machines

MDPI

Article

An Energy Management Strategy for FCHEVs Using Deep Reinforcement Learning with Thermal Runaway Fault Diagnosis Considering the Thermal Effects and Durability

Yongqiang Wang [1], Fazhan Tao [1], Longlong Zhu [1], Nan Wang [1] and Zhumu Fu [2,*]

[1] College of Information Engineering and Artificial Intelligence, Henan University of Science and Technology, Luoyang 471023, China; 220410040026@stu.haust.edu.cn (Y.W.); taofazhan@haust.edu.cn (F.T.); 9906654@haust.edu.cn (L.Z.); wangnan@haust.edu.cn (N.W.)

[2] School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China

* Correspondence: fuzhumu@haust.edu.cn; Tel.: +86-13513844662

Abstract

Temperature control plays a critical role in mitigating the lifespan degradation mechanisms and ensuring thermal safety of lithium-ion batteries (LIBs) and proton exchange membrane fuel cells (PEMFCs). However, current energy management strategies (EMS) for fuel cell hybrid electric vehicles (FCHEVs) generally lack comprehensive thermal effect modeling and thermal runaway fault diagnosis, leading to irreversible aging and thermal runaway risks for LIBs and PEMFCs stacks under complex operating conditions. To address this challenge, this paper proposes a thermo-electrical co-optimization EMS incorporating thermal runaway fault diagnosis actuators, with the following innovations: firstly, a dual-layer framework integrates a temperature fault diagnosis-based penalty into the EMS and a real-time power regulator to suppress heat generation and constrain LIBs/PEMFCs output, achieving hierarchical thermal management and improved safety; secondly, the distributional soft actor–critic (DSAC)-based EMS incorporates energy consumption, state-of-health (SoH) degradation, and temperature fault diagnosis-based constraints into a composite penalty function, which regularizes the reward shaping and guides the policy toward efficient and safe operation; finally, a thermal safe constriction controller (TSCC) is designed to continuously monitor the temperature of power sources and automatically activate when temperatures exceed the optimal operating range. It intelligently identifies optimized actions that not only meet target power demands but also comply with safety constraints. Simulation results demonstrate that compared to DDPG, TD3, and SAC baseline strategies, DSAC-EMS achieves maximum reductions of 39.91% in energy consumption and 29.38% in SoH degradation. With the TSCC implementation, enhanced thermal safety is achieved, while the maximum energy-saving improvement reaches 25.29% and the maximum reduction in SoH degradation attains 20.32%.

Keywords: energy management strategy; fuel cell vehicles; deep reinforcement learning; distributional soft actor-critic; lifespan durability; thermal stability; thermal runaway fault diagnosis

check for updates

Academic Editor: Zhen Jia

Received: 5 September 2025
Revised: 9 October 2025
Accepted: 16 October 2025
Published: 18 October 2025

Citation: Wang, Y.; Tao, F.; Zhu, L.; Wang, N.; Fu, Z. An Energy Management Strategy for FCHEVs Using Deep Reinforcement Learning with Thermal Runaway Fault Diagnosis Considering the Thermal Effects and Durability. *Machines* **2025**, *13*, 962. https://doi.org/10.3390/machines13100962

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the growing attention to environmental protection and energy security, fuel cell hybrid electric vehicles (FCHEVs) are playing an increasingly crucial role in promoting green

transportation and energy transformation [1]. A core challenge in FCHEV design lies in coordinating heterogeneous power sources—proton exchange membrane fuel cells (PEMFCs) and lithium-ion batteries (LIBs)—to simultaneously achieve high fuel economy, long component lifespan, and robust thermal safety. To this end, numerous energy management strategies (EMSs) have been developed, including rule-based methods, optimization-based approaches such as dynamic programming (DP) [2] and Pontryagin's minimum principle (PMP) [3], and real-time strategies like equivalent consumption minimization strategy (ECMS) [4,5], model predictive control (MPC) [6,7], fuzzy logic [8,9], and genetic algorithms [10–12]. More recently, intelligent control methods—particularly deep reinforcement learning (DRL)—have gained traction due to their adaptability and learning capabilities [13–21].

However, most existing EMSs prioritize fuel economy while treating durability and thermal safety as secondary considerations. This is problematic because aggressive power allocation can accelerate battery degradation and induce thermal runaway, especially under high-energy-density battery operation. Recognizing this, recent studies have begun integrating state-of-health (SoH) awareness into EMS design. For instance, MPC [22], optimized point-line methods [23], and DRL algorithms—including DQN [24], DDPG [25], TD3 [26], and SAC [27]—have been employed to jointly minimize fuel consumption and SoH degradation. Yet these approaches suffer from significant limitations: they often exhibit poor sample efficiency, unstable convergence, high computational demands, and sensitivity to hyperparameter tuning. Moreover, they typically assume idealized thermal conditions and lack mechanisms to enforce real-time thermal safety constraints, limiting their practical applicability in safety-critical automotive systems.

Concurrently, the rising energy density of LIBs has intensified thermal management challenges. Elevated temperatures not only hasten electrochemical aging but also increase the risk of thermal runaway—a catastrophic failure mode that compromises vehicle safety. Consequently, integrating thermal dynamics into EMS has become imperative. Pioneering efforts in this direction include Zhang et al. [28], who proposed a climate-adaptive EMS that coordinates cabin heating with battery thermal behavior to improve fuel economy in cold environments while preserving battery health. Khalatbarisoltani et al. [29] further advanced this idea by combining multi-agent DRL (MADRL) with LSTM-based driving and thermal prediction models to co-optimize energy use and thermal load. Han et al. [30] proposed a hybrid PMP-MPC framework enhanced with an ANN-based speed predictor that explicitly accounts for electric motor thermal dynamics to reduce operational costs and regulate temperature. Similarly, Abbasi et al. [31] used DRL to jointly optimize charging protocols and cooling strategies, demonstrating up to two years of extended battery life under controlled thermal conditions. Guo et al. [32] developed a thermal-aware EMS that links battery aging mechanisms directly to thermal regulation, achieving a balance between performance and longevity. Most recently, Zhang et al. [33] introduced a DRL-based EMS that embeds thermal effects into power-split decisions, dynamically managing temperature to enhance safety, efficiency, and lifespan.

Despite these advances, a critical research gap remains: existing integrated EMS either lack explicit safety enforcement under thermal runaway or rely on standard DRL frameworks that suffer from Q-value overestimation, high policy complexity, and insufficient robustness in multi-objective optimization. Crucially, none of the cited works [22–33] implement a coordinated, dual-layer architecture that proactively prevents thermal runaway while simultaneously optimizing economy, durability, and safety through a distributional reinforcement learning framework.

To address these shortcomings, this paper proposes a novel distributional soft actor-critic (DSAC)-based EMS with thermal runaway fault diagnosis for FCHEVs. Building upon the foundational insights from [28–33], our approach introduces three key innovations:

1. To mitigate thermal runaway under high-power operation, this paper proposes a dual-layer protection framework. The upper layer integrates a temperature fault diagnosis-based penalty into the EMS to proactively suppress heat generation. The lower layer employs a real-time safety-constrained power regulator that dynamically adjusts output limits of LIBs and PEMFCs. This strategy-to-execution coordination enables proactive thermal management, improving both system safety and operational reliability.
2. To overcome Q-value overestimation and policy complexity in DRL-based energy management, this work proposes a DSAC-based EMS for FCHEVs. By embedding power loss, SoH degradation, and temperature fault diagnosis-based constraints into a composite penalty function, the method reduces policy dimensionality and achieves multi-objective optimization for economy, durability, and safety. The DSAC algorithm enhances efficiency and accuracy by learning the return distribution via a single network using return variance.
3. To enhance safety under high-temperature conditions, this paper designs a DSAC-based safety-constrained controller for power action optimization. It continuously monitors source temperatures and activates automatically when thresholds are exceeded. The controller intelligently selects actions that meet power demand while satisfying thermal constraints, ensuring safe and efficient operation through real-time, adaptive decision-making.

The structure of this paper is as follows: Section 2 provides gives the SoH and thermal models of FCHEVs. Section 3 presents the optimization framework and control architecture of the proposed EMS, including its training procedure and a thermal safety-oriented soft constraint controller for the power sources. Section 4 presents the simulation results. A conclusion is given in Section 5.

2. Powertrain Model

2.1. FCHEV Powertrain System Overview

The architecture of the FCHEV powertrain system is shown in Figure 1. In this system, the FCHEV is powered by a permanent magnet synchronous motor (PMSM), which drives the rear wheels through a differential and final drive connected to the rear axle of the vehicle. The PEMFC system operates in parallel with a lithium-ion battery, and is connected to the DC bus via a unidirectional boost DC-DC converter. Other auxiliary electronic devices, such as the battery, are powered through a DC-DC converter linked to the DC bus. The power consumption of the thermal management system (TMS) and other auxiliary electronic devices is not considered in this paper. The detailed powertrain parameters provided in Table 1 are adopted from [33].

Table 1. FCHEV parameters.

Property	Parameter	Value	Unit
FCV mass	m_{fc}	2046	kg
Equivalent windward area	A_f	2.62	m^2
Wheel radius	r_w	0.35	m
Gear ratio of final drive	i_{trans}	9	–

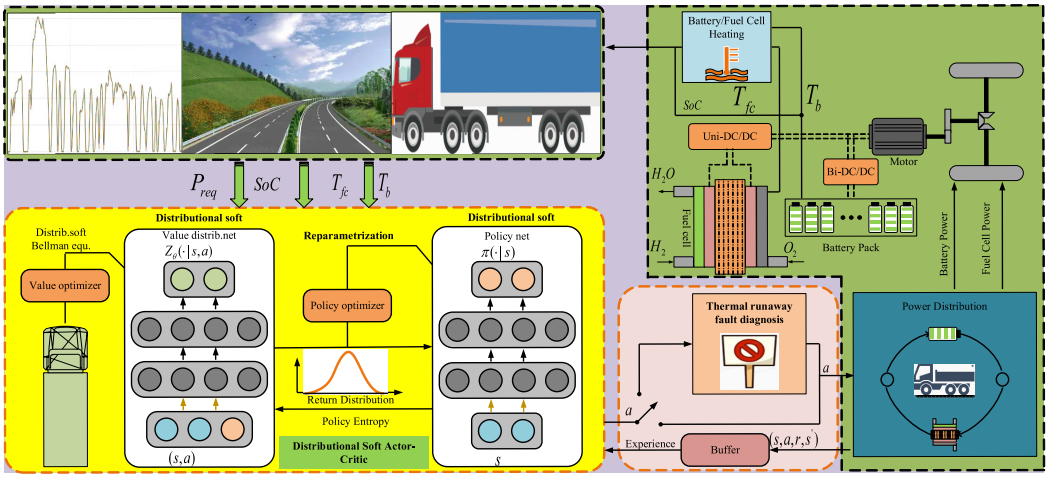


Figure 1. Schematic diagram of intelligent EMS and FCHEV powertrain.

