Optimal Energy Management for Multi-Stack Fuel Cell Vehicles based on Hybrid Quantum Reinforcement Learning

Wenzhuo Shi, Xianzhuo Sun, *Member, IEEE,* Zelong Zhang, Junyu Chen, *Member, IEEE,* Yuhua Du, *Member, IEEE,* Jiaqi Ruan, *Member, IEEE,* Yibo Ding, Lei Wang, Yigeng Huangfu, *Senior Member, IEEE,* Zhao Xu, *Senior Member, IEEE*

***Abstract*—This paper proposes a driving condition recognition-based (DCR-based) Hybrid Quantum Deep Deterministic Policy Gradient (HQDDPG) method for energy management in multi-stack fuel cell vehicle hybrid power systems (MFCV HPSs) and its quantum simulation setup on Digital Signal Processors (DSPs). Driving conditions are initially segmented into micro-trips and clustered into three types. The DCR method, using a Learning Vector Quantization Neural Network (LVQNN), is then developed, thus accurately and efficiently identifying driving condition types. Subsequently, quantum reinforcement learning is proposed to achieve optimal energy manaement of MFCV HPSS, i.e., power allocation among the multiple fuel cells to minimize the economic metrics based on the DCR results. Compared to classical large-scale neural networks, quantum reinforcement learning reduces parameters by combining a Parameterized Quantum Circuit (PQC) with a single-layer classical neural network. The PQC encodes and processes state information through quantum mechanics for enhanced computational expressiveness, while the classical neural network transforms the quantum measurement expectations into actionable outputs for energy management. The trained hybrid quantum circuits are implemented on DSPs through quantum simulations. The method is validated through Controller Hardware-in-the-Loop (CHIL) experiments, demonstrating superior performance in optimizing economic metrics compared to conventional methods.**

***Index Terms*—Multi-stack fuel cell vehicle, energy management, driving condition recognition, hybrid quantum reinforcement learning, DSP-based quantum simulation.**

I. Introduction

CURTAILING carbon emissions is propelling the search for sustainable transportation technologies [1]-[6]. The transportation sector, as a significant contributor to global greenhouse gas emissions, faces heightened demands to shift from fossil fuel dependency to more environmentally friendly fuel [7]-[10]. In this context, fuel cell vehicles (FCVs), which harness hydrogen to convert chemical energy into electrical energy, have attracted significant attention as a promising sustainable transportation technology driving the transition toward cleaner automotive solutions. Traditionally, the implementation of FCVs designed with a single fuel cell stack is limited by specific applications with high power demands [11]. This challenge has led to the exploration of more advanced configurations, specifically multi-stack fuel cell systems (MFCS). By integrating multiple fuel cell stacks, these systems not only enhance the power output but also bolster overall system reliability [12]-[13]. However, the development of MFCS will introduce high complexity stems from the integration of multiple energy sources and the real-time balancing of energy flows and pose significant challenges on the optimal energy management of such a hybrid power system. Therefore, an effective energy management strategy (EMS) is essential to optimizing system performance, extending vehicle range, and maximizing the lifespan of the fuel cells.

Traditional EMSs of HPS can be categorized mainly into rule-based, optimization-based and learning-based methods [14] [15]. Rule-based EMS, with predefined expert rules, is easy to implement but relies heavily on expert experience and struggles to adapt to unexpected conditions [16]-[17]. By contrast, optimization-based EMS utilizes optimization techniques, such as quadratic programming, dynamic programming, and Pontryagin's minimum principle, to achieve global optimal energy management for HPS [18]-[21]. However, optimization-based EMS requires sufficient computational resources and complete driving condition information, which hinders their practical and real-time applications in vehicles.

To address the limitations of model-based approaches, reinforcement learning (RL)-based energy management, with its model-free nature and efficiency in handling complex decision-making tasks, have been developed [22]-[27]. A hierarchical energy management strategy incorporating deep RL (DRL) was proposed to enhance fuel economy and extend the lifespan of power sources in FCVs, effectively optimizing power allocation between fuel cells and batteries while

This work was supported in part by National Natural Science Foundation of China under Grant 72331008, and 72271211, in part by Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under Grant CX2023066, and in part by the PolyU under Grant 1-YWCV. (Corresponding author, Yuhua Du and Zhao Xu).

W. Shi is with the School of Automation, Northwestern Polytechnical University, Xi'an, China, and also with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China. (e-mail: shiwenzhuo@mail.nwpu.edu.cn, 23036609r@connect.polyu.hk)

X. Sun, J. Chen, and J. Ruan are with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China. (e-mail: xianzsun@polyu.edu.hk, junychen@polyu.edu.hk, jiaqi.ruan@polyu.edu.hk)

Z. Zhang, L. Wang, Y. Du, and Y. Huangfu are with the School of Automation, Northwestern Polytechnical University, Xi'an, China. (e-mail: zhang_2017@mail.nwpu.edu.cn, l.wang@mail.nwpu.edu.cn, yuhua.du@ieee.org, yigeng@nwpu.edu.cn)

Y. Ding and Z. Xu are with the Department of Electrical and Electronic Engineering and the Research Institute of Smart Energy, The Hong Kong Polytechnic University, Hong Kong SAR, China. (e-mail: yibo0712.ding@connect.polyu.hk, zhao.xu@polyu.edu.hk)

mitigating peak power stress [24]. In [25], the authors proposed a knowledge-guided DRL approach for EMS in FCVs, integrating expert knowledge to enhance DRL's efficiency and adaptability, and superior performance in multi-objective optimization tasks across various scenarios was demonstrated. Ref. [26] introduced a DRL and PMP-based EMS for FCHEVs, significantly reducing fuel cell and battery degradation, and demonstrating potential for real-time applications. We also introduced an EMS for multiple fuel cell stacks FCVs HPS using the independent Q-learning algorithm, ensuring robust operation and efficient energy management, with capabilities for real-time implementation and validation under various conditions [27]. However, changes in driving condition categories, represented by state transition probabilities in RL, will compromise the performance of the employed EMS agent. To address this, some researchers have concentrated on adaptive EMS by driving conditions recognition (DCR) [28]-[30]. Ref. [28] introduced a hybrid method combining Markov chain prediction and fuzzy classification models to enhance the DCR accuracy in occurring micro-trips and EMS performance. In [29], a DCR method utilizing supervised learning demonstrated that integrating multiple algorithms within hierarchical frameworks significantly enhances recognition accuracy and fuel economy in vehicle operations. A fusion algorithm-based DCR method combining a Markov chain prediction model and a learning vector quantization neural network (LVQNN) was proposed to enhance real-time control by minimizing time delays, significantly improving the adaptability of energy management strategies in medium fuel cell trucks [30]. Therefore, we can utilize DCR methods to train distinct RL-based EMS agents for various driving condition categories. While this method tailors EMS to specific conditions, it significantly increases the total number of parameters due to the need for multiple neural networks, which consume excessive memory. Consequently, deploying such systems in real-world applications becomes impractical, as EMS should not excessively tax the computational resources of the broader vehicular systems.

Based on existing DRL schemes, few researchers have introduced quantum circuits, leveraging their superior representational capacities, to reduce parameters of neural networks for memory saving [31]-[33]. Ref. [31] proposed a classical-quantum hybrid algorithm called quantum circuit learning, which utilizes low-depth quantum circuits and classical computing for machine learning on near-term quantum processors, demonstrating the ability to approximate nonlinear functions and optimize performance through iterative parameter adjustments. In [32], the optimization of Parameterized Quantum Circuits (PQCs) is analyzed, identifying key descriptors such as expressibility and entangling capability. The study highlighted the superior performance of two-qubit gates in ring or all-to-all configurations over linear setups and found controlled X-rotation gates to outperform controlled Z-rotation gates in enhancing circuit expressibility. PQCs have shown to outperform classical neural networks in generative tasks, with enhanced capabilities when equipped with ancillary qubits for post-selection, demonstrating potential for broader applications in learning scenarios [33]. Therefore, incorporating quantum circuits into DRL enables more complex policies using fewer parameters. Theoretically, quantum neural networks have demonstrated enhanced capacity for learning, as proven by fisher information metrics, which highlight their superior expressibility with fewer parameters [34]. In the application, the quantum advantage actor-critic successfully formulates effective strategies in the Cart Pole environment using only 50-60 parameters, in contrast to the traditional actor-critic, which fails to converge [35]. It has also been verified in distributed frequency control of islanded microgrids, where multiagent QDRL not only improved frequency regulation but also demonstrated a quantum advantage in reducing parameters [36]. However, as the number of quantum gates increases, the fidelity of quantum operations decreases, resulting in a higher error rate in the real quantum computer [37]. Besides, the environmental conditions, such as ultra-low temperatures and vibration sensitivity, are not conducive to the dynamic and varied environments encountered in vehicular applications. This hinders the development of quantum computers in vehicular applications.

