vehicles

MDPI

Article

Hierarchical Energy Management for Fuel Cell Electric Vehicles with Adaptive-Modality Deep Deterministic Policy Gradient

Yantao Si, Zhuo Wang, Changqun Sun, Wen He and Yunge Zou *

Chongqing Changan Automobile Co., Ltd., No. 260, Jianxin East Road, Jiangbei District, Chongqing 400023, China; siyt@changan.com.cn (Y.S.); wangzhuo5@changan.com.cn (Z.W.); suncq@changan.com.cn (C.S.); hewen@changan.com.cn (W.H.)
* Correspondence: yunge_zou@cqu.edu.cn

Abstract

Fuel cell electric vehicles (FCEVs) require energy management strategies that can balance hydrogen economy, battery utilization, component protection, and real-time control under varying driving conditions. This paper proposes an Adaptive-Modality Deep Deterministic Policy Gradient and Model Predictive Control hierarchical energy management strategy (AMDDPG–MPC HEMS). In the proposed architecture, the upper-level AMDDPG controller identifies driving-condition patterns and generates adaptive weights for hydrogen consumption, battery power, and state-of-charge regulation, while the lower-level MPC controller performs constrained power allocation between the fuel cell and battery. To improve adaptability, the AMDDPG algorithm incorporates an adaptive modality perception mechanism that extracts driving-condition features and a multi-scale reward mechanism that coordinates short-term energy-saving objectives with long-term component-protection requirements. A dedicated weight-scheduling and switching mechanism is also introduced to ensure smooth transitions between operating conditions. The proposed strategy is evaluated under the World Light Vehicle Test Cycle and Urban Dynamometer Driving Schedule and compared with rule-based, equivalent consumption minimization, and fixed-weight MPC strategies. The results show that the AMDDPG–MPC HEMS achieves the lowest equivalent hydrogen consumption, with reductions of 18.853% and 11.732% relative to the rule-based strategy under the two driving cycles, respectively. It also improves fuel-cell operating efficiency and maintains feasible battery SOC regulation. These results demonstrate the effectiveness and engineering potential of the proposed hierarchical energy management strategy.

Keywords: fuel cell electric vehicles; energy management; Model Predictive Control; deep reinforcement learning; multi-scale reward

Check for updates

Academic Editors: Teresa Donateo, Peixing Zhang, Jin Chen, Yuande Jiang and Xin Cheng

Received: 29 June 2026
Revised: 20 July 2026
Accepted: 5 August 2026
Published: 28 August 2026

Copyright: © 2026 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.

1. Introduction

Fuel cell electric vehicles (FCEVs), characterized by zero tailpipe emissions, short hydrogen-refueling time, and long driving range, are regarded as a promising solution for sustainable transportation [1]. The energy management strategy (EMS) is a key element of the FCEV powertrain because it directly affects hydrogen consumption, operational safety, and the service life of powertrain components [2]. However, rapidly varying power demand and uncertain driving conditions make it difficult for conventional EMSs to maintain consistently efficient and stable performance. Therefore, an adaptive EMS

capable of coordinating energy efficiency, component protection, and real-time control under diverse driving conditions is required [3,4].

Research on FCEV energy management strategies (EMS) targets efficient, safe power allocation under complex, dynamic conditions. EMS approaches are commonly grouped into rule-based, optimization-based, and learning-based methods [5–7]. Rule-based strategies (e.g., logic-threshold control) offer simplicity, low computational cost, and fast real-time response [8–10] and were pivotal in early applications; however, their manually tuned parameters lack adaptability to time-varying conditions and component aging, limiting global optimality and often reducing control precision. Optimization-based energy management strategies can be categorized into global and instantaneous optimization approaches, both of which generally deliver superior fuel-economy performance compared with rule-based strategies. Global optimization methods mainly include dynamic programming (DP), Pontryagin's Minimum Principle (PMP), and various heuristic algorithms such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) [11–14]. These methods determine the optimal control sequence based on complete or predefined driving-cycle information. However, they typically require full-cycle previews or accurate forecasts and entail heavy computational burdens, which makes them sensitive to modeling/forecast errors and difficult to deploy online across highly variable conditions [15]. Instantaneous optimization methods—such as the Equivalent Consumption Minimization Strategy (ECMS) and Model Predictive Control (MPC)—exhibit satisfactory control performance and strong potential for real-time applications [16,17]. ECMS minimizes instantaneous equivalent fuel consumption by introducing an equivalence factor that converts battery energy consumption into equivalent hydrogen consumption [18]. However, its performance strongly depends on the calibration of this factor, and a fixed value cannot readily adapt to varying driving conditions. MPC, on the other hand, generates optimal control sequences by explicitly considering predicted state evolution within a moving time horizon, showing high robustness under various operating scenarios. Learning-based energy management strategies utilize artificial intelligence and machine learning techniques to optimize power allocation [19]. By learning from historical data and predicting system behavior and external conditions, these strategies can automatically adapt control policies to improve efficiency and overall performance. Learning-based approaches are generally divided into supervised-learning-based and reinforcement-learning-based EMS [20]. In supervised learning, each training sample consists of input–output pairs, and the objective is to establish a mapping function from inputs to outputs that enables accurate prediction of unknown states. Supervised-learning-based EMSs establish input–output mappings from labeled training data, but their performance depends strongly on the availability and representativeness of high-quality datasets. Such datasets are difficult to obtain comprehensively for real-world FCEV operating conditions. In contrast, reinforcement learning (RL) enables an agent to learn control policies through trial-and-error interaction with the environment, without requiring explicitly labeled input–output pairs, and seeks to maximize long-term cumulative rewards [21]. The core principle of RL resembles dynamic programming in that both rely on Bellman's optimality equation and the principle of optimal substructure [22–24]. However, unlike DP—which assumes complete environmental knowledge—RL acquires information through interaction and is therefore often referred to as adaptive dynamic programming. In summary, no single paradigm can simultaneously satisfy the fourfold requirements of real-time performance, feasibility, cross-scenario adaptability, and long-term durability [25,26]. MPC provides a constraint-aware framework for predictive power allocation and has strong potential for real-time implementation. However, its control performance depends strongly on the weighting coefficients assigned to hydrogen consumption, battery utilization, SOC regulation, and control smoothness. Fixed weights

are generally unable to maintain an appropriate balance among these objectives under rapidly changing driving conditions. Therefore, adaptive weight scheduling is necessary to improve the cross-condition performance of MPC-based FCEV energy management.

To address the optimization of MPC for FCEV energy management, various strategies have been proposed, including feedforward control, feedback control, and weight optimization. Among these, weight optimization can be effectively approached using deep reinforcement learning (DRL) strategies. DRL in this context encompasses several categories: value-based algorithms, policy-gradient algorithms, and Actor–Critic-based algorithms. Value-based algorithms (e.g., DQN and its variants) derive decision policies by learning the state–action value function. These methods are mature and easy to implement in discrete control scenarios [27]. However, since the continuous power-output space of FCEVs must be discretized, such methods often lose control precision, making it difficult to meet fine-grained energy-management requirements. Policy-gradient algorithms (e.g., REINFORCE, A2C) directly optimize the policy function to maximize cumulative returns and can naturally handle continuous action spaces, but they suffer from high variance, unstable convergence, and a tendency to fall into local optima [28]. Actor–Critic-based algorithms (e.g., DDPG, TD3, SAC) use an Actor network to generate actions and a Critic network to evaluate their value, combining the stability of value-function methods with the continuous-control capability of policy-gradient methods [29–33]. These approaches have become the mainstream research direction [34,35]. Unlike rule-based strategies, which depend heavily on empirical calibration, and optimization-based strategies, which rely on precise models, DRL methods reformulate the energy-management problem as a Markov Decision Process (MDP). By leveraging deep neural networks to approximate the policy or value function, DRL autonomously learns power-distribution patterns from large-scale driving data during offline training—without requiring explicit derivation of the system dynamics [36]. This property effectively mitigates the model-dependency and prediction-reliance issues of MPC, providing intrinsic adaptability to handle component aging and dynamic disturbances [37]. Specifically, the Deep Deterministic Policy Gradient (DDPG) algorithm, as a representative Actor–Critic method, is well suited to continuous decision-making problems and can therefore be used to generate continuously varying optimization weights for FCEV energy management [38]. MPC- and DRL-based controllers can be integrated using either centralized or hierarchical architectures [39]. A centralized architecture jointly processes all control decisions, but its computational burden and implementation complexity increase rapidly with the number of states, actions, and component constraints. A hierarchical architecture separates strategic decision-making from real-time execution, thereby improving modularity, scalability, and controller interpretability. This structure is particularly suitable for the present study, because the upper-level DRL agent can focus on adaptive weight generation while the lower-level MPC controller handles constrained power allocation.

In this context, this paper proposes an Adaptive-Modality Deep Deterministic Policy Gradient and Model Predictive Control hierarchical energy management strategy (AMDDPG–MPC HEMS) for FCEVs. The proposed strategy employs a two-level architecture in which the upper-level AMDDPG controller generates adaptive multi-objective weights, while the lower-level MPC controller performs constrained power allocation between the fuel cell and battery. The AMDDPG algorithm incorporates an adaptive modality perception mechanism to identify driving-condition patterns and a multi-scale optimization mechanism to coordinate short-term energy economy with long-term component-protection objectives. In addition, a dedicated weight-scheduling and switching mechanism is introduced to ensure smooth weight transitions under varying operating conditions. The

effectiveness of the proposed strategy is evaluated under WLTC and UDDS driving cycles through comparisons with rule-based, ECMS, and fixed-weight MPC strategies.

The main contributions of this paper are as follows:

1. A two-level AMDDPG–MPC hierarchical architecture is developed to decouple adaptive weight generation from constrained power allocation. The upper-level AMDDPG controller generates the optimization weights, while the lower-level MPC controller allocates power between the fuel cell and battery under physical constraints. This structure improves the modularity and scalability of the energy management system and facilitates real-time implementation.
2. A dynamic weight-scheduling and switching mechanism is introduced to update the hydrogen-consumption, battery-power-limitation, and SOC-regulation weights according to the current driving condition and component states. The mechanism enables smooth transitions among different weight combinations and improves the adaptability of the MPC controller compared with fixed-weight strategies.
3. An AMDDPG algorithm incorporating adaptive modality perception and multi-scale optimization is proposed. The adaptive modality perception mechanism provides driving-condition information for real-time weight generation, while the multi-scale optimization mechanism balances short-term energy-saving performance and long-term component-protection requirements.

The remainder of this paper is organized as follows. Section 2 presents the FCEV powertrain model and its main components. Section 3 describes the proposed AMDDPG–MPC HEMS, including the hierarchical architecture, MPC formulation, and adaptive weight-generation algorithm. Section 4 evaluates the proposed strategy under representative driving cycles and compares its performance with benchmark EMSs. Section 5 summarizes the main conclusions and outlines future research directions.

2. Modeling

This study considers an FCEV with a hybrid powertrain layout, as shown in Figure 1. The vehicle is driven by a front-axle motor whose torque is transmitted to the front wheels through a differential, while two independent in-wheel motors provide direct propulsion at the rear wheels. The required electrical energy is supplied by a combined energy system consisting of a hydrogen storage tank, a fuel cell system, and a battery pack. The specific parameters of the vehicle are detailed in Table 1.

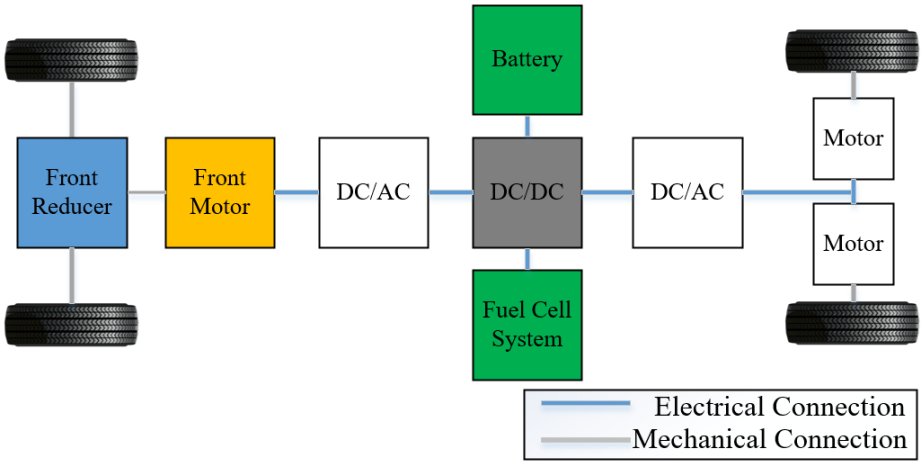


Figure 1. The schematic of the FCEV configuration.

