energies

MDPI

Article

Energy Management Strategy Based on Reinforcement Learning and Frequency Decoupling for Fuel Cell Hybrid Powertrain

Hongzhe Li [1], Jinsong Kang [2,*] and Cheng Li [3]

[1] College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China; 2011243@tongji.edu.cn
[2] Institute of Rail Transit, Tongji University, Shanghai 201804, China
[3] CRRC Zhuzhou Electric Locomotive Research Institute Co., Ltd., Zhuzhou 412001, China; licheng5@csrzic.com
* Correspondence: kjs@tongji.edu.cn; Tel.: +86-17721085566

Abstract: This study presents a Two-Layer Deep Deterministic Policy Gradient (TL-DDPG) energy management strategy for Hydrogen fuel cell hybrid train, that aims to solve the problem that traditional reinforcement learning strategies require high initial values and are difficult to optimize global variables. Augmenting the optimization capabilities of the inner layer, a frequency decoupling algorithm integrates into the outer layer, furnishing a fitting initial value for strategy optimization. This addition aims to bolster the stability of fuel cell output, thereby enhancing the overall efficiency of the hybrid power system. In comparison with the traditional reinforcement learning algorithm, the proposed approach demonstrates notable improvements: a reduction in hydrogen consumption per 100 km by 16.3 kg, a 9.7% increase in the output power stability of the fuel cell, and a 1.8% enhancement in its efficiency.

Keywords: energy management strategy (EMS); hybrid electric train; reinforcement learning; Two-Layer Deep Deterministic Policy Gradient (TL-DDPG); frequency decoupling

check for updates

Citation: Li, H.; Kang, J.; Li, C. Energy Management Strategy Based on Reinforcement Learning and Frequency Decoupling for Fuel Cell Hybrid Powertrain. *Energies* **2024**, *17*, 1929. https://doi.org/10.3390/en17081929

Academic Editor: Francesco Calise

Received: 20 February 2024
Revised: 27 March 2024
Accepted: 4 April 2024
Published: 18 April 2024

1. Introduction

The global community is currently grappling with escalating energy crises and heightened concerns over the worsening greenhouse effect. Hydrogen, acknowledged as a clean energy source, holds substantial promise for future development [1]. The application of hydrogen in rail transit is of considerable strategic significance, given its high environmental friendliness [2], and it is indicated by promising prospects in Hydrogen hybrid power system [3].

In contrast to traditional trains relying on internal combustion engines or conventional power sources like pantograph-catenary or third-rail currents, hydrogen fuel cell hybrid trains offer advantages such as heightened environmental friendliness [4], reduced construction costs [5], and enhanced resistance to interference, along with improved compatibility [6]. In a hybrid power system integrating multiple energy sources, the distribution of output power from these sources plays a pivotal role in ensuring the safety of power supply [7], the efficiency of power output, the economy of energy utilization, and the dynamic performance of the train [8]. Consequently, the Energy Management System (EMS) becomes an indispensable component for hydrogen fuel cell hybrid power system trains.Broadly categorized, the EMS of hybrid power systems comprises three main branches: rule-based, optimization-based, and data-based methods [9]. Rule-based EMS entails establishing a set of rules to govern power allocation, relying on engineering expertise or mathematical models [10]. However, this approach heavily depends on historical experience and model accuracy, lacking adaptability in real-world driving scenarios. Conversely, optimization-based methods can be classified into global and instantaneous optimization approaches. Dynamic programming (DP) emerges as a classic numerical

algorithm for achieving global optimization. Additionally, commonly employed strategies encompass Genetic Algorithms (GA) [11], Particle Swarm Optimization (PSO) [12], Pontryagin's Minimum Principle (PMP) [13], and Equivalent Consumption Minimization Strategy (ECMS) [14]. Nevertheless, optimization-based EMS has inherent drawbacks, including high computational demands and challenges in real-time updates, rendering it more suitable for offline planning or integration with other strategies.

The operational parameters of rail transit vehicles exhibit distinctive characteristics. In a single operation, the imperative to accomplish rapid acceleration, deceleration, and stopping within a condensed timeframe results in swift fluctuations in power demand [15]. Furthermore, the fixed running lines and schedules contribute to a high degree of repeatability in running conditions during multiple vehicle operations [16]. Consequently, data-driven methodologies have gained prominence in addressing energy management challenges in rail transport and have garnered considerable attention within the scholarly community [17]. Energy management strategies rooted in data-driven approaches, prominently exemplified by reinforcement learning and its refined algorithms, have witnessed increasing application. Reinforcement learning is delineated by four fundamental elements: State, Action, Policy, and Reward. Throughout the reinforcement learning process, agents engage in continual interaction with the environment, adapting their strategies based on rewards garnered from environmental feedback. Presently, a multitude of studies is dedicated to the energy management of hybrid systems utilizing reinforcement learning techniques [18]. Among these techniques, Q-learning emerges as the foundational and most extensively applied strategy [19]. Leveraging recent advancements in Deep Reinforcement Learning (DRL) [20], researchers have assimilated cutting-edge findings to propose energy management strategies for Hybrid Electric Vehicles (HEVs) based on Deep Q-Network (DQN) for tasks with discrete action spaces [21]. Similarly, approaches founded on Deep Deterministic Policy Gradients (DDPG) have been implemented for tasks with continuous action spaces, showcasing effective policy behavior [22].

However, these reinforcement learning algorithms confront specific challenges. Firstly, optimization results are significantly influenced by the specified initial value [23]. Inadequate specification may lead to slow convergence, diminished accuracy, and assigning a random initial value can impede the scalability of the model. Secondly, the value function Q is updated solely based on variables from adjacent time steps, such as the instantaneous hydrogen consumption of the system and the instantaneous speed of the train [24]. This limitation hinders the consideration of variables obtainable only after completing the entire working condition, such as the smoothness of the fuel cell's output power and the braking energy recovery efficiency of the lithium battery.To address the impact of the initial value on results, a method proposed in [25] employs the frequency decoupling approach. Initially, the total demand power is allocated, with the low-frequency part assigned to the fuel cell and the high-frequency part allocated to the storage battery. While this method effectively enhances the stability of fuel cell output, the manual specification of the frequency threshold remains a challenge [26]. In an alternative approach, Ref. [27] integrates expert experience to guide the selection of the training initial value. However, reliance solely on expert experience may not consistently prove effective. To surmount the limitation of single-step variables, Ref. [28] introduces a sliding time window, enabling a single training session to learn not only the state variable of a single step but also information from the entire window. Building on this, Ref. [29] enhances the method by utilizing a neural network to integrate information from the entire window for the agent to learn. Nonetheless, the sliding time window may encounter difficulty covering a broad range, and the selection of its length can introduce truncation effects on the data [30].

In order to solve the problem of initial value dependence of reinforcement learning and difficult global optimization of global variables for energy management strategies, an enhanced energy management framework incorporating the frequency decoupling algorithm into the Deep Deterministic Policy Gradient (DDPG) is proposed. Initially, a decoupling algorithm based on the frequency of a low-pass filter is implemented to decompose the

power demand signal of the train. Subsequently, a Two-Layer Deep Deterministic Policy Gradient (TL-DDPG) is employed to guide the power allocation of the fuel cell and battery. In this framework, the outer layer network integrates the frequency decoupling algorithm to provide an initial value for the optimization in the inner layer. Detailed elucidation of this approach will be provided in Section 3 of this paper. Test results under various operating conditions demonstrate that the proposed algorithm effectively reduces costs, enhances fuel cell efficiency and stability, and achieves superior braking energy recovery. Diverging from prior works, this paper comprehensively considers the attributes of the data-driven method, frequency decoupling, and rail transit operation. It deliberates on the energy management strategy from a data-driven perspective. The primary contributions of this work are summarized below:

(I) The impact of filter parameters on power flow in the dynamic system is unveiled through frequency decoupling, and optimal filter parameters are determined using the reinforcement learning method. This leads to enhanced operational efficiency, improved output power stability of the fuel cell, and the attainment of superior economic and energy-saving benefits.

(II) An two-layer reinforcement learning optimization framework is established for the iterative optimization of both single-step variables and global variables. This approach addresses the challenge of the reinforcement learning model struggling to assimilate all information from the data. Furthermore, the initial value for power distribution is acquired through frequency decoupling, presenting an intuitive relationship and a favorable trade-off between fuel cell and battery costs.

(III) Tailored to rail transit operational scenarios, the proposed methodology conducts the model training process on a local server rather than real-time training on the train controller. This approach alleviates the burden on computing resources and exhibits a favorable power distribution effect for typical situations characterized by fixed running tracks.

The remainder of this paper proceeds as follows. In Section 2, the train hybrid system modeling is detailed. Section 3 proposes an adaptive EMS combining frequency decoupling and two-layer DDPG. In Section 4, the simulation results of the proposed strategy are provided and analyzed, and the conclusions are given in Section 5.

2. Hybrid System Train Modeling

This article conducts research on a fuel cell hybrid train depicted in Figure 1. The train configuration comprises a fuel cell system, a lithium-ion battery pack, two DC-DC converters, and a traction system involving a DC-AC inverter and traction motor. Figure 1 also shows the input and output of the energy management module, as well as the output characteristics of key components of the hybrid system, including fuel cells, batteries, and traction motors. The model is established using MATLAB 2022a, and the intricate parameters of the system are outlined in Table 1, whose data is from IEEE VTS Motor Vehicles Challenge 2019 [31]. We have token the locomotive and line data from the race, and reconstructed their powertrain models.

Table 1. Detailed Parameters of the System.