In response to these challenges, this paper proposes a DCR-based Hybrid Quantum Deep Deterministic Policy Gradient (HQDDPG) method for EMS in MFCS HPS and its quantum simulation setup on DSP. Initially, micro-trip driving conditions are clustered into three types. Then, the appropriate PQC structure is selected and trained into multiple EMS agents based on clustering results. Subsequently, the mathematical formulations for the PQCs are derived from the trained EMS agents, which enables the implementation of quantum simulations on DSPs. It not only addresses computational limitations but also circumvents the need for quantum computers in vehicular environments. The contributions are discussed as follows:

- Given the complex transition, the DCR stratifies driving cycles into distinct types through clustering based on their features. By integrating DCR into the RL-based EMS, the complexity managed by one EMS agent can be decomposed and distributed across multiple EMS agents, each tailored to one distinct driving cycle type with homogeneous characteristics. This enhances the EMS's adaptability and performance across varied driving conditions.
- The integration of DCR exacerbates the substantial parameter burden of RL-based EMS agents, complicating the practical implementation in real-world controllers. To address this challenge, PQCs are firstly integrated into the EMS of MFCV HPS. This integration leverages the high expressiveness of PQCs in conjunction with DRL, substantially reducing the number of parameters required.
- While PQCs can reduce parameter requirements, integrating quantum computers into vehicular systems remains impractical in the near term. To bridge this gap, we innovatively propose the use of quantum simulation to implement PQCs on Digital Signal Processors (DSPs) or other Microcontroller Units (MCUs). This approach ensures the practical application of advanced quantum-enhanced EMS within existing vehicle technologies, cir-

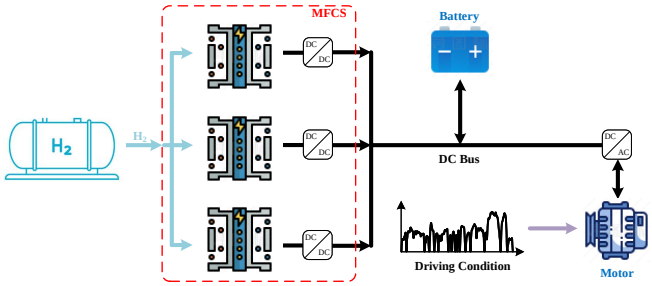

Fig. 1. The configuration of MFCV HPS.

cumventing the current limitations of quantum hardware deployment in automotive environments.

The remainder of this paper is organized as follows: Section II describes the configuration of the Multi-Stack Fuel Cell Vehicle Hybrid Power System and its energy management problem. Section III introduces the Driving Condition Recognition-based Quantum-Simulated Hybrid Quantum Deep Deterministic Policy Gradient method. Simulation results and discussions are presented in Section IV. Finally, conclusions and potential directions for future research are discussed in Section V.

II. MFCV HPS Configuration and Energy Management Problem

This section outlines the configuration of the MFCV HPS, as shown in Fig. 1, and introduces the economic metrics and problem modeling for evaluating the EMS. The HPS configuration comprises three identical Proton Exchange Membrane fuel cells and one battery, all connected in parallel. These fuel cells are identical in terms of design and operational parameters, including degradation factors [8] [27]. Each power source integrates with the HPS through individual DC/DC converters, enhancing system flexibility. Subsequently, the section delves into defining economic metrics that focus on hydrogen consumption and degradation. These metrics are vital for assessing the effectiveness of the EMS in improving the operational efficiency and extending the lifespan of the HPS. Furthermore, we integrate economic metrics to formulate an energy management problem model.

A. Economic Metrics of Fuel Cells Hydrogen Consumption

The economic metrics are calculated based on experimental data to reflect real-world efficiency under varying power outputs. The hydrogen efficiency of the fuel cell, according to studies like [38], can be approximated as a function of its power output under controlled temperature and humidity conditions. This approach simplifies the complex dynamics typically observed in fuel cell operations:

$$H_{fc,k} = \int \frac{P_{fc,k}}{\eta_{fc,k} \times LHV_{H_2}} dt \tag{1}$$

where $H_{fc,k}$ represents the hydrogen consumed by the k-th fuel cell, $P_{fc,k}$ denotes the power output to the DC bus, and LHV_{H_2} is the lower heating value of hydrogen, typically 120 MJ/kg. $\eta_{fc,k}$, the efficiency of the k-th fuel cell and its converter, is not a constant but a non-linear function of the power output $P_{fc,k}$. This functional relationship is derived from experimental data and polynomial fitting, as detailed in our previous work [27]. Equation (1) reflects the direct relationship between the efficiency of the energy conversion process and the amount of hydrogen needed. In analyzing the economic metrics associated with hydrogen consumption, both the price of the hydrogen and the operational efficiencies ought to be considered:

$$C_{H_2} = \sum_{k=1}^{n} H_{fc,k} \times Pr_{H_2} \tag{2}$$

where n is the number of fuel cells, C_{H_2} is the economic cost of hydrogen, and Pr_{H_2} is the price per unit of hydrogen, $\$2.50$ per kilogram based on the mid-range estimates provided by recent market analyses [39].

B. Economic Metrics of Fuel Cells Degradation

Dynamic operational stress and environmental conditions, particularly high fluctuations in power output, can exacerbate the degradation of fuel cell components in MFCVs. This degradation impacts the long-term economic viability of fuel cells. Due to the absence of a definitive mathematical model for fuel cell system degradation, we adopt an empirical approach that focuses on these power output fluctuations to assess their effects on system durability [40].

The degradation level of the fuel cell, denoted as $\Delta\Gamma_{fc}(P_{fc,k})$, is modeled as a function of the operational conditions [8]:

$$\Delta\Gamma_{fc}(P_{fc,k}) = \frac{\tau_s(P_{fc,k})}{10\%} \tag{3}$$

$$\tau_s(P_{fc,k}) = c_\tau \times \int \left| \frac{dP_{fc,k}}{dt} \right| \tag{4}$$

where c_τ represents the stress degradation factors, 3.51×10^{-13}. It influences the lifecycle and operational efficiency of the fuel cells, necessitating periodic replacements.

The economic impact of fuel cell degradation is directly tied to the cost of replacing degraded components. The replacement cost, $C_{fc,rep}$, is a function of the rated power and unit cost:

$$C_{fc,rep} = P_{fc,rat} \times P_{r_{fc}}^{unit} \tag{5}$$

where $P_{fc,rat}$ is the rated power, $30kW$, and $P_{r_{fc}}^{unit}$ is the price per kW, $\$75/kW$. This equation is key for estimating replacement costs from routine maintenance and degradation. To quantify the economic metrics of degradation, consider:

$$C_{deg} = \Delta\Gamma_{fc}(P_{fc,k}) \times C_{fc,rep} \tag{6}$$

C. Economic Metrics of Battery Equivalent Hydrogen Consumption

To quantify the equivalent hydrogen consumption of the battery H_{bat}, it is first calculated based on the power output of the battery P_{bat}:

$$H_{bat} = \begin{cases} \frac{P_{bat}}{\eta_{dis} LHV_{H_2}}, & \text{if } P_{bat} \geq 0 \\ \frac{\eta_{chg} P_{bat}}{LHV_{H_2}}, & \text{if } P_{bat} < 0 \end{cases} \tag{7}$$

where η_{dis} and η_{chg} are the discharge and charge efficiencies of the battery, respectively, as detailed in our previous work [27]. With the equivalent hydrogen consumption determined, the next step is to convert this figure into cost terms. The concept of battery equivalent hydrogen consumption, denoted as H_{bat}, is developed to quantify the energy output of the battery in terms analogous to the hydrogen consumption of fuel cells [41]. It allows us to integrate economic considerations directly with power output [11] [15]. The economic value of H_{bat} is calculated by considering the current market price of hydrogen Pr_{H_2}:

$$C_{H_{eq}} = H_{bat} \times Pr_{H_2} \tag{8}$$

D. Energy Management Problem Model

By integrating the economic metrics established in the previous section, the MFCV HPS energy management problem is formulated as a Markov Decision Process (MDP). The model is tailored to the inherent physical characteristics of the fuel cells and the battery and aims to optimize HPS economic metrics in multi-stage.

1) Multi-Stage Decision Problem Formulation of the Energy Management Problem: The MFCV HPS energy management problem involves state variables such as the vehicle's velocity (v), acceleration (a), and the battery's state of charge (SOC_{BAT}). They can be represented in a set for each timestep t:

$$S_t = \{v(t), a(t), SOC_{BAT}(t)\} \tag{9}$$

The energy management problem for the MFCV HPS is modeled as a multi-stage decision problem. The primary goal is to minimize the total economic cost accrued over the entire operation period. This objective is formulated as the sum of the cost functions across all stages, where the fuel cell power outputs are decided at each stage:

$$J = \sum_{t=0}^{T-1} L(S_t, \mathcal{P}_{fc}(t)) \tag{10}$$

$$\begin{aligned} L(S_t, \mathcal{P}_{fc}(t)) = {} & C_{H_2}(S_t, \mathcal{P}_{fc}(t)) \\ & + C_{deg}(S_t, \mathcal{P}_{fc}(t)) \\ & + C_{H_{eq}}(S_t, \mathcal{P}_{fc}(t)) \end{aligned} \tag{11}$$

$$S_{t+1} = f(S_t, \mathcal{P}_{fc}(t)), \quad t = 0, 1, 2, \ldots, T-1 \tag{12}$$

where, J represents the aggregate of the costs at each timestep; $L(S_t, \mathcal{P}_{fc}(t))$ is the cost function at each timestep, encompassing costs associated with hydrogen consumption C_{H_2}, $C_{H_{eq}}$ and system degradation C_{deg}. S_{t+1} updates based on the system dynamics influenced by the adopted decisions.