Table 1. Basic parameters of the studied FCEV.

Vehicle Body	Mass	1860 [kg]
	Tire Rolling Radius	0.35 [m]
Front Motor	Motor type	Permanent-magnet synchronous motor
	Speed	0~14,000 [rpm]
	Torque	−137~137 [Nm]
	Maximum efficiency	93%
Rear Motor	Motor type	Permanent-magnet synchronous in-wheel motor
	Speed	0~10,000 [rpm]
	Torque	−97~97 [Nm]
	Maximum efficiency	92%
Battery	nominal voltage	365 [V]
	capacity	40 [Ah]
	Discharging efficiency	95%
Fuel Cell	Idle Power	2 [kW]
	Maximum Output Power	50 [kW]
	Maximum system efficiency	41%

2.1. Vehicle Dynamic

The traction force generated by the powertrain is used to overcome rolling resistance, grade resistance, aerodynamic drag, and inertial resistance. Using standard SI units, the required longitudinal force at the wheels is expressed as

$$F_{\text{req}} = mgf\cos\alpha + mg\sin\alpha + \frac{1}{2}\rho C_D A v^2 + m\frac{\mathrm{d}v}{\mathrm{d}t} \tag{1}$$

According to Equation (2), the vehicle power demand can be shown as

$$\hat{m}_{FC} = \int P_{FC} \cdot R_{H_2}(P_{FC}) \tag{2}$$

where m is the vehicle mass in kg, g is the gravitational acceleration in m/s^2, f is the rolling-resistance coefficient, and α is the road-slope angle in rad; it is set to zero in the WLTC and UDDS simulations presented in Section 4. ρ is the air density in kg/m, C_D is the aerodynamic drag coefficient, A is the frontal area in m^2, v is the vehicle speed in m/s, F_{req} is the required traction force in N, and P_{req} is the vehicle power demand in W. The power demand formulation used in this study focuses on longitudinal vehicle motion and the corresponding power required from the powertrain, since these terms dominate the energy flow in the considered EMS problem. This simplified representation captures the main effects of traction, acceleration, rolling resistance, aerodynamic drag, and road slope and is therefore suitable for energy management design while avoiding unnecessary model complexity.

2.2. Fuel Cell

Since this study mainly evaluates the control performance of the proposed energy management strategy, environmental effects on the fuel-cell system, such as temperature and humidity variations, are not considered in the model. Figure 2 presents the relation-

ships among fuel-cell output power, system efficiency, and hydrogen-consumption rate. The fuel-cell efficiency curve was adapted from the empirical fuel-cell model reported in Ref. [40]. Based on this curve, the hydrogen-consumption rate RH2(PFC) was obtained and used to calculate the cumulative hydrogen consumption according to Equation (4).

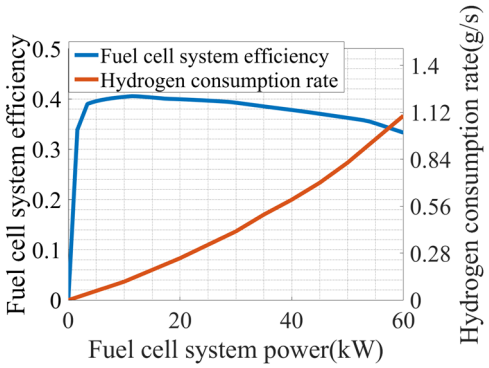


Figure 2. Fuel-cell system efficiency and hydrogen-consumption rate. Redrawn based on Ref. [40].

Theoretically, the hydrogen consumption of the fuel cell system can be calculated as

$$m_{FC} = \frac{1}{LHV_{H_2}} \int \frac{P_{FC}}{\eta_{FC}(P_{FC})} \tag{3}$$

where P_{FC} [w] is the fuel cell power, m_{FC} [J] is the hydrogen consumption of the fuel cell system, LHV_{H_2} is the hydrogen lower heating value, taken here as 120 MJ/kg, and η_{FC} is the dimensionless fuel cell system efficiency.

According to the polarization curve in Figure 2, the hydrogen consumption of the fuel cell system can be expressed as

$$\hat{m}_{FC} = \int P_{FC} \cdot R_{H_2}(P_{FC}) \tag{4}$$

where $\hat{m}_{FC}$ [g] is the hydrogen consumption; R_{H_2} [g/s] is the hydrogen consumption rate.

2.3. Battery

Analogous to the fuel-cell modeling approach, a simplified R-C equivalent-circuit model is adopted to describe the main electrical characteristics of the battery while reducing computational complexity. Transient electrochemical and thermal dynamics are not explicitly modeled; therefore, battery temperature is treated as a constant parameter rather than a dynamic state. In all simulations, the battery temperature is fixed at 25 °C. The terminal voltage of the battery is expressed as

$$U_{batt} = E_{batt} - I_{batt} R_{batt} \tag{5}$$

where U_{batt} [V] is the load voltage, E_{batt} [V] is the open-circuit voltage, R_{batt} [Ω] is the internal resistance of the battery, and I_{batt} [A] is the battery current.

The state of charge (SOC) of the battery is shown as

$$SOC = \begin{cases} SOC_{init} + \dfrac{1}{\eta_{batt} C_{batt}} \int_0^t I_{batt} dt \text{ for } I_{batt} > 0 \text{ (Charging)} \\ SOC_{init} - \dfrac{\eta_{batt}}{C_{batt}} \int_0^t (-I_{batt}) dt \text{ for } I_{batt} < 0 \text{ (Discharging)} \end{cases} \tag{6}$$

where SOC_{init} is the initial value of battery SOC, dimensionless, and C_{batt} is the battery capacity in Ah. $I_{batt} > 0$ means the battery is in the charging phase.

The battery efficiency η_{batt} is defined as the ratio of battery power to total power, and it can be expressed as

$$\eta_{batt} = \begin{cases} \frac{P_{batt}}{P_{batt}+P_{batt_loss}} = \frac{U_{batt}I_{batt}}{U_{batt}I_{batt}+I^2_{batt}R_{batt}}, P_{batt} < 0 \\ \frac{P_{batt}+P_{batt_loss}}{P_{batt}} = \frac{U_{batt}I_{batt}+I^2_{batt}R_{batt}}{U_{batt}I_{batt}}, P_{batt} > 0 \end{cases} \quad (7)$$

where P_{batt} [W] is the battery power; P_{batt_loss} [W] is the battery power loss due to internal resistance. $P_{batt} > 0$ means the battery is in the charging phase. The battery internal resistance is obtained from the SOC–temperature resistance maps. Because thermal dynamics are not included, the temperature input is fixed at $T_{\text{ref}} = 25\ ^\circ\text{C}$, and the resistance varies only with SOC during each simulation:

$$R_{\text{batt}} = \begin{cases} R_{\text{chg}}(SOC, T_{\text{ref}}), & I_{\text{batt}} > 0, \\ R_{\text{dis}}(SOC, T_{\text{ref}}), & I_{\text{batt}} < 0. \end{cases} \quad (8)$$

The resistance characteristic of the battery during charging and discharging is shown in Figure 3.

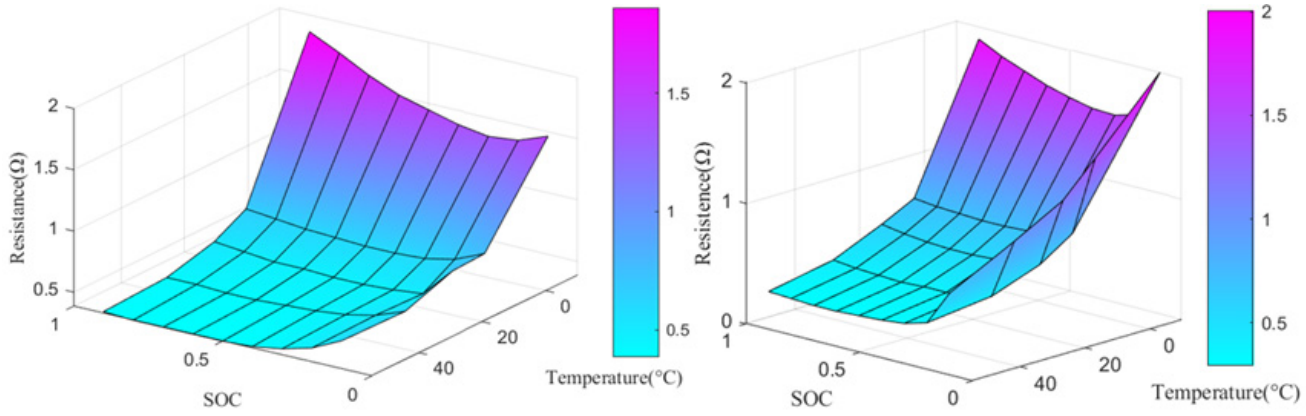


Figure 3. Presents the complete SOC–temperature resistance maps, whereas only the 25 °C slices of these maps are used in the simulations.

2.4. Motor

The studied FCEV employs a front-axle permanent-magnet synchronous motor (PMSM) and two rear permanent-magnet synchronous in-wheel motors. PMSMs offer high efficiency, high power density, fast torque response, and good regenerative-braking capability, making them suitable for electric-vehicle propulsion. The rear in-wheel motors provide direct drive and flexible wheel-level torque distribution while reducing mechanical transmission losses. To reduce computational complexity, the motor thermal and transient electromagnetic dynamics are neglected, and the motor efficiencies are obtained from torque–speed lookup maps. The electrical power of the front and rear motors is calculated as follows:

$$P_{elec_m} = \begin{cases} \frac{T_m\omega_m}{\eta_m}, T_m > 0 \\ T_m\omega_m\eta_{gen}, T_m \le 0 \end{cases} \quad (9)$$

where P_{elec_m} denotes the motor electrical power, T_m is the motor torque, ω_m is the motor rotational speed, η_m represents the motor efficiency in traction mode, and η_{gen} represents the efficiency in regenerative mode. When $T_m > 0$, the motor operates in traction mode; otherwise, it works in regenerative braking mode.

In this study, motor efficiency is obtained from a two-dimensional look-up table indexed by torque and speed. The efficiency relationship is expressed as

$$\eta_m = \eta(T_m, \omega_m) \tag{10}$$

The steady-state efficiency maps of the front PMSM and rear in-wheel PMSMs are shown in Figure 4.

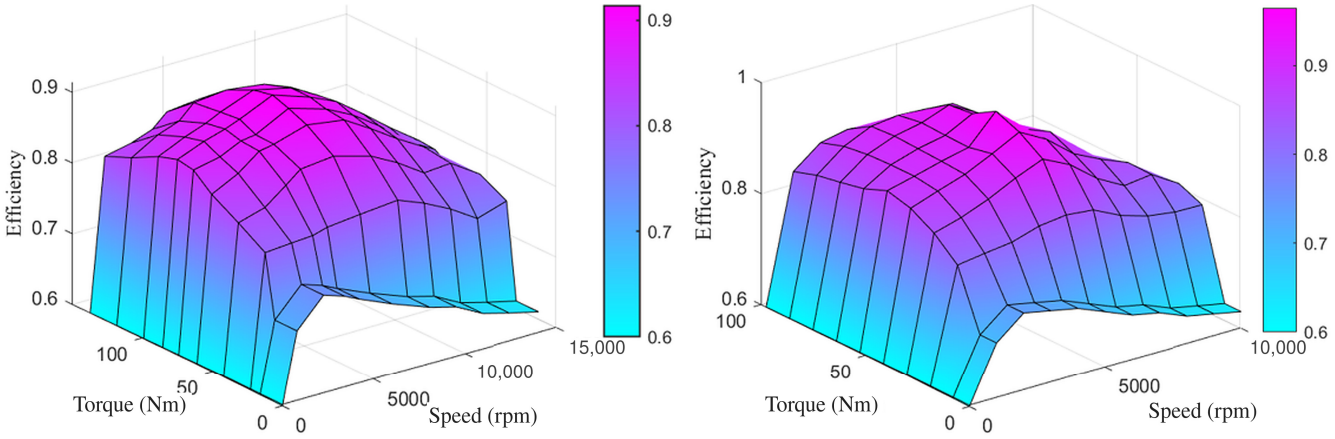


Figure 4. Motor efficiency map.