Locomotive	
Mass	140 t
Number of traction drive	4
Efficiency of the electric drives (considered as constant)	85%
Gearbox ratio	4.14
Gearbox efficiency	95%
Diameter of a wheel	0.92

Table 1. *Cont.*

Fuel Cell	
Type of fuel cell	PEMFC
Number of cells in series	350
Number of modules in parallel	2
Voltage range of a cell	0.3–0.75 V
Rated power of the fuel cell system	400 kW
Battery	
Type of battery	$LiFePO_4$
Total number of cells	345
Rated voltage of a cell	3.8 V
Minimal voltage of a cell (charge)	4.0 V
Maximal voltage of a cell (discharge)	2.8 V
Minimal state of charge	20%
Maximal current of a cell	2 C = 320 A
Minimal current of a cell	−0.5 C = −80 A

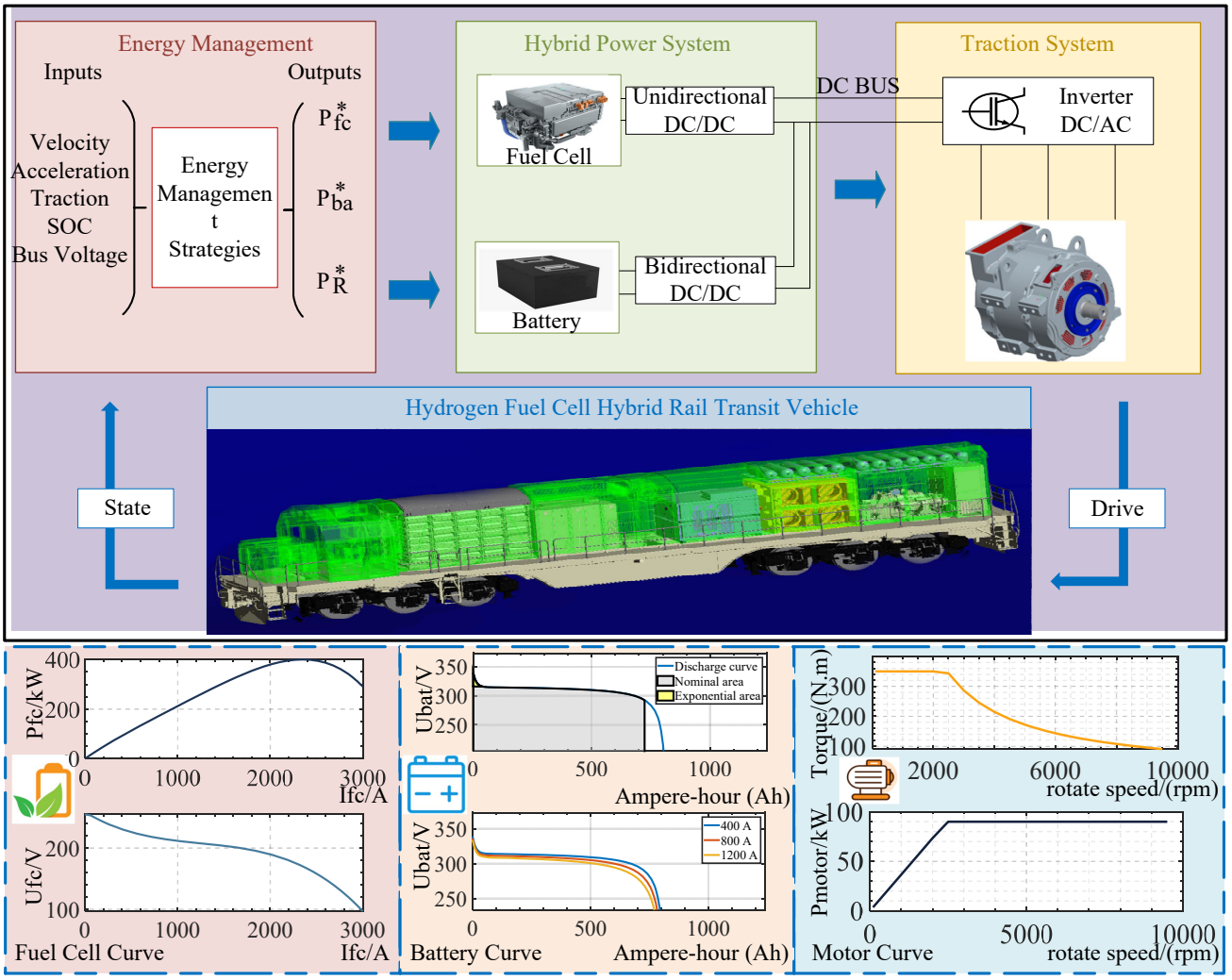


Figure 1. Hybrid system train modeling and the corresponding component characteristics.

(a) Train Modeling: Assuming the train drives on a flat road with only the longitudinal dynamic model considered, the force on the train can be expressed as Equation (1)

$$\begin{cases} F_i = \delta ma \\ F_a = 1/2 \cdot \rho \cdot A \cdot C_d \cdot v^2 \\ F_f = mgf \end{cases} \tag{1}$$

where F_i is the inertial force, F_a is the aerodynamic drag, F_f is the rolling resistance, δ is the correction coefficient of rotating mass, m is the mass of train, a is the acceleration, ρ is the air density, C_d is the aerodynamic coefficient, A is the fronted area, g is the gravity coefficient, and f is the rolling resistance coefficient. Then, the power demand of the train can be calculate by Equation (2).

$$P_{\text{req}} = \left(F_i + F_a + F_f\right)v \tag{2}$$

The purpose of EMS is to distribute the power to different energy source, as shown in Equation (3)

$$P_{\text{req}} = (P_{\text{fc}}\eta_{fc-dcdc} + P_{\text{bat}}\eta_{bat-dcdc})\eta_T \tag{3}$$

where P_{fc} is power of fuel cell, P_{bat} is power of battery, $\eta_{fc-dcdc}$ is the efficiency of the fuel cell converter, $\eta_{bat-dcdc}$ is the efficiency of the battery converter, and η_T is the transmission efficiency.

(b) Fuel Cell Modeling: The fuel cell can be equivalent to a controlled source connected to a fixed resistor. Specifically, the output voltage loss of fuel cell includes activation voltage loss, ohmic voltage loss and concentration voltage loss. At the same time, when the control quantity changes, the whole curve will have a certain delay characteristic [32]. The output voltage of the fuel cell is expressed as Equations (4) and (5)

$$V_{fc} = E - R_{ohm} \cdot i_{fc} \tag{4}$$

$$E = E_{oc} - NA\ln(\frac{i_{fc}}{i_0}) \cdot \frac{1}{sT_d\ /\ 3 + 1} \tag{5}$$

where E_{oc} is the open circuit voltage, N is the number of fuel cell monomer, A is tafel slope, i_0 is the exchange current, T_d is dynamic response time, $\frac{1}{sT_d/3+1}$ represents a delay, and s is the symbol of transfer function. R_{ohm} is internal resistance, i_{fc} is fuel cell current, and V_{fc} is fuel cell voltage. The characteristic curve is shown in the lower left of Figure 1.

(c) Battery Modeling: The model of the battery established in this paper mainly focuses on the change of its SOC and output characteristics. SOC in the battery can be calculated by [33]

$$\dot{\text{SOC}} = -\frac{I_{\text{bat}}(t)}{Q_{\text{bat}}} \tag{6}$$

where $I_{\text{bat}}(t)$ is the battery current and Q_{bat} is the battery nominal capacity. Furthermore, the Equation can be written as

$$\text{SOC} = -\frac{V_{\text{oc}}(SOC,t) - \sqrt{V_{\text{oc}}^2(SOC,t) - 4R_{\text{int}}(SOC,t)P_{\text{bat}}}}{2R_{\text{int}}(SOC,t)Q_{\text{bat}}} \tag{7}$$

$$V_{bat} = V_{\text{oc}} - R_{\text{int}}i - K\frac{Q_{\text{bat}}}{Q_{\text{bat}} - \int idt}(\int idt + i^*) + Ae^{(-B\int idt)} \tag{8}$$

$$V_{bat} = V_{\text{oc}} - R_{\text{int}}i - K\frac{Q_{\text{bat}}}{\int idt - 0.1Q_{\text{bat}}}i^* - K\frac{Q_{\text{bat}}}{Q_{\text{bat}} - \int idt} + Ae^{(-B\int idt)} \tag{9}$$

when voltage dynamics are neglected and the battery circuit do not exist RC branches. In Equation (7), P_{bat} is the battery output power, V_{oc} is the battery open circuit voltage, and R_{int} is the battery internal resistance. Note that the values of V_{oc} and R_{int} vary with SOC.

The output voltage of the battery is expressed as Equations (8) and (9), where A is the amplitude of the exponential region, B is the inverse amplitude of the exponential region, K is polarization constant, and i^* is the battery filtration current. Note that Equation (8) represents discharging state, and (9) represents charging state. The characteristic curve is shown in the lower middle of Figure 1.

(d) Traction Motor Modeling: For traction motor, this study focuses on its external characteristics. For traction motor, this study focuses on its external characteristics. The traction force supplied by the motor can be expressed by the Equation (10)

$$F = (T \times \eta)/R \tag{10}$$

where F is traction force, R is the radius of wheel and η is motor efficiency which is shown in Figure 2.

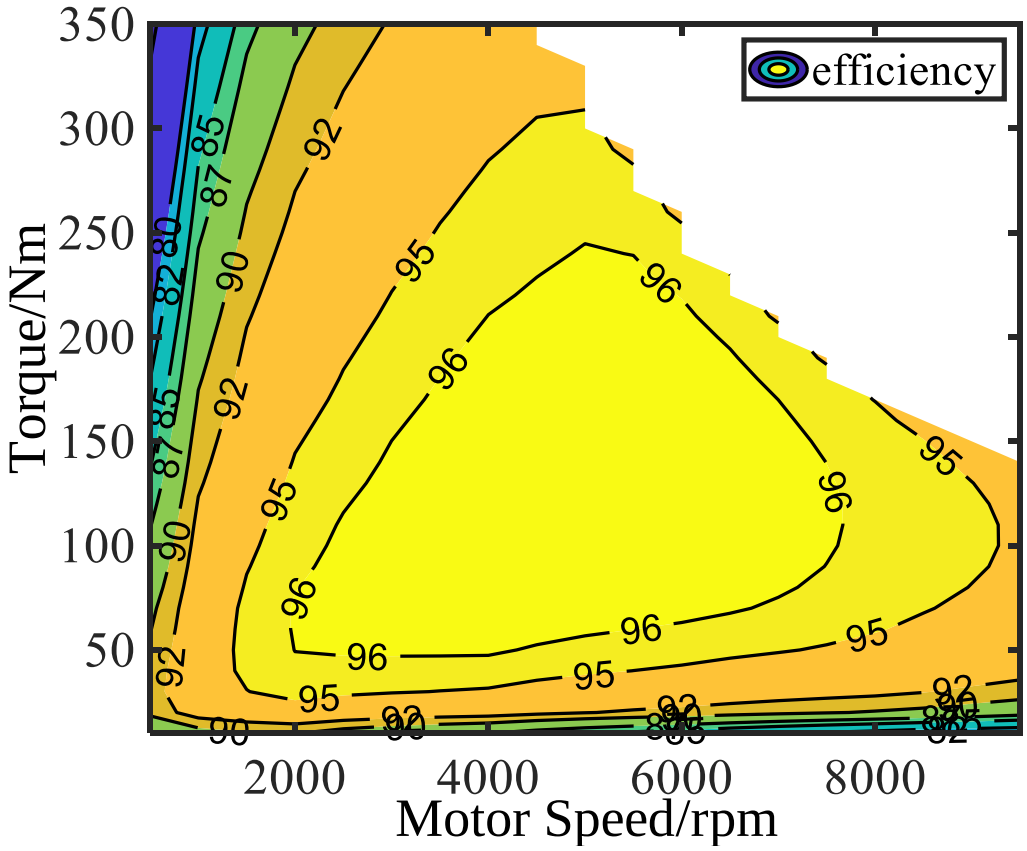


Figure 2. Motor efficiency diagram.