2) Fuel Cells Operation Constrains: As stated in Ref. [27], the battery is utilized to stabilize the DC bus voltage, enabling control over the power outputs of three fuel cells. Therefore, the power outputs of fuel cells serve as decision variables in the model:

$$\mathcal{P}_{fc}(t) = \{P_{fc,1}(t), P_{fc,2}(t), P_{fc,3}(t)\} \tag{13}$$

Due to the electrochemical process within the fuel cells, rapid changes in power output are restricted [40]. Therefore, the power output of each fuel cell is not only constrained by its maximum rated power but also by its inherent ramp rate limitations. These constraints accommodates the slow response characteristics of fuel cells and protects the HPS against potential operational risks:

- The power output for k-th fuel cell must remain within its maximum and minimum power limits:

$$P_{fc,\min} \leq P_{fc,k}(t) \leq P_{fc,\max}, \quad k \in \{1, 2, 3\} \tag{14}$$

- The ramp rate is regulated to ensure gradual power adjustments:

$$|P_{fc,k}(t) - P_{fc,k}(t-1)| \leq \Delta P_{fc,\max}, \quad k \in \{1, 2, 3\} \tag{15}$$

3) Battery Operation Constrains: As previously discussed, the battery is utilized to stabilize the bus voltage. Consequently, the power output of the battery is determined by the power balance constraints:

$$\sum_{k=1}^{3} P_{fc,k}(t) + P_{BAT}(t) = P_{\text{demand}}(t) \tag{16}$$

$$P_{BAT,\min} \leq P_{BAT}(t) \leq P_{BAT,\max} \tag{17}$$

$$\text{SOC}_{\min} \leq \text{SOC}_{BAT}(t) \leq \text{SOC}_{\max} \tag{18}$$

$$\Delta\text{SOC}_{BAT} = \frac{\Delta t}{3600} \times \left(\frac{P_{BAT}(t)}{E_{BAT}} \times \eta(t) \right) \tag{19}$$

$$\text{SOC}_{BAT}(t+1) = \text{SOC}_{BAT}(t) + \Delta\text{SOC}_{BAT} \tag{20}$$

where $P_{BAT,\min}$ and $P_{BAT,\max}$ represent the minimum and maximum power that the battery can discharge or charge, respectively. $\text{SOC}_{\min}$ and $\text{SOC}_{\max}$ are the minimum and maximum allowable states of charge for the battery, protecting against excessive depletion or overcharging. Δt is the time step in seconds, and E_{BAT} is the total energy capacity of the battery in watt-hours. The efficiency $\eta(t)$ varies, being $1/\eta_{\text{chg}}$ when charging (if $P_{BAT}(t) > 0$) and η_{dis} when discharging (if $P_{BAT}(t) < 0$), reflecting efficiency losses during energy conversion processes.

4) Vehicle Power Demand Calculation: The power demand of the vehicle, denoted by P_{demand}, is determined by the sum of the forces acting on the vehicle multiplied by its velocity (v), where v represents the longitudinal velocity of the vehicle unaffected by wind influences. The equation for P_{demand} is given as [27]:

$$P_{\text{demand}} = (F_a + F_r + F_f + F_g) \cdot v \tag{21}$$

where, $F_a = \frac{1}{2}\rho A C_d v^2$ is the aerodynamic drag force, with ρ being the air density, A the frontal area of the vehicle, and C_d the drag coefficient. The acceleration force is calculated as $F_r = M \cdot a$, where M is the mass of the vehicle and acceleration a. The rolling resistance force is $F_f = \mu M g$, with μ as the rolling resistance coefficient and g the acceleration due to gravity. $F_g = Mg\sin\theta$ represents the gravitational component along the slope of the road, which is considered to be zero in this study.

In summary, the energy management problem is modeled as a Markov Decision Process (MDP), which is a special

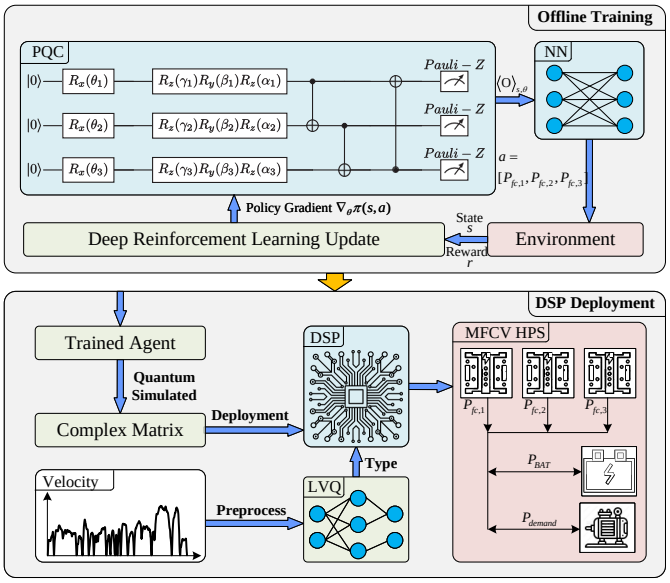

Fig. 2. The framework of HQDDPG.

case of multi-stage decision problems. It can capture the sequential decision-making process involved in energy management, where the system's state at each time step influences the decisions made at the next time step. Specifically, the state space S_t includes variables such as the vehicle's velocity (v), acceleration (a), and the battery's state of charge (SOC_{BAT}), which together describe the system's operational conditions at any given time. The action space $\mathcal{A}_t$, or decision variables, consists of the power outputs of the fuel cells $\mathcal{P}_{fc}(t) = \{P_{fc,1}(t), P_{fc,2}(t), P_{fc,3}(t)\}$. By formulating the problem as an MDP, we can effectively incorporate the nonlinear relationships between battery and fuel cell efficiencies, which are difficult to handle with traditional optimization methods. This approach avoids the need for binary variables and allows us to model the system dynamics and decision-making process in a way that reflects the complex, real-world interactions of the energy management system.

III. DCR-based Hybrid Quantum Deep Deterministic Policy Gradient Method and its Quantum Simulation Setup on DSP

This section introduces the DCR-based Hybrid Quantum Deep Deterministic Policy Gradient method and its quantum simulation setup on DSP, as shown in Fig. 2. The approach is divided into two parts. The first part involves the DCR method based on micro-trips, where classical driving conditions are decomposed into micro-trips then clustered. The classification results are used to train a LVQNN. The second part integrates a hybrid quantum DDPG method, which combines PQC with classical neural networks. This hybrid framework leverages the expressive power of PQCs while and classical neural networks are used to change the number of actions. Then, the trained hybrid method is implemented on DSPs or other MCUs through quantum simulations.

A. Micro-trip Clustering and Driving Condition Recognition

Generally, driving conditions are classified according to classical driving cycles such as the Urban Dynamometer Driving Schedule (UDDS), the New European Driving Cycle (NEDC) and so on. However, recent research has shown that using micro-trips as the classification criterion results in better performance than classical driving cycles [30]. Many researchers have explored various neural networks to be the structure of the DCR, among which the LVQNN is widely adopted due to simplicity and high recognition accuracy [42].

Specifically, the implementation of the DCR method begins by dividing driving cycles into smaller segments called micro-trips. Each micro-trip is characterized by distinct operational patterns. Once the micro-trips are defined, the next step involves clustering these segments to identify common driving conditions. The clustering process utilizes specific features that capture the critical characteristics of each micro-trip. These features include maximum velocity ($v_{\max}$), average velocity (v_{avg}), maximum acceleration ($a_{\max}$), average acceleration (a_{avg}), maximum de-acceleration ($d_{\max}$), and average de-acceleration (d_{avg}), as shown in Fig. 3. After the clustering, LVQ is employed to train a model on these categorized micro-trips. LVQ is a type of supervised machine learning algorithm that is particularly effective for classification problems where the training dataset includes labels. By iteratively adjusting the prototypes (vectors representing each category), LVQ minimizes classification errors. The trained LVQ model is employed in real-time to predict driving conditions. We utilize a 50-second sliding window to continuously update and analyze the vehicle's driving cycle. For the initial 50 seconds, any missing data points are filled with zeros to maintain a consistent input size for the model. After this initialization period, the sliding window is updated every second, refreshing the input with the most recent driving data. Within this window, we process the data to extract features as mentioned above. The extracted features are then fed into the LVQ model, which classifies the current driving condition based on learned patterns [43].