2.2. Electric–Thermal PEMFC Model

Thermal management is critical during the operation of fuel cells. The heat generated by electrochemical reactions within the fuel cell stack leads to a temperature rise, which can adversely impact the efficiency and lifespan of the fuel cell. Therefore, temperature regulation through the cooling system is essential. The rate of temperature change within the fuel cell stack can be expressed as:

$$\frac{dT_{fc}(t)}{dt} = \frac{P_{fc,loss}(t) - Q_{fc,coolant}(t) - Q_{fc,e}(t)}{m_{fc}C_{s,fc}} \tag{1}$$

where $P_{fc,loss}(t)$ represents the thermal power generated by electrochemical reactions, $Q_{fc,coolant}(t)$ is the thermal power transferred by the coolant, and $Q_{fc,e}(t)$ is the surface heatloss of the stack. m_{fc} and $C_{s,fc}$ are the mass and specific heat capacity of the PEMFCs stack, respectively.

The thermal power $P_{fc,loss}(t)$ can be calculated as:

$$P_{fc,loss}(t) = N_{cell} \times (V_{therm} - V_{fc}(t)) \times I_{fc}(t) \tag{2}$$

where N_{cell} represents the number of fuel cell units in the stack, $V_{therm}(t)$ is the thermoneutral voltage, while $V_{fc}(t)$ and $I_{fc}(t)$ denote the stack's voltage and current, respectively.

The thermal power transferred by the coolant, $Q_{fc,coolant}(t)$, is dependent on the temperature difference between the stack and the coolant, and is generally expressed by the following equation:

$$Q_{fc,coolant}(t) = h_{fc,heat} \cdot A_{fc} \cdot \left(T_{fc}(t) - T_{fc,coolant}(t)\right) \tag{3}$$

where $h_{fc,heat}$ represents the heat transfer coefficient, A_{fc} represents the contact area between the stack and the coolant, and $T_{fc}(t)$ and $T_{fc,coolant}(t)$ correspond to the temperatures of the stack and the coolant, respectively.

$Q_{fc,e}(t)$ represents the heat loss power from the PEMFCs stack to the environment. Heat is primarily dissipated when the stack temperature exceeds the environment temperature, while reverse heat transfer may occur under cooling conditions. It can be expressed by the following formula:

$$Q_{fc,e}(t) = \frac{T_{fc}(t) - T_e}{R_{fc}} \tag{4}$$

where $T_{fc}(t)$, R_{fc}, and T_e represent the PEMFC stack temperature, thermal resistance, and environment temperature, respectively.

2.3. Electric–Thermal LIB Model

In order to optimize the performance and lifespan of the LIBs within the powertrain system, this paper adopts a battery management system (BMS) model. This model not only monitors the state of charge (SoC) of the lithium-ion batteries (LIBs) but also controls the thermal dynamics to ensure that the battery operates within the optimal temperature range. The BMS dynamically adjusts the cooling system's power output based on real-time monitoring of the battery temperature, thereby regulating the thermal management process.

The temperature variation of the LIBs can be described by the following equation:

$$\frac{dT_{lib}(t)}{dt} = \frac{P_{loss}(t) - Q_{cooling}(t)}{m_{lib}C_{s,lib}} \tag{5}$$

where $P_{loss}(t)$ represents the heat loss due to current transmission, $Q_{cooling}(t)$ denotes the thermal power transferred by the cooling system, and m_{lib} and $C_{s,lib}$ are the mass and specific heat capacity of the LlB, respectively.

The heat loss power $P_{loss}(t)$ can be expressed as:

$$P_{loss}(t) = I_{lib}(t)^2 \cdot R_{lib}(t) \tag{6}$$

where $I_{lib}(t)$ is the battery current, and $R_{lib}(t)$ is the equivalent internal resistance of the battery, which varies dynamically with temperature and current rate.

The thermal power transferred by the cooling system, $Q_{cooling}(t)$, is a function of the temperature difference between the battery and the coolant and can be expressed as:

$$Q_{cooling}(t) = h_{cool} \cdot A_{cool} \cdot (T_{lib}(t) - T_{coolant}(t)) \tag{7}$$

where h_{cool} is the heat transfer coefficient, A_cool is the contact area between the coolingsystem and the battery, and $T_{lib}(t)$ and $T_{coolant}(t)$ represent the temperatures of the battery and the coolant, respectively.

$$\frac{dT_{b,s}(t)}{dt} = \frac{T_{b,coolant} - T_{b,s}(t)}{R_{b,ext}C_{b,s}} - \frac{T_{b,s}(t) - T_{b,c}(t)}{R_{b,in}C_{b,s}} \tag{8}$$

$$\frac{dT_{b,c}(t)}{dt} = \frac{T_{b,s}(t) - T_{b,c}(t)}{R_{b,in}C_{b,c}} + \frac{P_{b,loss}(t)}{C_{b,c}} \tag{9}$$

$$T_b(t) = \frac{T_{b,s}(t) + T_{b,c}(t)}{2} \tag{10}$$

where $T_{b,s}(t)$, $T_{b,c}(t)$, and $T_{b,coolant}$ denote the surface, core, and coolant temperatures of the LIB cell, respectively; $T_b(t)$ is defined as the average of $T_{b,s}(t)$ and $T_{b,c}(t)$; $R_{b,in}$ and $R_{b,ext}$ represent the internal and external thermal resistances of the LIB cell; and $C_{b,s}$ and $C_{b,c}$ are the surface and core heat capacities of the LIB cell.

2.4. SoH Models of PEMFCs and LIBs

The degradation of the PEMFCs stack is essentially the accumulation of degradation rates under idle, high power load, load cycling, and start–stop conditions. Since the power output of the PEMFCs stack is continuous during a driving cycle, the start–stop condition is not considered in the calculation of the PEMFCs stack degradation. Therefore, the

formula for the decline in the SoH of the PEMFCs stack, denoted as $\Delta SoH_{fc}(t)$, is expressed as follows:

$$\begin{aligned}\Delta SoH_{fc}(t) = \Delta SoH_{fc,low}(t) + \Delta SoH_{fc,high}(t) \\ + \Delta SoH_{fc,\Delta P}(t)\end{aligned} \tag{11}$$

where $\Delta SoH_{fc,low}(t)$, $\Delta SoH_{fc,high}(t)$, and $\Delta SoH_{fc,\Delta P}(t)$ represent the SoH degradation rates of the PEMFCs stack under low load operation, high load operation, and load variation operation, respectively.

$$\Delta SoH_{fc,low}(t) = \begin{cases} -k_{low}\Delta t / V_{fc}, & if \quad P_{fc} \leq 0.2 P_{fc,max} \\ 0, & else \end{cases} \tag{12}$$

$$\Delta SoH_{fc,high}(t) = \begin{cases} -k_{high}\Delta t / V_{fc}, & if \quad P_{fc} \geq 0.8 P_{fc,max} \\ 0, & else \end{cases} \tag{13}$$

$$\Delta SoH_{fc,chg}(t) = -k_{chg}|P_{fc}(t) - P_{fc}(t-1)|/(N_{cell}V_{fc}) \tag{14}$$

where k_{low}, k_{high}, and k_{chg} represent the voltage degradation rates during idling, high power load, and load change periods, respectively. Δt represents the sampling interval time. The specific values for these parameters can be found in [33].

The performance degradation of lithium batteries is typically manifested as a loss of their maximum remaining capacity, which is commonly characterized by a semi-empirical exponential model [34]. This model expresses the instantaneous capacity loss rate of lithium batteries as a function of current, as follows:

$$\Delta Q_{bat}(t) = M(c)exp(\frac{-E_a(c)}{RT})A(c)^{\phi} \tag{15}$$

where ΔQ_{bat} represents the capacity loss of the lithium battery, expressed as a percentage, c denotes the discharge rate of the lithium battery, $M(c)$ is the exponential prefactor associated with the discharge rate c, R is the ideal gas constant, T is the temperature of the lithium battery, $A(c)$ refers to the current throughput of the lithium battery (Ah), φ is the power-law factor, and $E_a(c)$ represents the activation energy related to the discharge rate c, expressed as:

$$\begin{cases} E_a(c) = 31700 - 163.3c \\ \varphi = 0.57 \end{cases} \tag{16}$$

it is commonly understood that a lithium battery reaches the end of its life when it experiences a 20% capacity loss. By defining $\Delta Q_{bat} = 20\%$, the total charge–discharge throughput A_{total} and the number of charge–discharge cycles N_{bat} from the beginning of usage to the end of the battery's life can be determined as follows:

$$A_{total}(c, T_c) = \left[\frac{20}{M(c)exp(\frac{-E_a(c)}{RT})}\right]^{\frac{1}{\varphi}} \tag{17}$$

$$N_{bat}(c, T_a) = \frac{A_{total}(c, T_c)}{Q_{bat}} \tag{18}$$

where Q_{bat} denotes the rated capacity of the lithium battery. Therefore, the remaining maximum usable capacity of the lithium battery, Q_{rem}, can be calculated as follows:

$$Q_{rem}(t) = \frac{\int_0^t |I_{bat}(t)|dt}{2 \times 3600 N_{bat}(c, T_c) Q_{bat}} \tag{19}$$

the SoH is defined to quantify the degradation of lithium battery performance. By differentiating and discretizing Equation (19), the following formula is derived:

$$\Delta SoH_{bat}(t) = \frac{T_s |I_{bat}(t)|}{2 \times 3600 N_{bat}(c, T_c) Q_{bat}} \tag{20}$$

where ΔSoH_{bat} represents the rate of change in the state of health of the lithium battery.