The power demand of motor is shown in the Equation (11):

$$P = (T \times n \times \eta)/9550 \tag{11}$$

where n is the motor speed, which positively correlated the vehicle running speed. In this equation, power is measured in kilowatts. The characteristic curve is shown in the lower left of Figure 1.

3. Adaptive Energy Management Strategy Combining Frequency Decoupling and Data-Driven Deep Reinforcement Learning

In this work, one of the methods suitable for fuel cells hybrid power system, namely, frequency decoupling, will be combined with DDPG, which is a state-of-the-art data-driven RL algorithm. Note that similar structures could also be applied to other methods and RL frameworks. The key ideas of the proposed strategy are described as follows.

3.1. Frequency Decoupling

To fulfill the requisite smoothness criteria for the output power in fuel cells, the energy management strategy based on frequency decoupling has been introduced and widely adopted in fuel cell hybrid systems [34]. The fundamental concept behind frequency decoupling is treating the power demand signal as a low-frequency signal combined with a high-frequency signal, with the two signals separated through signal processing. The isolated low-frequency signal is considered as the output power of the hydrogen fuel cell, while the high-frequency signal is deemed as the output power of the lithium battery.

The critical aspect of frequency decoupling involves the selection of the filtering algorithm and the determination of the frequency threshold. Currently employed filtering methods include Fourier filter, wavelet transform, Kalman filter, and others. The choice of the frequency threshold predominantly relies on expert experience, necessitating the integration of the frequency decoupling strategy with other approaches. The implementation workflow of the frequency decoupling strategy is illustrated in Figure 3.

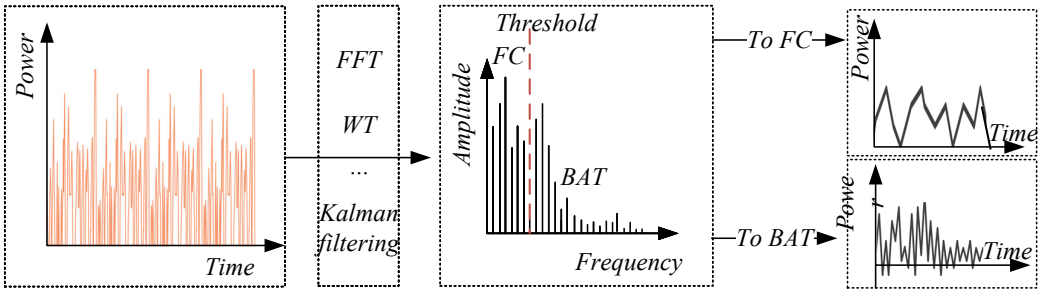


Figure 3. Diagram of the framework of the frequency decoupling.

3.2. Deep Deterministic Policy Gradient

DDPG is a model-free, off-policy reinforcement learning algorithm designed for learning policies θ in high-dimensional continuous action spaces. It stems from the DPG algorithm and incorporates a deep function approximator, hence earning its name DDPG [35]. The effectiveness of DDPG relies on two key techniques. Firstly, experience replay enables the algorithm to learn from a set of unrelated content. Secondly, akin to the hard fixed Q-target network utilized in DQN, DDPG employs a "soft" target update for the actor-critic. This technique enhances training stability as the evaluation network $(\theta^{\mu'}, \theta^{Q'})$ updates more swiftly than the target network $(\theta^{\mu}, \theta^{Q})$. The architecture of DDPG is depicted in Figure 4.

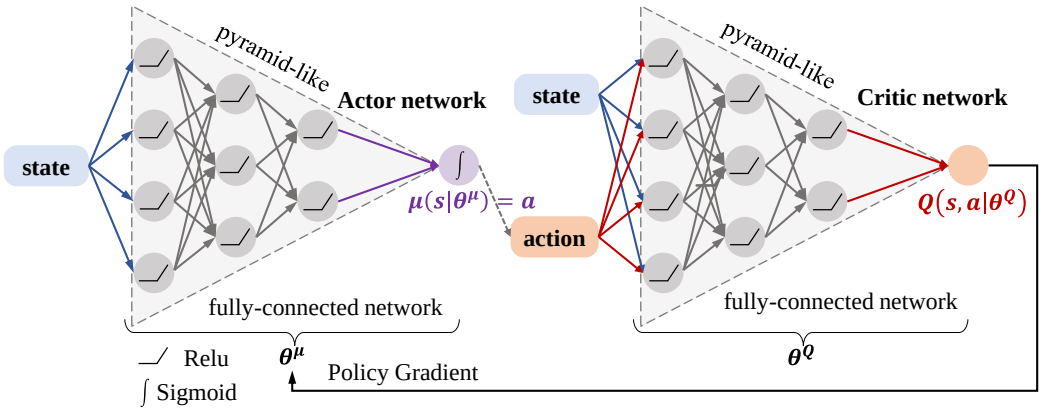


Figure 4. Framework of DDPG.

3.3. Our Energy Management Strategy for Hydrogen Fuel Cell Hybrid Train

In this work, There are two DDPG frameworks are nested to achieve energy management. The one in outer layer is combined with the frequency decoupling algorithm to

determine the threshold of frequency division by optimizing long-term variables (such as the degree of fuel cell power fluctuation and average efficiency), so as to initially allocate energy. The DDPG in the inner layer is further optimized by the instantaneous variables (such as instantaneous speed, SOC, etc.) based on the initial value given after frequency decoupling. Through the method proposed in this paper, we not only focus on long-term variables, but also have a good specification of the initial value for training. The block diagram of the method is shown in Figure 5, and more detailed discussion is given below.

$$Cost_{inner} = K_{H_2} m(H_2) + K_{FC_{open}} num(FC_{start}) + \int (K_{FC_{run}} P_{fc}(t) + K_{BAT_{run}} P_{bat}(t)) dt \quad (12)$$

$$r_{inner} = -(K_{cost_{inner}} Cost_{inner} + K_{soc} |(SOC_{now} - SOC_{init})|) \quad (13)$$

(a) Short-time optimization layer: In the inner DDPG, the state variable s_t comprises the State of Charge (SOC), the speed and acceleration of the train, and the busbar voltage. The control action a corresponds to the output power of the fuel cell. The immediate cost is defined by Equation (12), encompassing the costs of hydrogen, fuel cells, and lithium batteries. Given that the reward function is the negative counterpart of the cost, $r_{inner} = -cost_{inner}$. Specifically, if the battery's SOC is excessively low, a penalty term is incorporated, and r is defined by Equation (13). Here, K_{H_2} represents the cost of hydrogen (\$/kg), $m(H_2)$ denotes the mass of hydrogen consumed in two adjacent time steps, $K_{FC_{open}}$ represents the cost of activating the fuel cell once (\$), while $K_{FC_{run}}$ and $K_{BAT_{run}}$ signify the operating costs of fuel cells and lithium batteries (\$). The scale factors $K_{cost_{inner}}$ and K_{soc} are set to ensure a fundamental balance between the two components during training.

(b) Long-time optimization layer: In the outer layer, there exist frequency decoupling modules and an additional DDPG framework. The filter employed in the frequency decoupling segment is defined as in Equation (14). The state variables s_T accepted by this DDPG encompass the standard deviation of the fuel cell output power in the last entire operational condition, the average efficiency of the fuel cell, and the difference value between the SOC at the end and the beginning. These variables are notably challenging to measure within a single time step. The threshold f_c for frequency decoupling is determined by training the loss function expressed in Equation (15). The loss function comprises three components representing the smoothness of the fuel cell, the system's efficiency, and the braking energy recovery of the lithium battery. K_a, K_b, and K_c are scale factors. It's important to note that to minimize the impact of sampling frequency on results, this work doesn't calculate the true value of f_c but computes the ratio of f_c and the maximum frequency of the power signal to derive the filtering ratio f. The relationship between the two is expressed in Equation (16), where f_{max} denotes the maximum frequency of power demand. Notably, the value of f ranges between 0 and 1, representing the fraction of low-frequency signals. For instance, with a sampling frequency of 10 Hz, indicating a maximum signal frequency of 5 Hz, if the optimized f is 0.6, it signifies that signals from 0–3 Hz are considered low frequency, while signals above 3 Hz are considered high frequency.