B. Hybrid Quantum DDPG Framework

This subsection introduces a hybrid learning framework that integrates he DDPG algorithm and quantum circuits. By harnessing the expressive power of quantum circuits, we combine PQC with classical neural networks to reduce the parameters of DDPG-trained actor agent. The critic network uses the classical neural network because the hybrid quantum framework's train is processing in the classical computer. Then, the basic principles of quantum circuits and PQCs will be introduced, followed by the integration of hybrid quantum circuit within the DDPG.

1) Quantum Circuits and PQCs: Quantum circuits represent the foundational framework within which quantum computing operates, utilizing qubits as the basic units of quantum information. Unlike classical bits, which are restricted to binary states, qubits exist in multiple states simultaneously due to the quantum phenomenon known as superposition. This capability, along with the entanglement, allows quantum

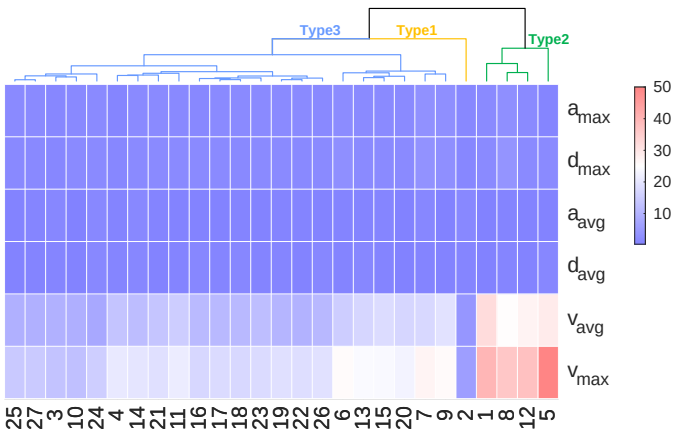

Fig. 3. Cluster Analysis of Micro-Trips.

circuits to perform complex computations. The the qubit representation, quantum gate operations, and the entanglement in interlinking qubit states are introduced as follows.

A qubit can be represented as a linear combination of the basis states $|0\rangle$ and $|1\rangle$:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{22}$$

where α and β are complex numbers that describe the probability amplitudes of the qubit being in each state. The condition $|\alpha|^2 + |\beta|^2 = 1$ must be satisfied, ensuring the sum of probabilities for all possible states equals one.

Quantum gates manipulate the states of qubits through unitary operations. These gates are the quantum analogues of classical logic gates but can perform operations that affect multiple states simultaneously due to the property of entanglement. For example, a commonly used quantum gate is the Hadamard gate, which transforms the base states into superposition states:

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{23}$$

When applied to a qubit initially in the state $|0\rangle$, the Hadamard gate yields:

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{24}$$

The entanglement, which allows two or more qubits to exist in a linked state, meaning the state of one directly affects the state of the other. An example of an entangled state is the Bell state:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{25}$$

This state shows that if one qubit is observed to be in the state $|0\rangle$, the other qubit will instantaneously be in the state $|0\rangle$ as well, and similarly for $|1\rangle$.

PQCs, as a specialized class of quantum circuits designed with tunable parameters, are a satisfied choice for applications in quantum machine learning. PQCs comprise multiple layers of parameterized gates, whose settings can be adjusted based on the outcomes of quantum measurements, thus enabling the circuits to adapt and learn from data. A PQC is typically constructed by applying a sequence of gate operations repeatedly to a set of qubits. The structure can be represented as:

$$U(\theta) = U_L(\theta_L)\cdots U_1(\theta_1) \tag{26}$$

where $U_i(\theta_i)$ represents the i-th unitary gate operation parameterized by θ_i, and L is the number of layers in the circuit. Each U_i is a combination of different quantum gates, such as Pauli-X, Y, Z, Hadamard, and rotation gates, and is parameterized by angles that adjust the quantum state. Generally, the first unitary gate operation, or layer, is state embedding layer, where classical data is encoded into the quantum system. This embedding layer transforms classical information into quantum states that are then processed by subsequent layers of quantum gates.

Recent research, as highlighted in [32], has revealed expressibility differences among these combinations of quantum gates. The expressibility of gate combinations affects how complex quantum phenomena are represented. Consequently, researchers are exploring the expressibility of different combinations of quantum gates [44]. The structure shown in Fig. 2 is recognized for its expressibility [45]. It consists of layers of parameterized rotational gates $R_z(\gamma)R_y(\beta)R_z(\alpha)$, which entangle qubits through a sequence of CNOT gates applied from one qubit to the next. Following the entanglement, Pauli-Z measurements are performed to assess the final state of each qubit, yielding the observable values $\langle O\rangle_{s,\theta}$. To address the variable number of actions, a single-layer classical neural network of size $n{\times}n_a$ is employed, where n represents the number of qubits in the PQC, and n_a denotes the number of actions. Then, a sigmoid activation function is employed because of its high efficiency, allowing the policy to be expressed as follows:

$$\pi_{\theta,\omega}(a|s) = \frac{1}{1 + e^{-\sum_{i=1}^{n} \boldsymbol{\omega_i}\langle O_i\rangle_{s,\theta}}} \tag{27}$$

where $\boldsymbol{\omega_i} \in \mathbb{R}^{n_a}$ represents the weight parameters of the classical network layer. Then, we can provide the theorem.

Theorem 1. *Given the hybrid quantum-classical architecture where a classical neural network layer with sigmoid activation function is appended to a PQC, the policy output discrepancy between the pure quantum and hybrid models can be effectively bounded. If the expectation error for each qubit measurement is assumed to be ϵ, then the error in the policy output, denoted by $\varepsilon(\widetilde{\pi}_{\theta,\omega}(a|s))$, satisfies the following inequality:*

$$\varepsilon(\widetilde{\pi}_{\theta,\omega}(a|s)) < \frac{\sqrt{n}\cdot \max_{\boldsymbol{\omega_i}} |\boldsymbol{\omega_i}|}{4}\epsilon$$

Proof. There are the error propagation formulas:

(i) $F = ae^{bX} \implies \varepsilon(F) \approx F|b|\varepsilon(X)$

(ii) $F = aX + bY \implies \varepsilon(F) \approx \sqrt{a^2\varepsilon(X)^2 + b^2\varepsilon(Y)^2}$

(iii) $F = \dfrac{X}{Y}$

$$\implies \varepsilon(F) \approx F\sqrt{\frac{\varepsilon(X)^2}{X^2} + \frac{\varepsilon(Y)^2}{Y^2} - 2\frac{\varepsilon(X,Y)^2}{XY}}$$

where X, Y and F are random variables, a and b are real constants. Therefore, we can get:

$$\begin{aligned}
&\varepsilon(1+e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}}) \\
&\approx |-1|e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}}\varepsilon(\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}) \\
&\approx e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}}\sqrt{\sum_{i=1}^{n}\boldsymbol{\omega_i^2}}\epsilon \\
&< \sqrt{n}\cdot\max_{\boldsymbol{\omega_i}}|\boldsymbol{\omega_i}|e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}}\epsilon
\end{aligned}$$

Then, for $\widetilde{\pi}_{\theta,\omega}(a|s)=\frac{1}{1+e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}}}$:

$$\begin{aligned}
\varepsilon(\widetilde{\pi}_{\theta,\omega}(a|s)) &\approx \widetilde{\pi}_{\theta,\omega}(a|s)\sqrt{\frac{\varepsilon(1+e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}})^2}{(1+e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}})^2}} \\
&< \frac{\sqrt{n}\cdot\max_{\boldsymbol{\omega_i}}|\boldsymbol{\omega_i}|e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}}\epsilon}{(1+e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\widetilde{\langle O_i\rangle}_{s,\theta}})^2} \\
&\leq \frac{\sqrt{n}\cdot\max_{\boldsymbol{\omega_i}}|\boldsymbol{\omega_i}|}{4}\epsilon
\end{aligned}$$

It completes the proof. □

C. Training of HQDDPG

The training of HQDDPG starts from the parameter update of the critic network, which is illustrated in [46]. The output layer of the actor network under hybrid quantum framework is a single-layer classical neural network, which has also been introduced in [46]. Consequently, we focus on the PQC parameters training. According to the actor update methodology, we can get:

$$\Delta\theta=\alpha\cdot\nabla_{\theta,\omega}\pi_\theta(S_t)\nabla_a Q(S_t,a;\phi)\Big|_{a=\pi_\theta(s)} \tag{28}$$

The gradient of the policy, $\nabla_\theta\pi_{\theta,\omega}(a|s)$, with respect to the PQC parameters θ is [47]:

$$\nabla_\theta\pi_{\theta,\omega}(a|s)=\frac{e^{-2\sum_{i=1}^{n}\boldsymbol{\omega_i}\langle O_i\rangle_{s,\theta}}}{\left(1+e^{-\sum_{i=1}^{n}\boldsymbol{\omega_i}\langle O_i\rangle_{s,\theta}}\right)^2}\sum_{i=1}^{n}\boldsymbol{\omega_i}\nabla_\theta\langle O_i\rangle_{s,\theta} \tag{29}$$

where the gradients of quantum observables, $\nabla_\theta\langle O_i\rangle_{s,\theta}$, can be computed using the parameter-shift rule [48]. This method estimates the derivatives by applying small shifts to the parameters and measuring the resultant changes in the quantum states:

$$\nabla_\theta\langle O_i\rangle_{s,\theta}\approx\frac{\langle O_i\rangle_{s,\theta+\delta}-\langle O_i\rangle_{s,\theta-\delta}}{2\delta} \tag{30}$$

where δ is a small shift. This numerical approach to derivative estimation is particularly suited to quantum systems, where analytical gradients can not be obtained.

