on the action and generates the corresponding reward for the action.

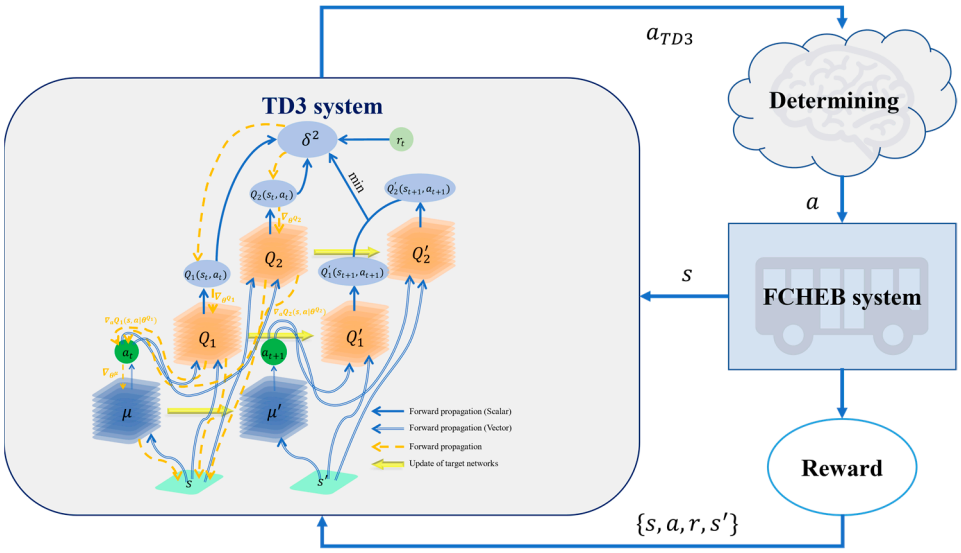

Figure 5. The architecture of TD3-based strategy.

Table 2. Training results of each strategy.

	TD3 Algorithm
1:	Initialize critic networks Q_i and actor network μ with parameters θ^{Q_i} and θ^{μ}
2:	Initialize target networks $\theta^{Q_1'} \leftarrow \theta^{Q_1}$, $\theta^{Q_2'} \leftarrow \theta^{Q_2}$, $\theta^{\mu'} \leftarrow \theta^{\mu}$, Initialize replay buffer β
3:	Initialize learning rate τ, a random process N for action exploration
4:	for episode = 1: M do
5:	Initialize a random noise $\mathcal{N}(0, \lambda^2) \sim \frac{1}{2\lambda}\exp(-\frac{\lvert x\rvert}{\lambda})$ for action exploration
6:	Get initial states: v, acc, SOC, SOH
7:	for $t = 1 : T$ (time step of driving cycle) do
8:	Select action with exploration noise $a = \mu(s_t\lvert\theta^{\mu}) + \mathcal{N}$
9:	Execute action a and observe reward r and new state s_{t+1}
10:	Store transition tuple (s_t, a, r, s_{t+1}) in β
11:	Randomly sample mini-batch of N transitions (s_t, a, r, s_{t+1}) from β
12:	$\hat{a} \leftarrow \mu'(s_{t+1}\lvert\theta^{\mu'}) + \xi$, $\xi \sim \mathrm{clip}(\mathcal{N}(0, \sigma^2), -c, c)$
13:	$y \leftarrow r + \gamma \min_{i=1,2} Q_i'(s_{t+1}, \hat{a}\lvert\theta^{Q_i'})$
14:	Update critics $\theta^{Q_i} \leftarrow \min_{\theta^{Q_i}} \frac{1}{M}\sum (y_{\mathrm{target}}(t) - Q_i(s_t, a_t\lvert\theta^{Q_i}))^2$
15:	if t mod d then
16:	Update μ by the deterministic policy gradient: $\nabla_{\theta^{\mu}} J(\theta^{\mu}) = \frac{1}{M}\sum[\nabla_a Q_1(s_t, a_t\lvert\theta^{Q_1})\nabla_{\theta^{\mu}}\mu(s_t\lvert\theta^{\mu})]$
17:	Update target networks: $\theta^{Q_i'} \leftarrow \tau\theta^{Q_i} + (1-\tau)\theta^{Q_i'}$, $\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$
18:	end if
19:	end for
20:	End

4. Results and Discussion

This section presents a comprehensive comparative analysis of the TD3-based energy management strategy, then discusses and evaluates the simulation result and performance.