$$H(s) = \frac{1}{\frac{s}{w_c} + 1}, \; w_c = 2\pi * f_c \quad (14)$$

$$Cost_{outer} = K_a Std(P_{fc}) + K_b \Delta SOC + K_c \eta_{fc} \quad (15)$$

$$f = \frac{f_c}{f_{max}} \quad (16)$$

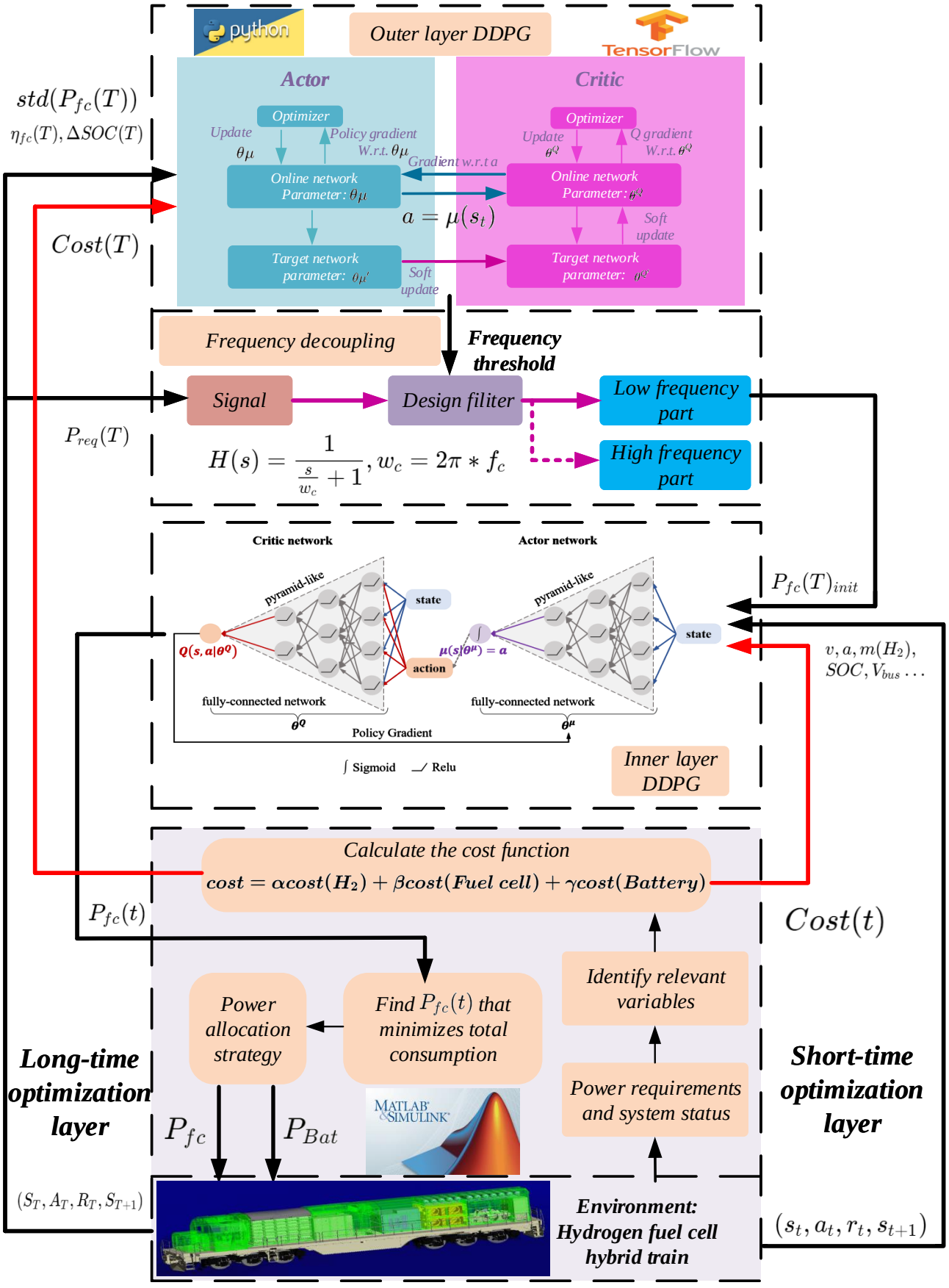


Figure 5. Adaptive hierarchical EMS combining frequency decoupling and two-layer DDPG.

In this study, the optimization of short-term variables is achieved through classical DDPG, which focuses solely on optimizing the energy flow within a single-step time scale in the hybrid system. The algorithm based on frequency decoupling furnishes an initial value for optimization, and the outer DDPG is capable of addressing long-term variables, thereby optimizing the system's performance on a more rational scale. The DDPG networks of both the inner and outer layers are updated according to the following Equations (17) and (18).

$$L = \frac{1}{N}\sum_{i}\left(y_i - Q(s_i, a_i|\theta^Q)\right)^2 \tag{17}$$

$$\nabla_{\theta^\mu} \mathrm{J} \approx \frac{1}{N}\sum_{i} \nabla_{\mathrm{a}} \mathrm{Q}(s, \mathrm{a}|\theta^Q)|_{a=s_i, a=\mu(s_i)} \cdot \nabla_{\theta^\mu} \mathrm{\mu}(\mathrm{s}|\theta^\mu)|_{s_i} \tag{18}$$

4. Simulation Validation and Discussion

This section aims to validate the proposed adaptive hierarchical strategy. Initially, the study explores the impact of the proposed algorithm on hydrogen consumption and State of Charge (SOC) maintenance. Subsequently, to comprehend the influence of different reward formulations on learning efficiency and the final policy, a comparison is made between the proposed strategy and other method. Notably, in contrast to the majority of related studies, the proposed energy management strategy takes into account the output stability of fuel cells. Lastly, the robustness of the method is verified under different driving conditions.

To assess the experimental outcomes, the proposed method undergoes comparison with other benchmark algorithms through simulation experiments conducted under the driving conditions illustrated in Figures 6 and 7 [31]. The proposed algorithm is implemented in Python 3.9, utilizing Tensorflow as the primary machine learning library, while the hybrid system model is constructed using MATLAB R2022a. ONNX is employed for transforming neural network models between Tensorflow and Simulink frameworks. The simulation experiment is conducted on a server equipped with two NVIDIA RTX 3060 graphics cards. The algorithms involved in the comparison encompass rule-based, filter-based, and traditional DDPG, with their profiles and settings detailed in Table 2. The train's running condition spans a 1200 s time series, and training continues until the network converges. After each epoch completion (totaling 120 epochs), relevant results such as the average efficiency of the fuel cell and the standard deviation of the output power are recorded and averaged.

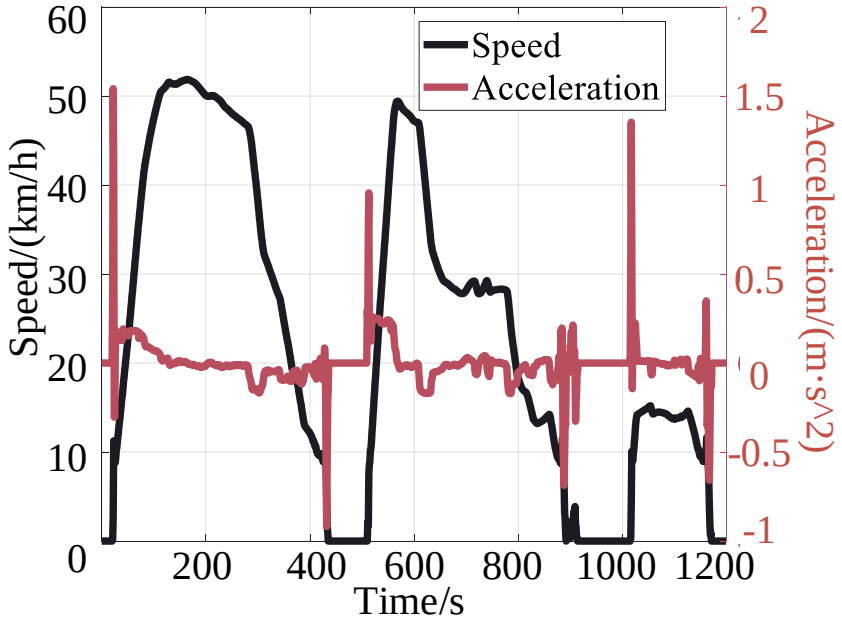


Figure 6. Train driving cycle of our training condition.

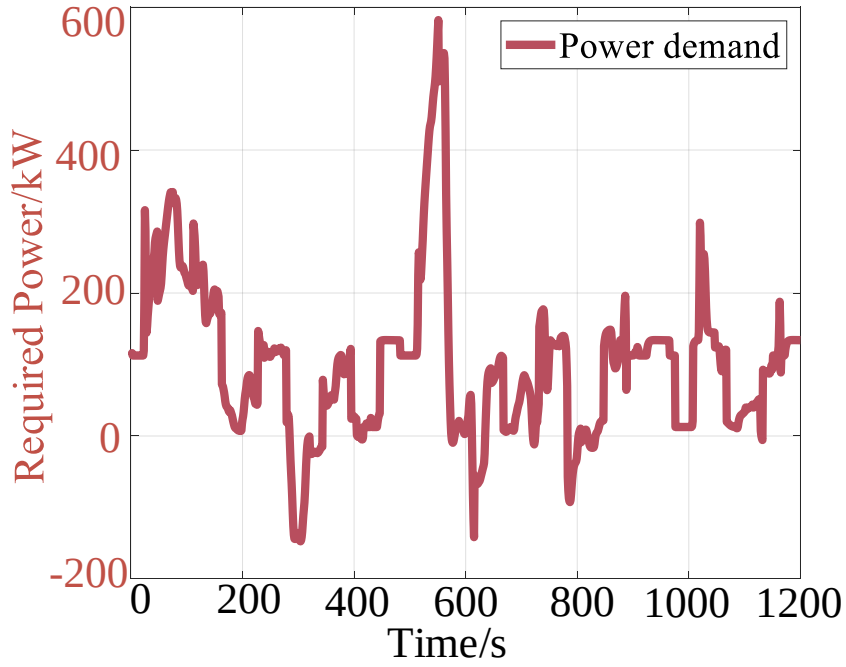


Figure 7. Train power demand of our training condition.

Table 2. Introduction to the reference algorithm.

Algorithm	Brief Introduction	Parameter Set
Rule-based	Design power distribution according to expert experience	$P_{req} < 0$, $P_{bat} = P_{req}$; $SOC < SOC_{min}$ $P_{fc} = \max$; $SOC > SOC_{max}$ $P_{fc} = 0$; else . . .
DDPG	Reinforcement learning	The loss function is shown in Equation (13)
Frequency Decoupling	low frequency to the fuel cell and high frequency to the battery	The filtering algorithm is Fourier transform, filter frequency is shown in Equations (14) and (16)

4.1. Validation of the Proposed Strategy

Figure 8 illustrates the results of the proposed algorithm after 120 training epochs under the specified driving conditions (each condition spanning 1200 s as in Figure 6). The results exhibit convergence tendencies. The loss curves of the inner and outer layers are depicted in Figure 8a, showing similar change trends and confirming the rationality of the loss function settings. At the 100th epoch, the filtering threshold is determined to be 0.376 times the maximum frequency (as depicted in Figure 8b). This implies, under a sampling frequency of 1 Hz, that power changes below 0.188 Hz are considered as filtering threshold for the fuel cell, while the remaining power demand is supplied by the lithium battery. A visual representation of the method is presented in Figure 8c. The time-frequency diagram of the power is obtained by wavelet transform of the required power-time. The brighter the colour, the greater the amplitude of a signal at that frequency at that point. The horizontal lines in the figure are derived from the results in Figure 8b). The section below the horizontal line will be allocated to the fuel cell and the section above will be allocated to the battery.