The EMS of MFCV HPS is developed based on the proposed HQDDPG algorithm. Then, the EMS agent state, action, and reward function will be delineated. The state, denoted by S_t, is defined as equation (9). It encompasses the vehicle's speed, acceleration, and the battery's SOC. The action, represented by $\mathcal{P}_{fc}$, comprises a set of fuel cells power output. The reward function can be expressed as:

$$r=L(S_t,\mathcal{P}_{fc}(t))+\gamma\cdot\tanh(\beta\cdot\delta SC_k)\cdot SC_k, \tag{31}$$

where $L(S_t,\mathcal{P}_{fc}(t))$ is defined as equation (11), SC_k represents the SOC deviation from a reference point, defined as $SC_k = SOC_k - SOC_{\text{ref}}$, δSC_k indicates the change in SOC from the previous timestep, expressed as $\delta SC_k = SOC_k - SOC_{k-1}$, as for the coefficients γ and β, which modulate the sensitivity of the reward function to changes in SOC, are chosen as 500 and 7, respectively [27]. It ensures that the reward reflects the immediate economic metrics and drives optimal decision-making in real-time.

D. DSP Implementation of HQDDPG

To facilitate the practical application of the trained actor, we employ quantum simulation techniques for deployment on DSP. Quantum circuits involve complex computations and their simulation complexity grows exponentially with the increase in the number of qubits. However, we primarily leverages the minimal parameter requirements and superior expressive capabilities of quantum circuits. Thus, the selected small-scale quantum circuits can be aligned with the computational capabilities of DSPs. By simulating quantum circuits using complex number computations, the high expressibility of quantum is harnessed within a classical computing framework.

We will provide a conceptual overview of quantum simulation on DSPs. It involves the matrix forms of various quantum rotation gates such as RX, RY, and RZ. Additionally, we discuss the tensor product operations involved in composing the overall matrix for a quantum system comprising multiple qubits. For the detailed content of the SP Implementation of HQDDPG, please refer to the supplementary material [47].

$$RX(\theta)=\begin{bmatrix}\cos(\theta/2) & -i\sin(\theta/2)\\ -i\sin(\theta/2) & \cos(\theta/2)\end{bmatrix}, \tag{32}$$

$$RY(\theta)=\begin{bmatrix}\cos(\theta/2) & -\sin(\theta/2)\\ \sin(\theta/2) & \cos(\theta/2)\end{bmatrix}, \tag{33}$$

$$RZ(\theta)=\begin{bmatrix}e^{-i\theta/2} & 0\\ 0 & e^{i\theta/2}\end{bmatrix} \tag{34}$$

As for a system with multiple qubits, the overall transformation applied to the quantum state can be represented as the tensor product of individual gate matrices. For example, if each qubit i in a three-qubit system undergoes a rotation described by the gate matrix U_i, the total operation across all qubits is given by:

$$U_{\text{total}}=U_1\otimes U_2\otimes U_3 \tag{35}$$

Quantum simulation in DSPs allows for precise quantum circuits and avoids the error-proneness of real quantum computers [49].

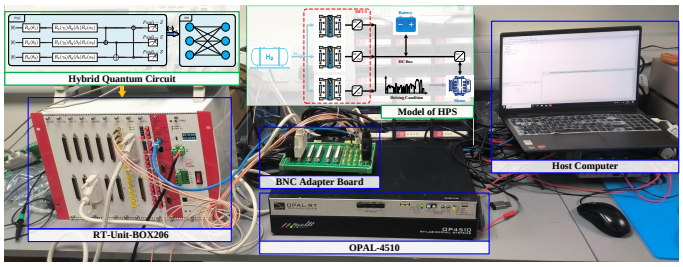

Fig. 4. CHIL experimental platform for the MFCV HPS.

IV. Results and Discussions

In this section, we validate the proposed HQDDPG against three methods: the classical DDPG, the DCR-based fuzzy logic control [43] and Dimensionality-reduced DP [27]). The HQDDPG and classical DDPG algorithms were both programmed in PyCharm and solved using PyTorch on an Intel i7-13700K processor with 64GB RAM. The validation is processed in a Controller Hardware-in-the-Loop (CHIL) platform, as shown in Fig. 4, involving the RT-Unit-BOX206, equipped with TI DSP TMS320C28377, 32-bit floating point processor, to run the EMS and OPAL-4510 to run the MFCV HPS model. We compare the EMSs under three driving conditions, which are randomly combined with different micro-trips.

A. System Configuration and Comparative Methods

This study uses the MFCV HPS, whose structure is shown in Fig. 1 and parameters are listed in [27].

Method #1 (Classical Composite DDPG): It features an actor network composed of fully connected layers. The activation functions for the hidden layers are ReLU, while the output layer employs a Sigmoid function. The network architecture includes layers with $n_s = 3$, 100, 60, 40, $n_a = 3$ neurons, where n_s and n_a represent the number of states and actions, respectively. This model is trained over 1200 episodes, using the same reward function as the HQDDPG. It is trained on a composite driving condition formed by all micro-trips, without DCR.

Method #2 (HQDDPG without DCR): This method employs the HQDDPG but without the use of DCR. The actor and critic architectures in this method are identical to those in the proposed method. However, unlike the proposed method which trains on categorized driving cycles, this method trains the HQDDPG across a unified dataset combining all micro-trips, similar to Method #1.

Method #3 (DCR-based Fuzzy Logic Control [43]): It is adapted for three distinct categories of driving conditions. For each category, its fuzzy membership functions are optimized offline through genetic algorithms.

Method #4 (Dimensionality-reduced DP [27]): It can obtain the global optimal power distribution of the MFCV HPS energy management problem because the global information must be provided. Therefore, it cannot be used as a real time EMS and serves as a benchmark to provide the optimal reference for comparative analysis.

B. Performance for Economic Metrics

In Fig. 5(a1)-(a3), the black lines indicate the MFCV's velocity under three test conditions, with real and DCR types represented by red and blue lines, respectively. The DCR accuracies for the three conditions are 86.62%, 79.31%, and 75.35%, respectively. Fig. 5(b1)-(b3) depict the fuel cell output power of the proposed HQDDPG algorithm against three methods concerning the fuel cell output power under three driving conditions. The HQDDPG algorithm, trained under three driving condition types, shows distinct characteristics: For the first type, characterized by high power demands, all three fuel cells contribute to meet the demand power. In the second type, although the demand power is less than that of the first type, it still exceeds the capacity of a single fuel cell, with two or three fuel cells output together. The third type, marked by lower and more fluctuating power demands, is predominantly managed by a single fuel cell. As for method #1 and #2, they have similar state-value functions, so their overall fuel cells' power outputs are similar to HQDDPG. Method #3, operating on an average power distribution under the fuzzy logic optimized via genetic algorithms, maintained a more stable output across varying demands, ensuring timely power support during peak demand power. Method #4, representing a global optimum, achieves the most effective power distribution and serves as a benchmark for other methods. Fig. 5(c1)-(c3) showcases the SOC across different EMS. It is observed that Method #3 exhibits the greatest fluctuations in SOC, showing the largest deviations from Method #4. In contrast, HQDDPG and Method #1 and #2 consistently maintain sufficient SOC margins for charging and discharging at any time. Unlike Method #4, which has complete driving condition information allowing it to deplete battery power early and recharge before ending without the risk of sudden power demands, HQDDPG and Method #1 and #2 had to manage without this foresight.