Based on the SoH models of the two energy sources, the lifespan degradation trends of the lithium-ion battery and fuel cell are accurately quantified. This not only provides a theoretical foundation for the coordinated management of dual-energy systems, but also establishes a solid basis for the design of energy management strategies aimed at lifespan optimization.

3. Design of DSAC-Based EMS with Thermal Flexibility Constraints Considering the Thermal Effects and Durability

The implementation architecture of the DSAC-based EMS framework incorporating thermal constraints is illustrated in Figure 2. First, a DSAC-based EMS considering thermal effects and durability is designed to optimize power distribution in FCHEVs. Second, a flexible-constrained thermal safety controller is developed to ensure the thermal safety of energy sources in FCHEVs.

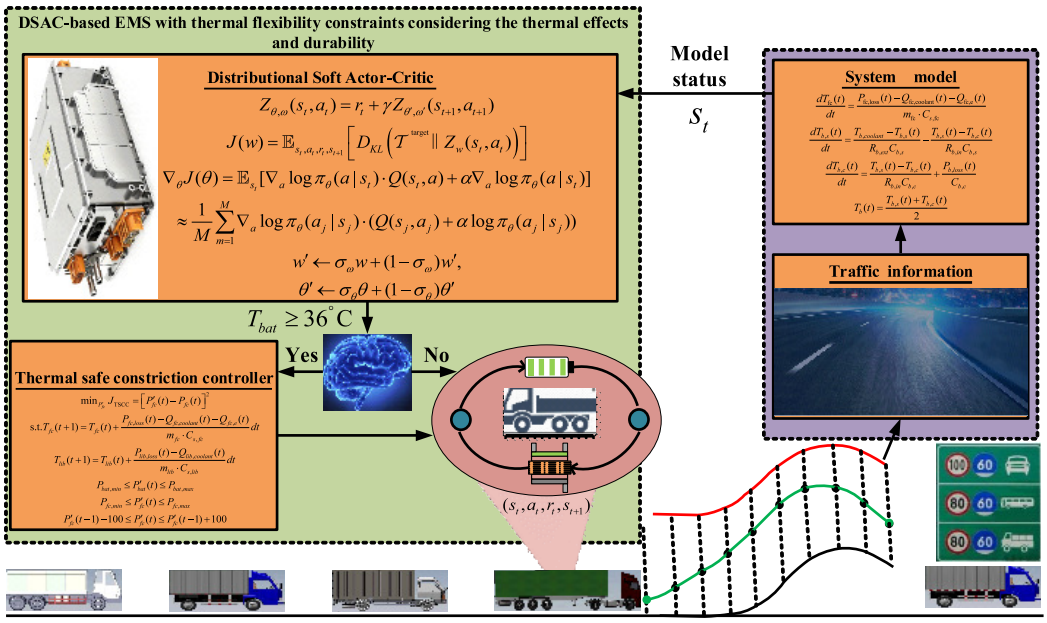


Figure 2. DSAC-based energy management framework with thermal constraints.

3.1. Training of EMS

In this paper, the DSAC algorithm is employed to optimize the EMS for a FCHEVs. The core capability of DSAC lies in its ability to learn an optimal policy that maximizes long-term rewards through interaction with the environment, effectively addressing uncertainty and delayed feedback, while maintaining a balance between exploration and exploitation, and demonstrating strong generalization and adaptation capabilities [35]. The iterative process of DSAC consists of four core components: (1) the environment model; (2) the expected value function $Q^\pi(s_t, a_t)$ based on the distributional soft return; (3) the sample transition $e_t = (s_t, a_t, r_t, s_{t+1})$; (4) the stochastic policy $\pi_\theta(a_t|s_t)$ parameterized by θ. In this context, the design objective of the control policy π_t^* is to learn a distributional model of returns for each state-action pair, thereby reducing estimation errors in Q-value approximation and improving both the stability and performance of long-term reward maximization. The

optimization goal is to maximize the entropy-regularized total discounted reward while accurately capturing the uncertainty in return distributions, as formulated below:

$$\pi_t^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t + \mathcal{H}(\pi(\cdot|s_t))\right] \tag{21}$$

where $\mathbb{E}[\cdot]$ denotes the expectation over environment dynamics and the policy; $\gamma \in [0, 1]$ is the discount factor; $\mathcal{H}(\pi(\cdot|s_t))$ is the entropy of the action distribution under the policy at state s_t, which encourages exploration; the reward function r_t is designed to incorporate multiple key performance indicators, such as fuel economy, component durability, and thermal management effectiveness, to reflect the system's overall efficiency and reliability at each time step t.

The state space S_t of the FCHEVs EMS agent is formulated to represent the key variables influencing energy flow and system performance. It includes two primary parameters: (1) vehicle power demand, and (2) SoC. These variables are essential for making intelligent decisions regarding power allocation between the fuel cell and energy storage system. Specifically, the state vector at time step t is defined as:

$$S_t = [P_{req}, SoC_t] \tag{22}$$

where P_{req} denotes the total power demand of the vehicle at time t, determined by the driver's torque request and vehicle dynamics; SoC_t represents the current state of charge (%) of the lithium-ion battery, which directly affects the allowable discharging/charging behavior and overall energy availability.

This compact yet informative state representation ensures that the DRL agent can effectively learn the trade-off between hydrogen consumption and battery usage, particularly under varying driving conditions. To enable flexible and efficient energy distribution, the action space is designed as a continuous structure, reflecting the adjustable output capability of the fuel cell system. Specifically, the action variable a_t corresponds to the fuel cell power output:

$$a_t = P_{fc}(t) \in [0, P_{fc,max}] \tag{23}$$

where $P_{fc}(t)$ denotes the power provided by the fuel cell at time t; $P_{fc,min}$ and $P_{fc,max}$ are the minimum and maximum allowable fuel cell power output, respectively, constrained by system efficiency and operational safety.

The remaining power required to meet the vehicle demand is supplied by the battery, calculated as:

$$P_{bat}(t) = P_{req} - P_{fc}(t) \tag{24}$$

This formulation ensures that the total power delivered to the drivetrain satisfies the propulsion requirement, while maintaining flexibility in energy source allocation.

By integrating this structured state-action mapping into the DRL framework, the proposed EMS enables intelligent coordination between the fuel cell and battery systems, leading to improved fuel economy and thermal performance under diverse driving conditions.

The reward terms include hydrogen consumption cost, battery SoC deviation penalty, degradation costs for both fuel cell and lithium-ion battery, and penalties related to temperature deviations from optimal operating ranges. All cost items are assigned negative values to represent energy expenditure or system loss, while weighting coefficients are used

to balance the relative importance of each sub-objective. The reward function is defined as follows:

$$\begin{aligned} r_t = -\Big[&\Gamma_{H_2} m_{eq}(t) + \Gamma_{bat} SoC(t) + \Gamma_{fc} \Delta SoH_{fc}(t) \\ &+ \Gamma_b \Delta SoH_{bat}(t) + \Gamma_{ba} T_{bat}(t) + \Gamma_f T_{fc}(t)\Big] \end{aligned} \tag{25}$$

where Γ_{H_2} represents the unit price of hydrogen; $m_{eq}(t)$ denotes the equivalent hydrogen consumption of the entire vehicle, incorporating the battery energy consumption converted into hydrogen demand to account for its impact on overall energy use; $SoC(t)$ refers to the current battery state of charge (%), with a corresponding penalty term modeled in quadratic form to encourage maintaining SoC within a target range; $\Delta SoH_{fc}(t)$ and $\Delta SoH_{bat}(t)$, respectively, denote the state of health degradation per percentage for the fuel cell and lithiumion battery; $T_{bat}(t)$ and $T_{fc}(t)$ represent the temperature penalty functions for the lithiumion battery and fuel cell, aiming to guide the control policy to operate within optimal temperature ranges, thereby improving training efficiency and reducing component aging rates.

The detailed formulations of all the terms involved are adapted from the theoretical framework of [33] as follows:

$$\begin{cases} \Gamma_{H_2} m_{eq}(t) = \Gamma_{H_2} \left(\frac{1}{\eta_{fc} \rho_{H_2}} P_{fc}(t) + \lambda_{bat}(t) P_{bat}(t) \right) \\ \Gamma_{bat} SoC(t) = -\mu_1 (SoC(t) - SoC_{int})^2 \\ \Gamma_{fc} \Delta SoH_{fc}(t) = \Delta SoH_{fc}(t) \cdot \Gamma_{fc} / \Gamma_{H_2} \\ \Gamma_b \Delta SoH_{bat}(t) = \Delta SoH_{bat}(t) \cdot \Gamma_b / \Gamma_{H_2} \\ \Gamma_f T_{fc}(t) = \begin{cases} 0.05 \times \frac{|T_{fc} - 67|}{50}, & if \quad 65 < T_{fc} < 70 \\ -0.5 \times \frac{|T_{fc} - 67|}{50}, & otherwise \end{cases} \\ \Gamma_{ba} T_{bat}(t) = \begin{cases} 0.05 \times \frac{|T_b - 33|}{50}, & if \quad 29 < T_b < 36 \\ -0.5 \times \frac{|T_b - 33|}{50}, & otherwise \end{cases} \end{cases} \tag{26}$$

where η_{fc} denotes the global efficiency of the fuel cell system, which varies with fuel cell power; ρ_{H_2} represents the lower heating value of hydrogen; Γ_{fc} and Γ_{bat} denote the unit prices of the fuel cell stack and battery pack, respectively; T_b and T_{fc} are the current temperatures (in °C) of the lithium-ion battery and fuel cell, respectively; SoC_{int} is the ideal initial value of the battery SoC; $P_{fc}(t)$ and $P_{bat}(t)$ are the output powers of the fuel cell and battery; μ_1 is the weighting factor for the SoC penalty term. The detailed parameter values are shown in Table 2.

Notably, for the temperature fault diagnosis-based penalty functions, a small positive penalty is applied when the temperature lies within the optimal operating range to encourage stable operation; However, once the temperature exceeds safety thresholds, a larger negative penalty is introduced to avoid performance degradation or component damage due to extreme temperatures. In the context of DRL-based EMS for FCHEVs, the design of the state and action spaces plays a critical role in enabling optimal energy distribution and system efficiency. A well-defined state space allows the agent to perceive real-time vehicle operating conditions, while an effective action space facilitates precise control over powertrain components.