4.1. Training Validation Condition

The TD3-based EMS is validated using real-world collected driving conditions as training data. The driving condition data were collected on a typical city bus route in Zhengzhou, China, and can realistically represent the driving information of the vehicle during urban road driving, which is important for the development and evaluation of energy management strategies of FCHEB. The data for real driving conditions vehicle velocity are shown in Figure 6. The mean value of the training velocity driving condition

was 5.8117 m/s with a standard deviation of 5.2190. The mean value of the test velocity of driving conditions was 6.1730 m/s with a standard deviation of 5.2238.

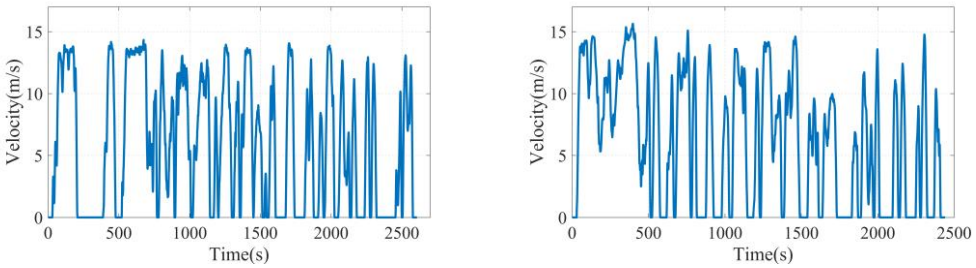

Figure 6. Velocity of working conditions: (**a**) training velocity working conditions; (**b**) testing velocity working conditions.

To fully validate the performance of the proposed strategy, the TD3-based EMS is compared with the DDPG-based EMS and the DDPG-based EMS without considering the battery *SOH* (DDPG-NOSOH-EMS). DDPG-NOSOH-EMS does not consider the battery life constraint in the setting of the reward function. In order to verify the generality of the proposed strategy, the proposed strategy is compared with the DDPG-based EMS and the DDPG-NOSOH EMS under the standard Urban Dynamometer Driving Schedule (UDDS). Figure 7 shows the UDDS conditions. Meanwhile, in order to guarantee the reliability of the training and the credibility of the results, we take the results of four training sessions and take the mean value as the final simulation result. All simulations were performed on a computer with a processor of i5-7300HQ CPU @ 2.50GHz.

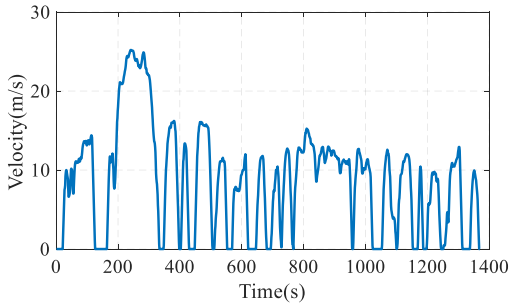

Figure 7. Urban Dynamometer Driving Schedule.

In order to reasonably compare the hydrogen consumption of each EMS, the power difference between the terminal *SOC* and the target *SOC* of each EMS is involved in the calculation of hydrogen consumption based on the hydrogen calorific conversion, considering that the fuel cell provides all the energy:

$$m_{equ} = \frac{\Delta SOC \cdot Q_{bat} \cdot 3600}{\eta_{FC} \cdot H_{heat}} \tag{32}$$

where m_{equ} is the equivalent hydrogen consumption mass which reflects the *SOC* variation; ΔSOC represents the difference between the final *SOC* and the target *SOC*; Q_{bat} is the capacity of the power battery; H_{heat} is the is the heating value of hydrogen, which is taken as 120,000,000 J/kg.

