# Deep Reinforcement Learning of Energy Management Strategy for Aging Fuel Cell Power Systems

Qilin Shuai[*1,a1], Qingsong Hua[1,b], Junchao Ma[2,c], Yiheng Wang[1], Zhenxiong Jiang[1]
[1]College of Nuclear Science and Technology , Beijing Normal University, Beijing, China;
[2] State Grid Zhejiang Electric Power Research Institute, Hangzhou 310014, China;

**ABSTRACT.**

While the combustion vehicle promote the economic development, they also bring environmental pollution and energy crisis issues. As a new type of new energy vehicle, Fuel Cell Hybrid Electric Vehicles (FCHEV) attracted widespread attentions due to its characteristics of zero-emission and high energy density. At present, fuel economy and fuel cell aging are the important factors affecting the development of FCHEV. To this end, A deep reinforcement learning based energy management strategy (DQN) is proposed in this paper for energy management system of FCHEV, and improve the life-time of fuel cell by reducing frequent fluctuation in fuel cell output power. The standard CycleRun are used to train the energy management strategy, and compared the training result with two benchmark energy management strategies. The result shows that the proposed energy management strategy can achieve lower computation cost, and reduce the power fluctuations.

**Keywords:** FCHEV; DQN; fluctuation; aging; fuel economy.

## 1. INTRODUCTION

### 1,1 Background

With the development of economy and industry, the amount of vehicles increases rapidly, which has aroused concern of energy unsustainability and enviroment pollution. FCHEV, powered by hydrogen fuel cells,offer a promising solution for sustainable and zero-emission transportation.

The development of Fuel Cell Hybrid Electric Vehicles (FCHEVs) is crucial in the face of global challenges such as climate change and dwindling fossil fuel reserves. In this context, powertrain powered by fuel cells are considerd a promising alternative to combustion engines. Fuel cells convert chemical energy to electrical energy with high efficiency and performance. This result in bettery fuel economy and extended cruising ranges. FCHEV development presents economic opportunities and industrial growth potential. This results in better fuel economy and extended driving ranges. Additionally, hybrid powertrain improves power management and energy storage. FCHEV development presents economic opptunities and industrial growth potential, and can also enhance energy security by reducing reliance on imported fossil fuels.

In conclusion, FCHEVs are essential for addressing climate change, reducing fossil fuel dependency, and achieving sustainable transportation.

### 1,2 Literature review

Recently, energy management is a key constraint in the advancement of FCHEV. The energy management strategies for FCHEV can be categorized into rule-based strategies and optimization-based strategies[1]. Rule-based strategies are characterized by their simplicity and ease of implementation, relying on predetermined knowledge and design control principles. However, these strategies often fall short of achieving optimal fuel economy. On the other hand, optimization-based strategies employ techniques such as dynamic programming, Pontryagin's minimum value principle, and model predictive control. Dynamic programming[2], while providing optimal fuel economy, requires pre-determined global operating conditions and extensive computational resources, making real-time optimization challenging. It is often used as a benchmark for comparing other strategies. The Pontryagin minimum value principle-based strategy achieves optimization by minimizing the Hamilton function and requires relatively lower computational resources compared to

[1*a] 202131220021@mail.bnu.edu.cn, [b] q.hua@bnu.edu.cn, [c] mjc_zju@sina.com

The International Conference Optoelectronic Information and Optical Engineering (OIOE2024), edited by Yang Yue, Lu Leng, Proc. of SPIE Vol. 13513, 1351323 · © 2025 SPIE · 0277-786X · © The Authors.

dynamic programming[3]. Some researchers have also utilized real-time optimization algorithms like model predictive control for online rolling optimization of control variables, which depends on the length of the control and prediction domains for real-time performance. In summary, optimization-based strategies demand significant computational resources and exhibit limited adaptability to unknown conditions.

In recent years, with the development of artificial intelligence technology, energy management strategies based on learning strategies have gradually become a research hotspot in the field of FCHEV energy management. At present, they can be further divided into neural networks and reinforcement learning. The strategy uses the powerful nonlinear modelling ability and classification prediction ability of the neural network to obtain the optimal fuel economy through working condition classification, speed prediction, parameter optimization, etc; the reinforcement learning algorithm directly learns the optimal control strategy from the data, enabling intelligent agent explores the external environment, obtains feedback signals and updates policies, and realizes model-free, end-to-end control[4], [5].