This paper leverages DSAC to train the energy management agent FCHEVs, where the state space is expanded to encompass powertrain dynamics. The proposed framework integrates real-time power demand, SoH, and thermal states into the control system, enabling an EMS with near-global optimality.

Table 2. Parameters in Equation (26) and their definitions.

Parameter	Physical Meaning	Typical Value/Range	Unit
η_{fc}	Global efficiency of the fuel cell system	0.45–0.60	—
$\rho_{\mathrm{H_2}}$	Lower heating value (LHV) of hydrogen	33.3	—
Γ_{fc}	Fuel cell degradation per percentage point of SOH loss	750	\$
$\Gamma_{\mathrm{H_2}}$	Unit price of hydrogen	16	\$/kg
Γ_{bat}	Unit price of the lithiumion battery pack	200	\$/kWh
Γ_{b}	Lithiumion battery degradation per percentage point of SOH loss	375	\$
Γ_{ba}	Lithiumion battery temperature penalty weight	-10	—
Γ_{f}	Fuel cell temperature penalty weight	-10	—
μ_1	Weighting factor for the SoC penalty term	10	—
T_{fc}	Current operating temperature of the fuel cell	Optimal: 65–70 °C	°C
T_b	Current operating temperature of the lithium-ion battery	Optimal: 29–36 °C	°C
SoC_{init}	Ideal initial state-of-charge of the battery	0.5–0.6	—
$P_{\mathrm{fc}}(t)$	Real-time output power of the fuel cell	Varies with demand	kW
$P_{\mathrm{bat}}(t)$	Real-time output power of the battery	Varies with demand	kW

DSAC learns a stochastic policy that maximizes both expected reward and entropy, thereby achieving more robust exploration and better adaptability under complex driving conditions. The actor network is trained to output a probability distribution over the fuel cell power output $P_{fc} \in [0, P_{\max}]$, denoted as $\pi_\theta(a|s_t)$, from which actions are sampled without explicit noise injection. Before training begins, the replay memory D with capacity M is initialized. Transition tuples (s_t, a_t, r_t, s_{t+1}) collected during interactions are stored in the experience replay buffer for off-policy learning. When the buffer reaches its maximum capacity, older data are replaced by newer transitions.

Instead of estimating the expected Q-value, DSAC models the entire distribution of returns associated with each (s_t, a_t) pair. Specifically, the critic network outputs a categorical distribution over a set of discrete atoms representing possible Q-values. This distribution is updated using the distributional Bellman operator:

$$Z_{\theta,w}(s_t, a_t) \leftarrow r_t + \gamma Z_{\theta',w'}(s_{t+1}, a_{t+1}) \tag{27}$$

where $a_{t+1} \sim \pi_{\theta'}(\cdot|s_{t+1})$, and $Z_{\theta',w'}$ denotes the target distribution parameterized by the target networks.

The critic loss is defined as the Kullback–Leibler (KL) divergence between the predicted distribution and the projected target distribution:

$$J(w) = \mathbb{E}_{s_t,a_t,r_t,s_{t+1}}\left[D_{KL}\left(\mathcal{T}^{target} \| Z_w(s_t, a_t)\right)\right] \tag{28}$$

where $\mathcal{T}^{target}$ refers to the distribution obtained by projecting the target values onto the support of the learned distribution.

The actor network is updated by minimizing the KL divergence between the current action distribution and the exponentiated Q-values:

$$\begin{aligned} \nabla_\theta J(\theta) &= \mathbb{E}_{s_t}[\nabla_a \log \pi_\theta(a|s_t) \cdot Q(s_t, a) + \alpha \nabla_a \log \pi_\theta(a|s_t)] \\ &\approx \frac{1}{M} \sum_{m=1}^{M} \nabla_a \log \pi_\theta(a_j|s_j) \cdot \left(Q(s_j, a_j) + \alpha \log \pi_\theta(a_j|s_j)\right) \end{aligned} \tag{29}$$

The target networks are updated using a slow-moving average mechanism parameterized by σ_θ and σ_w, such as in the rule:

$$\begin{aligned} w' &\leftarrow \sigma_w w + (1-\sigma_w) w', \\ \theta' &\leftarrow \sigma_\theta \theta + (1-\sigma_\theta)\theta'. \end{aligned} \tag{30}$$

This means that the new target actor parameters (w') are a weighted average of the current evaluate actor parameters (w') and the previous target actor parameters (w'). The weight σ_w determines how much the target network "follows" the evaluate network. Similarly, the new target critic parameters (θ') are a weighted average of the current evaluate critic parameters (θ) and the previous target critic parameters (θ').

3.2. Safe Constriction Control System

To mitigate the safety issues of FCHEVs caused by the unconstrained exploration of RL agents (such as battery overcharging/overdischarging and overheating), a thermal safe constriction controller (TSCC) system has been designed between the RL agent and the control system. This system ensures the reliable operation of the energy sources. The specific design process is as follows: Define the state variables P_{bat}, T_F, T_{bat}, the output variables P_{bat}, T_F, T_{bat}, the control variables $u_2 = P_{fc}$, and the external disturbances $\omega_2 = P_{req}$ of the energy management system. A flexible constraint controller is designed based on the state variables, as follows:

$$\begin{aligned} \min_{P'_{fc}} \quad & J_{Tscc} = \left[P'_{fc}(t) - P_{fc}(t)\right]^2 \\ s.t. & T_{fc}(t+1) = T_{fc}(t) + \frac{P_{fc,loss}(t) - Q_{fc,coolant}(t) - Q_{fc,e}(t)}{m_{fc} \cdot C_{s,fc}}\, dt \\ & T_{lib}(t+1) = T_{lib}(t) + \frac{P_{lib,loss}(t) - Q_{lib,coolant}(t)}{m_{lib} \cdot C_{s,lib}}\, dt \\ & P_{bat,min} \le P'_{bat}(t) \le P_{bat,max} \\ & P_{fc,min} \le P'_{fc}(t) \le P_{fc,max} \\ & P'_{fc}(t-1) - 100 \le P'_{fc}(t) \le P'_{fc}(t-1) + 100 \end{aligned} \tag{31}$$

In this equation, P'_{fc} represents the output of the TSCC system under the safety intervention of the car-following control system. lt is important to note that when the system state satisfies the safety constraints under the action $P_{fc}(t)$ of the vehicle control agent, then $P'_{fc} = P_{fc}(t)$, indicating that TSCC does not interfere with the agent's autonomous exploration. The TSCC becomes active only when the system state fails to meet the safety constraints under the action $P_{fc}(t)$. In such cases, TSCC projects $P_{fc}(t)$ onto the nearest action P'_{fc} that satisfies the system's safety constraints through quadratic programming, thereby minimizing the impact on the agent's exploration process and ensuring the global optimality of the solution. Furthermore, the action $P'_{fc}(t)$ after safety intervention, along with the corresponding subsequent state, will be stored in the experience replay buffer, which in turn guides the agent in learning the system's safety constraints.

The pseudo-code of the DSAC-based energy management strategy with energy source thermal constraints is presented in Algorithm 1.

Algorithm 1 DSAC-based energy management and energy source thermal constraint algorithm

Input: Required power P_{req}, State of Charge (SoC) of lithium battery
Output: Fuel cell output power P_{fc}
Initialize parameters: maximum training episodes $\mathcal{E}_{\max}$, maximum steps per episode $n_{\max}$, actor learning rate α_a, critic learning rate α_c, temperature parameter learning rate α_τ, reward discount factor γ, target network soft update ratio τ, memory pool capacity $\mathcal{M}$, batch size B, target entropy H_{target}
Initialize actor and critic neural networks for both the vehicle control agent and the energy management agent
Initialize corresponding target networks and experience replay buffer D_1
Load $n_{\max}$ time steps of leading vehicle driving condition data
for $e = 1$ to $\mathcal{E}_{\max}$ **do**
    Initialize the initial observation state of the fuel cell electric hybrid vehicle (FCEHV)
    for $k = 1$ to $n_{\max}$ **do**
        Determine the energy management action using the DSAC algorithm:
            $P_{fc}(t) = P_{fc}(t-1) + 100 \times a(t)$
        if $T_{bat} \leq T_{bat_max}$ **then**
            Retain the original control action: $P_{fc}(t)$
        else
            Compute the safe control action $P'_{fc}(t)$ according to Equation (31)
            Deploy safe action: $P_{fc}(t) = P'_{fc}(t)$
        end if
        Execute $P_{fc}(t)$ in the EMS, observe next state $s(t+1)$, reward $R(t)$
        Store transition tuple $[s(t), P_{fc}(t), s(t+1), R(t)]$ into replay buffer D_1
        Update critic network using the defined loss function
        Update actor network via gradient descent with learning rate α_a
        Perform soft update on target networks with ratio τ
    end for
    Compute and record average reward: $\frac{1}{n_{\max}} \sum_{k=1}^{n_{\max}} R(k)$
end for

4. Results and Discussion

This paper employs the 2WLTC driving cycle as the benchmark test condition to evaluate the performance of DRL-based EMSs, including the DDPG, TD3, SAC, and DSAC approaches. Each strategy incorporates a multi-objective reward function that considers fuel economy optimization, fuel cell and lithium battery pack lifespan, and thermal management. Particularly, the implementation of TSCC achieves coordinated optimization between system thermal safety and energy distribution efficiency under complex operating conditions.

4.1. Verification Preparation

To support the presentation of subsequent results, Table 3 summarizes the key differences among various DRL algorithms, while Table 4 lists the DRL-based energy management strategies (EMSs) considered in this paper. Specifically, the proposed strategy in this paper is the DSAC-based EMS, which takes into account the thermal, thermal safety, and durability characteristics of the powertrain, along with flexible thermal safety constraints. The DRL-based algorithms, such as DDPG, TD3, SAC, and DSAC, focus on fuel consumption, battery life, and self-heating of lithium-ion batteries and hydrogen fuel cells in the reward function. The definitions of the DRL-based algorithms DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCC are similar to the previous EMSs, with the main difference being the use of the thermal safety flexible constraints. The driving scenario dataset is shown in Figure 3, where the 2WLTC driving cycle is used for training.