4.2. Training Situation Discussion

The mean reward during training reflects the convergence efficiency and the learning ability of the algorithm. Specifically, the convergence can be judged by the distance between the mean reward of two adjacent episodes. Meanwhile, according to the setting of the

reward function in this study, the closer the mean reward is to zero, the better the strategy is optimized. Figure 8 compares the mean reward of the proposed strategy and the baseline strategies, and Table 3 compares the training effects of these three strategies. The results show that the convergence efficiency of TD3-based EMS is significantly improved, and the training time is reduced by 54.69% and 56.63% compared with DDPG-based EMS and DDPG-NOSOH EMS, respectively. This result is due to the addition of a delayed update mechanism for the target-networks and actor-network, as well as the addition of noise in the target actions in the TD3 algorithm, which improves the stability of the algorithm and increases the convergence efficiency. Since battery aging is not considered in the reward function setting of the DDPG-NOSOH EMS, its mature mean reward is minimal. Meanwhile, because the TD3 algorithm uses Clipped Double Q-learning, it can effectively avoid the problem of overestimation of Q values by the DDPG algorithm, The learning ability of TD3-based EMS is then 36.82% higher than that of DDPG-based EMS.

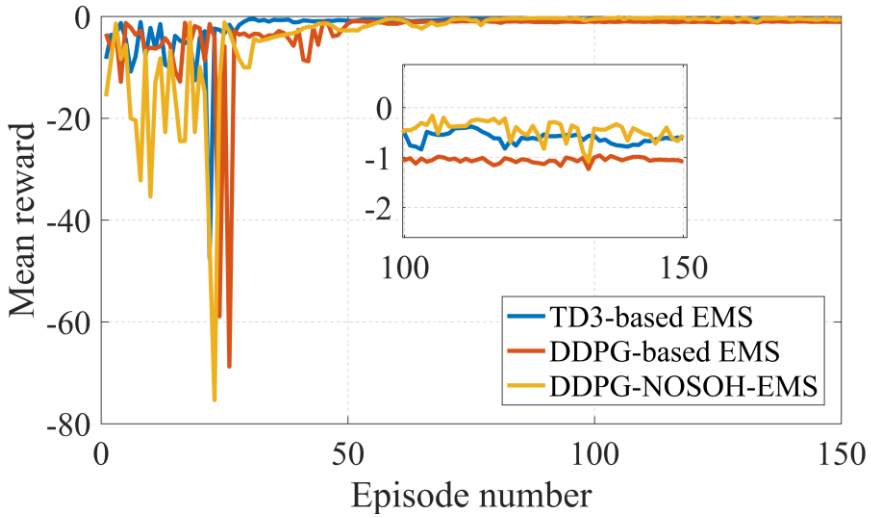

Figure 8. Mean reward of each strategy.

Table 3. Training Results of Each Strategy.

Strategy	Episodes before Convergence	Time Consumption (min)	Mature Mean Reward
TD3-based EMS	30	6.67	−0.6741
DDPG-based EMS	51	14.72	−1.0670
DDPG-NOSOH EMS	53	15.38	−0.6476

4.3. Training Situation Discussion

Figure 9a shows the *SOC* trajectory of each strategy. The *SOC* of each strategy fluctuates around the target value, indicating that the strategies are able to strictly enforce the constraint of *SOC* maintenance in the reward function. Figure 9b–d illustrate the PEMFC power curves for each strategy. The power curves of the TD3-based EMS and DDPG-NOSOH EMS showed less fluctuation, while the *SOC* curves showed more fluctuation, indicating their tendency to adjust the battery output to meet the vehicle power variation. The power curve of the DDPG-based EMS is more volatile, while the *SOC* curve is less volatile, indicating that it prefers to adjust the PEMFC output power to satisfy the variation of vehicle power. Additionally, the power variation rate of the fuel cell for each strategy is less than 3 kW, which satisfies the limit on power variation in the reward function.