Table I provides a detailed comparison of economic metrics under three different driving conditions. The metrics include hydrogen consumption (C_{H_2}), equivalent hydrogen cost (C_{eqH}), degradation cost (C_{deg}), total cost (L), and the percentage of optimality (% Optimal), defined as $\frac{L(\text{Method \#4})}{L(\text{Candidate Method})} \times 100\%$, achieved by each method. HQDDPG consistently demonstrates the highest optimality percentages, achieving 86.76%, 88.60%, and 88.41% across the three driving conditions respectively, underscoring its efficient use of resources. Method #1 and #2, sharing a similar framework with HQDDPG but without the DCR integration, shows optimality percentages of 81.68%, 85.82%, 85.51% and 80.68%, 84.87%, 84.96%, respectively. The absence of DCR in Method #1 and #2 explains its suboptimal performance compared to HQDDPG. Method #3 demonstrates reduced optimality that equals 77.54%, 83.17%, and 84.67% respectively, indicating a lower efficiency in more complex or demanding scenarios. Method #4 serves as a baseline with 100% optimality.

C. Parameter and Training Performance Comparison

In this subsection, we evaluate and compare the parameter and training performance of various methods implemented in our study, including the HQDDPG with and without DCR, the

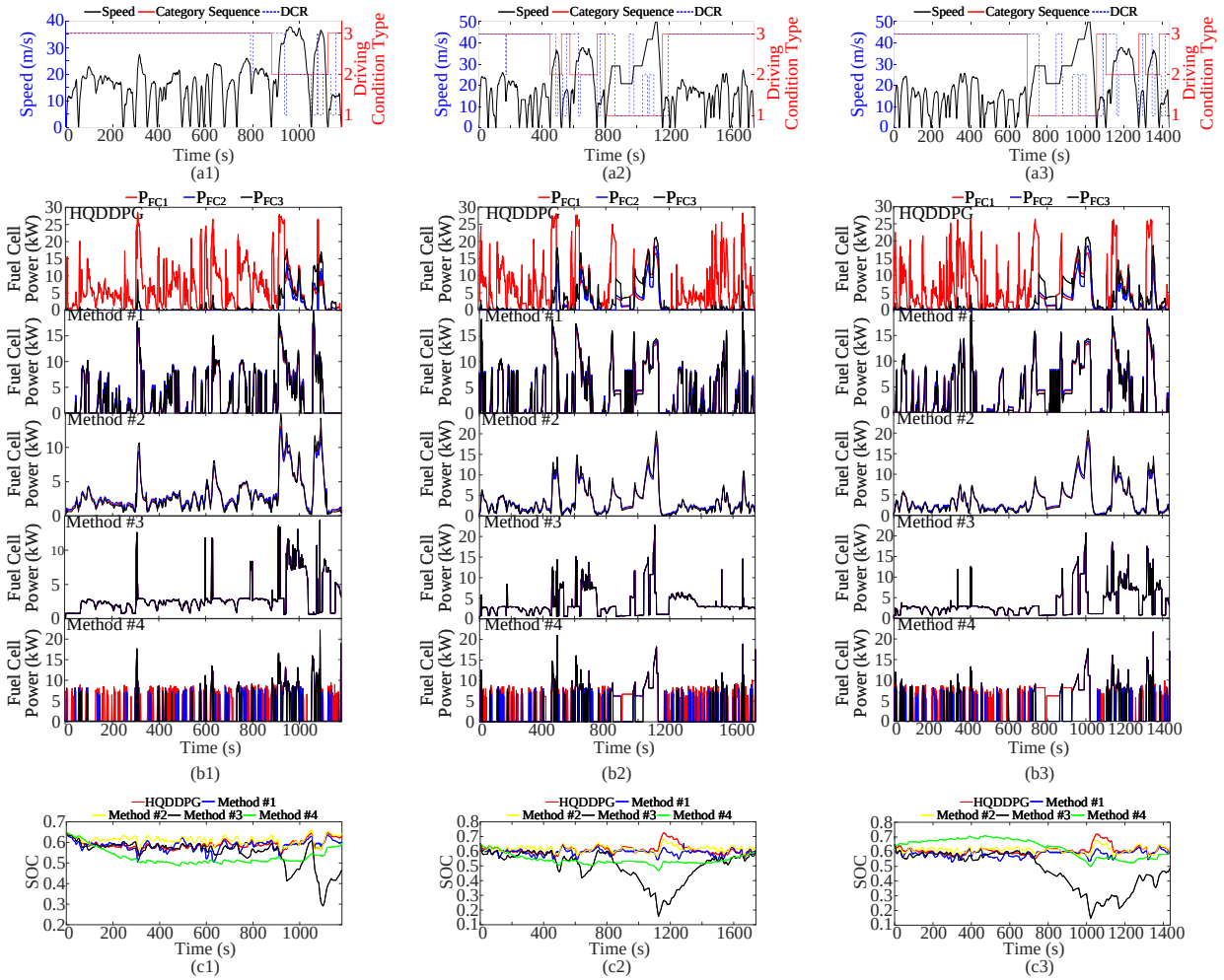

Fig. 5. Detailed Analysis of EMS Performance Under Varied Driving Conditions: (a1, a2, a3) Recognition results for Driving Conditions 1, 2, and 3 respectively; (b1, b2, b3) FC output power for each condition; (c1, c2, c3) SOC variations corresponding to each driving condition.

DDPG with the same number of parameters as HQDDPG, and a DDPG with a larger parameter set (Method #1).

To illustrate the parameter number of different methods, we present a detailed comparison in Table II. The parameter counts for each model are calculated based on their network architectures.

As demonstrated in Table II, the HQDDPG models significantly reduce the number of parameters. The HQDDPG without DCR requires only 21 parameters. This model is structured around three quantum bits, each connected to a strongly entangled layer comprising three rotational transformations, Z-Y-Z rotations, per qubit, amounting to 3×3 parameters for these rotations [47]. Additionally, the model includes a classical linear layer, contributing an extra 3×3 parameters for the weights and 3 parameters for biases, thus totaling 21 parameters. When DCR is applied, this model architecture is replicated across three different driving cycle types then tripling the parameters, leading to a total of 63 parameters.

In comparison, the parameter-matched DDPG uses 24 parameters. It is configured with a single hidden layer containing 3 neurons. This setup includes 3×3 parameters for the weights connecting the input layer to the hidden layer, 3 bias terms for the hidden layer, 3×3 parameters for the weights connecting the hidden layer to the output layer, and 3 biases for the output

TABLE I
EMS PERFORMANCE COMPARISON

Method	Driving Condition 1					Driving Condition 2					Driving Condition 3				
	C_{H_2}	C_{eqH}	C_{deg}	L	% Optimal	C_{H_2}	C_{eqH}	C_{deg}	L	% Optimal	C_{H_2}	C_{eqH}	C_{deg}	L	% Optimal
HQDDPG	379.1	-8.042	83.23	454.29	**86.76%**	641.6	-9.347	140.3	772.55	**88.60%**	558.7	-10.50	122.1	670.30	**88.41%**
Method #1	402.0	-9.859	90.41	482.55	81.68%	663.4	-14.76	149.0	797.64	85.82%	575.2	-11.35	129.2	693.05	85.51%
Method #2	412.8	-12.01	87.76	488.55	80.68%	677.0	-16.06	145.6	806.54	84.87%	562.2	16.95	120.8	697.5	84.96%
Method #3	398.0	26.170	84.16	508.33	77.54%	676.9	0.226	145.9	823.03	83.17%	585.3	-14.3	126.5	699.95	84.67%
Method #4	318.0	2.956	73.20	394.16	100.00%	555.0	1.796	127.7	684.50	100.00%	480.5	1.532	110.6	592.63	100.00%

TABLE II
COMPARISON OF PARAMETER COUNTS AND ESTIMATED MEMORY USAGE ACROSS DIFFERENT METHODS

Method	Configuration	Total Parameters	Memory Usage (Bytes)
HQDDPG without DCR	$1 \times 3 + 3 \times 3 + 3$	21	84
HQDDPG with DCR	$3 \times (1 \times 3 + 3 \times 3 + 3)$	63	252
DDPG (Parameter-Matched)	$3 \times 3 + 3 + 3 \times 3 + 3$	24	96
Classical DDPG (Method #1)	$100 \times 3 + 100 + 60 \times 100 + 60 + 40 \times 60 + 40 + 3 \times 40 + 3$	9023	36092

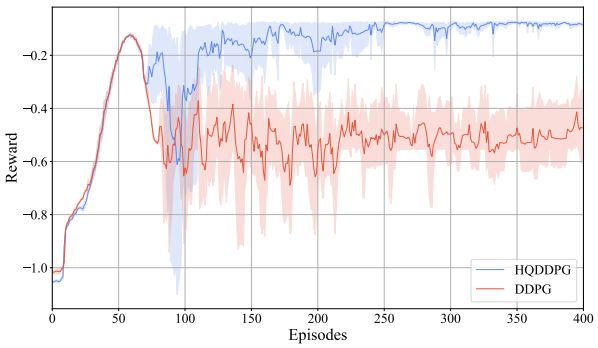

Fig. 6. Comparison of training curves for HQDDPG and parameter-matched DDPG.

layer, totaling 24 parameters. The classical DDPG (Method #1) utilizes a network architecture with large hidden layers of 100, 60, and 40 neurons, resulting in a significant total of 9023 parameters.