Table 3. Comparison of DRL algorithms for EMS.

Algorithm	Distributional Learning	Maximum Entropy	Double Q-Network	Continuous Action	Suitable for EMS
DDPG	no	no	no	yes	yes
TD3	no	no	yes	yes	yes
SAC	no	yes	yes	yes	yes
DSAC	yes	yes	no	yes	yes

Table 4. EMSs algorithm table.

Algorithm	Fuel Consumption	Temperature and SoH	Thermal Safety Constraints
DDPG	Considering	Considering	Neglecting
TD3	Considering	Considering	Neglecting
SAC	Considering	Considering	Neglecting
DSAC	Considering	Considering	Neglecting
DDPG_TSCC	Considering	Considering	Considering
TD3_TSCC	Considering	Considering	Considering
SAC_TSCC	Considering	Considering	Considering
DSAC_TSCC	Considering	Considering	Considering

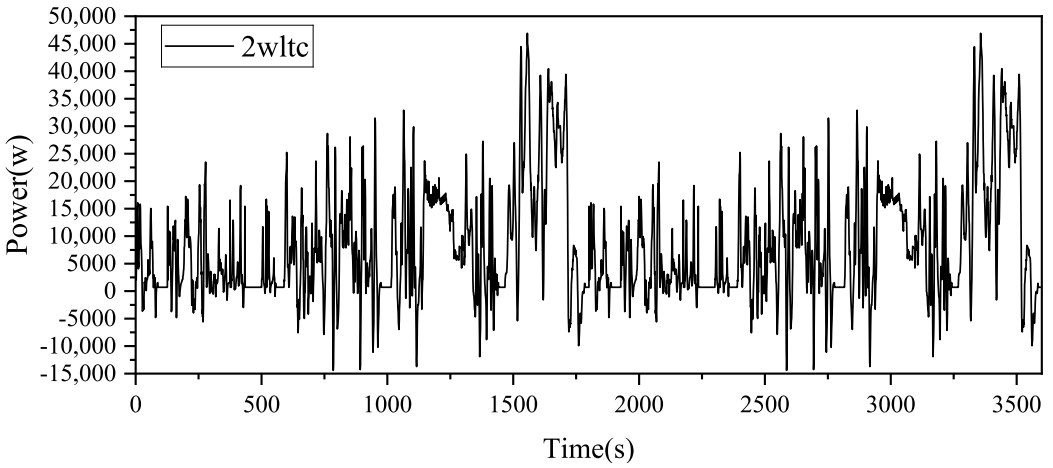


Figure 3. 2wltc working condition.

4.2. Energy Distribution Control Verification

Figures 4–8 compare the power distribution performance between the LIBs and PEMFCs stack for the online control of DRL-based EMS. As shown in Figure 4, by using the DSAC-based EMS, the power distribution of the PEMFCs stack is smoother. This indicates that the implementation of the DSAC algorithm, along with the consideration of the powertrain's thermal and durability characteristics, enables the power distribution to remain more concentrated within the safe operating range of the PEMFCs system, thereby reducing the degradation of the fuel cell's lifespan due to power fluctuations. As seen in Figures 5 and 7, when the powertrain's thermal factors are incorporated, DDPG, TD3, SAC, and DSAC algorithms make certain inefficient decisions in power distribution to reduce the generation of heat in the fuel cell. Figures 6 and 8 show that by using the TSCC-based EMS, the power distribution points of the PEMFCs stack become more concentrated, meaning that the implementation of the TSCC algorithm helps focus the power distribution in the higher efficiency region of the PEMFCs system, with the DSAC algorithm showing the most

prominent effect. Figure 9 illustrates the frequency distribution of battery power demand across different strategies, highlighting that DSAC-EMS significantly reduces occurrences of high-magnitude power transients—thereby mitigating thermal runaway and aging. Figure 10 shows the battery power trajectories when the TSCC is activated. Compared with Figure 9, the power trajectories of the four algorithms are relatively more concentrated, illustrating how the TSCC actively restricts extreme power outputs to maintain thermal safety. Figure 11 demonstrate that all DRL-based EMSs are capable of maintaining the SoC trajectory and returning the SoC to a level close to the initial value of the LIBs.

Furthermore, the terminal hydrogen consumption for each EMS is listed in Table 5. Due to the exceptional efficiency of the PEMFCs system, the proposed EMS results in lower hydrogen consumption. In Table 5, the corrected hydrogen consumption, expressed as the overall fuel economy of the FCV powertrain, is composed of the hydrogen consumption of the PEMFCs stack and the equivalent hydrogen consumption caused by the varying terminal SoC values of the LIBs. The formula for the corrected hydrogen consumption C_{H_2} is given as follows:

$$C_{H_2} = C_{fc,H_2} + C_{bat,H_2} \tag{32}$$

The equivalent hydrogen costs of the PEMFCs stack and LIBs are represented by C_{fc,H_2} and C_{bat,H_2}, respectively. The total cost of the PEMFCs stack and LIBs corresponding to the DSAC-based EMS is 421.776 g, which is lower compared to other EMS based on DRL.

Table 5. Performance indicators of 2wltc driving cycle under different EMS strategies.

Strategy/Driving Cycle	Algorithm	Effective Fuel Consumption (kg)	Energy Loss ($)
RL WLTC	DDPG	0.503	0.494
	TD3	0.689	0.514
	SAC	0.613	0.502
	DSAC	0.414	0.363
TSCC-RL WLTC	DDPG	0.503	0.464
	TD3	0.675	0.495
	SAC	0.458	0.400
	DSAC	0.422	0.357

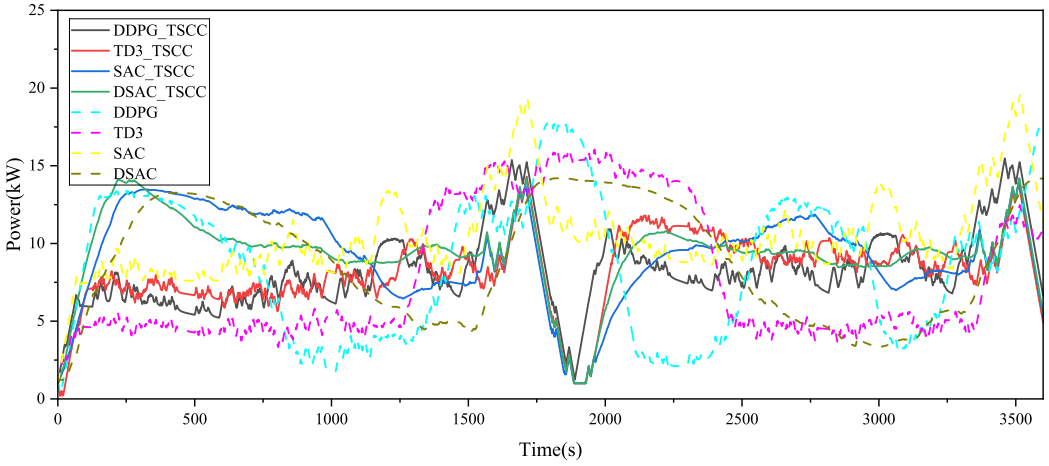


Figure 4. The power distribution results of the fuel cell when using the DDPG, TD3, SAC, DSAC, DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCC algorithms.

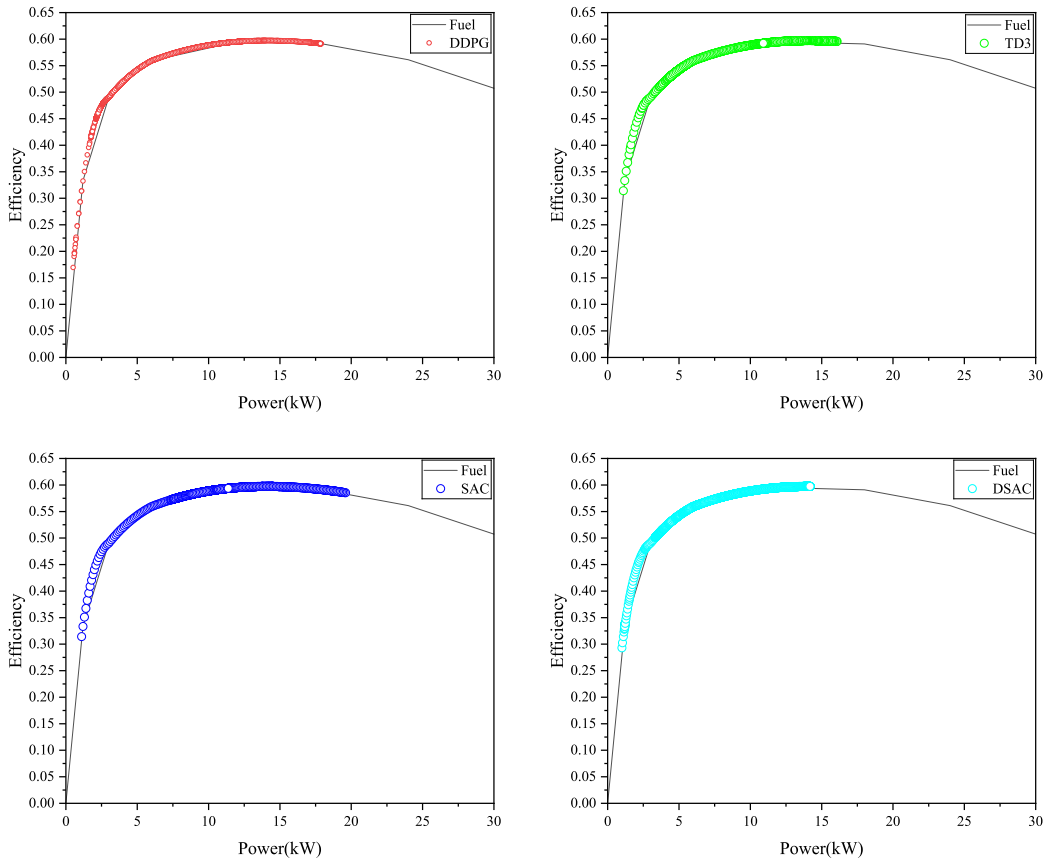


Figure 5. The efficiency distribution diagram of the fuel cell when using the DDPG, TD3, SAC, and DSAC algorithms.