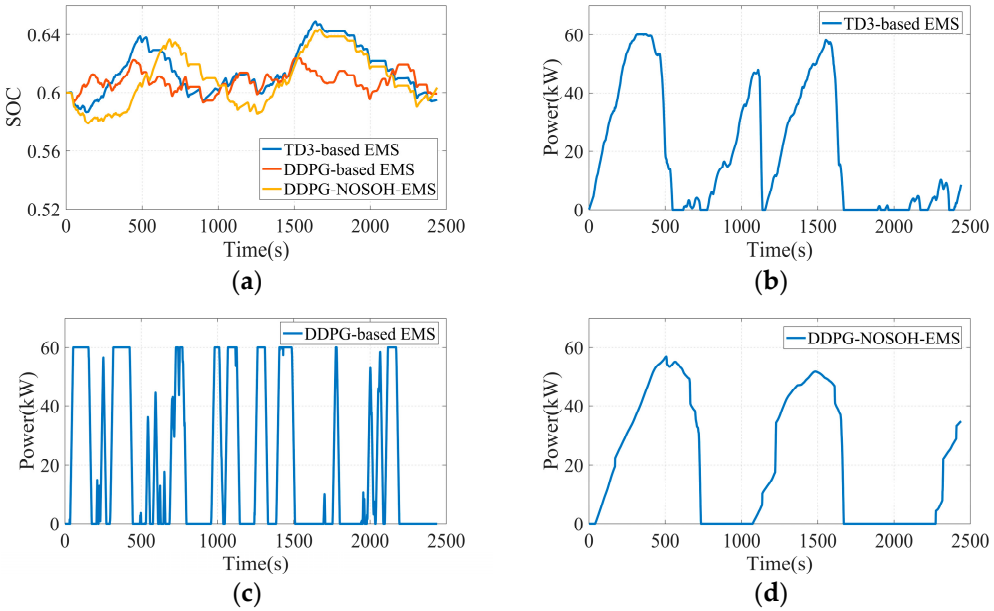

Figure 9. The results of three strategies: (**a**) the *SOC* trajectory of each strategy; (**b**) the PEMFC power of TD3-based EMS; (**c**) the PEMFC power of DDPG-based EMS; (**d**) the PEMFC power of DDPG-NOSOH EMS.

Figure 10 shows the *SOH* curves for each strategy. The TD3-based EMS achieves the least capacity degradation, while the DDPG-NOSOH-based EMS has the highest capacity degradation. These points illustrate the importance of adding a battery life constraint to the reward function.

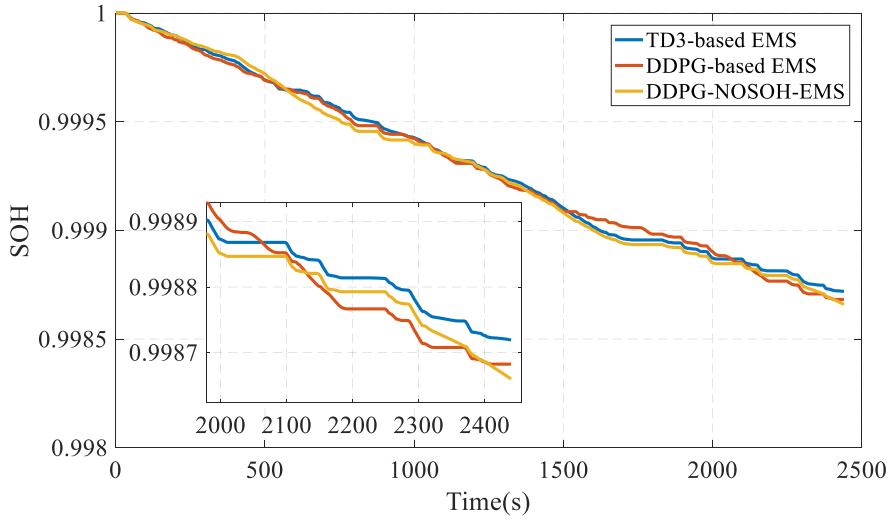

Figure 10. *SOH* curves of each strategy.