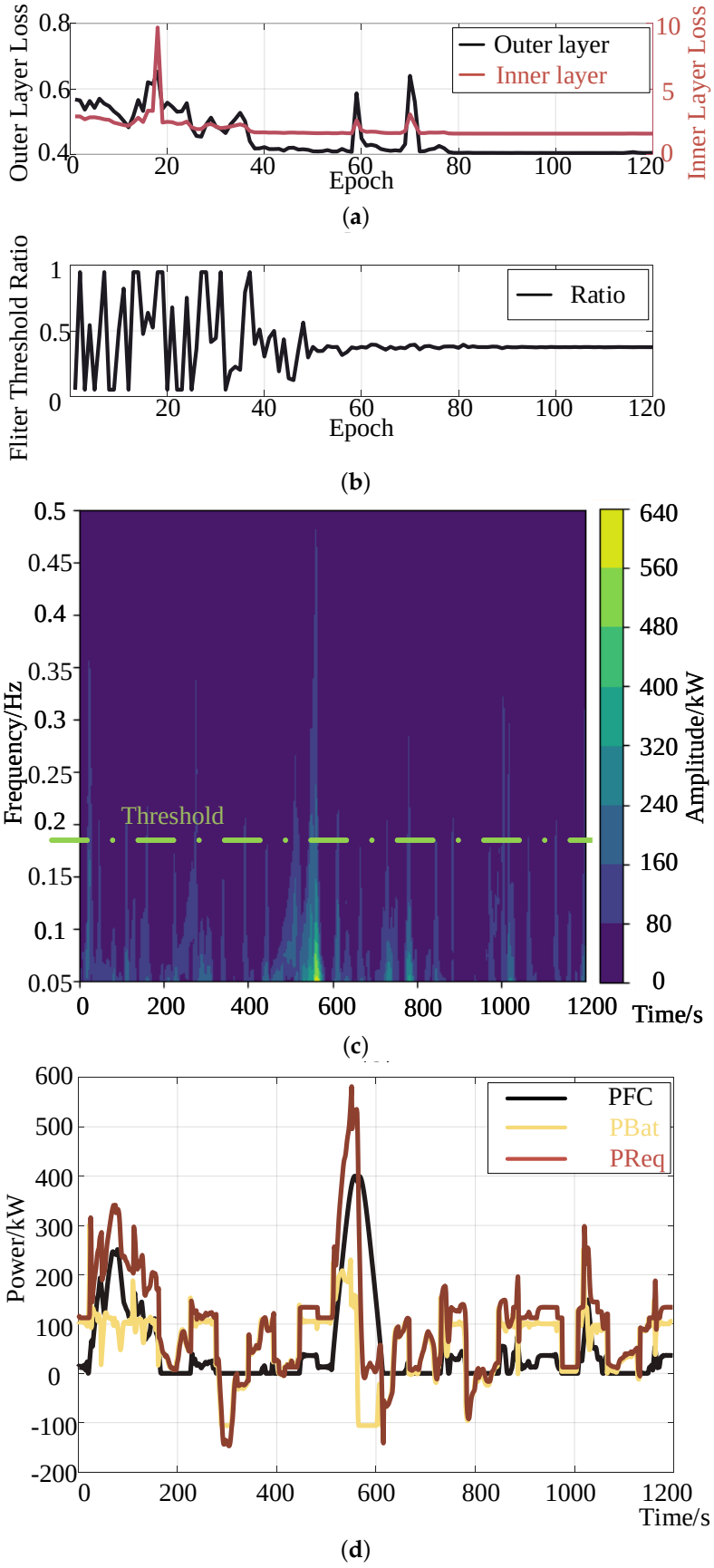


Figure 8. Result of our method. (**a**) Inner and outer layer loss curves. (**b**) Frequency threshold ratio. (**c**) Frequency decoupling visualization. (**d**) Power allocation.

Additionally, Figure 9 showcases the superiority of the proposed strategy across various metrics. Figure 9a displays the output power curves of the fuel cell in the first training epoch for the proposed method, frequency decoupling, and traditional DDPG. It is evident that traditional DDPG generates initial values randomly during training, leading to volatility, longer adjustment times, and difficulty in finding the optimal solution. When using the frequency decoupling method, almost all power output comes from fuel cells, underutilizing lithium batteries. Our method employs the result of frequency decoupling as the initial value, enabling better learning of operating conditions' characteristics. Notably, this represents only the first training session, demonstrating the effectiveness of our method. The final result is depicted in Figure 8d. Figure 9b illustrates the total cost of a single run using different energy management strategies, encompassing fuel cell start and stop costs, hydrogen consumption costs, and storage battery usage costs, as defined by Equation (12). Once training stabilizes, the total cost of our strategy is \$18.86, lower than other strategies. The comparison also reveals that the proposed strategy outperforms traditional DDPG in terms of speed and convergence due to the consideration of initial value optimization. Figure 9c demonstrates the impact of different energy management strategies on the SOC trajectory, with the demand power curve attached at the bottom to illustrate the effect of braking energy recovery for each strategy. Our proposed strategy effectively enables the storage battery to absorb braking energy, showcasing the largest SOC variation range, indicating the strategy's ability to stimulate the potential of lithium batteries while maintaining their performance. Figure 10a,b present the working efficiency of the fuel cell across different strategies, showcasing the advantages of the proposed strategy through two aspects: the average efficiency change of each epoch during training and the proportion of different efficiency intervals after training. The proposed strategy ensures the fuel cell operates mostly in the high-efficiency interval. Finally, Table 3 provides a comprehensive comparison of different strategies. Notably, while the rule-based strategy is relatively close to our strategy in indicators such as hydrogen consumption and SOC, it neglects the switching loss of fuel cells, resulting in a significantly larger final cost compared to other strategies.

Table 3. Comparison between different EMSs.

Algorithm	Fuel Consumption (kg)	Terminal SOC (%)	Average Efficiency of Fuel Cell (%)	Total Cost (\$)	Training Time (s)
Rule-based	2.78	62.40	54.78	56.49	40.12
DDPG	3.84	63.8	53.38	21.45	123.56 + 33.54
Frequency Decoupling	5.02	69.52	51.49	34.29	66.38
Proposed	**2.21**	**60.36**	**55.20**	**18.90**	**206.71 + 45.40**

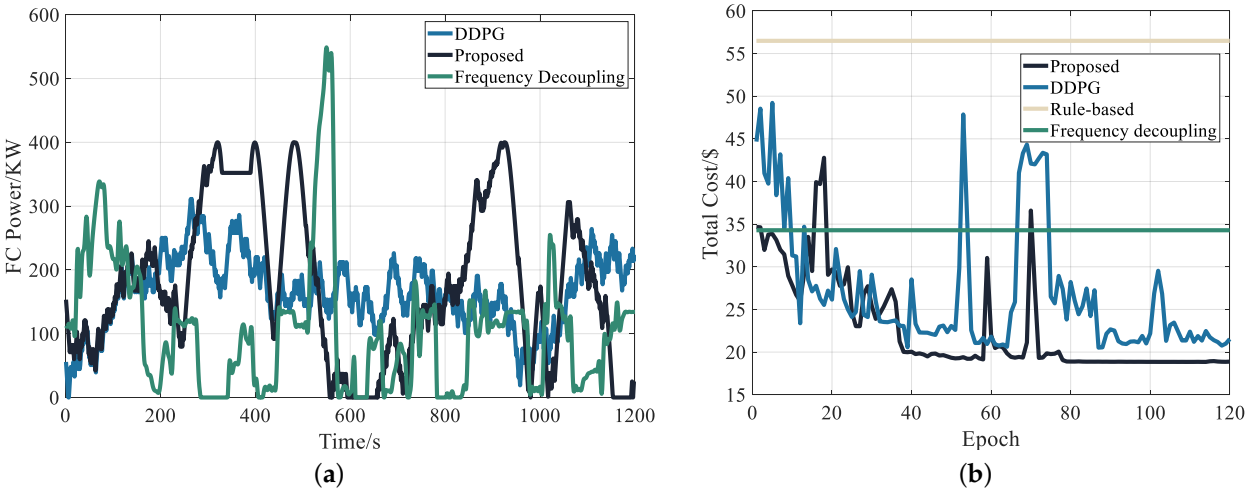
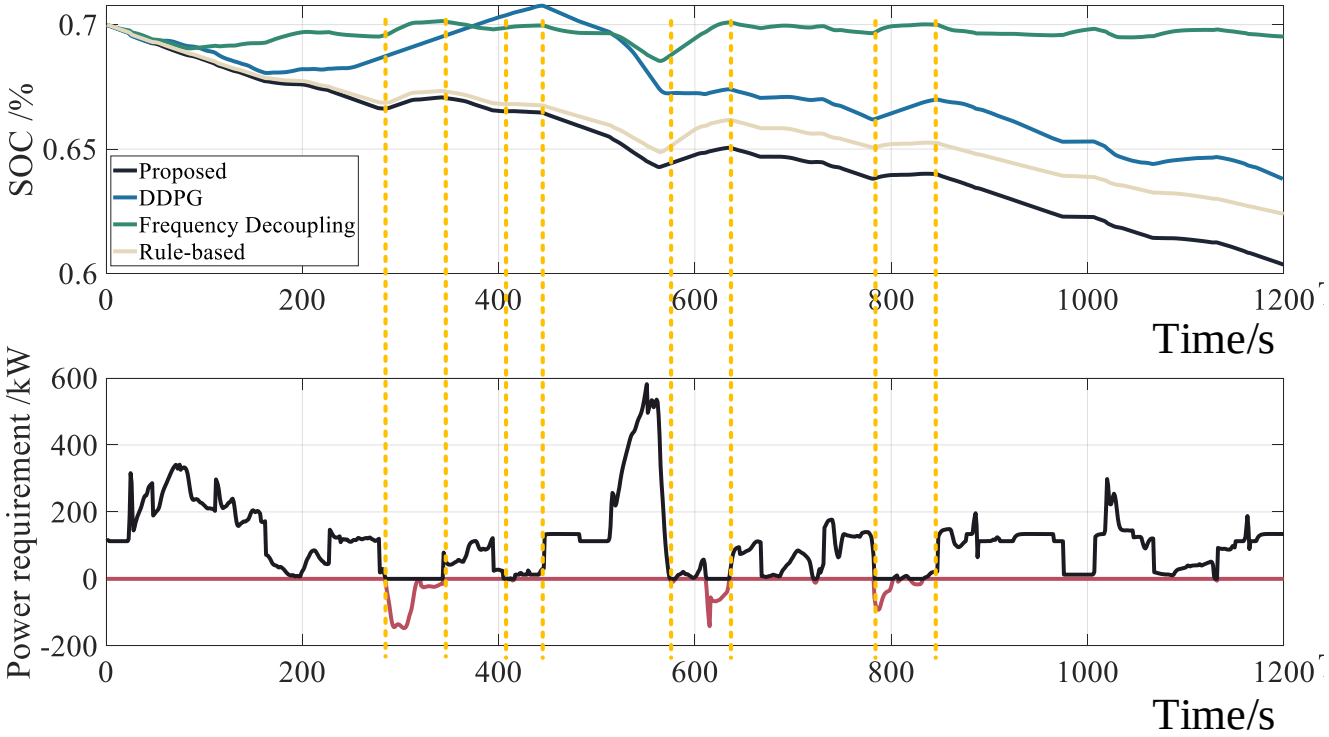
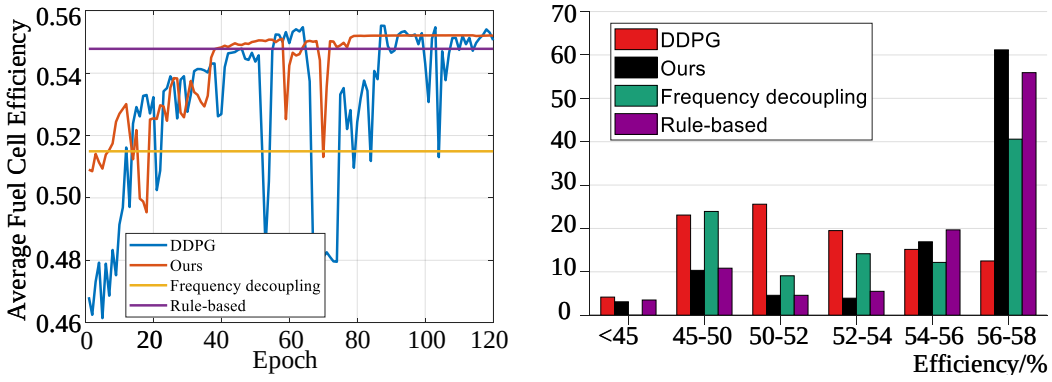