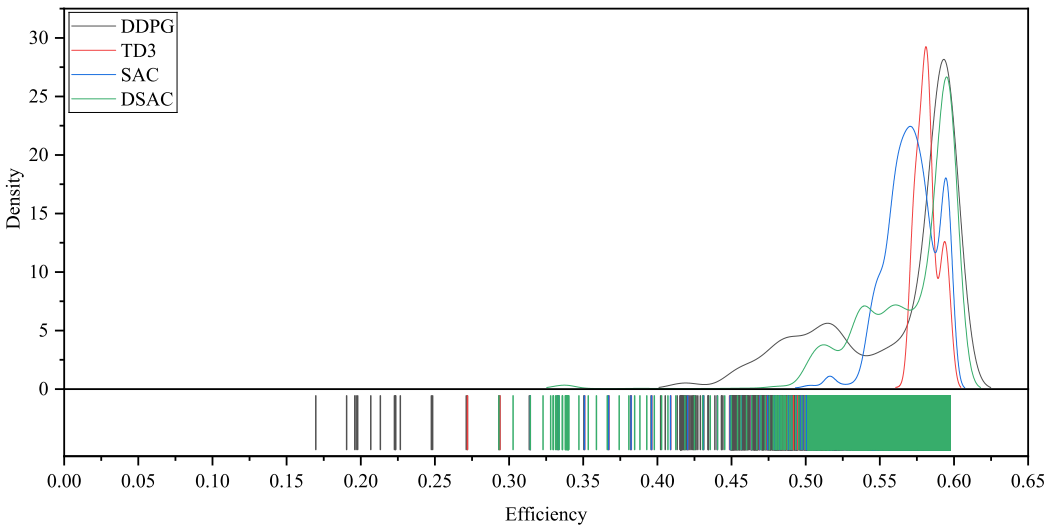


Figure 6. The efficiency distribution statistics of the fuel cell when using the DDPG, TD3, SAC, and DSAC algorithms.

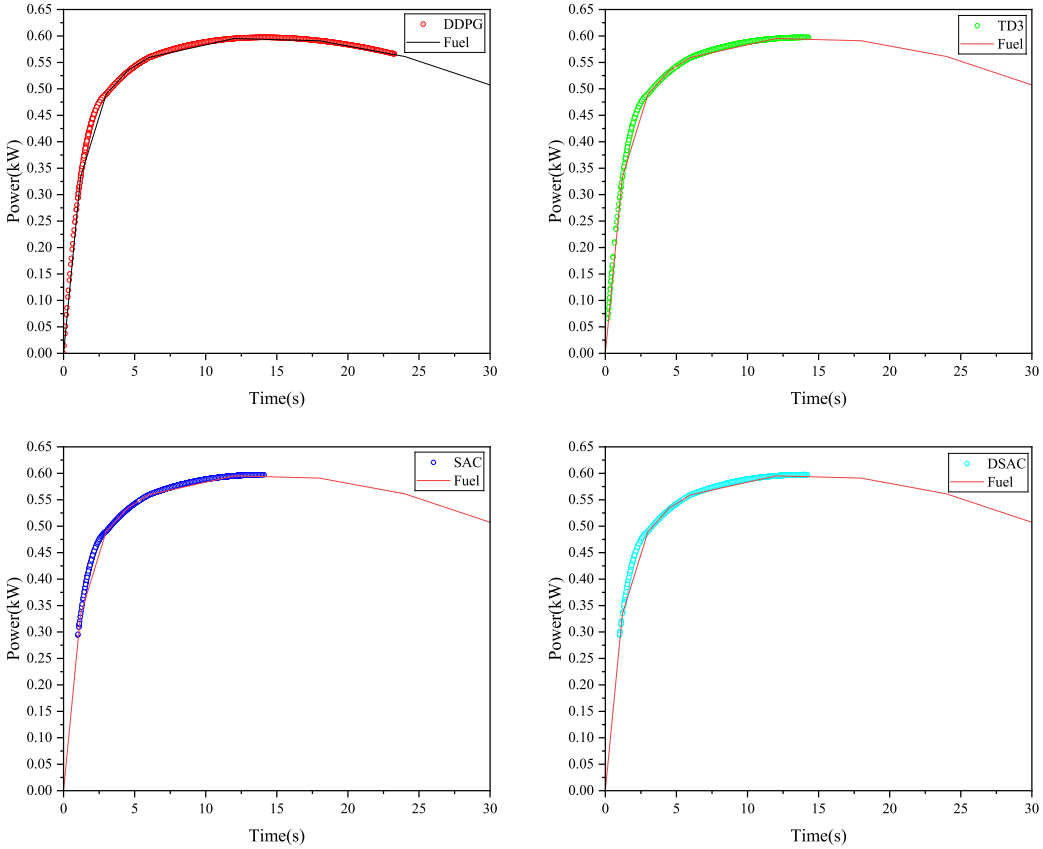


Figure 7. The efficiency distribution diagram of the fuel cell when using the DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCC algorithms.

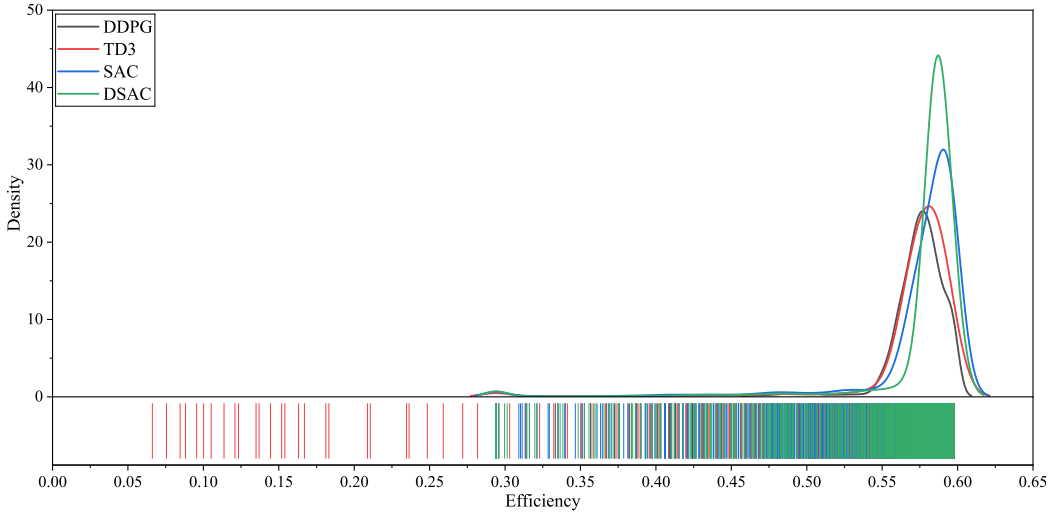


Figure 8. The efficiency distribution statistics of the fuel cell when using the DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCC algorithms.

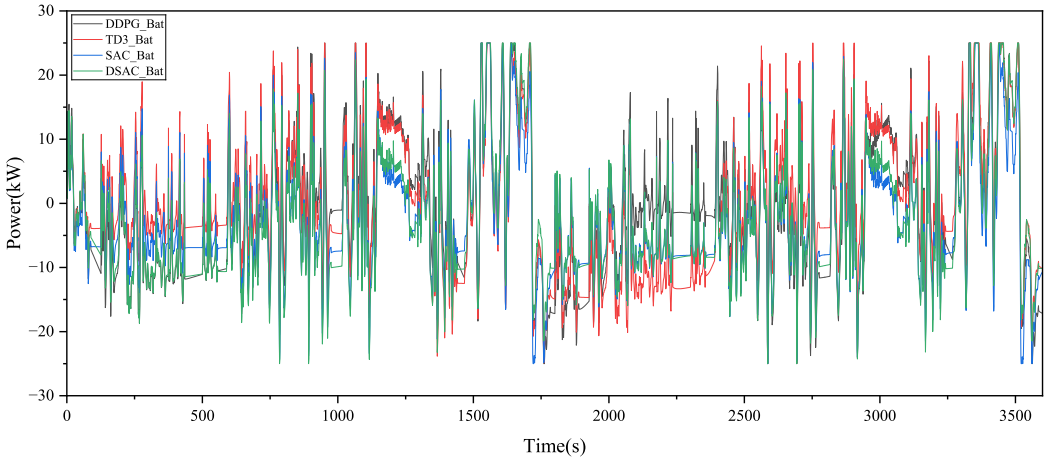


Figure 9. The power distribution results of the lithium-ion battery when using the DDPG, TD3, SAC, and DSAC algorithms.

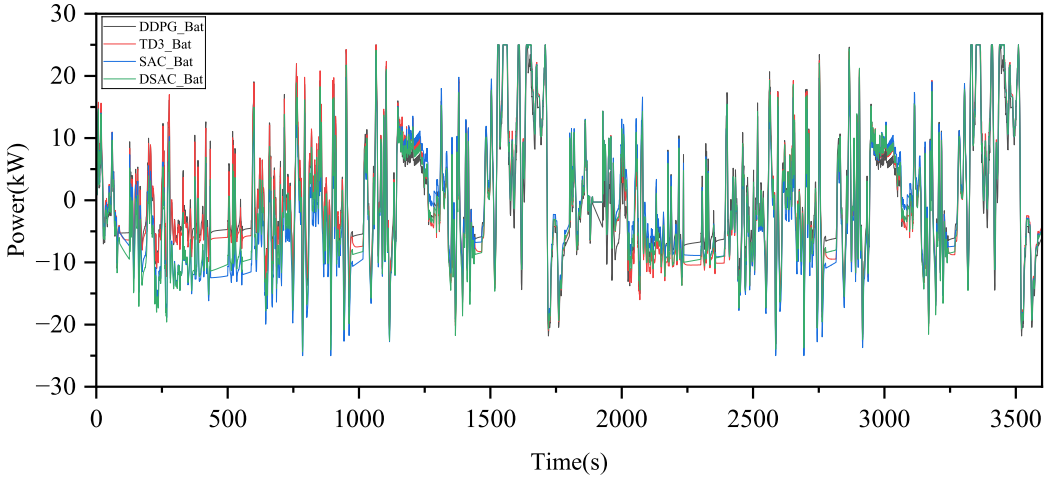


Figure 10. The power distribution results of the lithiumion battery when using the DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCC algorithms.