4.4. Verification of Total Cost Optimization

The operating costs for each strategy under a Zhengzhou driving working condition cycle are shown in shown in Table 4. The total driving costs include hydrogen consumption cost and battery degradation cost. The hydrogen consumption cost takes into account the equivalent hydrogen consumption caused by the different terminal *SOC* of each strategy. Considering the table data, it can be concluded that for each strategy, the cost of battery aging is over 75% of the total cost. The results show that, compared with the DDPG-based EMS and the DDPG-NOSOH-based EMS, the TD3-based EMS indicated a 1.40% and 4.88% decrease in hydrogen consumption cost, a 2.91% and 4.51% decrease in battery deterioration

cost, and a 2.45% and 4.60% decrease in total cost, respectively. Table 5 shows the operating costs for each strategy under a UDDS cycle, where the hydrogen consumption costs are corrected. The comparison results show that the proposed strategy is optimized in terms of hydrogen consumption cost and battery deterioration cost.

Table 4. Costs under Zhengzhou working conditions.

Strategy	Hydrogen Consumption Cost ($)	Battery Deterioration Cost ($)	Total Cost ($)
TD3-based EMS	3.51	10.81	14.32
DDPG-based EMS	3.56	11.12	14.68
DDPG-NOSOH EMS	3.69	11.32	15.01

Table 5. Costs under UDDS working conditions.

Strategy	Hydrogen Consumption Cost ($)	Battery Deterioration Cost ($)	Total Cost ($)
TD3-based EMS	3.54	9.16	12.70
DDPG-based EMS	3.56	9.70	13.26
DDPG-NOSOH EMS	3.62	9.95	13.57

5. Conclusions

In this paper, a fuel cell hybrid electric bus energy management strategy based on TD3 algorithm is proposed to optimize the driving cost of fuel cell hybrid electric buses. Using the real collection data of Zhengzhou buses as training data, the effectiveness of the TD3-based EMS is verified by comparing it with other strategies in terms of hydrogen consumption costs, battery degradation costs, and total cost. The major conclusions of this paper are as follows:

(1) The cost of battery degradation accounts for a large proportion of the vehicle operating cost, and it is undesirable to neglect the battery degradation.

(2) By adding new mechanisms to the DDPG algorithm, the training effect of TD3-based EMS is enhanced. Compared with the DDPG-based EMS using the same reward function, the training efficiency and learning ability of the TD3-based EMS are 54.69% and 36.82% higher, respectively.

(3) Under Zhengzhou bus driving conditions, the overall cost of TD3-based EMS was decreased by 2.45% and 4.60% compared to DDPG-based EMS and DDPG-NOSOH EMS, respectively.

In the future, we will consider integrating a cyber physical system (CPS) into the EMS framework when designing an energy management strategy. Information on vehicle operation is obtained through cyber physical systems to optimize the controller's control decisions. In addition, in terms of overall vehicle operating costs, an analysis of fuel cell aging costs will be added. Meanwhile, we will actively focus on the latest deep reinforcement learning algorithms for their application in the field of energy management.