**1.3 Motivation**

With the development of fuel cell research, some researchers have proposed fuel cell EMS based on degradation perception. Ref [6] designed a two-stage EMS, the first stage does not consider the SOH of the fuel cell, the second stage considers the health status of the fuel cell and adds fuel cell aging constraints into the controller, so as to ensure the improvement of fuel economy while reducing fuel cell degradation. Ref [7] proposes a long-term fuel cell EMS, the proposed EMS changes the SOC boundary of the battery, and uses simulation to show that the proposed EMS can slow down the degradation of fuel cells and lithium batteries. Ref [8] proposed a degradation-aware multi-objective EMS based on the fuzzy control method, and used the fuel cell SOH evaluator to divide the fuel cell SOH into four levels. Ref [9] proposed the optimal cost EMS, and by calculating the empirical degradation rate of fuel cells under different operating conditions, but most studies used empirical functions to estimate the degradation of fuel cells over time or classify the state of fuel cells into several stages. Therefore, this study adopts a data-driven approach to optimize the fuel cell mechanism model, and studies the fuel cell energy management strategy based on the model.

The rest of the paper is divided into four main sections as follows: Section 2 presents the powertrain model of the studied vehicle and the aging model of fuel cells and lithium batteries. Section 3 proposes an EMS based fuel cell aging model, and presents training and simulation results in Section 4. In Section 5, the research results are summarized.

## 2. POWERTRAIN MODEL

**2.1 System configuration**

The vehicle model investigated is and the topology of its powertrain is shown in Fig 1.

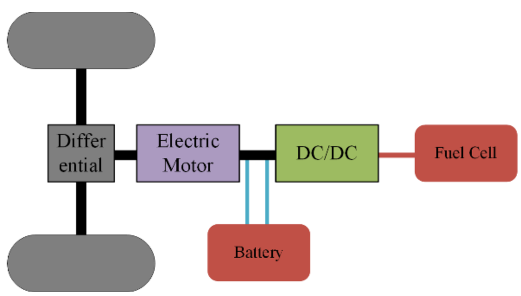


Figure.1. Powertrain Topology of FCHEV Vehicle

The powertrain configuration composed of a driving Motor, the output power source of Hydrogen FC, and a power source of li-ion battery. Besides, the li-ion battery is chosen as the energy storage system.

**2.2. The power request model**

The primary power source for FCHEV is FC. Unidirectional DC/DC conveter serves as an intermediate layer for linking FC to DC bus, and as a regulator for maintaining the state of SOC of battery at a proper level, on the premise that FC is working in high efficiency field. By applying the bi-directional DC/DC converter, the battery has ability to provide or absorb the rest power through DC bus. Additionally, DC/AC inverter generates any desired power for driving motor to drive the vehicle for meeting the driver’s demand.

The vehicle model is used to obtain the required power on the power bus. The required power $P_{req}$ is function of the vehicle speed $v$, which is given by (1), as follows:

$$P_{req} = \left( mg\sin\alpha + \delta m\frac{du}{dt} + \frac{1}{2}\cdot C_D A\rho v^2 + mgf\ \cos\alpha \right)\bullet v \tag{1}$$

Where $m$ is the mass of vehicle, $g$ is the gravity acceleration, $C_D$ is the cofficient of air resistance, $\rho$ is the density of the air, $f$ is the rolling resistance of vehicle.

### 2.3 Powertrain system model

#### 2.3.1 Fuel cell model

The proton exchange membrane Fuel-Cell as major power source for FCHEV convertsthe chemical energy into electric energy through the reaction betweent hydrogen and oxygen. The output voltage of FC is expressed as:

$$V = (E_{\text{nernst}} - V_{act} - V_{\text{ohm}} - V_{\text{con}}) \times n_{cell} \tag{2}$$

The maximum power of the studied fuel cell is 30kW, and the rated power is 25kW. It consists of 240 single cells. The hydrogen consumption rate of the studied fuel cell system measured by experiment.