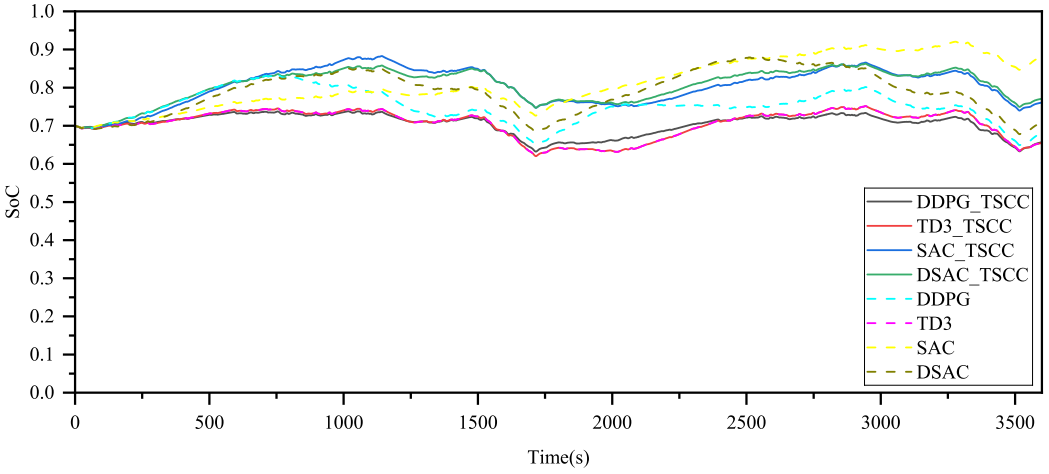


Figure 11. The SoC variation curves of lithiumion batteries when using the DDPG, TD3, SAC, DSAC, DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCCalgorithms.

4.3. Temperature Control Verification

Figures 12 and 13 present the comparative temperature distributions of the PEMFCs stack and LIBs. The results indicate that the operating temperature ranges for the PEMFCs stack and LIBs are 60–76 °C and 25–45 °C, respectively. Since both the PEMFCs stack and

LIBs operate within the temperature ranges of 60–80 °C and 20–40 °C, respectively, the temperature curves of the PEMFCs stack and LIBs controlled by the EMS based on DDPG, TD3, SAC, and DSAC are reasonable. In practical engineering, if the battery temperature exceeds 50 °C, the BMS will force a shutdown, halting further operation. The EMS based on DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCC can effectively regulate the temperature of the LIBs and PEMFCs stack, limiting their maximum temperatures. The temperature curves are smoother, preventing excessive temperatures that could lead to additional losses, while also accounting for the thermal characteristics of the LIBs and PEMFCs stack to avoid overheating.

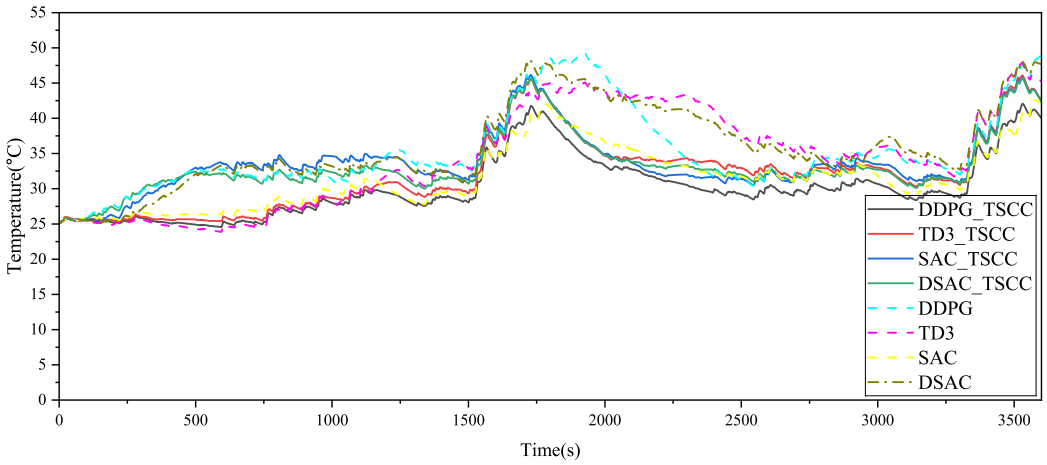


Figure 12. The temperature distribution results of the lithiumion battery when using the DDPG, TD3, SAC, DSAC, DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCC algorithms.

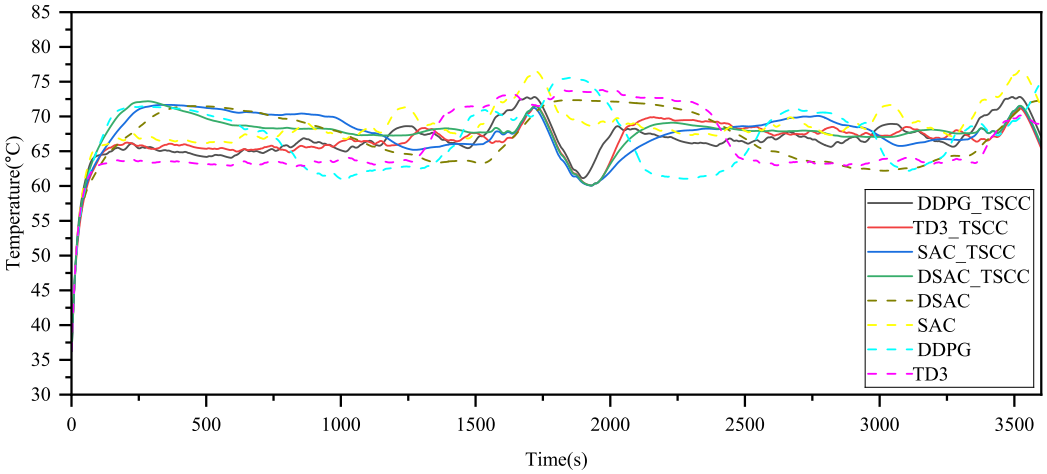


Figure 13. The temperature distribution results of the PEMFCs stack when using the DDPG, TD3, SAC, DSAC, DDPG_TSCC, TD3_TSCC, SAC_TSCC, and DSAC_TSCC algorithms.

5. Conclusions

A DSAC-based thermo-electrical co-optimization EMS for FCHEVs is proposed, which explicitly integrates thermal runaway fault diagnosis and a TSCC to address thermal management and safety challenges by jointly considering fuel cell powertrain thermal characteristics, durability, and real-time fault constraints. The overheating penalty of the LIBs and PEMFCs stack, as well as the equivalent SoH degradation cost, are added to the reward space alongside conventional metrics, which include hydrogen consumption and SoC of LIBs penalties used for optimization and training. Additionally, the DSAC-based EMS introduces a single Q-learning technique to prevent Q-value overestimation in the soft Q-network. The validation of the proposed EMS is conducted in several long-term

driving scenarios, covering both the training process and online control, with the online control aspects involving fuel economy, thermal stability, and SoH for the PEMFCs stack and lithiumion battery. Simulation results show the following:

1. Under the same operating conditions, considering the thermal characteristics and durability of the powertrain system, the energy consumption saved by the EMS based on DSAC is 17.69%, 39.91%, and 32.46% compared to those based on DDPG, TD3, and SAC, respectively. The SoH consumption is reduced by 26.52%, 29.38%, and 27.69%, respectively. After adding TSCC, the EMS based on DDPG, TD3, SAC, and DSAC achieves energy savings of 0%, 2.03%, 25.29%, and −1.93%, respectively, while SoH consumption is reduced by 6.07%, 3.7%, 20.32%, and 1.65%, respectively.
2. The proposed flexible temperature TSCC effectively regulates the temperature of the LIBs and PEMFCs stack, minimizing their peak temperatures. However, without the introduction of a thermal management system, the energy source temperature cannot be more effectively maintained within the optimal range, which may prevent reductions in energy consumption and mitigate potential lifespan degradation.
3. The method has good versatility and can be extended to other types of FCVs to achieve optimal fuel economy, durability, and thermal safety of the fuel cell vehicle power system in unpredictable driving scenarios.

Despite the promising results achieved in this paper, several limitations warrant further investigation. First, the strategy is developed and evaluated using standard driving cycles (e.g., WLTC, UDDS) under static conditions, focusing on thermal management. However, real-world driving is highly dynamic, with varying traffic, weather, terrain, and driver behavior, which can challenge the DSAC algorithm under thermal constraints. To address this, future work should integrate multi-source data—such as traffic, weather, terrain, and in-vehicle CAN signals (e.g., energy use, temperatures, and driving dynamics)—to develop a more robust and generalizable energy management strategy. This will support validation in complex real-world conditions. To advance toward practical deployment, we plan to conduct hardware-in-the-Loop simulation by deploying the DSAC strategy on a real-time controller and conducting co-simulation with physical models of the fuel cell system, battery management system, and thermal management system. In the long term, the strategy will be further validated on a real vehicle platform to evaluate its energy efficiency, thermal safety, and durability performance. Second, the training of deep reinforcement learning algorithms presents several challenges, including the risk of overfitting to specific driving conditions. Advanced transfer learning and domain adaptation techniques could be investigated to enhance the generalization capability of the strategy across different vehicle models and operating scenarios. Finally, extending this framework to a broader range of new energy vehicle platforms (such as battery electric vehicles and plug-in hybrid electric vehicles) to validate its universality represents an important research direction. Furthermore, integrating our energy management model with intelligent fleet management systems (e.g., the multi-agent collaborative thermal management approach proposed by Khalatbarisoltani et al. [29]) may enable more efficient energy co-optimization.