Figure 9. *Cont.*



(**c**)

Figure 9. Comparison of each index of the algorithm. (**a**) Fuel cell power of different algorithms at the first epoch. (**b**) the total cost of one complete run. (**c**) SOC trajectory and energy recovery.



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

Figure 10. Comparison of fuel cell efficiency under different strategies. (a) The average efficiency curve of each training epoch. (b) Finished training, the proportion of efficiency intervals under different strategies.

4.2. Impact of Different Reward Expressions on the Strategy

In this section, the hyperparameters setting of the reward function are discussed. For the reward function, K_{H_2}, $K_{FC_{open}}$, $K_{FC_{run}}$ and $K_{BAT_{run}}$ are determined by economics, and they are given by Table 4.

Table 4. Economic Parameters Setting.

Parameter	Value
K_{H_2} ($/kg)	5
$K_{FC_{open}}$	40
$K_{FC_{run}}$	159
$K_{BAT_{run}}$	105

Figure 11 explores the impact of different reward expressions on the strategy's learning efficiency and final policy when $K_{cost_{inner}}$ and K_{soc} in Equation (13) assume different values (essentially determining their ratio value). As observed, a reward function with greater emphasis on SOC sustenance tends to exhibit conservative behavior, underutilizing the battery buffer. Conversely, if more weight is assigned to minimizing hydrogen consumption, the final policy may breach the battery SOC constraint, causing the battery to enter the risky region in such instances. Simultaneously, adjustments to the weight of the standard deviation of the fuel cell output power, the difference between the initial and final SOC, and the average fuel cell efficiency should be considered. Given that the ranges of $Std(P_{fc})$, ΔSOC, and η_{fc} are all confined within the 0–1 range (the fuel cell standard deviation has been normalized), the ratio values of K_a, K_b, and K_c in Equation (15) are set to be equal.

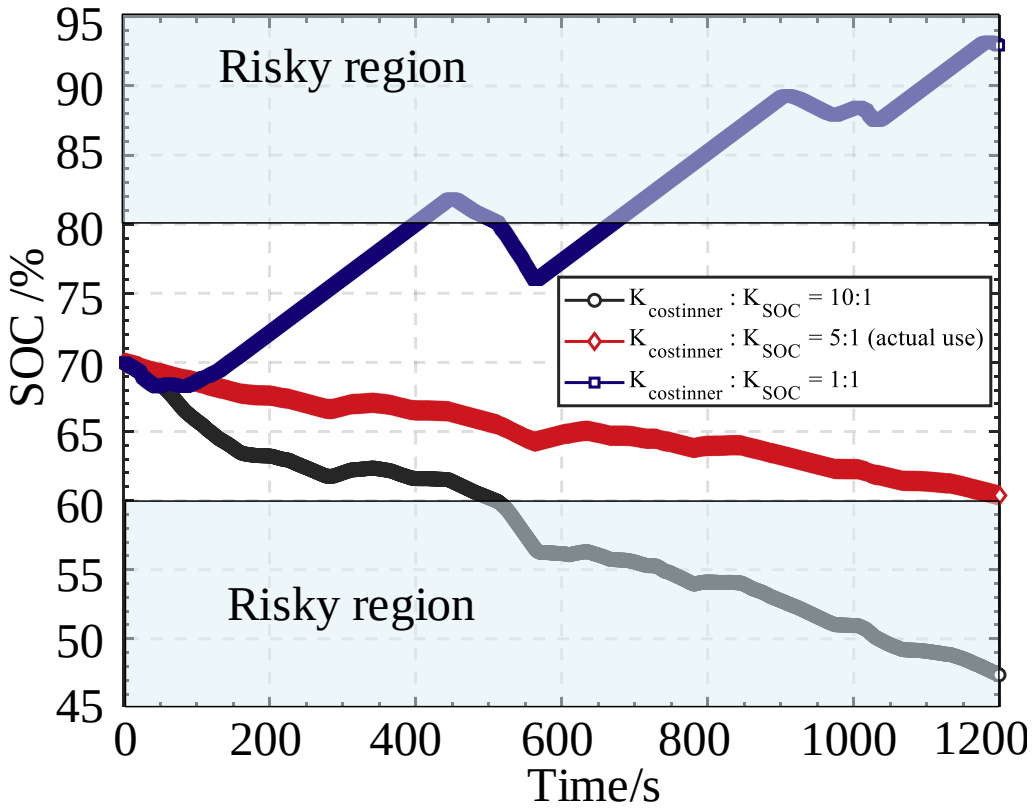


Figure 11. SOC Variation Under Different Hyperparameters.

4.3. Discussion of the Performance about Fuel Cell and Battery

This section delves into further details concerning fuel cells and storage batteries. The frequent start and stop as well as the output stationarity of fuel cells are crucial factors influencing their lifespan and hydrogen consumption. However, traditional strategies often overlook the switching times of fuel cells and the smoothness of output power. Additionally, a significant portion of the energy generated by train braking is typically dissipated through braking resistance, leading to inadequate braking energy recovery in traditional strategies.

Table 5 presents the standard deviation of switching times and output power of the fuel cell under different energy management strategies. It is evident that the application of reinforcement learning (DDPG and our strategy) reduces unnecessary switching losses and effectively maintains the smoothness of the fuel cell's output power. Figure 12 illustrates the braking energy recovery ratio under different energy management strategies, calculated as the ratio of absorbed power by the battery to the total braking power. The proposed algorithm demonstrates the ability to maximize braking energy recovery. In contrast, the DDPG strategy and the frequency decoupling strategy encounter issues such as the battery absorbing energy from the hydrogen fuel cell, leading to overcharging.

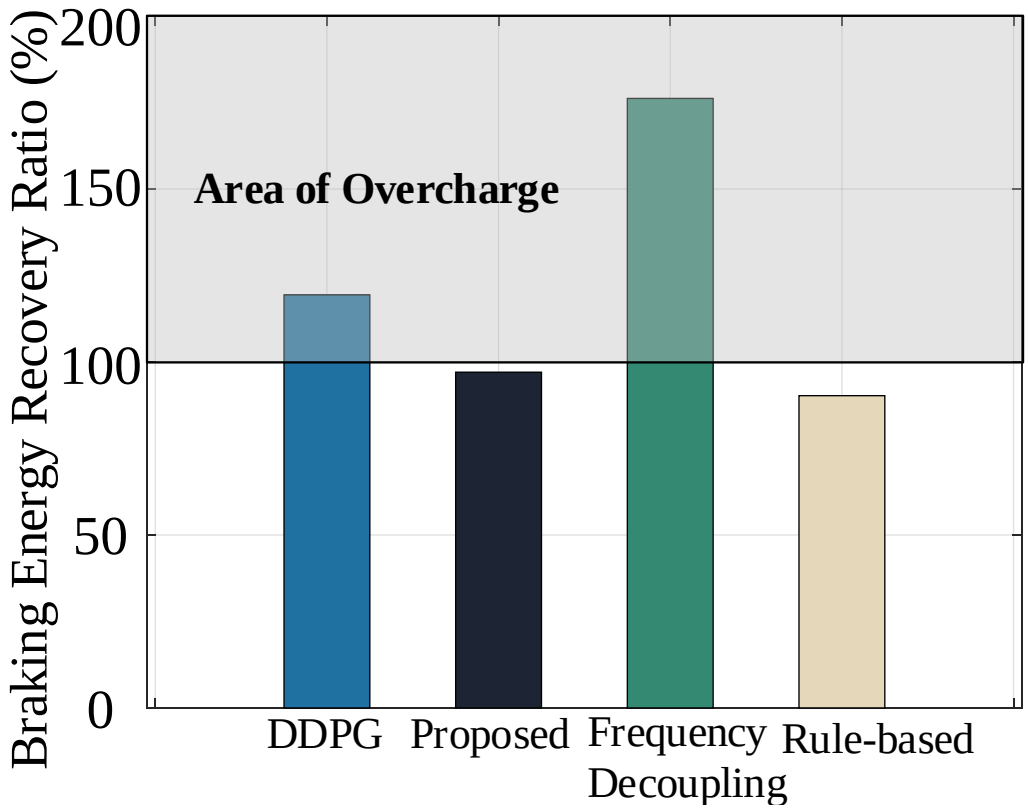


Figure 12. Comparison of braking energy recovery ratio under different strategies.