3. Hierarchical Energy Management Strategy

To endow the energy management strategy (EMS) of fuel cell electric vehicles (FCEVs) with fast processing capability, robustness, and engineering application potential, this study proposes a hierarchical energy management strategy (AMDDPG–MPC HEMS) integrating Adaptive-Modality Deep Deterministic Policy Gradient (AMDDPG) and Model Predictive Control (MPC). The overall workflow of the proposed strategy can be summarized as driving condition feature extraction and power allocation optimization. A novel fast computation algorithm is adopted in the EMS to realize energy management based on predefined logic for driving condition identification. Figure 5 presents the implementation framework of the AMDDPG–MPC HEMS. This hierarchical control architecture is designed to enhance adaptability to diverse driving conditions and improve control performance. The upper level integrates driving mode perception and weight generation logic, processes driver demand and component states, and then generates adaptive optimization weights according to the current driving condition. The lower level serves as an advanced executive control layer, processes the observed driver demand and component states, and determines the power allocation in the powertrain through the integrated MPC method. Figure 6 illustrates the logic flow of weight generation and optimization.

To improve real-time performance, the MPC implementation incorporates offline training strategies and online inference calls, significantly reducing online computational requirements. To further enhance the control effect of the MPC-based energy management strategy, a multi-dimensional weight generation process is developed in this study, which adaptively adjusts according to the state of charge (SOC), vehicle power demand, and prediction horizon length to cope with different driving scenarios.

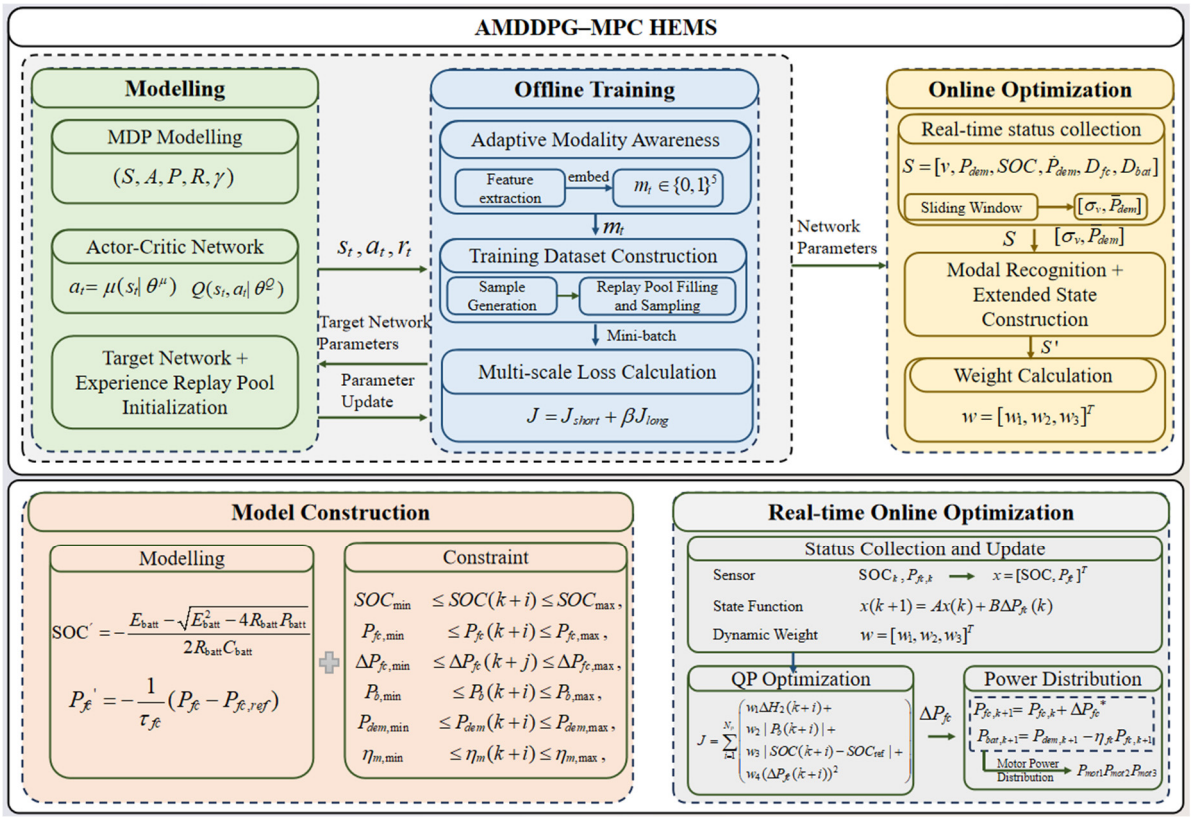


Figure 5. Schematic diagram of the AMDDPG–MPC HEMS.

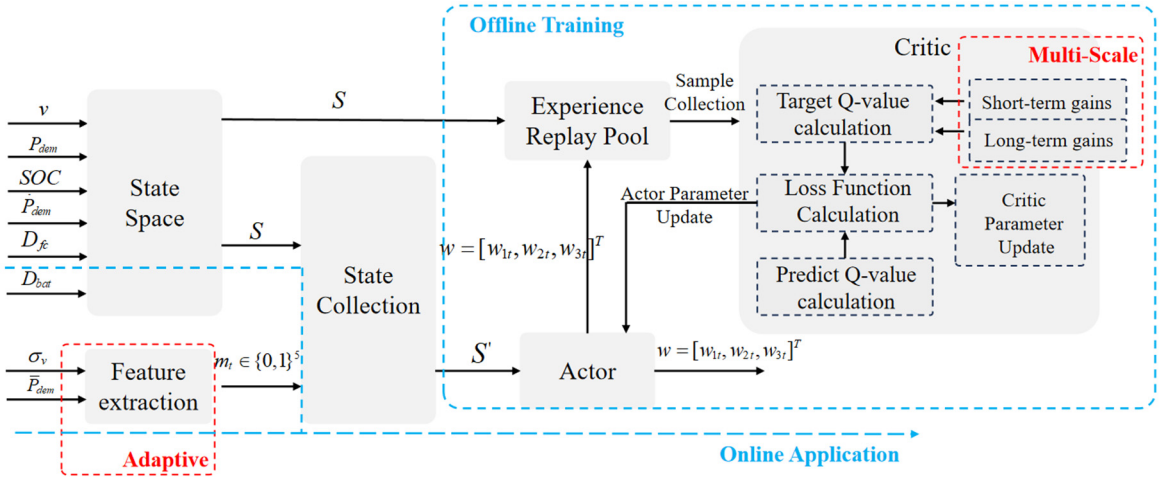


Figure 6. Logic flow of weight generation and optimization.

3.1. Hierarchical Control Architecture

Given variable driving conditions, powertrain component interactions strongly influence FCEVs' energy efficiency and service life. To maximize powertrain energy-saving potential and optimal energy management, a hierarchical control architecture is critical. This architecture decomposes the control strategy, simplifies the problem, and enables precomputation of strategic decisions, alleviating real-time MPC's computational load. Consequently, it accelerates and optimizes power allocation, while boosting MPC adaptability via real-time updates of target weights and constraints. It allows MPC to focus on specific operational goals, yielding precise control and enhanced vehicle performance.

3.1.1. Upper-Level Controller

To optimize the adaptability and control performance of the AMDDPG–MPC HEMS under dynamic driving conditions, the upper-level controller defines specific operating modes of FCEVs (steady cruising, acceleration, and congestion modes) and presets engineering threshold parameters offline to assist in mode identification and boundary constraints (these thresholds ensure stability without participating in training). It categorizes modes based on energy source activation states and driving condition characteristics: in steady cruising mode, the fuel cell meets power demand to minimize hydrogen consumption; in acceleration mode, the fuel cell and battery coordinate to respond to transient loads; and in congestion mode, the battery acts as a buffer to reduce fuel cell start–stop losses. Simultaneously, the upper level outputs a dynamic weight vector $w_t = [w_{1,t}, w_{2,t}, w_{3,t}]^T$ (hydrogen consumption weight $w_{1,t}$, battery power limitation weight $w_{2,t}$, and SOC balance weight $w_{3,t}$) in real time, transmitting it to the lower level via the CAN bus every 100 ms.

Figure 5 illustrates the control algorithms assigned to different operating modes within the hierarchical framework, whereas the adaptive adjustment of optimization weights is further detailed through the weight-generation logic. Since mode transitions in practical driving are continuous and coupled with powertrain states, the upper-level controller must provide a clear rule for updating the MPC weights. In the proposed AMDDPG–MPC HEMS, this process is carried out by jointly considering driver power demand and key component states. Specifically, vehicle power demand and speed are used to identify the current driving mode in real time, while SOC, battery power, and fuel-cell power provide additional constraints for weight adjustment. Based on these variables, the upper layer generates matched optimization weights for the lower-level MPC, as shown in Figure 6, thereby improving energy allocation efficiency and supporting component durability.

In Figure 6, $w_{1,t}$, $w_{2,t}$ and $w_{3,t}$ denote the hydrogen consumption weight threshold, battery power limitation weight threshold, and SOC balance weight threshold, respectively, with values of 0.4, 0.3, and 0.3 as shown in Table 2. These values are derived based on multiple factors, including ensuring the balance between component life and performance, maintaining sufficient reserves under dynamic driving conditions, and achieving efficient energy recovery in various driving scenarios. v_{th} [km/h] is the maximum constrained vehicle speed for mode switching, $P_{dem,th}$ [kW] is the power demand threshold, $P_{bat,th}$ [kW] is the battery power threshold, and $P_{fc,th}$ [kW] is the fuel cell output threshold, with values of 80, 60, 40, and 50, respectively, which are related to the specific energy requirements and operating efficiency under different driving conditions.

Table 2. Upper-level controller threshold parameter.

Parameter Symbol	Description	Value	Unit/Remarks
$w_{1,th}$	Hydrogen consumption weight threshold	0.4	Used for cruising mode switching boundary
$w_{2,th}$	Battery power limitation weight threshold	0.3	Used for acceleration mode response threshold
$w_{3,th}$	SOC balance weight threshold	0.3	Used for congestion mode buffer priority
λ_{adapt}	Adaptive adjustment coefficient	0.1	Mode encoding fusion adjustment factor
v_{th}	Vehicle speed threshold for mode switching	80	Maximum constrained vehicle speed
$p_{dem,th}$	Power demand threshold	60	Demand trigger point
$p_{bat,th}$	Battery power threshold	40	Charging/discharging boundary
$p_{fc,th}$	Fuel cell output threshold	50	Maximum output limitation

3.1.2. Lower-Level Controller

At the lower level, the MPC determines powertrain power allocation, taking upper-level weights w_t, current state $x_t = [SOC_t, P_{fc,t}]^T$, and prediction horizon $N_p = 10$ as inputs. It outputs the fuel cell power change rate ΔP_{fc}, further calculating battery power $P_{bat,t+1} = P_{dem,t+1} - \eta_{fc} P_{fc,t+1}$ and motor power allocation (residual power to front/rear motors) with a 100 ms rolling execution frequency. In specific modes, steady cruising

mode relies on the fuel cell to meet main power demand for hydrogen consumption optimization; acceleration mode uses MPC to coordinate battery and fuel cell outputs (detailed in subsequent chapters); and congestion mode operates the fuel cell at its high-efficiency point, where its output serves both buffering and response needs.