Author Contributions: Methodology, Y.W.; Software, Y.W.; Data curation, Y.W.; Writing—original draft, Y.W.; Writing—review & editing, Y.W., F.T., L.Z., N.W. and Z.F.; Visualization, Y.W.; Funding acquisition, F.T. and Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (Grant No. 62301212, 62371182), the Program for Science and Technology Innovation Talents in the University of Henan Province (Grant No. 23HASTIT021), the Major Science and Technology Projects of Longmen Laboratory (Grant No. 231100220200), Aeronautical Science Foundation of China (Grant No. 20220001042002), the Science and Technology Development Plan of Joint Research Program of

Henan (Grant No. 222103810036), the Scientific and Technological Project of Henan Province (Grant No. 222102210056, 222102240009, 252102241051).

Data Availability Statement: The data are available from the corresponding author on reasonable request.

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

References

1. Mazouzi, A.; Hadroug, N.; Alayed, W.; Hafaifa, A.; Iratni, A.; Kouzou, A. Comprehensive optimization of fuzzy logic-based energy management system for fuel-cell hybrid electric vehicle using genetic algorithm. *Int. J. Hydrogen Energy* **2024**, *81*, 889–905. [CrossRef]
2. Sun, W.; Zou, Y.; Zhang, X.; Guo, N.; Zhang, B.; Du, G. High robustness energy management strategy of hybrid electric vehicle based on improved soft actor-critic deep reinforcement learning. *Energy* **2022**, *258*, 124806. [CrossRef]
3. Ma, Y.; Ma, Q.; Liu, Y.; Gao, J. Adaptive optimization control strategy for electric vehicle battery thermal management system based on pontryagin's minimal principle. *IEEE Trans. Transp. Electrif.* **2023**, *10*, 3855–3869. [CrossRef]
4. Rudolf, T.; Schürmann, T.; Schwab, S.; Hohmann, S. Toward holistic energy management strategies for fuel cell hybrid electric vehicles in heavy-duty applications. *Proc. IEEE* **2021**, *109*, 1094–1114. [CrossRef]
5. Pu, Z.; Jiao, X.; Yang, C.; Fang, S. An adaptive stochastic model predictive control strategy for plug-in hybrid electric bus during vehicle-following scenario. *IEEE Access* **2022**, *8*, 13887–13897. [CrossRef]
6. Wu, Y.; Huang, Z.; Hofmann, H.; Liu, Y.; Huang, J.; Hu, X.; Song, Z. Hierarchical predictive control for electric vehicles with hybrid energy storage system under vehicle-following scenarios. *Energy* **2022**, *251*, 123774. [CrossRef]
7. Jia, C.; Qiao, W.; Cui, J.; Qu, L. Adaptive model-predictive-control-based real-time energy management of fuel cell hybrid electric vehicles. *IEEE Trans. Power Electron.* **2022**, *38*, 2681–2694. [CrossRef]
8. Shen, Y.; Xie, J.; He, T.; Yao, L.; Xiao, Y. CEEMD-fuzzy control energy management of hybrid energy storage systems in electric vehicles. *IEEE Trans. Energy Convers.* **2023**, *39*, 555–566. [CrossRef]
9. Velimirović, L.Z.; Janjić, A.; Vranić, P.; Velimirović, J.D.; Petkovski, I. Determining the optimal route of electric vehicle using a hybrid algorithm based on fuzzy dynamic programming. *IEEE Trans. Fuzzy Syst.* **2022**, *31*, 609–618. [CrossRef]
10. Da Silva, S.F.; Eckert, J.J.; Corrêa, F.C.; Silva, F.L.; Silva, L.C.; Dedini, F.G. Dual HESS electric vehicle powertrain design and fuzzy control based on multi-objective optimization to increase driving range and battery life cycle. *Appl. Energy* **2022**, *324*, 119723. [CrossRef]
11. Wang, C.; Liu, R.; Tang, A. Energy management strategy of hybrid energy storage system for electric vehicles based on genetic algorithm optimization and temperature effect. *J. Energy Storage* **2022**, *51*, 104314. [CrossRef]
12. Xu, Y.; Zhang, H.; Yang, Y.; Zhang, J.; Yang, F.; Yan, D.; Wang, Y. Optimization of energy management strategy for extended range electric vehicles using multi-island genetic algorithm. *J. Energy Storage* **2023**, *61*, 106802. [CrossRef]
13. Liu, T.; Tan, K.; Zhu, W.; Feng, L. Computationally efficient energy management for a parallel hybrid electric vehicle using adaptive dynamic programming. *IEEE Trans. Intell. Veh.* **2023**, *9*, 4085–4099. [CrossRef]
14. Min, D.; Song, Z.; Chen, H.; Wang, T.; Zhang, T. Genetic algorithm optimized neural network based fuel cell hybrid electric vehicle energy management strategy under start-stop condition. *Appl. Energy* **2022**, *306*, 118036. [CrossRef]
15. Ganesh, A.H.; Xu, B. A review of reinforcement learning based energy management systems for electrified powertrains: Progress, challenge, and potential solution. *Renew. Sustain. Energy Rev.* **2022**, *154*, 111833. [CrossRef]
16. Sotoudeh, S.M.; HomChaudhuri, B. A deep-learning-based approach to eco-driving-based energy management of hybrid electric vehicles. *IEEE Trans. Transp. Electrif.* **2023**, *9*, 3742–3752. [CrossRef]
17. 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]
18. Deng, K.; Liu, Y.; Hai, D.; Peng, H.; Löwenstein, L.; Pischinger, S.; Hameyer, K. Deep reinforcement learning based energy management strategy of fuel cell hybrid railway vehicles considering fuel cell aging. *Energy Convers. Manag.* **2022**, *251*, 115030. [CrossRef]
19. Jia, C.; Liu, W.; He, H.; Chau, K.T. Deep reinforcement learning-based energy management strategy for fuel cell buses integrating future road information and cabin comfort control. *Energy Convers. Manag.* **2024**, *321*, 119032. [CrossRef]
20. Wang, Y.; Wu, J.; He, H.; Wei, Z.; Sun, F. Data-driven energy management for electric vehicles using offline reinforcement learning. *Nat. Commun.* **2025**, *16*, 2835. [CrossRef]
21. Shi, D.; Xu, H.; Wang, S.; Hu, J.; Chen, L.; Yin, C. Deep reinforcement learning based adaptive energy management for plug-in hybrid electric vehicle with double deep Q-network. *Energy* **2024**, *305*, 132402. [CrossRef]
22. Engel, J.; Schmitt, T.; Rodemann, T.; Adamy, J. Hierarchical MPC for building energy management: Incorporating data-driven error compensation and mitigating information asymmetry. *Appl. Energy* **2024**, *372*, 123780. [CrossRef]

23. Cheng, J.; Yang, F.; Zhang, H.; Yang, A.; Xu, Y. Multi-objective adaptive energy management strategy for fuel cell hybrid electric vehicles considering fuel cell health state. *Appl. Therm. Eng.* **2024**, *257*, 124270. [CrossRef]
24. Qi, Y.; Xu, X.; Liu, Y.; Pan, L.; Liu, J.; Hu, W. Intelligent energy management for an on-grid hydrogen refueling station based on dueling double deep Q network algorithm with NoisyNet. *Renew. Energy* **2024**, *222*, 119885. [CrossRef]
25. Ouyang, T.; Jin, S.; Xie, X.; Gong, Y.; Zhang, Z. Adaptive Energy Management in Dual-Motor Electric Vehicles Using Deep Deterministic Policy Gradient. *IEEE Trans. Transp. Electrif.* **2025**, *11*, 12647–12656. [CrossRef]
26. Chen, B.; Wang, M.; Hu, L.; Zhang, R.; Li, H.; Wen, X.; Gao, K. A hierarchical cooperative eco-driving and energy management strategy of hybrid electric vehicle based on improved TD3 with multi-experience. *Energy Convers. Manag.* **2025**, *326*, 119508. [CrossRef]
27. Wang, J.; Du, C.; Yan, F.; Duan, X.; Hua, M.; Xu, H.; Zhou, Q. Energy Management of a Plug-In Hybrid Electric Vehicle Using Bayesian Optimization and Soft Actor-Critic Algorithm. *IEEE Trans. Transp. Electrif.* **2024**, *11*, 912–921. [CrossRef]
28. Zhang, H.; Chen, B.; Lei, N.; Li, B.; Li, R.; Wang, Z. Integrated thermal and energy management of connected hybrid electric vehicles using deep reinforcement learning. *IEEE Trans. Transp. Electrif.* **2023**, *10*, 4594–4603. [CrossRef]
29. Khalatbarisoltani, A.; Han, J.; Liu, W.; Liu, C.Z.; Hu, X. Health-consciousness integrated thermal and energy management of connected hybrid electric vehicles using cooperative multi-agent deep reinforcement learning. *IEEE Trans. Intell. Veh.* **2024**, 1–12. [CrossRef]
30. Han, J.; Shu, H.; Tang, X.; Lin, X.; Liu, C.; Hu, X. Predictive energy management for plug-in hybrid electric vehicles considering electric motor thermal dynamics. *Energy Convers. Manag.* **2022**, *251*, 115022. [CrossRef]
31. Abbasi, M.H.; Arjmandzadeh, Z.; Zhang, J.; Xu, B.; Krovi, V. Deep reinforcement learning based fast charging and thermal management optimization of an electric vehicle battery pack. *J. Energy Storage* **2024**, *95*, 112466. [CrossRef]
32. Guo, Z.; Wang, Y.; Zhao, S.; Zhao, T.; Ni, M. Modeling and optimization of micro heat pipe cooling battery thermal management system via deep learning and multi-objective genetic algorithms. *Int. J. Heat Mass Transf.* **2023**, *207*, 124024. [CrossRef]
33. Zhang, Y.; Zhang, C.; Fan, R.; Deng, C.; Wan, S.; Chaoui, H. Energy management strategy for fuel cell vehicles via soft actor-critic-based deep reinforcement learning considering powertrain thermal and durability characteristics. *Energy Convers. Manag.* **2023**, *283*, 116921. [CrossRef]
34. Han, L.; Yang, K.; Ma, T.; Yang, N.; Liu, H.; Guo, L. Battery life constrained real-time energy management strategy for hybrid electric vehicles based on reinforcement learning. *Energy* **2022**, *259*, 124986. [CrossRef]
35. Duan, J.; Guan, Y.; Li, S.E.; Ren, Y.; Sun, Q.; Cheng, B. Distributional soft actor-critic: Off-policy reinforcement learning for addressing value estimation errors. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6584–6598. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.