#### 2.3.2 Battery model

The most classical method to estimate $SOC$ of battery can be expressed as:

$$\dot{SOC} = -\frac{E - \sqrt{E^2 - 4RP_b}}{2RQ} \tag{3}$$

The output voltage $V_{out}$ can be obtained by:

$$V_{out} = E - IR \tag{4}$$

$$I_b = \frac{E - \sqrt{E^2 - 4R_b P_b}}{2R_b} \tag{5}$$

Where $E$ is the Open circuit voltage, $R_b$ is the charge and discharge resistance of the lithium battery, and $P_b$ represents the output power of the lithium battery.

#### 2.3.3 Fuel cell aging model

The degradation of stack is one of the main factors affecting the development of fuel cell vehicles. The key of fuel cell degradation is mainly due to changes in external working conditions, which lead to severe fluctuations in reaction conditions such as temperature, humidity, and pressure inside the stack. The most influentia l working conditions mainly include start-stop, rapid load change and other working conditions. Therefore, this paper uses the fuel cell decay rate model to measure the life extension performance of the energy management strategy.

$$\Delta\phi_{\text{dr}} = Kp\left(\left(k_1 n_1 + k_2 t_2\right) + \psi\right) \tag{6}$$

Where $\Delta\phi_{dr}$ represents the aging rate after cyclic working conditions, $Kp$ represents the correction coefficient, which is used to measure the difference between the experimental situation and the simulation situation, $n_1$, $t_2$ represent the number of start-stops and load-changing time, respectively, $k_1$, $k_2$ are two kinds The attenuation rate corresponding to the working condition, $\psi$ is the natural attenuation rate.

## 3. ENERGY MANAGEMENT STRATEGIES

Relying on the EMS, this module determines the power distribution between the fuel ell system and the battery system. Considering the constraints of the system, the EMS in this work aims to achieve best fuel economy, maintain battery SOC and slow down the fuel cell aging. To obtain an adaptive strategy, the deep Q reinforcement learning is applied.

Generally, the energy management problem in the form of RL is represented as Markov decision process (MDP). In the following subsections, the DQN agent is presented and the settings for training are explained. After the training process, the trained policy obtained from the DQN agent is utilized as the EMS. In Fig, an overview of the DQN based energy management for the fuel cell vehicle is displayed.

### 3.1. DQN Algorithm

In the area of RL, an MDP is applied to represent the interaction between an agent and its enviromnet. According to the current state ***s*** of the enviroment, an agent performs an action ***a*** that follows a policy for the enviroment. Meanwhile, the agent receives a reward ***R*** for performing the action and a new state ***s'*** from the enviroment. Based on this feedback, the agent updates the policy. Its target is to find the policy $\pi$ which maxminzes the action-value function. Hereby, the action-value function, also known as the ***Q*** function is specified as the expected discounted sum of rewards:

$$Q^{\pi}(s,a)=\sum_{t=0}^{\infty}\left\{\gamma^{t}R\left(s_{t},a_{t}\right) \| \, s_{0}=s,a_{0}=a\right\} \tag{7}$$

Where $R$ is the single-step reward and $\gamma$ in the range of [0,1], represents the discount factor to future reward value. The Q function represent the accumulated value of the long term expected reward and is exploited to measure the advantages of taking action a under the states. Optimal Q function Q* can be defined as:

$$Q^{*}(s,a)=\max_{\pi} Q^{\pi}(s,a) \tag{8}$$

And thus the strategy can choose action ***a*** by:

$$a=\underset{a\in A}{\operatorname{argmax}}\, Q^{*}(s,a) \tag{9}$$

We define an experience pool to store the data that nerual network needed, the form of experience pool in this research is quadruplet **(*s, a, r, s'*)** that contains the current state ***s***, the agent action ***a***, the immediate reward after executing the action ***r***, and the state at the next moment ***s'***. In order to avoid correlations of data getting from driving cycle, the experience replay method is employed to smooth overchanges in the data distribution and reduce training difficulty.

### 3.2 DQN based energy management

As in the DQN configuration shown in Fig.2 the training enveiroment including the entire vehicle model interacts with the DQN-based agent. The settings of agent actions, enviroment states, and rewards are critical to the interaction and learning of the agent. In this subsection, the setup of the DQN-based energy management is explained in detail.