Power allocation is determined by optimizing a multi-objective cost function. The optimization problem is formulated as follows:

$$J = \sum_{i=1}^{N_p} \left(w_{1,t} \Delta H_2(k+i) + w_{2,t} |P_b(k+i)| + w_{3,t} |SOC(k+i) - SOC_{\text{ref}}| + w_{4,t} (\Delta P_{fc}(k+i))^2 \right) \tag{11}$$

where $w_t = [w_{1,t}, w_{2,t}, w_{3,t}]^T$ is the dynamic weight vector generated by the upper level, $w_{4,t}$ is an intrinsic control-smoothing weight native to the MPC layer, $\Delta H_2(k+i)$ is the hydrogen consumption rate, $P_b(k+i)$ is the battery power, $SOC(k+i)$ is the battery state of charge, and ρ is the control smoothing coefficient, which can be expressed as

$$\Delta H_2 = \frac{P_{fc,k}}{LHV \cdot \eta_{fc}} \tag{12}$$

$$P_{b,k} = P_{dem,k} - \eta_{fc} P_{fc,k} \tag{13}$$

where $P_{fc,k}$ denotes the fuel-cell power variation between two consecutive sampling instants. Equation (13) represents the high-level source-side power balance used in the MPC optimization, in which the fuel cell provides the commanded power and the battery supplies the residual power demand. The lower-level MPC does not independently optimize the power outputs of the front, rear-left, and rear-right motors. Instead, the required traction power is distributed among the three motors according to the predefined motor power-allocation rule, and the corresponding electrical power is subsequently calculated using the individual motor-efficiency maps introduced in Equation (19). Therefore, Equation (13) serves as a computationally efficient abstraction within the MPC cost function, whereas Equation (19) is used in the detailed powertrain model to account for the efficiency differences among the three motors.

The MPC controller is implemented in a receding-horizon manner to ensure smooth and feasible control actions. At each sampling instant, the optimal fuel-cell power variation is calculated, and only the first control action is applied before the prediction horizon is shifted forward. This procedure enables the fuel cell and battery to coordinate under rapidly varying power demand while limiting abrupt fuel-cell power changes.

In Section 4, identical driving-cycle inputs, initial conditions, component constraints, and operating-mode definitions are applied to all benchmark EMSs. This setting ensures consistent simulation conditions and enables a fair comparison of their energy-management performance.

3.2. Energy Management Based on Implicit MPC

In the proposed hierarchical energy management strategy for FCEVs, the lower-level controller adopts implicit MPC for real-time optimal power allocation. Unlike explicit MPC with offline precomputed affine control laws, implicit MPC solves optimization online at each sampling step, using efficient QP solvers to address powertrain nonlinearities and time-varying parameters. It adapts to dynamic weights from the upper-level AMDDPG module, ensuring robust energy distribution across operating conditions. The formulation resolves fuel cell–battery–motor power coupling via a cost function balancing hydrogen consumption, battery degradation, and FC stress, under component physical constraints.

3.2.1. Basic Implicit MPC

Model Predictive Control (MPC) is a receding-horizon optimization framework that computes an optimal sequence of control actions over a prediction horizon by solving a finite-time optimal control problem at each sampling step, while only applying the first control input before shifting the horizon forward. For linear time-invariant (LTI) systems, the core process involves (1) modeling the system dynamics via a discrete-time state-space representation; (2) defining a quadratic cost function to penalize deviations from a reference trajectory and control effort; (3) enforcing input, state, and output constraints; and (4) reformulating the problem as a QP for efficient online solution.

Consider a discrete-time LTI system described by

$$x(t+1) = Ax(t) + Bu(t) \tag{14}$$

where $x(t) \in R^n$ is the state vector, $u(t) \in R^m$ is the control input vector, and A, B are the system matrices. The MPC optimization problem over a prediction horizon N_p and control horizon N_c (with $N_c \leq N_p$) is formulated as

$$\min_{U} \sum_{i=1}^{N_p} \left(\| y(k+i|k) - r(k+i) \|_Q^2 + \sum_{j=0}^{N_c-1} \| \Delta u(k+j|k) \|_R^2 \right) \tag{15}$$

subject to the system dynamics, constraints on states $x \in \mathcal{X}$, inputs $u \in \mathcal{U}$ and outputs $y \in \mathcal{Y}$. Here, $U = [\Delta u(k \mid k)^T, \ldots, \Delta u(k+N_c-1 \mid k)^T]^T$ stacks the control increments, $y(k+i \mid k)$ denotes the predicted output at step i based on current measurements at k, $r(k+i)$ is the reference, and Q, R are weighting matrices.

By stacking predictions and substituting the state-space model, the problem is recast as a QP:

$$\min_{U} \frac{1}{2} U^T H U + f^T U, \; GU \leq w \tag{16}$$

where H is the Hessian, f is the linear term, and $GU \leq w$ encapsulates the constraints affine in U. The optimal U^* is obtained via QP solvers and only the first element $\Delta u(k \mid k)$ is applied, yielding the control law $u(k) = u(k-1) + \Delta u(k|k)$.

This implicit formulation, solved online, contrasts with explicit MPC by avoiding offline region partitioning, thus reducing preprocessing complexity at the expense of mild computational overhead, which is negligible for embedded solvers in automotive applications.

3.2.2. Application of MPC in EMS

In the hierarchical energy management strategy for fuel cell electric vehicles (FCEVs), the Model Predictive Control (MPC) component enables efficient real-time power allocation through online optimization. At each sampling instant, MPC solves a constrained optimization problem leveraging quadratic programming (QP) solvers, which accommodates the dynamic weights from the upper-level Adaptive-Modality Deep Deterministic Policy Gradient (AMDDPG) module while ensuring computational efficiency. The states are selected as the battery state of charge (SOC) and fuel cell output power P_{fc}, with the control input defined as the fuel cell power rate ΔP_{fc}, facilitating gradual power adjustments to minimize component degradation. This implementation is grounded in the FCEV powertrain model established in Section 2, balancing real-time adaptability and computational feasibility.

The state-space equations are derived to model energy allocation dynamics. The discrete SOC update, based on battery electrochemistry, is given by

$$\mathrm{SOC}' = -\frac{E_{\mathrm{batt}} - \sqrt{E_{\mathrm{batt}}^2 - 4R_{\mathrm{batt}}P_{\mathrm{batt}}}}{2R_{\mathrm{batt}}C_{\mathrm{batt}}} \tag{17}$$

where E_{batt} is the battery open-circuit voltage, R_{batt} is the internal resistance, C_{batt} is the capacitance, and $P_{\mathrm{batt}} = P_b$ is the battery power.

The electrical-mechanical power coupling is

$$\begin{cases} P_b + P_{fc} = \frac{P_{mF}}{\eta_{mF}} + \frac{P_{mRL}}{\eta_{mRL}} + \frac{P_{mRR}}{\eta_{mRR}} \\ P_{mF} + P_{mRL} + P_{mRR} = P_{\mathrm{req}} \\ P_{\mathrm{elec,req}} = P_{fc} + P_b \end{cases} \tag{18}$$

where P_{mF}, P_{mRL}, and P_{mRR} are the front, rear-left, and rear-right motor powers, respectively; η_{mF}, η_{mRL}, and η_{mRR} are their efficiencies; and P_{req} is the vehicle power demand.

From Equation (18), the battery power is expressed as

$$P_b = \frac{P_{\mathrm{req}}}{\eta_{mF}} + \left(\frac{1}{\eta_{mRL}} - \frac{1}{\eta_{mF}}\right)P_{mRL} + \left(\frac{1}{\eta_{mRR}} - \frac{1}{\eta_{mF}}\right)P_{mRR} - P_{fc} \tag{19}$$

Substituting Equation (19) into Equation (17) yields the coupled SOC dynamics:

$$\mathrm{SOC}' = -\frac{E_{\mathrm{batt}} - \sqrt{E_{\mathrm{batt}}^2 - 4R_{\mathrm{batt}}\left[\frac{P_{\mathrm{req}}}{\eta_{mF}} + \left(\frac{1}{\eta_{mRL}} - \frac{1}{\eta_{mF}}\right)P_{mRL} + \left(\frac{1}{\eta_{mRR}} - \frac{1}{\eta_{mF}}\right)P_{mRR} - P_{fc}\right]}}{2R_{\mathrm{batt}}C_{\mathrm{batt}}} \tag{20}$$

Discretizing these with sampling time T_s using zero-order hold, and incorporating fuel cell first-order dynamics,

$$\dot{P}_{fc} = -\frac{1}{\tau_{fc}}(P_{fc} - P_{fc,ref}) \tag{21}$$

with τ_{fc} as the time constant results in the linear time-varying (LTV) state-space model:

$$x(k+1) = Ax(k) + B\Delta P_{fc}(k) \tag{22}$$

where $x = [\mathrm{SOC}, P_{fc}]^T$

$$A = \begin{bmatrix} 1 & -\frac{T_s\eta_m}{V_bQ_b} \\ 0 & 1 - \frac{T_s}{\tau_{fc}} \end{bmatrix},\ B = \begin{bmatrix} 0 \\ \frac{T_s}{\tau_{fc}} \end{bmatrix} \tag{23}$$

This linear time-invariant (LTI) approximation, linearized around nominal operating points, incorporates measurable disturbances like P_{dem} and η_m.

The finite-horizon optimal control problem is formulated with a quadratic cost:

$$J = \sum_{i=1}^{N_p}\left(w_{1,t}\Delta H_2(k+i) + w_{2,t}|P_b(k+i)| + w_{3,t}|SOC(k+i) - SOC_{\mathrm{ref}}| + w_{4,t}(\Delta P_{fc}(k+i))^2\right) \tag{24}$$

where weights $w_t = [w_{1,t}, w_{2,t}, w_{3,t}]^T$ are parameterized offline based on AMDDPG mappings ($w_1, w_2, w_3 \geq 0$ with $w_1 + w_2 + w_3 = 1$) and w_4 is an intrinsic control-smoothing weight native to the MPC layer ($w_4 > 0$); it enforces physically feasible and smooth control actions. $\Delta H_2(k+i) = \frac{P_{fc}(k+i)T_s}{\eta_{fc}LHV}$ quantifies hydrogen consumption at time step $k+i$, $P_b(k+i)$ penalizes the magnitude of battery power to mitigate degradation from extreme charging/discharging, $SOC(k+i)$ is the battery State of Charge, and

$SOC_{\text{ref}} = 0.5$ is the reference SOC selected to balance battery longevity and buffer capacity; $\Delta P_{fc}(k+i) = P_{fc}\,(k+i) - P_{fc}\,(k+i-1)$ represents the fuel cell power variation between consecutive time steps.

To safeguard component integrity and vehicle performance, the optimization incorporates box and linear constraints on states, controls, and derived variables, ensuring feasible operation within physical limits. These constraints, affine in the control sequence U, are expressed as

$$\begin{cases} SOC_{\min} \le SOC(k+i) \le SOC_{\max} \\ P_{fc,\min} \le P_{fc}(k+i) \le P_{fc,\max} \\ \Delta P_{fc,\min} \le \Delta P_{fc}(k+j) \le \Delta P_{fc,\max} \\ P_{b,\min} \le P_b(k+i) \le P_{b,\max} \\ P_{dem,\min} \le P_{dem}(k+i) \le P_{dem,\max} \\ \eta_{m,\min} \le \eta_m(k+i) \le \eta_{m,\max} \end{cases}, \; i = 1, \ldots, N_p; j = 0, \ldots, N_c - 1 \tag{25}$$

where $SOC_{\min/\max}$ prevents overcharge/discharge; $P_{fc,\min/\max}$ bounds fuel cell output; $\Delta P_{fc,\min/\max}$ limits slew rates for durability; $P_{b,\min/\max}$ reflects battery ratings; and $P_{dem,\min/\max}$ enforces demand and efficiency bounds. Voltage and resistance constraints are embedded via $P_b = V_b I_b - R_b I_b^2$, linearized around nominals.

Stacking predictions over horizons N_p and N_c reformulates J as an mp-QP:

$$\min_U \frac{1}{2} U^T H U + f^T U, \; GU \le w \tag{26}$$

where $U = [\Delta P_{fc}(k \mid k)^T, \ldots, \Delta P_{fc}(k+N_c-1 \mid k)^T]^T$ stacks the control increments; H, f, G, w are derived from A, B, the weights and constraints. The resulting optimization is solved online using a QP solver. The optimal U^* yields $\Delta P_{fc}(k|k)$ for the P_{fc} update, and the residual battery power is $P_b(k) = P_{dem}(k) - \eta_m(k) P_{fc}(k)$.

This MPC integration with AMDDPG enables adaptive, predictive control, surpassing static methods by anticipating demands and handling uncertainties, thus enhancing FCEV efficiency, durability, and robustness to extremes.

3.3. Design and Implementation of the AMDDPG Algorithm

As the core weight generator in the hierarchical architecture, the Adaptive-Modality Deep Deterministic Policy Gradient (AMDDPG) learns the coupling between driving condition features and component states offline and outputs online real-time adaptive MPC multi-objective weights (w_1—hydrogen consumption; w_2—battery power limitation; w_3—SOC balance). This section details its framework, offline training, and online application, highlighting innovations in adaptive mode perception and multi-scale optimization versus traditional DDPG.