A comparative analysis of the training curves for HQDDPG and a parameter-matched DDPG across all driving conditions are conducted. Nine random seeds for training are utilized to obtain a distribution of training rewards. Fig. 6 displays the resulting reward distribution, demonstrating that HQDDPG consistently converges to effective EMS across different seeds, achieving stable reward values. In contrast, the DDPG model with equivalent parameters fails to converge, further substantiating the superior expressiveness of HQDDPG.

D. Quantum Circuit Behavior Analysis Across Different Driving Conditions

In this section, we compare the training parameters of the PQC under diverse driving conditions. To intuitively compare parameters, the Bloch sphere visualization is employed, mapping the PQC qubits' quantum states [50]. Fig. 7 showcases the quantum states of three qubits under different driving condition types. This approach involves uniformly sampling states across the entire feasible input range and subsequently processing them through the PQC for accurate state preparation and measurement. The color progression from green to blue in the figure symbolizes the sequence of input states processed by the PQC. It facilitates a visual observation of how the inputs map onto the quantum states, akin to observing how input data traverses through a network, influencing neuron activations at each layer. Furthermore, the degree of similarity observed among the quantum states across different driving conditions can be interpreted as an indicator of analogous behavior in response to these conditions. Essentially, the closer the quantum states appear on the Bloch sphere under varying inputs, the more similar their operational responses are. This resemblance suggests that the PQC maintains a consistent energy management strategy across similar driving scenarios, changing when faced with significantly different conditions. Such insights are crucial for refining the PQC's parameters to enhance energy management efficiency, ensuring optimal performance and sustainability of the system under diverse operational demands.

In Fig. 7, we find that the distribution of different qubits under various types appear similar, illustrating commonalities in the energy management problem of MFCV HPS. For Types 1 and 2, where the demand power typically ranges from medium to high, the similarity is more pronounced. In contrast, Type 3, which deviates from the first two types in terms of demand characteristics, displays the most distinct quantum state distributions across the three qubits. This distinction is particularly marked in the second qubit, where the distribution transitions from a mixed state to a more progressive spread.

V. CONCLUSION

This paper proposes a DCR-based HQDDPG method tailored for energy management in Multi-Stack Fuel Cell Hybrid Vehicles. The proposed method harnesses the high expressiveness of PQC, thereby reducing the number of parameters required for the EMS. By using quantum simulations, it can be deployed on DSPs for high economic metrics and small memory requirements. For validation, experiments are conducted on the CHIL platform, composed of RT-Unit-BOX206 and OPAL-4510. It is concluded that the HQDDPG method outperforms other methods under three test driving conditions, achieving efficiencies of 86.76%, 88.60%, and 88.41% respectively. Future work will focus on refining PQC structures to enhance their expressiveness and improve the economic metrics in MFCV HPS energy management.

REFERENCES

[1] Q. Li, H. Wang, T. Wang, X. Li, Y. Liu, W. Chen, and Z. You, "Online diagnosis method of water management faults based on hybrid-frequency electrochemical impedance spectroscopy for pemfc," *IEEE Transactions on Transportation Electrification*, 2024.

[2] A. Fan, J. Yang, Y. Du, Z. Li, F. Gao, and Y. Huangfu, "Secondary control for distributed converter interfaced generation with prescribed transient-state performance in dc microgrid," *Journal of Modern Power Systems and Clean Energy*, 2024.

[3] X. Sun and J. Qiu, "A customized voltage control strategy for electric vehicles in distribution networks with reinforcement learning method," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6852–6863, 2021.

[4] C. Yuan, Y. Du, Y. Huangfu, C. Gong, F. Gao, and Z. Li, "Consensus-based powered cruise and yaw controls for unmanned aerial vehicle with distributed electric propulsion system," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2024.

[5] J. Chen, Y. Zhao, M. Wang, X. Lyu, K. Yang, Y. Ge, X. Jiang, P. Pan, K. Jia, and Z. Xu, "Power management and control for flexible traction substation in railway hub integrated with ess, pv, and ev," *IEEE Transactions on Smart Grid*, 2025.

[6] X. Li, W. Yang, Y. Liao, S. Zhang, Y. Zheng, Z. Zhao, M. Tang, Y. Cheng, and P. Liu, "Short-term risk-management for hydro-wind-solar hybrid energy system considering hydropower part-load operating characteristics," *Applied Energy*, vol. 360, p. 122818, 2024.

[7] J. Chen, H. Hu, M. Wang, Y. Ge, K. Wang, Y. Huang, K. Yang, Z. He, Z. Xu, and Y. R. Li, "Power flow control-based regenerative braking energy utilization in ac electrified railways: Review and future trends," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[8] S. Li, P. Zhao, C. Gu, S. Bu, X. Pei, X. Zeng, J. Li, and S. Cheng, "Hybrid power system topology and energy management scheme design for hydrogen-powered aircraft," *IEEE Transactions on Smart Grid*, vol. 15, no. 2, pp. 1201–1212, 2023.

[9] X. Sun, J. Qiu, Y. Yi, and Y. Tao, "Cost-effective coordinated voltage control in active distribution networks with photovoltaics and mobile energy storage systems," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 1, pp. 501–513, 2021.

[10] Z. Hu, B. Gao, Z. Xu, and S. Jiang, "Optimized temporary frequency support for wind power plant considering expanded operational region of wind turbines," *CSEE Journal of Power and Energy Systems*, 2025.

[11] Q. Li, H. Tan, T. Wang, X. Li, and W. Chen, "An energy management strategy for multi-stack fuel cell hybrid locomotives with integrated optimization of system service life," *IEEE Transactions on Transportation Electrification*, 2024.

[12] X. Li, Q. Li, T. Wang, W. Chen, and S. Zhang, "Adaptive power transient smoothing control considering performance degradation for multi-stack fuel cell hybrid power systems," *IEEE Transactions on Transportation Electrification*, 2023.

[13] Q. Li, T. Su, L. Yin, S. Xie, Y. Tan, W. Chen, and F. Gao, "Health management for pemfc system long-term operation based on optimal temperature trajectory real-time optimization," *IEEE Transactions on Industrial Electronics*, 2024.

[14] R. Ma, X. Chai, R. Geng, L. Xu, R. Xie, Y. Zhou, Y. Wang, Q. Li, K. Jiao, and F. Gao, "Recent progress and challenges of multi-stack fuel cell systems: Fault detection and reconfiguration, energy management strategies, and applications," *Energy Conversion and Management*, vol. 285, p. 117015, 2023.

[15] L. Xu, Y. Huangfu, R. Ma, R. Xie, Z. Song, D. Zhao, Y. Yang, Y. Wang, and L. Xu, "A comprehensive review on fuel cell uav key technologies: propulsion system, management strategy, and design procedure," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 4, pp. 4118–4139, 2022.

[16] B. Somaiah and V. Agarwal, "Distributed maximum power extraction from fuel cell stack arrays using dedicated power converters in series and parallel configuration," *IEEE Transactions on Energy Conversion*, vol. 31, no. 4, pp. 1442–1451, 2016.