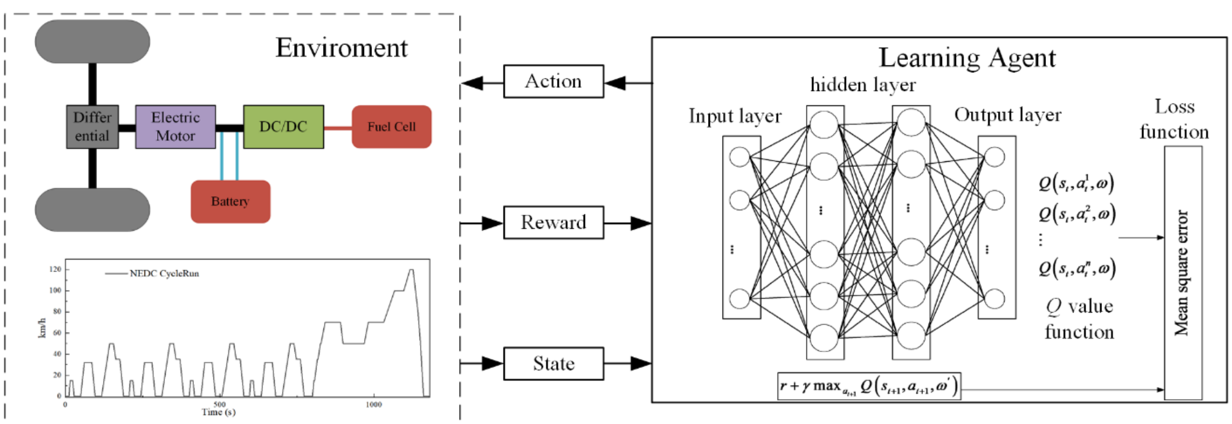


Figure.2 DQN-based strategy EMS diagram.

#### 3.2.1 DQN-based agent actions

The agent controls the power output of the fuel cell system. Thus, the desired fuel cell power is a direct control input to the enviroment, where the fuel cell system is automatically controlled by its own DC/DC converter to achieve the desired power output. Therefore, the action is defined as

$$a = \left[ P_{\mathrm{fc}} \right] \tag{10}$$

### 3.2.2 Enviroment states

In this research, the agent needs proper states information to reasonably manage the power. The power demand of the vehicle $\boldsymbol{P_{req}}$, the output power of the fuel cell system $\boldsymbol{P_{fc}}$, the battery state of charge ***SOC***, Motor current velocity $\boldsymbol{V}$ and acceleration ***acc*** are set as state variables.

$$\mathrm{s}=\{\mathrm{P}_{req}, P_{\mathrm{fc}}, v, a, \mathrm{SOC}\} \tag{11}$$

### 3.2.3 Rewards

In order to guarantee the regular iteration of the network, the reforcement learning single step reward function $R$ needs to be defined, reward function is one of the key factors in determining the performance of the DQN. Fuel consumption, $SOC$ stability, and the remaining life of fuel cell should be take into consider by defining the reward function.

$$R = -\left( \alpha \frac{dfuel}{dt} + \beta \left( SOC_{\mathrm{ref}} - SOC \right)^2 + \xi f\left( t_{\mathrm{life}} \right) \right) \tag{12}$$

Where $\alpha$, $\beta$, $\xi$ is the cofficient of fuel consumption, SOC deviation and the degradation rate of fuel cell. The parameters in rewards is listed in Table.1.

Table.1 Parameters in Reward

| Coefficient | Value |
|---|---|
| $\alpha$ | 1 |
| $\beta$ | 350 |
| $\xi$ | 2.5 |

### 3.3 Enviroment settings for training

The number of hidden layers are set to 4 and all of them are fully-connected sigmoid layer consisting of 500 rectifier units. The output layer is a fully-connected linear layer with a single output for each action, where the number of actions as described above is 6. The exact procedures of DQN algorithm in this research is demonstrated in Table and the parameters in neural network is listed in Table.2.