3.3.1. Basic Formulas and Framework of the Algorithm

The AMDDPG is extended based on the traditional Deep Deterministic Policy Gradient (DDPG) framework, incorporating an adaptive mode perception module and a multi-scale optimization mechanism, and is formalized in the paradigm of the Markov Decision Process (MDP). The core of this framework lies in transforming the weight generation problem into policy optimization in the continuous action space, ensuring robust decision-making response to the dynamic driving conditions of fuel cell electric vehicles (FCEVs). The core formulas and network structure are detailed as follows.

The energy management weight generation problem is modeled as an MDP five-tuple (S, A, P, R, γ). The state space S is designed to capture current driving conditions and component states, defined as

$$S = [v, P_{dem}, SOC, \dot{P}_{dem}, D_{fc}, D_{bat}] \tag{27}$$

where v denotes vehicle speed (km/h), P_{dem} is the driver's power demand (kW), SOC is the battery state of charge (0–1), $\dot{P}_{dem}$ is the power demand change rate (kW/s), D_{fc} is the fuel cell degradation degree (0–1, with 0 indicating a new state and 1 indicating failure), and D_{bat} is the battery degradation degree (0–1). The action space A corresponds to the output MPC weight vector:

$$w = [w_{1t}, w_{2t}, w_{3t}]^T \tag{28}$$

where w_1 prioritizes hydrogen consumption reduction, w_2 suppresses battery power fluctuations to mitigate degradation, and w_3 balances SOC at $SOC_{\text{ref}} = 0.5$ for optimal buffering and longevity. The state transition probability $P(s_t + 1|s_t, a_t)$ is governed by FCEV powertrain dynamics, while the reward function $R(s_t, a_t)$ quantifies the immediate benefit of weight decisions, formulated as

$$R(s_t, a_t) = -(w_1 \cdot \Delta H_2 + w_2 \cdot |P_b| + w_3 \cdot |SOC - SOC_{ref}|) + \lambda_t \cdot \left(1 - \max(D_{fc,t}, D_{bat,t})\right) \tag{29}$$

In this, the hydrogen consumption rate ΔH_2 is calculated as

$$\Delta H_2 = P_{fc} / (LHV \cdot \eta_{fc}) \tag{30}$$

where $LHV = 120$ MJ/kg is hydrogen's lower heating value and η_{fc} is fuel cell efficiency; $|P_b|$ is the absolute battery power; $|SOC - SOC_{ref}|$ is the SOC deviation; and $\lambda_t \in [0, 1]$ is a dynamically adjusted coefficient based on driving modes. The discount factor $\gamma \in (0, 1)$ balances short-term and long-term rewards.

To improve the clarity of the proposed control formulation, the physical meanings, units or ranges, and selection or acquisition basis of the main variables and parameters involved in the MPC and AMDDPG controllers are summarized in Table 3.

Table 3. Physical meanings, units, and selection basis of the main variables and parameters.

Symbol	Physical Meaning	Unit/Range	Selection or Acquisition Basis
v	Vehicle speed	km/h	Obtained from the vehicle-state measurement
P_{dem}	Driver-demand power	kW	Calculated from the longitudinal vehicle dynamics
$\dot{P}_{\text{dem}}$	Variation rate of driver-demand power	kW/s	Calculated from consecutive power-demand samples
SOC	Battery state of charge	0–1	Estimated using the battery model
SOC_{ref}	Reference battery SOC	Dimensionless; 0.5	Selected near the middle of the allowable SOC range to retain traction and regenerative-braking capacity
P_{fc}	Fuel-cell output power	kW	Constrained by the fuel-cell operating range in Table 1
P_{bat}	Battery power	kW	Determined from the power balance and limited by battery ratings
ΔP_{fc}	Fuel-cell power variation between consecutive control steps	kW/control step	Limited to suppress rapid fuel-cell load changes
$P_{\text{bat,min}}$, $P_{\text{bat,max}}$	Battery charging and discharging power limits	kW	Determined from battery voltage, capacity, allowable current, and SOC limits

Table 3. *Cont.*

Symbol	Physical Meaning	Unit/Range	Selection or Acquisition Basis
w_1	Hydrogen-consumption weight	0–1	Generated online by the trained AMDDPG Actor network
w_2	Battery-power-limitation weight	0–1	Generated online by the trained AMDDPG Actor network
w_3	SOC-regulation weight	0–1	Generated online by the trained AMDDPG Actor network
w_4	Fuel-cell power-smoothing weight	Dimensionless	Selected through offline calibration under the fuel-cell slew-rate constraint
D_{fc}	Normalized fuel-cell degradation indicator	0–1	Obtained from the estimated fuel-cell health state
D_{bat}	Normalized battery degradation indicator	0–1	Obtained from the estimated battery health state
N_{p}	MPC prediction horizon	10 steps	Selected by balancing control performance and online computational burden
T_{s}	Controller update period	0.1 s	Selected according to the real-time hierarchical control cycle
λ	Mode-dependent reward adjustment coefficient	0–1	Determined according to the identified driving mode
γ	Reward discount factor	0–1	Selected during offline training to balance current and future rewards
β	Multi-scale loss balance coefficient	Dimensionless	Calibrated offline to balance short-term and long-term losses

The physical constraints are selected according to the component specifications and allowable operating ranges, whereas the learning-related parameters are determined through offline calibration by balancing energy economy, component protection, training stability, and computational efficiency.

The Actor–Critic network structure comprises two subnetworks. The Actor network takes state s_t as input and outputs a deterministic weight decision $a_t = \mu(s_t|\theta^{\mu})$, with θ^{μ} as network parameters. It adopts a three-layer fully connected (FC) structure (input layer → hidden layer 128 → hidden layer 64 → output layer 3), with a Softmax activation function at the output to ensure weight normalization. The Critic network takes s_t and a_t as inputs and outputs the action value $Q(s_t, a_t|\theta^{Q})$ to evaluate decision quality, utilizing a three-layer FC structure (state–action concatenation → hidden layer 128 → hidden layer 64 → output layer 1) optimized via mean squared error (MSE) loss. To enhance training stability, target Actor $\mu\prime(s|\theta^{\mu\prime})$ and Critic $Q\prime(s,a|\theta^{Q\prime})$ networks are introduced, with parameters synchronized via a soft update mechanism:

$$\theta^{\mu\prime} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu\prime}, \theta^{Q\prime} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q\prime} \tag{31}$$

where τ is the soft update coefficient. The replay buffer stores samples (s_t, a_t, r_t, s_{t+1}), and random sampling is performed during training to avoid sample correlation.

3.3.2. Offline Training

The offline training phase enables AMDDPG to learn the optimal "state → weight" mapping, with its "Adaptive-Modality" (AM) essence manifested in adaptive mode perception and multi-scale optimization mechanisms, as elaborated below.

Adaptive Mode Perception Mechanism

To achieve precise weight adaptation to dynamic driving conditions, an adaptive mode perception module based on fuzzy clustering is designed, classifying driving scenarios into

five distinct modes (Table 4) and integrating mode-specific features into the network input to guide context-aware weight decisions.

Table 4. Classification of driving modes and corresponding weight priorities.

Driving Mode	Classification Criteria	Core Objective of Weight Decision
High-speed Cruising	$v > 80$ km/h, $\sigma_v < 0.1$	Prioritize hydrogen consumption reduction
Medium-speed Steady	$50 \leq v \leq 80$ km/h $0.1 \leq \sigma_v < 0.3$	Balance efficiency and responsiveness
Low-speed Congestion	$v < 50$ km/h $\sigma_v \geq 0.3$	Suppress battery power fluctuation to reduce fuel cell start-stop loss
Rapid Acceleration	$\dot{P}_{dem} > 5$ kW/s	Coordinate fuel cell and battery for transient load response
Rapid Deceleration	$\dot{P}_{dem} < -5$ kW/s	Prioritize SOC balance for energy recovery

The mechanism operates through two key steps: first, driving condition features are extracted by calculating vehicle speed volatility $\sigma_v = std(v)/mean(v)$ and average power demand $\overline{P}_{dem}$ over a 5 s sliding window, capturing the dynamic nature of scenarios. Second, a mode embedding network is integrated before the Actor network's input layer, converting clustering results into one-hot vectors $m_t \in \{0,1\}^5$ that are concatenated with state features s_t. This fusion enables the network to focus on mode-specific core needs and dynamically adjusts the coefficient λ in the reward function via

$$\lambda_t = \sum_{i=1}^{5} m_{t,i} \cdot \lambda_{base,i} \tag{32}$$

where $\lambda_{base,i}$ is a preset base value for each mode, ensuring the reward function's sensitivity to driving contexts. The "adaptive" optimization here lies in real-time scenario identification and targeted weight tuning guided by mode encoding, allowing the network to prioritize objectives that align with current driving conditions.

Multi-Scale Optimization Mechanism

To address the inherent trade-off between short-term driving dynamics and long-term component aging, a dual-scale optimization framework is proposed, integrating both immediate and extended operational horizons into the training process.

The short-term scale aligns with the MPC prediction horizon $N_p = 10$, focusing on the Critic network's evaluation of weight impacts on instantaneous hydrogen consumption ΔH_2 and power fluctuations $|P_b|$. This ensures rapid responsiveness to abrupt changes in power demand $\dot{P}_{dem}$, stabilizing transient performance. Conversely, the long-term scale incorporates a "long-term reward pool" within the replay buffer, storing cumulative rewards over 100 consecutive steps $\sum_{t=0}^{99} \gamma^t R_t$. Long-term samples are sampled every 100 training steps to emphasize suppression of component degradation D_{fc} , D_{bat}, guiding the network to prioritize life-cycle optimization.

These scales are fused via a multi-scale loss function that balances their contributions:

$$J = J_{short} + \beta J_{long} \tag{33}$$

where J_{short} denotes the standard temporal difference (TD) error for short-term performance, J_{long} represents the long-term reward error, and β is a balance coefficient. This "multi-scale" optimization ensures the learned weights not only adapt to immediate driving demands

but also mitigate long-term component wear, achieving robust performance across both transient and extended operational periods.

3.3.3. Online Application

In the online phase, trained network parameters are frozen to ensure efficient inference, seamlessly integrating with the 100 ms vehicle control cycle and communicating via the Controller Area Network (CAN) bus with end-to-end latency < 10 ms. Real-time state $s_t = [v, P_{\text{dem}}, \text{SOC}, \dot{P}_{\text{dem}}, D_{\text{fc}}, D_{\text{bat}}]$ is acquired every 100 ms from on-board sensors, with D_{fc} and D_{bat} estimated via the On-Board Diagnostics (OBD) system. Based on σ_v and $\overline{P}_{\text{dem}}$ within the current 5 s window, the driving mode is identified using a pre-trained fuzzy clustering model (solidified during offline training), generating mode encoding m_t. State features s_t and mode encoding m_t are input into the Actor network (with only forward computation enabled) to output normalized weights $w = [w_1, w_2, w_3]^T$, which are transmitted to the MPC controller via CAN. For robustness, if the MPC constraint violation rate exceeds 5% over five consecutive cycles (500 ms), weight fine-tuning is triggered—temporarily increasing w_3 (or w_2) until constraints are satisfied, then restoring normal output. This emergency mechanism is retained as a supervisory safety layer for extreme or previously unseen operating conditions. During the WLTC and UDDS simulations reported in Section 4, no constraint violation persisted for five consecutive control cycles; therefore, the emergency weight-tuning mechanism was not activated. This design leverages AMDDPG's adaptive mode perception (to handle diverse scenarios) and multi-scale optimization (to balance short- and long-term goals), enabling dynamic weight generation across full driving conditions and enhancing hierarchical EMS performance.

4. Comparison of Simulation Results

The AMDDPG–MPC HEMS for the FCEV is developed based on the vehicle model in Section 2 and the proposed control algorithm in Section 3. This section evaluates the economic (energy-saving) performance of this AMDDPG-MPC strategy. For this purpose, the proposed strategy is tested under two representative driving cycles: the World Light Vehicle Test Cycle (WLTC) and the Urban Dynamometer Driving Schedule (UDDS). These cycles cover both mixed driving conditions and urban stop-and-go scenarios, providing representative evaluation scenarios for assessing the adaptability and energy-management performance of different EMSs. The WLTC SOC trajectory used in the comparison is shown in Figure 7.