Table 5. Fuel Cell Related Indicators.

Algorithm	Number of Fuel Cell Starts	Standard Deviation
Rule-based	17	100.04
DDPG	9	95.63
Frequency Decoupling	13	88.93
Proposed	**6**	**86.32**

4.4. Robustness Verification against Different Driving Cycle

The working conditions vary between different scenarios. In order to evaluate the robustness of the proposed strategy, another different driving cycles are tested. The new test condition is shown in Figure 13 [31]. It can be seen that the overall speed of the vehicle under the new condition is higher than that under the old condition, and the requirements for the power system are more stringent. Therefore, we changed the cells of the battery from 345 in Table 1 to 1580. Figure 14 and Table 6 show the energy allocation results of different strategies under this condition. The results indicate that the cost under the proposed strategy is lower than that of other strategies validating the robust performance

of the proposed strategy. It should be noted that this result is pre-trained on the previous condition, so that the common characteristics between different working conditions can be better learned.

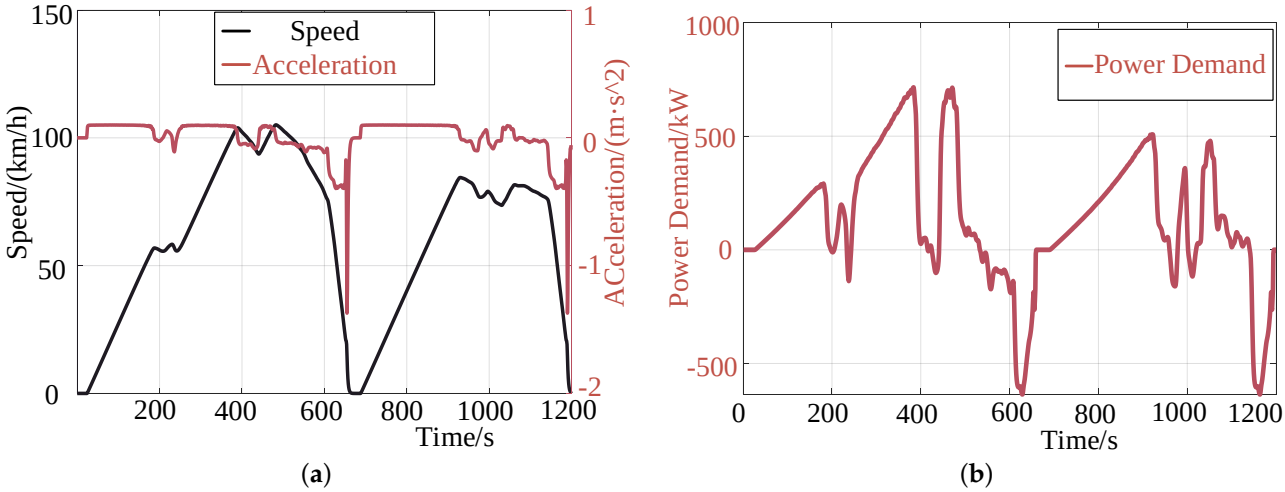


Figure 13. New test condition. (**a**) Train driving cycle. (**b**) Train power demand.

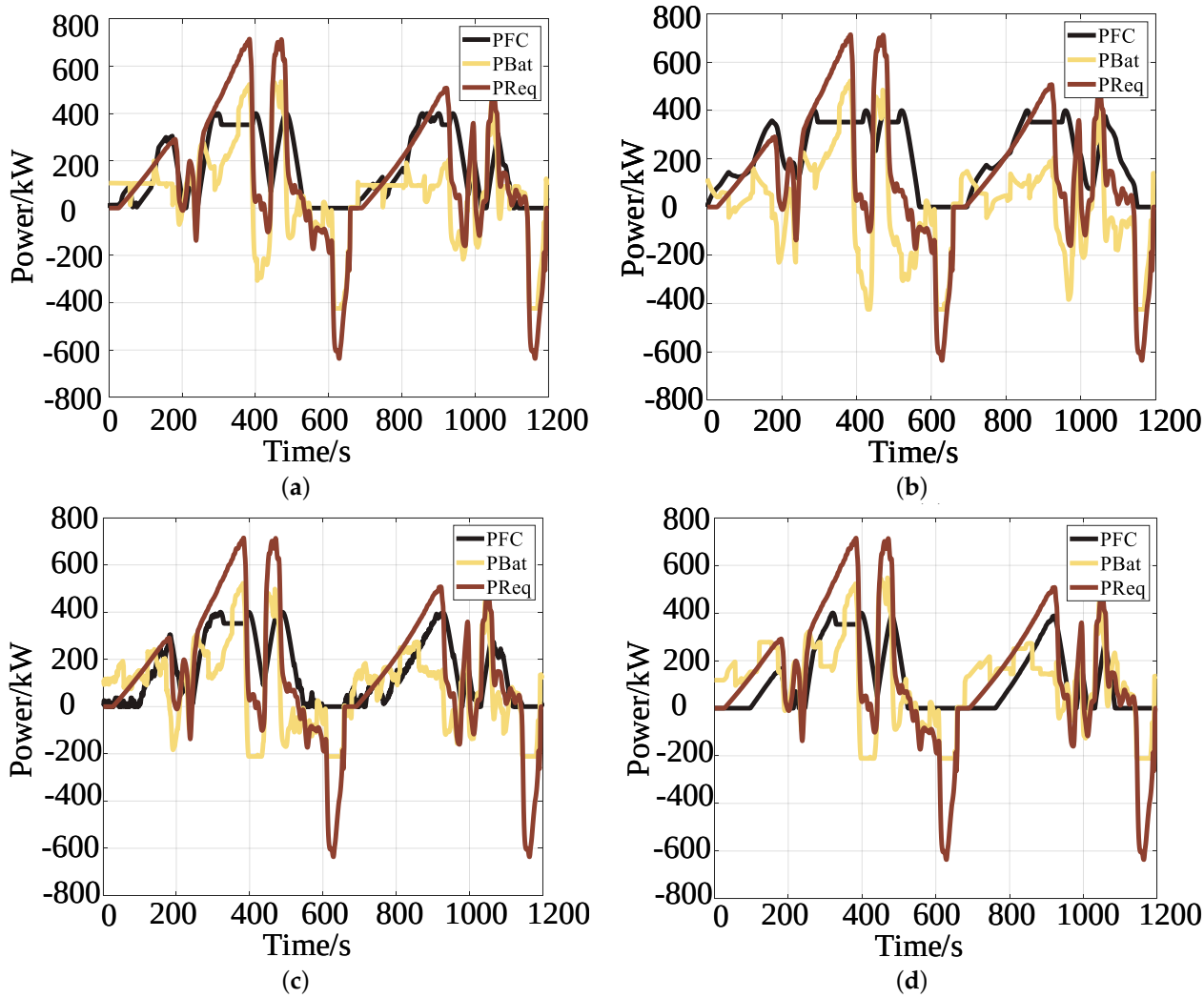


Figure 14. Power allocation under different EMS. (**a**) Rule-based EMS. (**b**) Frequency Decoupling. (**c**) DDPG. (**d**) Proposed method.

Table 6. Comparison between different EMSs in new condition.

Algorithm	Fuel Consumption (kg)	Terminal SOC (%)	Average Efficiency of Fuel Cell (%)	Total Cost ($)	Training Time (s)
Rule-based	8.85	67.24	48.53	86.14	52.14
DDPG	7.70	66.87	49.86	45.96	178.55 + 55.54
Frequency Decoupling	12.44	69.73	44.87	56.91	105.47
Proposed	**5.96**	**65.53**	**51.69**	**37.70**	**252.84 + 68.40**

5. Conclusions

In this research, An adaptive hierarchical EMS for trains of hydrogen fuel cell hybrid power system is proposed by combining frequency decoupling and data-driven DDPG. The goal is to address the difficulty of reinforcement learning based EMS to consider multiple time steps, and training initial values, and directly facilitate online training in real-world Settings. The main findings are summarized as follows.

Our proposed adaptive hierarchical EMS represents a significant advancement in addressing the intricacies of hydrogen fuel cell hybrid power systems within the rail transit context. The integration of frequency decoupling and data-driven DDPG not only showcases reduced hydrogen consumption but also offers a strategic advantage in battery SOC management over extended timeframes. This, in turn, enhances the overall potential and lifespan of the battery while ensuring a stable and optimized fuel cell power output. Unlike conventional RL-based EMS strategies, where random initialization may lead to prolonged training times and potential pitfalls in fuel cell stability, our approach leverages the synergies between reinforcement learning techniques and frequency decoupling methods. This unique combination allows for efficient and safe exploration in real-world environments, positioning our strategy as a robust contender for future RL-based algorithms.

It is crucial to note that our study primarily focuses on energy management under fixed working conditions, presenting an offline training strategy. Future endeavors will delve into the dynamic coupling of speed trajectory optimization and energy management, paving the way for real-time trajectory optimization and adaptive energy distribution for trains. This ongoing research aims to bridge the gap between theoretical advancements and practical applications, contributing to the sustainable and efficient integration of hydrogen fuel cell technology in rail transit systems.

The study still has the following limitations: First, we design a double-layer nested reinforcement learning framework, which results in larger model size and longer training time. Secondly, due to the optimization of global variables, this study only trains the known conditions of the working conditions, rather than online real-time training. When the working conditions change, the model needs to be further fine-tuned to achieve better results.

Author Contributions: Conceptualization, Methodology, Validation, Writing—original draft, Writing—review & editing, H.L.; Project administration, Funding acquisition, Resources, Supervision, J.K.; Data curation, Software, Investigation, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities.

Data Availability Statement: The driving data for this study came from the IEEE VTS Motor Vehicles Challenge 2019 [31], and the vehicle modeling data came from the author Cheng Li.