Table.2. Parameters in deep nerual network

| Description | Value |
|---|---|
| Training episodes | 1500 |
| Memory capacity N | 10000 |
| Batchsize | 64 |
| Learning rate | 0.0001 |
| Discount factor | 0.99 |

The computation process of the whole DQN-based energy management strategy presented in Fig, and the whole work is programmed and realized in Python.

## 4. TRAINING AND SIMULATION RESULT

### 4.1. Training settings

For the proposed DQN energy management strategy (DQN-EMS), a training procedure should be performed before testing. To fully explore the policy, the original exploration rate of DQN agent will set as 1 and when the exploration rate of DQN agent reach 0.1, it will be stopped decrease. At the beginning of the training process, the replay buffer acquires the memory data in the first ten training episode. The speed profiles are randomly generated for each episode.

### 4.2. Impact of the reward settings on convergence

The reward setting is critical for training convergence. In this work, the SOC of the battery can very greatly due to the large power dynamics of the vehicle. When the SOC reaches the system limits shown in Table, the reward of this training episode will be given a large penalty. To achieve adaptive maintenance of the SOC and fast convergence of the training, a penalty term is introduced.

### 4.3. Simulation results and analysis

In order to evaluate the proposed DQN-EMS, a test speed profile of the CycleRun from European shown in Fig.3 is utilized. A typical fault in the reinforcement learning is the overfitting, which results in promising performance on the training cycles while not on the test cycles. To investigate the transferability and adaptability of the proposed DQN-EMS, another test profile of the CycleRun shown in Fig is used. In the following subsections, results on battery's charge-sustraining, fuel cell aging and optimality of the operational cost are presented.

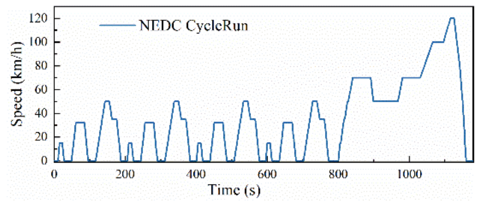


Figure.3. NEDC driving cycle

#### 4.3.1. Battery' s charge-sustaining

To investigate the performance of DQN-EMS in terms of battery's charge-sustaining, the simulations are performed with different EMS strategies while the reference SOC remains at 0.6. The results of fuel cell power and SOC trajectories are displayed in Fig.4

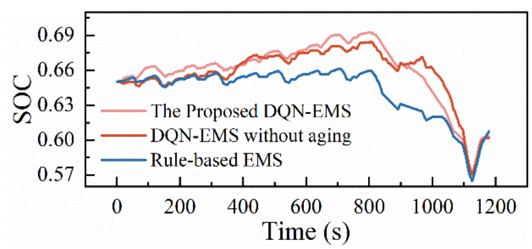


Figure.4. SOC trajectory in DQN and Rule-based strategy

#### 4.3.2 Fuel cell aging

To evaluate the aging behavior of fuel cells under the proposed DQN-EMS, a reference DQN-EMS without the fuel cell aging term in the reward function is uesed for comparison. From the fuel cell power trajectories in Fig 5a and b, it can be seen that the proposed DQN-EMS controls the power output of fuel cell system less dynamically than the reference Rule-based EMS.

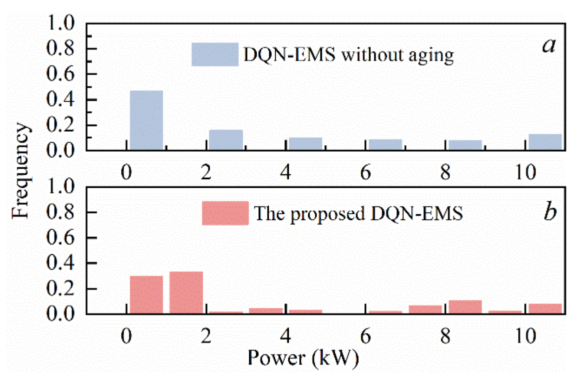


Figure.5. Fuel consumption distribution of fuel cell working points in both strategies

#### 4.3.3. Optimality of operational cost

The operational cost of the vehicle consists of the hydrogen cost. To verify the optimality in the opearational cost of the DQN-EMS, the Rule-based EMS strategy is used as the benchmark. To show the improvement of the proposed DQN-EMS strategy, the DQN-EMS without considering aging is used for comparison. The simulation results are presented in Table.