Author Contributions: Conceptualization, H.H.; Data curation, C.J.; Formal analysis, K.L.; Investigation, K.L., C.J. and X.H.; Methodology, K.L.; Software, K.L. and C.J.; Supervision, X.H. and H.H.; Visualization, K.L.; Writing—original draft, K.L.; Writing—review & editing, K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was founded by the National Natural Science Foundation of China [U1864202].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. İnci, M.; Büyük, M.; Demir, M.H.; İlbey, G. A review and research on fuel cell electric vehicles: Topologies, power electronic converters, energy management methods, technical challenges, marketing and future aspects. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110648. [CrossRef]
2. Zhou, J.; Liu, J.; Su, Q.; Feng, C.; Wang, X.; Hu, D.; Yi, F.; Jia, C.; Fan, Z.; Jiang, S. Heat Dissipation Enhancement Structure Design of Two-Stage Electric Air Compressor for Fuel Cell Vehicles Considering Efficiency Improvement. *Sustainability* **2022**, *14*, 7259. [CrossRef]
3. Cheng, S.; Hu, D.; Hao, D.; Yang, Q.; Wang, J.; Feng, L.; Li, J. Investigation and analysis of proton exchange membrane fuel cell dynamic response characteristics on hydrogen consumption of fuel cell vehicle. *Int. J. Hydrogen Energy* **2022**, *47*, 15845–15864. [CrossRef]
4. Zhang, K.; Liang, X.; Wang, L.; Sun, K.; Wang, Y.; Xie, Z.; Wu, Q.; Bai, X.; Hamdy, M.S.; Chen, H. Status and perspectives of key materials for PEM electrolyzer. *Nano Res. Energy* **2022**, *1*, e9120032. [CrossRef]
5. Xu, J.; Zhang, C.; Wan, Z.; Chen, X.; Chan, S.H.; Tu, Z. Progress and perspectives of integrated thermal management systems in PEM fuel cell vehicles: A review. *Renew. Sustain. Energy Rev.* **2022**, *155*, 111908. [CrossRef]
6. Sun, J.; Ye, L.; Zhao, X.; Zhang, P.; Yang, J. Electronic Modulation and Structural Engineering of Carbon-Based Anodes for Low-Temperature Lithium-Ion Batteries: A Review. *Molecules* **2023**, *28*, 2108. [CrossRef] [PubMed]
7. Hu, D.; Liu, J.; Yi, F.; Yang, Q.; Zhou, J. Enhancing heat dissipation to improve efficiency of two-stage electric air compressor for fuel cell vehicle. *Energy Convers. Manag.* **2022**, *251*, 115007. [CrossRef]
8. Yi, F.; Su, Q.; Feng, C.; Wang, X.; Yang, L.; Zhou, J.; Fan, Z.; Jiang, S.; Zhang, Z.; Yu, T.; et al. Response analysis and stator optimization of ultra-high-speed PMSM for fuel cell electric air compressor. *IEEE Trans. Transp. Electrif.* **2022**. [CrossRef]
9. Jia, C.; Zhou, J.; He, H.; Li, J.; Wei, Z.; Li, K.; Shi, M. A novel energy management strategy for hybrid electric bus with fuel cell health and battery thermal-and health-constrained awareness. *Energy* **2023**, *271*, 127105. [CrossRef]
10. Zhao, X.; Wang, L.; Zhou, Y.; Pan, B.; Wang, R.; Wang, L.; Yan, X. Energy management strategies for fuel cell hybrid electric vehicles: Classification, comparison, and outlook. *Energy Convers. Manag.* **2022**, *270*, 116179. [CrossRef]
11. He, H.; Jia, C.; Li, J. A new cost-minimizing power-allocating strategy for the hybrid electric bus with fuel cell/battery health-aware control. *Int. J. Hydrogen Energy* **2022**, *47*, 22147–22164. [CrossRef]