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

Nomenclature

Abbreviations

EMS	Energy Management Strategy
SOC	State of Charge
RL	Reinforcement Learning
DDPG	Deep Deterministic Policy Gradient
DP	Dynamic programming
GA	Genetic Algorithm
PSO	Particle Swarm Optimization
ECMS	Equivalent Consumption Minimization Strategy
DRL	Deep Reinforcement Learning
DQN	Deep Q-Network

Parameters

F_i	inertial force
F_a	aerodynamic drag
F_f	rolling resistance
δ	correction coefficient of rotating mass
m	mass of train
a	acceleration
ρ	air density
C_d	aerodynamic coefficient
A	fronted area
g	gravity coefficient
f	rolling resistance coefficient
P_{fc}	power of fuel cell
P_{bat}	power of battery
$\eta_{fc-dcdc}$	efficiency of the fuel cell converter
$\eta_{bat-dcdc}$	efficiency of the battery converter
η_T	transmission efficiency
E_{oc}	open circuit voltage
N	number of fuel cell monomer
i_0	exchange current
T_d	dynamic response time
R_{ohm}	internal resistance
i_{fc}	fuel cell current
V_{fc}	fuel cell voltage
P_{bat}	battery output power
V_{oc}	battery open circuit voltage
R_{int}	battery internal resistance
B	inverse amplitude of the exponential region
K	polarization constant
i^*	battery filtration current
F	traction force
R	radius of wheel
η	motor efficiency
K_{H_2}	cost of hydrogen ($/kg)
$m(H_2)$	mass of hydrogen consumed
$K_{FC_{open}}$	cost of turning on the fuel cell once ($)
$K_{FC_{run}}$	operating costs of fuel cells
$K_{BAT_{run}}$	operating costs of lithium batteries ($)

References

1. Zou, Z.; Kang, J.; Hu, J. Analysis of Energy Management Strategies For Hydrogen Fuel Cell Hybrid Rail Transit. In Proceedings of the 2023 IEEE PELS Students and Young Professionals Symposium (SYPS), Shanghai, China, 27–28 August 2023; pp. 1–6. [CrossRef]
2. Sharma, S.; Ghoshal, S.K. Hydrogen the future transportation fuel: From production to applications. *Renew. Sustain. Energy Rev.* **2015**, *43*, 1151–1158. [CrossRef]

3. Abdelrahman, A.S.; Attia, Y.; Woronowicz, K.; Youssef, M.Z. Hybrid Fuel Cell/Battery Rail Car: A Feasibility Study. *IEEE Trans. Transp. Electrif.* **2016**, *2*, 493–503. [CrossRef]
4. Hanley, E.S.; Deane, J.; Gallachóir, B.Ó. The role of hydrogen in low carbon energy futures—A review of existing perspectives. *Renew. Sustain. Energy Rev.* **2018**, *82*, 3027–3045. [CrossRef]
5. Shuguang, L.; Zhenxing, Y. Modeling and Analysis of Contactless Traction Power Supply System for Urban Rail Transit. In Proceedings of the 2020 Chinese Control and Decision Conference (CCDC), Hefei, China, 22–24 August 2020; pp. 5653–5657. [CrossRef]
6. Weirong, C.; Qingquan, Q.; Qi, L. Research status and development trend of fuel cell hybrid electric train (in Chinese). *J. Southwest Jiaotong Univ.* **2009**, *44*, 6.
7. Yuan, X.H.; Yan, G.D.; Li, H.T.; Liu, X.; Su, C.Q.; Wang, Y.P. Research on energy management strategy of fuel cell–battery–supercapacitor passenger vehicle. *Energy Rep.* **2022**, *8*, 1339–1349. [CrossRef]
8. Sun, Y.; Anwar, M.; Hassan, N.M.S.; Spiryagin, M.; Cole, C. A review of hydrogen technologies and engineering solutions for railway vehicle design and operations. *Railw. Eng. Sci.* **2021**, *29*, 212–232. [CrossRef]
9. Fathy, H.K. Hybrid Electric Vehicles: Energy Management Strategies [Bookshelf]. *IEEE Control. Syst. Mag.* **2018**, *38*, 97–98. [CrossRef]
10. Li, G.; Chen, J.; Zheng, X.; Xiao, C.; Zhou, S. Research on Energy Management Strategy of Hydrogen Fuel Cell Vehicles. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; pp. 7604–7607. [CrossRef]
11. Peng, H.; Xie, J. Energy Management Strategy for Plug-In Hybrid Electric Vehicles Based on Genetic-Fuzzy Control Strategy. In Proceedings of the 2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC), Dalian, China, 19–21 December 2017; pp. 1053–1056. [CrossRef]
12. Mesbahi, T.; Rizoug, N.; Bartholomeüs, P.; Sadoun, R.; Khenfri, F.; Le Moigne, P. Optimal Energy Management for a Li-Ion Battery/Supercapacitor Hybrid Energy Storage System Based on a Particle Swarm Optimization Incorporating Nelder–Mead Simplex Approach. *IEEE Trans. Intell. Veh.* **2017**, *2*, 99–110. [CrossRef]
13. Schmid, R.; Buerger, J.; Bajcinca, N. Energy Management Strategy for Plug-in-Hybrid Electric Vehicles Based on Predictive PMP. *IEEE Trans. Control. Syst. Technol.* **2021**, *29*, 2548–2560. [CrossRef]
14. Xie, S.; Hu, X.; Xin, Z.; Brighton, J. Pontryagin's Minimum Principle based model predictive control of energy management for a plug-in hybrid electric bus. *Appl. Energy* **2019**, *236*, 893–905. [CrossRef]
15. Li, G.; Or, S.W.; Chan, K.W. Intelligent Energy-Efficient Train Trajectory Optimization Approach Based on Supervised Reinforcement Learning for Urban Rail Transits. *IEEE Access* **2023**, *11*, 31508–31521. [CrossRef]
16. Ning, L.; Zhou, M.; Hou, Z.; Goverde, R.M.; Wang, F.Y.; Dong, H. Deep Deterministic Policy Gradient for High-Speed Train Trajectory Optimization. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11562–11574. [CrossRef]
17. Gan, J.; Li, S.; Wei, C.; Deng, L.; Tang, X. Intelligent Learning Algorithm and Intelligent Transportation-Based Energy Management Strategies for Hybrid Electric Vehicles: A Review. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 10345–10361. [CrossRef]
18. 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. [CrossRef]
19. Zhang, Y.; Ma, R.; Zhao, D.; Huangfu, Y.; Liu, W. A Novel Energy Management Strategy Based on Dual Reward Function Q-learning for Fuel Cell Hybrid Electric Vehicle. *IEEE Trans. Ind. Electron.* **2022**, *69*, 1537–1547. [CrossRef]
20. Lin, X.; Zhou, B.; Xia, Y. Online Recursive Power Management Strategy Based on the Reinforcement Learning Algorithm with Cosine Similarity and a Forgetting Factor. *IEEE Trans. Ind. Electron.* **2021**, *68*, 5013–5023. [CrossRef]
21. Cong, J.; Li, B.; Guo, X.; Zhang, R. Energy Management Strategy based on Deep Q-network in the Solar-powered UAV Communications System. In Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. [CrossRef]
22. Hu, B.; Li, J. An Adaptive Hierarchical Energy Management Strategy for Hybrid Electric Vehicles Combining Heuristic Domain Knowledge and Data-Driven Deep Reinforcement Learning. *IEEE Trans. Transp. Electrif.* **2022**, *8*, 3275–3288. [CrossRef]
23. 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. [CrossRef]
24. Han, R.; Lian, R.; He, H.; Han, X. Continuous Reinforcement Learning-Based Energy Management Strategy for Hybrid Electric-Tracked Vehicles. *IEEE J. Emerg. Sel. Top. Power Electron.* **2023**, *11*, 19–31. [CrossRef]
25. Tao, F.; Zhu, L.; Fu, Z.; Si, P.; Sun, L. Frequency Decoupling-Based Energy Management Strategy for Fuel Cell/Battery/Ultracapacitor Hybrid Vehicle Using Fuzzy Control Method. *IEEE Access* **2020**, *8*, 166491–166502. [CrossRef]
26. Huangfu, Y.; Yu, T.; Zhuo, S.; Shi, W.; Zhang, Z. An Optimization Energy Management Strategy Based on Dynamic Programming for Fuel Cell UAV. In Proceedings of the IECON 2021—47th Annual Conference of the IEEE Industrial Electronics Society, Toronto, ON, Canada, 13–16 October 2021; pp. 1–6.
27. 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]
28. Zhou, Q.; Li, J.; Shuai, B.; Williams, H.; He, Y.; Li, Z.; Xu, H.; Yan, F. Multi-step reinforcement learning for model-free predictive energy management of an electrified off-highway vehicle. *Appl. Energy* **2019**, *255*, 113755. [CrossRef]

29. Zhou, Q.; Zhao, D.; Shuai, B.; Li, Y.; Williams, H.; Xu, H. 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.* **2021**, *32*, 5298–5308. [CrossRef]
30. Noh, S.; Shim, D.; Jeon, M. Adaptive Sliding-Window Strategy for Vehicle Detection in Highway Environments. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 323–335. [CrossRef]
31. Lhomme, W.; Letrouve, T.; Boulon, L.; Jemei, S.; Bouscayrol, A.; Chauvet, F.; Tournez, F. IEEE VTS Motor Vehicles Challenge 2019—Energy Management of a Dual-Mode Locomotive. In Proceedings of the 2018 IEEE Vehicle Power and Propulsion Conference (VPPC), Chicago, IL, USA, 27–30 August 2018; pp. 1–6. [CrossRef]
32. Njoya M., S.; Tremblay, O.; Dessaint, L.A. A generic fuel cell model for the simulation of fuel cell vehicles. In Proceedings of the 2009 IEEE Vehicle Power and Propulsion Conference, Dearborn, MI, USA, 7–10 September 2009; pp. 1722–1729.
33. Tremblay, O.; Dessaint, L.A. Experimental Validation of a Battery Dynamic Model for EV Applications. *World Electr. Veh. J.* **2009**, *3*, 289–298. [CrossRef]
34. Ao, Y.; Laghrouche, S.; Depernet, D.; Chen, K. Proton Exchange Membrane Fuel Cell Prognosis Based on Frequency-Domain Kalman Filter. *IEEE Trans. Transp. Electrif.* **2021**, *7*, 2332–2343. [CrossRef]
35. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.