Table.3. The fuel consumption of three strategies

| Algorithm | Fuel consumption (g/100km) | Fuel economy (%) |
|---|---|---|
| Rule-based strategy | 876.96 | - |
| DQN (considering with aging) | 648.12 | 26.09 |
| DQN (considering without aging) | 641.2 | 26.906 |

It can be seen that the hydrogen cost with the proposed DQN-EMS is generally lower to the benchmark. It is because of the lack of global optimization, the proposed DQN-EMS reacts to the timely power demands and manages higher power dynamics than the DQN-EMS without considering aging, resulting in larger voltage degradation, especially in voltage cycle (VC). Compared to the strategy without considering aging, the proposed DQN-EMS indicates up to 0.9% improvement in the total operational cost.

# 5. CONCLUSIONS

This research reports a deep reinforcement learning-based energy management strategy for FCHEV that takes fuel cell aging into account. In order to improve the prognosis for the aging of the fuel cell system, an operation mode-oriented estimation approach is proposed. Based on the vehicle model, an energy management issue is created using a reinforcement learning structure, and it is then solved using DQN, one of the most advanced DRL algorithm. A stochastic training environment based on CycleRun is used to produce more realistic simulations and prevent overfitting. Through the DQN agent's contact with the environment, training is accomplished. Following training, the suggested DQN-EMS results demonstrate a promising capacity for SOC maintenance.After training, the result of proposed DQN-EMS shows promising ability for SOC maintenance. The proposed DQN-EMS reduce the voltage degradation compared to the reference DQN-EMS without considering aging. Moreover, the overall operational cost is investigated for the two strategies and a benchmark strategy. The proposed DQN-EMS show results has lower fuel consumption to the benchmark with an improvement compared to another strategy. In future work, more available information that facilitates approaching the benchmark optimum will be considered to futher reduce overall operational costs.

# 6. ACKNOWLEDGMENTS

This work was financially supported by “the National Key Research and Development Program of ChinaResearch and Application Demonstration of Intelligent loT and Control Technology for UrbanIntegrated Enerty, China (grant number: 2020YFB2104500)’.

# 7. REFERENCES

[1] Xu N, Kong Y, Chu L, et al. Towards a smarter energy management system for hybrid vehicles: A comprehensive review of control strategies[J]. Applied Sciences, 2019, 9(10): 2026.

[2] Song Z, Hofmann H, Li J, et al. Optimization for a hybrid energy storage system in electric vehicles using dynamic programing approach[J]. Applied Energy, 2015, 139: 151-162.

[3] Yuan Z, Teng L, Fengchun S, et al. Comparative study of dynamic programming and Pontryagin’s minimum principle on energy management for a parallel hybrid electric vehicle[J]. Energies, 2013, 6(4): 2305-2318.

[4] He D, Zou Y, Wu J, et al. Deep Q-learning based energy management strategy for a series hybrid electric tracked vehicle and its adaptability validation[C]//2019 IEEE Transportation Electrification Conference and Expo (ITEC). IEEE, 2019: 1-6.

[5] Hu Y, Li W, Xu K, et al. Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning[J]. Applied Sciences, 2018, 8(2): 187.

[6] Geng B, Mills J K, Sun D. Two-stage energy management control of fuel cell plug-in hybrid electric vehicles considering fuel cell longevity[J]. IEEE Transactions on vehicular technology, 2011, 61(2): 498-508..

[7] Zhou Y F, Huang L J, Sun X X, et al. A Long - term Energy Management Strategy for Fuel Cell Electric Vehicles Using Reinforcement Learning[J]. Fuel Cells, 2020, 20(6): 753-761..

[8] Shen D, Lim C C, Shi P. Fuzzy model based control for energy management and optimization in fuel cell vehicles[J]. IEEE Transactions on Vehicular Technology, 2020, 69(12): 14674-14688.

[9] Li H, Zhou Y, Gualous H, et al. Optimal cost minimization strategy for fuel cell hybrid electric vehicles based on decision-making framework[J]. IEEE Transactions on Industrial Informatics, 2020, 17(4): 2388-2399.