12. Zheng, C.; Zhang, D.; Xiao, Y.; Li, W. Reinforcement learning-based energy management strategies of fuel cell hybrid vehicles with multi-objective control. *J. Power Sources* **2022**, *543*, 231841. [CrossRef]
13. 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* **2016**, *185 Pt 2*, 1633–1643. [CrossRef]
14. Dong, P.; Zhao, J.; Liu, X.; Wu, J.; Xu, X.; Liu, Y.; Wang, S.; Guo, W. Practical application of energy management strategy for hybrid electric vehicles based on intelligent and connected technologies: Development stages, challenges, and future trends. *Renew. Sustain. Energy Rev.* **2022**, *170*, 112947. [CrossRef]
15. Lu, D.; Hu, D.; Yi, F.; Li, J.; Yang, Q. Optimal selection range of FCV power battery capacity considering the synergistic decay of dual power source lifespan. *Int. J. Hydrogen Energy* **2023**, *48*, 13578–13590. [CrossRef]
16. Zhang, H.; Li, X.; Liu, X.; Yan, J. Enhancing fuel cell durability for fuel cell plug-in hybrid electric vehicles through strategic power management. *Appl. Energy* **2019**, *241*, 483–490. [CrossRef]
17. Ravey, A.; Blunier, B.; Miraoui, A. Control strategies for fuel-cell-based hybrid electric vehicles: From offline to online and experimental results. *IEEE Trans. Veh. Technol.* **2012**, *61*, 2452–2457. [CrossRef]
18. Kandidayeni, M.; Trovão, J.; Soleymani, M.; Boulon, L. Towards health-aware energy management strategies in fuel cell hybrid electric vehicles: A review. *Int. J. Hydrogen Energy* **2022**, *47*, 10021–10043. [CrossRef]
19. Li, C.; Hu, G.; Zhu, Z.; Wang, X.; Jiang, W. Adaptive equivalent consumption minimization strategy and its fast implementation of energy management for fuel cell electric vehicles. *Int. J. Energy Res.* **2022**, *46*, 16005–16018. [CrossRef]
20. Lin, X.; Xu, X.; Lin, H. Predictive-ECMS based degradation protective control strategy for a fuel cell hybrid electric vehicle considering uphill condition. *ETransportation* **2022**, *12*, 100168. [CrossRef]
21. Xu, N.; Kong, Y.; Yan, J.; Zhang, Y.; Sui, Y.; Ju, H.; Liu, H.; Xu, Z. Global optimization energy management for multi-energy source vehicles based on "Information layer-Physical layer-Energy layer-Dynamic programming"(IPE-DP). *Appl. Energy* **2022**, *312*, 118668. [CrossRef]
22. Yi, F.; Lu, D.; Wang, X.; Pan, C.; Tao, Y.; Zhou, J.; Zhao, C. Energy management strategy for hybrid energy storage electric vehicles based on pontryagin's minimum principle considering battery degradation. *Sustainability* **2022**, *14*, 1214. [CrossRef]
23. Xu, L.; Ouyang, M.; Li, J.; Yang, F. Dynamic programming algorithm for minimizing operating cost of a PEM fuel cell vehicle. In Proceedings of the 2012 IEEE International Symposium on Industrial Electronics, Hangzhou, China, 28–31 May 2012; IEEE: Washington, DC, USA, 2012; pp. 1490–1495.
24. Bao, S.; Sun, P.; Zhu, J.; Ji, Q.; Liu, J. Improved multi-dimensional dynamic programming energy management strategy for a vehicle power-split hybrid powertrain. *Energy* **2022**, *256*, 124682. [CrossRef]
25. Hu, D.; Cheng, S.; Zhou, J.; Hu, L. Energy Management Optimization Method of Plug-In Hybrid-Electric Bus Based on Incremental Learning. *IEEE J. Emerg. Sel. Top. Power Electron.* **2023**, *11*, 7–18. [CrossRef]