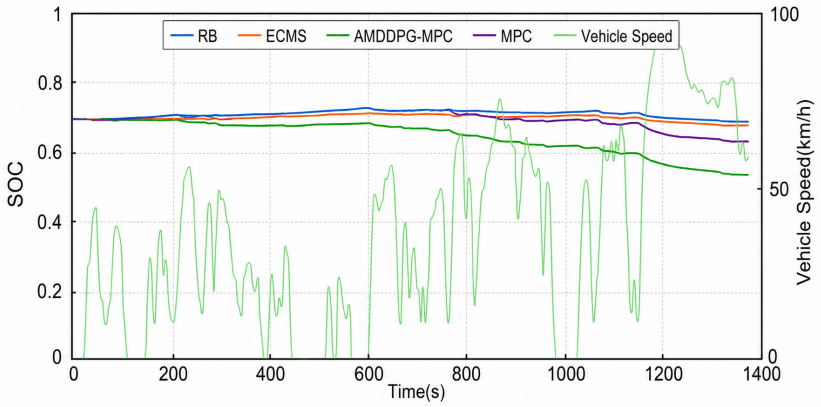


Figure 7. The changes in SOC in WLTC.

The overall control performance of the AMDDPG–MPC HEMS is evaluated by comparing it with several benchmark energy management strategies and the corresponding reference results, as described below:

Rule-based EMS (RB): The rule-based EMS (RB) is utilized as a benchmark strategy in this study [41]. Its operation is governed by predefined logic thresholds that determine the vehicle operating mode. In this study, the powertrain is divided into three modes: electric, hybrid, and charging. In electric mode, the required traction power is supplied only by the battery. In hybrid mode, the fuel cell and battery work together to meet the power demand. In charging mode, the fuel cell provides the propulsion power while producing additional energy to recharge the battery.

ECMS: A prominent instantaneous optimization-based EMS extensively explored in the literature [42] is also included. The central element of ECMS is the equivalent factor, which serves to unify the measurement of disparate energy sources by converting battery energy consumption into its hydrogen equivalent. For this study, a constant equivalent factor of $\lambda = 0.55$ is adopted. This value is derived from the average efficiencies of the fuel cell and the battery. Employing a fixed λ reduces real-time computational complexity while maintaining a balance between the energy sources.

Model Predictive Control (MPC) is also benchmarked, a control algorithm renowned for its inherent advantages of rolling optimization and feedback correction, which has led to its widespread adoption in practical engineering problems [43]. In this application, the MPC objective function, as formulated in Equation (28), is structured to manage the optimal trade-off between hydrogen consumption and battery usage. The respective weighting factors for this balance are set to $\alpha = 0.7$ and $\beta = 0.3$. The selected parameters used for the benchmark EMSs are summarized in Table 5.

All comparative simulations are conducted with a fixed time step of 0.01 s. The WLTC and UDDS simulations are performed on a level road, with the road slope angle set to $\alpha = 0$, corresponding to a road grade of 0%. To ensure consistent initial conditions among the tested EMSs, the vehicle starts in electric mode and the initial battery SOC is set to 0.85. The simulations are carried out on a computer equipped with an Intel i5-12400 CPU and 32 GB of memory. No emergency weight-tuning event was triggered during either the WLTC or UDDS simulations, because no state or input constraint violation persisted for five consecutive control cycles.

4.1. Assessment of Energy Consumption by Different EMSs in WLTC Driving Cycles

The vehicle simulation results for the four EMSs under four WLTCs are illustrated in this section. The WLTC provides a comprehensive, real-world test environment, covering diverse segments (urban, suburban, highway) and behaviors (acceleration, deceleration, etc.). This ensures a thorough evaluation of the strategies' adaptability and robustness across different scenarios. Table 6 summarizes the key performance metrics: final SOC, total equivalent hydrogen consumption (EHC), and EHC optimization percentage. The SOC trajectory during the simulation is shown in Figure 7, and the cumulative EHC is presented in Figure 8. To eliminate the influence of different terminal SOC values, all strategies are evaluated using an SOC-corrected equivalent hydrogen consumption metric. After each driving cycle, the electrical energy required to restore the battery from the terminal SOC to the common initial value of 0.85 is calculated as

$$E_{\mathrm{rec}} = \frac{E_{\mathrm{bat,nom}}(SOC_{\mathrm{ini}} - SOC_{\mathrm{end}})}{\eta_{\mathrm{chg}}} \tag{34}$$

where $E_{\mathrm{bat,nom}}$ is the nominal battery energy, $SOC_{\mathrm{ini}} = 0.85$, SOC_{end} is the terminal SOC, and η_{chg} is the overall charging efficiency, including the assumed charging losses. The SOC-corrected equivalent hydrogen consumption is then calculated as

$$m_{\mathrm{H_2,eq}} = m_{\mathrm{H_2,cycle}} + \lambda_{\mathrm{eq}} E_{\mathrm{rec}} \tag{35}$$

where $m_{H_2,cycle}$ is the actual hydrogen consumed during the driving cycle and λ_{eq} is the electrical-energy-to-hydrogen equivalence coefficient. In this study, $\eta_{chg} = 0.90$ and $\lambda_{eq} = 75$ g/kWh are applied consistently to all EMSs. Therefore, the total EHC values reported in the subsequent performance comparison tables already include the energy required to restore the terminal SOC to 0.85.

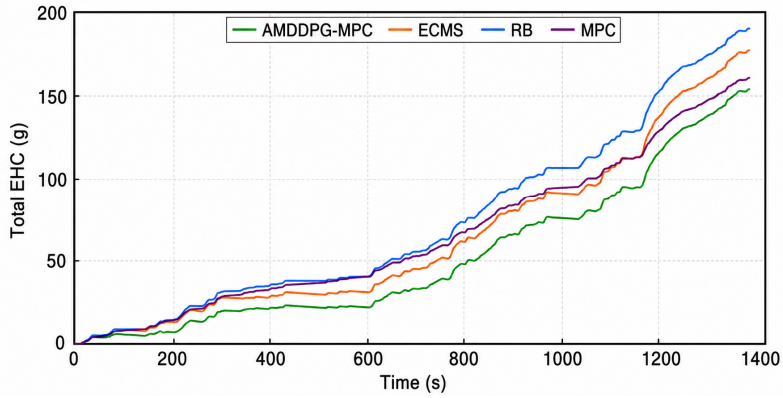


Figure 8. The changes in total EHC in WLTC.

The RB strategy maintains a relatively smooth SOC trajectory during the driving cycle, indicating its ability to preserve the battery charge within a narrow range and provide a stable energy reserve. Nevertheless, this behavior is achieved with limited optimization capability, since the rule-based controller depends on preset thresholds and cannot adjust its decisions to changing driving conditions. As a result, RB produces the highest total equivalent hydrogen consumption among the compared methods. Compared with RB, ECMS leads to more pronounced SOC variations because it continuously adjusts the power split between the fuel cell and the battery to minimize instantaneous equivalent fuel consumption. Although this mechanism improves hydrogen economy, it increases battery charge–discharge activity, which may be unfavorable for battery durability. MPC achieves a further reduction in total EHC relative to RB and ECMS. By incorporating predicted state evolution into the optimization process, MPC can select more effective control actions and improve fuel economy over the driving cycle.

By generating adaptive MPC weights in real time based on perceived driving modes, the AMDDPG-MPC achieves the lowest total EHC 155.957 g among all analyzed EMSs. It leverages a trained deep reinforcement learning agent to continuously output optimal weights (λ_H, λ_{Pbatt}, λ_{SOC}) for the underlying MPC controller. This agent uses its adaptive modal awareness to adjust weights for transient phases and its multi-scale optimization to balance short-term economy with long-term component health. The results indicate that AMDDPG-MPC can adapt effectively to the changing power demand in the WLTC and maintain favorable control performance over the whole test. Its terminal SOC reaches 0.541, showing that the strategy makes fuller use of the battery energy while coordinating the fuel cell to operate more efficiently. This behavior contributes to lower hydrogen consumption and demonstrates the practical potential of the proposed method for improving fuel economy without causing uncontrolled SOC deviation.

For a fair energy-consumption comparison, the final SOC is then restored to the initial value of 0.85 by stationary charging, and the additional charging energy is converted into equivalent hydrogen consumption.

4.2. Assessment of Energy Consumption by Different EMSs in UDDS Driving Cycles

The performance of the four energy management strategies is further examined under the UDDS cycle. Unlike WLTC, which contains a wider mixture of driving segments, UDDS

places greater emphasis on urban operating conditions, including repeated acceleration, braking, and idle periods. Therefore, using UDDS together with WLTC helps evaluate whether the proposed strategy can maintain stable energy-management performance in both mixed and city-oriented driving scenarios. Table 7 reports the terminal SOC, total equivalent hydrogen consumption (EHC), and energy-saving improvement of each EMS under the UDDS condition. The corresponding SOC evolution and cumulative EHC profiles are presented in Figures 9 and 10, respectively.

Table 5. The selected parameters.

Parameters	Value	Description
λ	0.55	Equivalent factor
α	0.7	Weight for fuel cell energy in the objective function
β	0.3	Weight for battery energy in the objective function
E_{batt}	Varies based on demand	Energy consumed from the battery
E_{FC}	Varies based on load	Energy consumed from the fuel cell

Table 6. Energy consumption performance by different EMSs in WLTC driving cycles.

EMSs	End SOC	Total Equivalent Hydrogen Consumption (g)	Energy-Saving Optimality (%)
RB	0.695	192.191	-
ECMS	0.687	178.687	7.026
MPC	0.638	162.833	15.275
AMDDPG-MPC	0.541	155.957	18.853

Table 7. Energy consumption performance by different EMSs in UDDS driving cycles.

EMSs	End SOC	Total Equivalent Hydrogen Consumption (g)	Energy-Saving Optimality (%)
RB	0.709	155.922	-
ECMS	0.698	144.269	7.474
MPC	0.683	139.770	10.359
AMDDPG-MPC	0.597	137.630	11.732

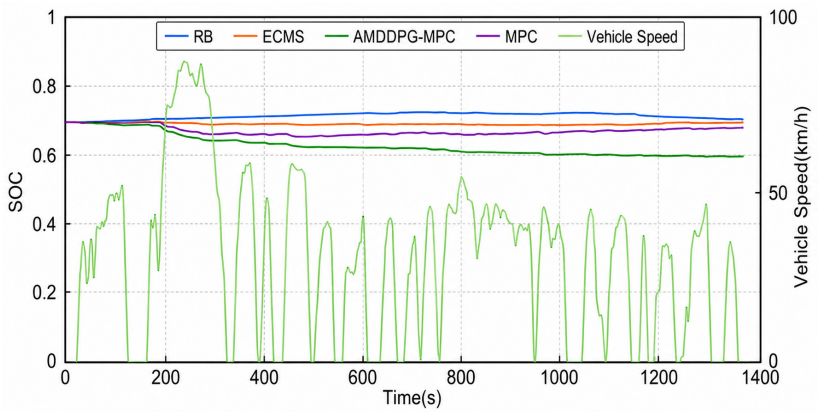


Figure 9. The changes in SOC in UDDS.

Following the WLTC analysis, the UDDS cycle is used to further assess the behavior of AMDDPG-MPC in urban traffic. As shown in Figure 9, the SOC trajectory fluctuates under the repeated start–stop events of UDDS, but the proposed strategy still regulates the battery charge within an acceptable range. Compared with MPC, AMDDPG-MPC maintains a higher SOC level, which helps preserve sufficient energy reserve for city driving. The

cumulative EHC curve in Figure 10 also shows that AMDDPG-MPC achieves a low energy-consumption level under the UDDS condition. According to Table 7, the proposed method obtains an energy-saving improvement of 11.732%. Although this improvement is smaller than that under WLTC, it remains evident compared with the benchmark EMSs. This benefit mainly comes from the adaptive modal awareness mechanism of AMDDPG, which identifies urban operating patterns such as low-speed congestion and rapid acceleration and then provides suitable weight commands for the MPC controller.