[17] J. Shen and A. Khaligh, "A supervisory energy management control

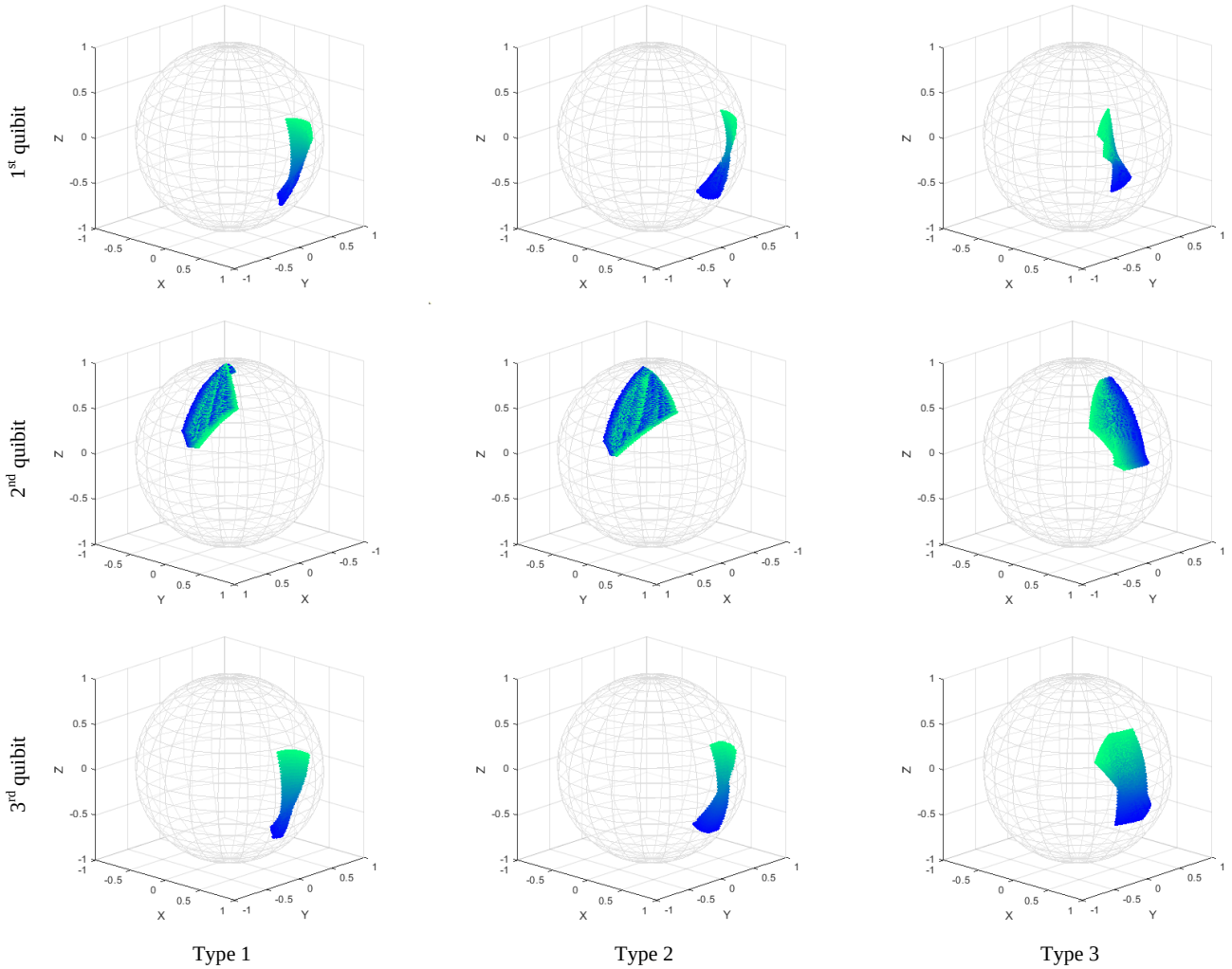

Fig. 7. Comparison of different driving conditions bloch sphere.

strategy in a battery/ultracapacitor hybrid energy storage system," *IEEE Transactions on transportation electrification*, vol. 1, no. 3, pp. 223–231, 2015.

[18] Y. Huangfu, P. Li, S. Pang, C. Tian, S. Quan, Y. Zhang, and J. Wei, "An improved energy management strategy for fuel cell hybrid vehicles based on pontryagin's minimum principle," *IEEE Transactions on Industry Applications*, vol. 58, no. 3, pp. 4086–4097, 2022.

[19] X. Sun and J. Qiu, "Hierarchical voltage control strategy in distribution networks considering customized charging navigation of electric vehicles," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 4752–4764, 2021.

[20] X. Sun, J. Qiu, Y. Tao, Y. Yi, and J. Zhao, "Distributed optimal voltage control and berth allocation of all-electric ships in seaport microgrids," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2664–2674, 2022.

[21] Y. Ding, X. Sun, J. Ruan, W. Shi, H. Wu, and Z. Xu, "Customized decentralized autonomous organization based optimal energy management for smart buildings," *Applied Energy*, vol. 376, p. 124223, 2024.

[22] J. Chen, Y. Zhao, M. Wang, K. Yang, Y. Ge, K. Wang, H. Lin, P. Pan, H. Hu, Z. He, and Z. Xu, "Multi-timescale reward-based drl energy management for regenerative braking energy storage system," *IEEE Transactions on Transportation Electrification*, 2025.

[23] X. Sun and J. Qiu, "Two-stage volt/var control in active distribution networks with multi-agent deep reinforcement learning method," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 2903–2912, 2021.

[24] Z. Fu, H. Wang, F. Tao, B. Ji, Y. Dong, and S. Song, "Energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles using deep reinforcement learning with action trimming," *IEEE transactions on vehicular technology*, vol. 71, no. 7, pp. 7171–7185, 2022.

[25] X. Li, H. He, and J. Wu, "Knowledge-guided deep reinforcement learning for multi-objective energy management of fuel cell electric vehicles," *IEEE Transactions on Transportation Electrification*, 2024.

[26] H. Hu, W.-W. Yuan, M. Su, and K. Ou, "Optimizing fuel economy and durability of hybrid fuel cell electric vehicles using deep reinforcement learning-based energy management systems," *Energy Conversion and Management*, vol. 291, p. 117288, 2023.

[27] W. Shi, Y. Huangfu, L. Xu, and S. Pang, "Online energy management strategy considering fuel cell fault for multi-stack fuel cell hybrid vehicle based on multi-agent reinforcement learning," *Applied Energy*, vol. 328, p. 120234, 2022.

[28] H. Xie, G. Tian, G. Du, Y. Huang, H. Chen, X. Zheng, and T. H. Luan, "A hybrid method combining markov prediction and fuzzy classification for driving condition recognition," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10411–10424, 2018.

[29] B. Xu, J. Shi, S. Li, and H. Li, "A study of vehicle driving condition recognition using supervised learning methods," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 2, pp. 1665–1673, 2021.

[30] Z. Ji, T. Wang, X. Sun, Z. Wu, J. Shi, D. Chen, and H. Ma, "Driving condition recognition combined with stochastic prediction and machine learning and its application in energy management of medium fuel cell trucks," *IEEE Transactions on Vehicular Technology*, 2023.

[31] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning," *Physical Review A*, vol. 98, no. 3, p. 032309, 2018.

[32] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019.

[33] Y. Du, M. Hsieh, T. Liu, and D. Tao, "The expressive power of parameterized quantum circuits. arxiv 2018," *arXiv preprint arXiv:1810.11922*, 2019.

[34] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, "The power of quantum neural networks," *Nature Computational Science*, vol. 1, no. 6, pp. 403–409, 2021.

[35] M. Kölle, M. Hgog, F. Ritz, P. Altmann, M. Zorn, J. Stein, and C. Linnhoff-Popien, "Quantum advantage actor-critic for reinforcement learning," *arXiv preprint arXiv:2401.07043*, 2024.

[36] R. Yan, Y. Wang, Y. Xu, and J. Dai, "A multiagent quantum deep reinforcement learning method for distributed frequency control of islanded microgrids," *IEEE Transactions on Control of Network Systems*, vol. 9, no. 4, pp. 1622–1632, 2022.

[37] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara, *et al.*, "High-fidelity parallel entangling gates on a neutral-atom quantum computer," *Nature*, vol. 622, no. 7982, pp. 268–272, 2023.

[38] C. Zheng, C. Oh, Y. Park, and S. W. Cha, "Fuel economy evaluation of fuel cell hybrid vehicles based on equivalent fuel consumption," *International Journal of Hydrogen Energy*, vol. 37, no. 2, pp. 1790–1796, 2012.

[39] K. Farhana, A. S. F. Mahamude, and K. Kadirgama, "Comparing hydrogen fuel cost of production from various sources-a competitive analysis," *Energy Conversion and Management*, vol. 302, p. 118088, 2024.

[40] P. Pei, Q. Chang, and T. Tang, "A quick evaluating method for automotive fuel cell lifetime," *International Journal of Hydrogen Energy*, vol. 33, no. 14, pp. 3829–3836, 2008.

[41] P. Rodatz, G. Paganelli, A. Sciarretta, and L. Guzzella, "Optimal power management of an experimental fuel cell/supercapacitor-powered hybrid vehicle," *Control engineering practice*, vol. 13, no. 1, pp. 41–53, 2005.

[42] C. Wei, Y. Chen, X. Li, and X. Lin, "Integrating intelligent driving pattern recognition with adaptive energy management strategy for extender range electric logistics vehicle," *Energy*, vol. 247, p. 123478, 2022.

[43] Z. Zhang, Y. Huangfu, L. Xu, J. Zhao, W. Shi, and T. Yu, "Research on optimized energy management strategy based on micro-trip recognition," in *2021 IEEE 1st International Power Electronics and Application Symposium (PEAS)*, pp. 1–6, IEEE, 2021.

[44] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, p. 226, 2020.

[45] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, "Circuit-centric quantum classifiers," *Physical Review A*, vol. 101, no. 3, p. 032308, 2020.

[46] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[47] W. Shi, "Supplementary material." https://drive.google.com/file/d/1pcwLMZE8S00D4HQqodhvzoZjorom9VJf/view?usp=drive_link.

[48] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, "General parameter-shift rules for quantum gradients," *Quantum*, vol. 6, p. 677, 2022.

[49] A. Skolik, S. Mangini, T. Bäck, C. Macchiavello, and V. Dunjko, "Robustness of quantum reinforcement learning under hardware errors," *EPJ Quantum Technology*, vol. 10, no. 1, pp. 1–43, 2023.

[50] H. Wiseman and G. Milburn, "Interpretation of quantum jump and diffusion processes illustrated on the bloch sphere," *Physical Review A*, vol. 47, no. 3, p. 1652, 1993.