26. Zeng, T.; Zhang, C.; Zhang, Y.; Deng, C.; Hao, D.; Zhu, Z.; Ran, H.; Cao, D. Optimization-oriented adaptive equivalent consumption minimization strategy based on short-term demand power prediction for fuel cell hybrid vehicle. *Energy* **2021**, *227*, 120305. [CrossRef]
27. Chen, H.; Chen, J.; Lu, H.; Yan, C.; Liu, Z. A modified MPC-based optimal strategy of power management for fuel cell hybrid vehicles. *IEEE/ASME Trans. Mechatron.* **2020**, *25*, 2009–2018. [CrossRef]
28. Li, W.; Ye, J.; Cui, Y.; Kim, N.; Cha, S.W.; Zheng, C. A speedy reinforcement learning-based energy management strategy for fuel cell hybrid vehicles considering fuel cell system lifetime. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2022**, *9*, 859–872. [CrossRef]
29. Hsu, R.C.; Chen, S.-M.; Chen, W.-Y.; Liu, C.-T. A reinforcement learning based dynamic power management for fuel cell hybrid electric vehicle. In Proceedings of the 2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS), Sapporo, Japan, 25–28 August 2016; IEEE: Washington, DC, USA, 2016; pp. 460–464.
30. Yang, D.; Wang, L.; Yu, K.; Liang, J. A reinforcement learning-based energy management strategy for fuel cell hybrid vehicle considering real-time velocity prediction. *Energy Convers. Manag.* **2022**, *274*, 116453. [CrossRef]
31. Zheng, C.; Li, W.; Li, W.; Xu, K.; Peng, L.; Cha, S.W. A Deep Reinforcement learning-based energy management strategy for fuel cell hybrid buses. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2022**, *9*, 885–897. [CrossRef]
32. Guo, X.; Yan, X.; Chen, Z.; Meng, Z. Research on energy management strategy of heavy-duty fuel cell hybrid vehicles based on dueling-double-deep Q-network. *Energy* **2022**, *260*, 125095. [CrossRef]
33. Huang, Y.; Hu, H.; Tan, J.; Lu, C.; Xuan, D. Deep reinforcement learning based energy management strategy for range extend fuel cell hybrid electric vehicle. *Energy Convers. Manag.* **2023**, *277*, 116678. [CrossRef]
34. Hu, H.; Lu, C.; Tan, J.; Liu, S.; Xuan, D. Effective energy management strategy based on deep reinforcement learning for fuel cell hybrid vehicle considering multiple performance of integrated energy system. *Int. J. Energy Res.* **2022**, *46*, 24254–24272. [CrossRef]
35. Zheng, C.; Li, W.; Xiao, Y.; Zhang, D.; Cha, S.W. A Deep Deterministic Policy Gradient-Based Energy Management Strategy for Fuel Cell Hybrid Vehicles. In Proceedings of the 2021 IEEE Vehicle Power and Propulsion Conference (VPPC), Gijon, Spain, 25–28 October 2021; IEEE: Washington, DC, USA, 2021; pp. 1–6.
36. Lu, H.; Tao, F.; Fu, Z.; Sun, H. Battery-degradation-involved energy management strategy based on deep reinforcement learning for fuel cell/battery/ultracapacitor hybrid electric vehicle. *Electr. Power Syst. Res.* **2023**, *220*, 109235. [CrossRef]
37. Zhou, J.; Feng, C.; Su, Q.; Jiang, S.; Fan, Z.; Ruan, J.; Sun, S.; Hu, L. The Multi-Objective Optimization of Powertrain Design and Energy Management Strategy for Fuel Cell–Battery Electric Vehicle. *Sustainability* **2022**, *14*, 6320. [CrossRef]
38. Zhang, Y.; Zhang, C.; Fan, R.; Huang, S.; Yang, Y.; Xu, Q. Twin delayed deep deterministic policy gradient-based deep reinforcement learning for energy management of fuel cell vehicle integrating durability information of powertrain. *Energy Convers. Manag.* **2022**, *274*, 116454. [CrossRef]
39. Xia, X.; Hashemi, E.; Xiong, L.; Khajepour, A. Autonomous Vehicle Kinematics and Dynamics Synthesis for Sideslip Angle Estimation Based on Consensus Kalman Filter. *IEEE Trans. Control Syst. Technol.* **2022**, *31*, 179–192. [CrossRef]
40. Meng, J.; Ricco, M.; Luo, G.; Swierczynski, M.; Stroe, D.-I.; Stroe, A.-I.; Teodorescu, R. An overview and comparison of online implementable SOC estimation methods for lithium-ion battery. *IEEE Trans. Ind. Appl.* **2017**, *54*, 1583–1591. [CrossRef]
41. Wang, J.; Liu, P.; Hicks-Garner, J.; Sherman, E.; Soukiazian, S.; Verbrugge, M.; Tataria, H.; Musser, J.; Finamore, P. Cycle-life model for graphite-LiFePO4 cells. *J. Power Sources* **2011**, *196*, 3942–3948. [CrossRef]
42. Lian, R.; Peng, J.; Wu, Y.; Tan, H.; Zhang, H. Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle. *Energy* **2020**, *197*, 117297. [CrossRef]