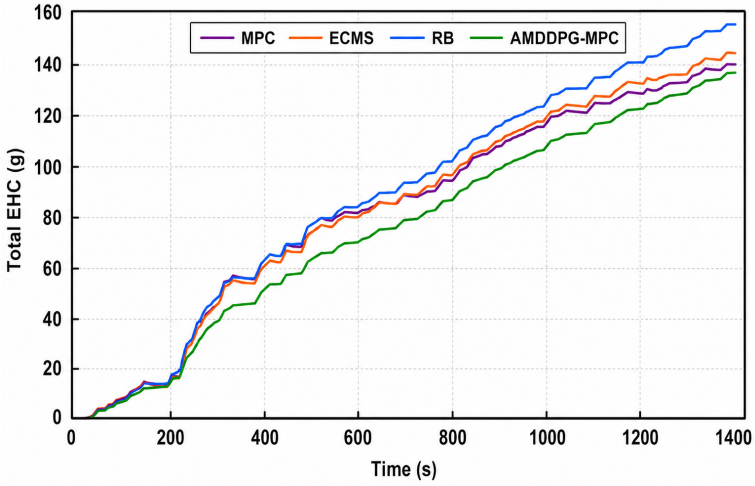


Figure 10. The changes in total EHC in UDDS.

The results under WLTC and UDDS indicate that AMDDPG-MPC can adapt to different driving characteristics. In WLTC, the strategy improves overall fuel economy under mixed-speed conditions, while in UDDS it responds effectively to frequent power-demand changes caused by acceleration, braking, and idling. By recognizing the current driving pattern and updating the MPC weights accordingly, the AMDDPG layer enables the lower-level controller to balance hydrogen consumption reduction and SOC regulation. Therefore, the proposed EMS shows stable control behavior across both mixed and urban scenarios, supporting its applicability to practical FCEV operation.

Figures 11 and 12 compare the fuel-cell operating-point distributions obtained by the different EMSs. In the WLTC case, AMDDPG-MPC places a larger share of operating points in the high-efficiency region of the fuel-cell system. The corresponding pie chart shows that only 7.4% of its operating points are located in the lowest-efficiency interval, which is lower than those of the other compared strategies. This result indicates that the proposed hierarchical strategy can guide the fuel cell away from inefficient operating regions. Since hydrogen consumption and fuel-cell degradation are considered during training, the AMDDPG agent tends to generate weights that encourage more economical and durability-aware operation.

For AMDDPG-MPC, the fuel-cell operating points are concentrated more reasonably within the efficient power range. Through adaptive adjustment of the hydrogen-consumption weight, the upper-level AMDDPG controller helps MPC respond to changing power demand while keeping the fuel cell near favorable efficiency regions. Compared with the benchmark methods, the proposed strategy reduces low-efficiency operation, thereby decreasing energy loss and improving fuel-cell utilization. This behavior not only contributes to hydrogen saving but also helps relieve unnecessary stress on the fuel-cell system, which is consistent with the long-term component-protection objective introduced in the multi-scale optimization design.

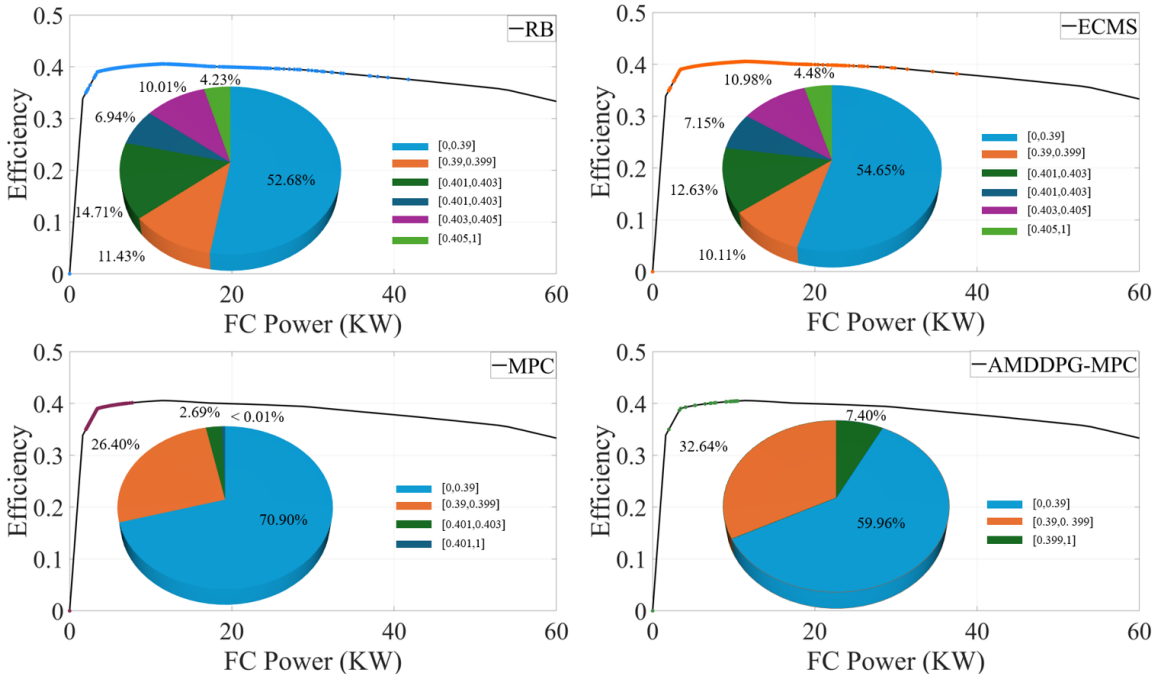


Figure 11. Fuel cell operating points by four EMSs in WLTC.

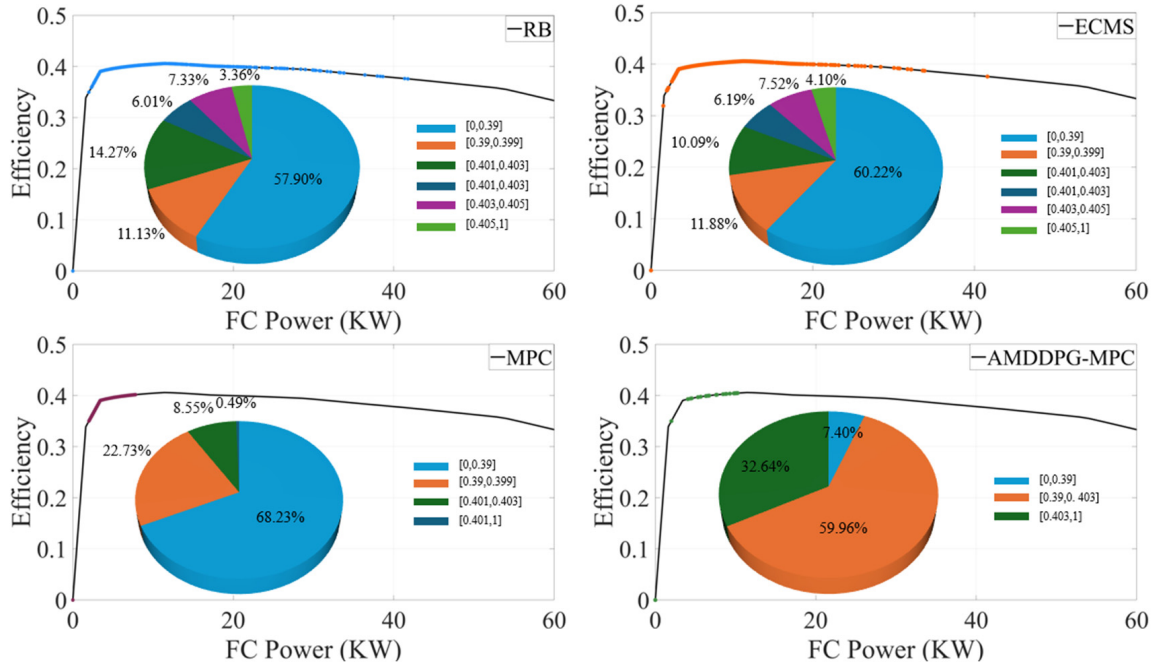


Figure 12. Fuel cell operating points by four EMSs in UDDS.

Under the UDDS cycle, AMDDPG-MPC still keeps a considerable share of fuel-cell operation within the high-efficiency region. As indicated by the pie chart, 32.64% of the operating points fall in this efficient range, showing that the proposed strategy can maintain favorable fuel-cell utilization even under urban conditions with repeated stops and low-speed operation. Such scenarios involve frequent acceleration and braking events, which impose rapid power-demand variations and therefore place higher requirements on EMS adaptability. AMDDPG-MPC's ability to sustain high efficiency under these conditions demonstrates its robustness, which stems from its adaptive modal awareness mechanism identifying and adapting to these challenging urban modes. AMDDPG-MPC minimizes the proportion of low-efficiency operating points in the UDDS cycle. By shortening the

operating duration in low-efficiency regions, AMDDPG-MPC improves the overall quality of energy distribution and makes the control response more stable. This advantage is especially important for urban driving, where rapid variations in vehicle speed and power demand make fixed-weight strategies less effective. The reduced proportion of low-efficiency fuel-cell operation further shows that the real-time weight-generation mechanism enables AMDDPG-MPC to better cope with stop-and-go traffic conditions.

As shown in Figure 13, the battery and fuel cell jointly meet the vehicle power demand during the driving cycle. The fuel cell output remains comparatively smooth, which helps avoid frequent load variations and thereby reduces load-cycling stress associated with membrane degradation and catalyst aging. Meanwhile, the battery is mainly used to absorb transient power demand and high-frequency fluctuations. Although the battery power changes more actively, AMDDPG-MPC keeps the SOC within a reasonable range, preventing excessive discharge and overcharge. In this way, the proposed strategy not only aims to reduce hydrogen consumption but also considers component protection through coordinated power allocation, helping improve the durability of both the fuel cell and the battery.

In summary, AMDDPG-MPC stands out for its remarkable adaptability and robustness across different driving conditions, as demonstrated by the fuel cell operating points in both WLTC and UDDS simulations. The strategy's ability to maintain high efficiency under varying power demands and driving scenarios underscores its versatility. This adaptability ensures that the vehicle can achieve optimal performance whether it is cruising on a highway or navigating through city traffic, as the hierarchical AMDDPG layer is designed to identify these distinct driving modes and provide tailored objectives for the MPC controller.

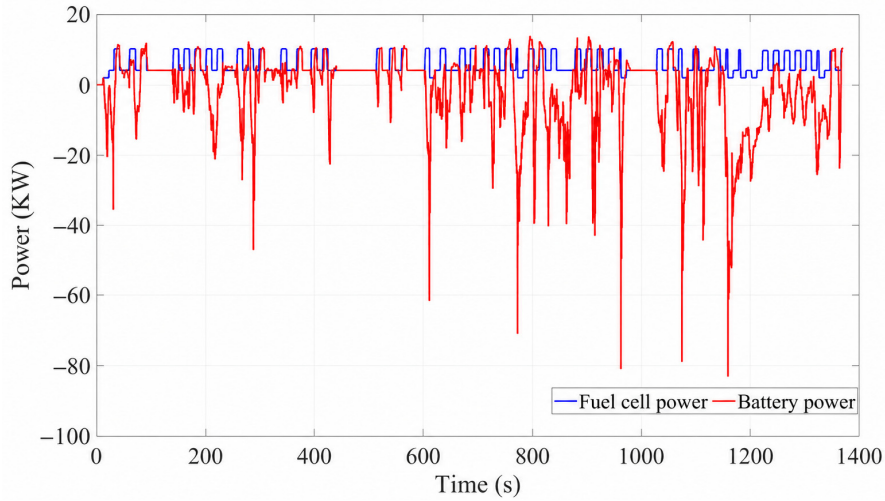


Figure 13. Fuel cell and battery power of AMDDPG-MPC in WLTC.

5. Conclusions

In this paper, a hierarchical AMDDPG–MPC energy management strategy is proposed for an FCEV, strengthening hydrogen-economy potential and improving the operating efficiency and durability awareness of the powertrain. Firstly, the adoption of a hierarchical control architecture enhances cross-scenario adaptability by decoupling upper-level dynamic weight generation from lower-level power-split optimization. Secondly, an implicit MPC with offline design and online inference reduces real-time computational burden while retaining constraint handling and receding-horizon optimality. Thirdly, the proposed AMDDPG-based multifaceted weight-generation process—featuring adaptive mode perception and a multi-scale reward—improves the anti-interference ability of MPC and

further enhances adaptability to time-varying driving conditions. Finally, compared with rule-based, ECMS, and fixed-weight MPC baselines, the proposed strategy achieves notable reductions in equivalent hydrogen consumption on standard cycles (about 18.9% on WLTC and 11.7% on UDDS), while maintaining stable SOC regulation and concentrating fuel-cell operation in high-efficiency regions. The simulation results verify that the hierarchical AMDDPG–MPC can effectively release economic potential with real-time feasibility.

However, to guarantee fair comparison, the required power in this study is determined from given a priori driving cycles, and the EHC evaluation employs a fixed equivalence factor with simplified component aging/thermal models. Future work will focus on integrating short-horizon power/speed prediction, embedding degradation-aware models, enabling online adaptation to forecast errors, and conducting hardware-in-the-loop validations.

Author Contributions: Conceptualization, Y.S. and Z.W.; methodology, Y.S. and Z.W.; software, Y.S.; validation, Y.S. and C.S.; formal analysis, Y.S. and W.H.; investigation, Y.S.; resources, Z.W. and Y.Z.; data curation, Y.S. and C.S.; writing—original draft preparation, Y.S.; writing—review and editing, Z.W., W.H. and Y.Z.; visualization, Y.S.; supervision, Z.W. and Y.Z.; project administration, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Key Project of Chongqing Municipal Science and Technology Bureau (Grant No. CSTB2022TIAD-KPX0038), National Key Research and Development Program of China (Grant No. 2021YFE0193800), Project of Postdoctoral Innovative Talent Support Program of Chongqing (Grant No. CQBX202516), and the Development and Industrial Application of Key Control Software Based on Energy Flow of New Energy Vehicles Project (Grant No. CSTB2024TIAD-STX0031). The authors thank them for their support.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request. The simulation models and datasets generated during the current study are not publicly available due to research confidentiality considerations.

Conflicts of Interest: Authors Yantao Si, Zhuo Wang, Changqun Sun, Wen He and Yunge Zou were employed by the company Chongqing Changan Automobile Co Ltd. All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yilmaz, T.; Yilmaz, A.; Erdinç, M.T. Energy recovery from hydrogen combustion at elevated pressures. *Int. J. Hydrogen Energy* **2017**, *42*, 23361–23365. [CrossRef]
2. Togun, H.; Aljibori, H.S.S.; Abed, A.M.; Biswas, N.; Alshamkhani, M.T.; Niyas, H.; Mohammed, H.I.; Rashid, F.L.; Dhabab, J.M.; Paul, D. A review on recent advances on improving fuel economy and performance of a fuel cell hybrid electric vehicle. *Int. J. Hydrogen Energy* **2024**, *89*, 22–47. [CrossRef]
3. Showers, S.O.; Raji, A.K. State-of-the-art review of fuel cell hybrid electric vehicle energy management systems. *AIMS Energy* **2022**, *10*, 458–485. [CrossRef]
4. Zhang, F.; Hu, X.; Langari, R.; Cao, D. Energy management strategies of connected HEVs and PHEVs: Recent progress and outlook. *Prog. Energy Combust. Sci.* **2019**, *73*, 235–256. [CrossRef]
5. Khalatbarisoltani, A.; Zhou, H.; Tang, X.; Kandidayeni, M.; Boulon, L.; Hu, X. Energy management strategies for fuel cell vehicles: A comprehensive review of the latest progress in modeling, strategies, and future prospects. *IEEE Trans. Intell. Transp. Syst.* **2023**, *25*, 14–32. [CrossRef]
6. He, H.; Jia, C.; Li, J. A new cost-minimizing power-allocating strategy for the hybrid electric bus with fuel cell/battery health-aware control. *Int. J. Hydrogen Energy* **2022**, *47*, 22147–22164. [CrossRef]
7. Sulaiman, N.; Hannan, M.A.; Mohamed, A.; Ker, P.J.; Majlan, E.H.; Daud, W.W. Optimization of energy management system for fuel-cell hybrid electric vehicles: Issues and recommendations. *Appl. Energy* **2018**, *228*, 2061–2079. [CrossRef]
8. Wang, Y.; Sun, Z.; Chen, Z. Rule-based energy management strategy of a lithium-ion battery, supercapacitor and PEM fuel cell system. *Energy Procedia* **2019**, *158*, 2555–2560. [CrossRef]

9. Li, C.Y.; Liu, G.P. Optimal fuzzy power control and management of fuel cell/battery hybrid vehicles. *J. Power Sources* **2009**, *192*, 525–533. [CrossRef]
10. Chen, J.; Xu, C.; Wu, C.; Xu, W. Adaptive fuzzy logic control of fuel-cell-battery hybrid systems for electric vehicles. *IEEE Trans. Ind. Inform.* **2016**, *14*, 292–300. [CrossRef]
11. Tao, S.; Chen, W.; Gan, R.; Li, L.; Zhang, G.; Han, Y.; Li, Q. Energy management strategy based on dynamic programming with durability extension for fuel cell hybrid tramway. *Railw. Eng. Sci.* **2021**, *29*, 299–313. [CrossRef]
12. Ou, K.; Yuan, W.W.; Choi, M.; Yang, S.; Jung, S.; Kim, Y.B. Optimized power management based on adaptive-PMP algorithm for a stationary PEM fuel cell/battery hybrid system. *Int. J. Hydrogen Energy* **2018**, *43*, 15433–15444. [CrossRef]
13. Ferrara, A.; Okoli, M.; Jakubek, S.; Hametner, C. Energy management of heavy-duty fuel cell electric vehicles: Model predictive control for fuel consumption and lifetime optimization. *IFAC-PapersOnLine* **2020**, *53*, 14205–14210. [CrossRef]
14. Pramuanjaroenkij, A.; Kakaç, S. The fuel cell electric vehicles: The highlight review. *Int. J. Hydrogen Energy* **2023**, *48*, 9401–9425. [CrossRef]
15. Jiang, H.; Xu, L.; Li, J.; Hu, Z.; Ouyang, M. Energy management and component sizing for a fuel cell/battery/supercapacitor hybrid powertrain based on two-dimensional optimization algorithms. *Energy* **2019**, *177*, 386–396. [CrossRef]
16. Choi, K.; Byun, J.; Lee, S.; Jang, I.G. Adaptive equivalent consumption minimization strategy (A-ECMS) for the HEVs with a near-optimal equivalent factor considering driving conditions. *IEEE Trans. Veh. Technol.* **2021**, *71*, 2538–2549. [CrossRef]
17. Chen, H.; Chen, J.; Lu, H.; Yan, C.; Liu, Z. A modified MPC-based optimal strategy of power management for fuel cell hybrid vehicles. *IEEE/ASME Trans. Mechatron.* **2020**, *25*, 2009–2018. [CrossRef]
18. Gharibeh, H.F.; Yazdankhah, A.S.; Azizian, M.R. Energy management of fuel cell electric vehicles based on working condition identification of energy storage systems, vehicle driving performance, and dynamic power factor. *J. Energy Storage* **2020**, *31*, 101760. [CrossRef]
19. Quan, S.; Wang, Y.X.; Xiao, X.; He, H.; Sun, F. Real-time energy management for fuel cell electric vehicle using speed prediction-based model predictive control considering performance degradation. *Appl. Energy* **2021**, *304*, 117845. [CrossRef]
20. Han, L.; Jiao, X.; Zhang, Z. Recurrent neural network-based adaptive energy management control strategy of plug-in hybrid electric vehicles considering battery aging. *Energies* **2020**, *13*, 202. [CrossRef]
21. Venkatasatish, R.; Dhanamjayulu, C. Reinforcement learning based energy management systems and hydrogen refuelling stations for fuel cell electric vehicles: An overview. *Int. J. Hydrogen Energy* **2022**, *47*, 27646–27670. [CrossRef]
22. 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]
23. 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.* **2021**, *69*, 1537–1547. [CrossRef]
24. Zhang, B.; Zou, Y.; Zhang, X.; Du, G.; Jiao, F.; Guo, N. Online updating energy management strategy based on deep reinforcement learning with accelerated training for hybrid electric tracked vehicles. *IEEE Trans. Transp. Electrif.* **2022**, *8*, 3289–3306. [CrossRef]
25. Zhang, B.; Wu, J.; Zou, Y.; Zhang, X. Reinforcement Learning Energy Management for Hybrid Electric Tracked Vehicle with Deep Deterministic Policy Gradient. In *Proceedings of the China SAE Congress 2020: Selected Papers*; Springer Nature: Singapore, 2022; pp. 879–893.
26. Qi, X.; Luo, Y.; Wu, G.; Boriboonsomsin, K.; Barth, M. Deep reinforcement learning enabled self-learning control for energy efficient driving. *Transp. Res. Part C Emerg. Technol.* **2019**, *99*, 67–81. [CrossRef]
27. He, H.; Cao, J.; Cui, X. Energy optimization of electric vehicle's acceleration process based on reinforcement learning. *J. Clean. Prod.* **2020**, *248*, 119302. [CrossRef]
28. Huang, Y.; Hu, H.; Tan, J.; Lu, C.; Xuan, D. Deep reinforcement learning based energy management strategy for range extend fuel cell hybrid electric vehicle. *Energy Convers. Manag.* **2023**, *277*, 116678. [CrossRef]
29. Li, W.; Ye, J.; Cui, Y.; Kim, N.; Cha, S.W.; Zheng, C. A speedy reinforcement learning-based energy management strategy for fuel cell hybrid vehicles considering fuel cell system lifetime. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2022**, *9*, 859–872. [CrossRef]
30. Chen, Z.; Gu, H.; Shen, S.; Shen, J. Energy management strategy for power-split plug-in hybrid electric vehicle based on MPC and double Q-learning. *Energy* **2022**, *245*, 123182. [CrossRef]
31. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]
32. 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]
33. Yuan, J.; Yang, L.; Chen, Q. Intelligent energy management strategy based on hierarchical approximate global optimization for plug-in fuel cell hybrid electric vehicles. *Int. J. Hydrogen Energy* **2018**, *43*, 8063–8078. [CrossRef]
34. Tan, H.; Zhang, H.; Peng, J.; Jiang, Z.; Wu, Y. Energy management of hybrid electric bus based on deep reinforcement learning in continuous state and action space. *Energy Convers. Manag.* **2019**, *195*, 548–560. [CrossRef]

35. Huang, Y.; Kang, Z.; Mao, X.; Hu, H.; Tan, J.; Xuan, D. Deep reinforcement learning based energy management strategy considering running costs and energy source aging for fuel cell hybrid electric vehicle. *Energy* **2023**, *283*, 129177. [CrossRef]
36. Tang, X.; Zhou, H.; Wang, F.; Wang, W.; Lin, X. Longevity-conscious energy management strategy of fuel cell hybrid electric Vehicle Based on deep reinforcement learning. *Energy* **2022**, *238*, 121593. [CrossRef]
37. Sun, H.; Fu, Z.; Tao, F.; Zhu, L.; Si, P. Data-driven reinforcement-learning-based hierarchical energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles. *J. Power Sources* **2020**, *455*, 227964. [CrossRef]
38. Xu, D.; Cui, Y.; Ye, J.; Cha, S.W.; Li, A.; Zheng, C. A soft actor-critic-based energy management strategy for electric vehicles with hybrid energy storage systems. *J. Power Sources* **2022**, *524*, 231099. [CrossRef]
39. 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]
40. Dong, H.; Zhao, L.; Zhou, H.; Li, H. Hierarchical Optimization Based on Deep Reinforcement Learning for Connected Fuel Cell Hybrid Vehicles through Signalized Intersections. *Processes* **2023**, *11*, 2689. [CrossRef]
41. Wang, Y.; Advani, S.G.; Prasad, A.K. A comparison of rule-based and model predictive controller-based power management strategies for fuel cell/battery hybrid vehicles considering degradation. *Int. J. Hydrogen Energy* **2020**, *45*, 33948–33956. [CrossRef]
42. Li, H.; Ravey, A.; N'Diaye, A.; Djerdir, A. Online adaptive equivalent consumption minimization strategy for fuel cell hybrid electric vehicle considering power sources degradation. *Energy Convers. Manag.* **2019**, *192*, 133–149. [CrossRef]
43. 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]

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.