IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 29, 2020, accepted December 7, 2020, date of publication January 5, 2021, date of current version January 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3049162

Intelligent Controller Based on Distributed Deep Reinforcement Learning for PEMFC Air Supply System

JIAWEN LI AND TAO YU
College of Electric Power, South China University of Technology, Guangzhou 510640, China
Corresponding author: Tao Yu (taoyul@scut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51777078.

ABSTRACT Air supply system is an important subsystem in a PEMFC engine system. Research on the control strategy of air supply system is of great importance and significance in engineering. In this paper an intelligent controller based on distributed deep reinforcement learning which exerts better control over the air flux of a proton exchange membrane fuel cell (PEMFC) air supply system is proposed. In addition, a collective intelligence exploration distributed multi-delay deep deterministic policy gradient (CIED-MD3) algorithm is presented for the controller. This improved algorithm is developed on the basis of deep deterministic policy gradient (DDPG) and adopted the collective intelligence exploration policy which enables full exploration of the environment. This classification experience replay mechanism is introduced to improve training efficiency. A number of techniques are employed in an effort to address the Q-value overestimation problem of the DDPG, including clipped multiple Q-learning, delayed update of policy and smooth regularization of target policy. Finally, the application of CIED-MD3 (with its better global search ability and optimization speed) is demonstrated to the model-free PEMFC air flux intelligent controller. The simulation results show that the proposed controller exerts greater control of the PEMFC air supply system. Compared with other control methods, the proposed intelligent controller exhibits better control performance and robustness. The control algorithm proposed in this paper is of significance to future PEMFC air flux control research.

INDEX TERMS Air supply systems, collective intelligence exploration distributed multi-delay deep deterministic policy gradient (CIED-MD3), proton exchange membrane fuel cell (PEMFC), air flux control, intelligent controller.

I. INTRODUCTION

There has been growing application of fuel cell (FC) across multiple industries in recent years in response to declining fossil fuel reserves, as well as worsening trends in environmental pollution and climate change. Among various types of fuel cells, the proton exchange membrane fuel cell (PEMFC) has the advantages of low operating temperature, high power density, fast response, great stability and environmental friendliness, making it suitable for several kinds of power generation applications (including mobile generator, static power, and distributed generation systems) [1]–[3].

However, the PEMFC's ability to respond to the load changes is severely restricted as its subsystems may have different dynamic characteristics. The cathode air flux of the PEMFC requires quick and accurate control in order for it to respond to a change of load in a timely manner according to different power demands (such as during the start, acceleration, or deceleration of a vehicle, each of which can cause a continuous sudden change of load) [4]. Lower air flux can result in insufficient oxygen supply, which will result in a reduced stack output voltage; a larger cathode air flux will lead to an increase in parasitic power consumption in the air supply system [5].

In terms of air supply systems, strategies for control and modeling have been proposed. Among them, many control strategies based on model construction, such as inner model control (IMC) [6], model predictive control (MPC) [7], adaptive control, nonlinear model predictive control (NMPC) [9], model-based sliding-mode control (SMC) [10], nonlinear

The associate editor coordinating the review of this manuscript and approving it for publication was Canbing Li.

multivariable control [11], and time delay control [12], have been applied to PEMFC air supply systems. Compared with one-order SMC methods, high-order ones [13], [14] are reported to have stronger robustness [15]. The improved algorithms with regard to Proportional Integral Derivative (PID) have attracted research attention, resulting in such innovations as Particle swarm optimization (PSO)-optimized PID used in the control system of PEMFC [16], a neural PID controller [17], the fuzzy PID control [18], a controller based on fuzzy controls combined with PID [19], the application of feedback linearization for the conversion of a nonlinear to linear control PEMFC model [20] and a fractional-order PID (FOPID) controller supported by a nonlinear observer [21].

The stability of operations is crucial for the accuracy and effectiveness of control actions, especially for air supply systems. Though the effective control of PEMFC, a complex nonlinear system, can be partially guaranteed by the strategies mentioned above, the control performance is not at the highest level due to the failure to accurately identify the air supply system. The neural network controller, the fuzzy logic controller (FLC) and other model-free controllers have wide applications in the field of air supply systems. For instance, a PEMFC air supply system can be well supported by the fuzzy logic control method [22]. Nonetheless, sufficient simulation and offline fine-tuning are required during the use of fuzzy control methods. Moreover, their applications are narrowed by the lack of a systematic formulation.

Known as the deep deterministic policy gradient (DDPG), the deep reinforcement learning algorithm is model-free [23], [24] and characterized by accurate control and fast response. Hence, it has been extensively applied to the field of control. Different from traditional control methods, DDPG achieves the combination of reinforcement learning's decision ability with deep learning's perception ability [25], thus formulating a control strategy through interacting with the environment sufficiently [26], [27]. It is able to adapt to a nonlinear system's uncertainty since model recognition is not required. Despite its wide application in the field of control [28]–[30], DDPG has not been used to control PEMFCs. It is not suitable for precision-sensitive objects because of its low exploration efficiency and Q value overestimation. To improve PEMFC's performance in controlling air flux, this paper proposes the design of a controller which is based on CIED-MD3. Hydrogen is used as the anode reaction gas in the PEMFC.

The innovations of our proposed method are as follows:

(1) This paper proposes an intelligent controller framework on the basis of deep reinforcement learning, which has better adaptivity, control performance and robustness. The output of the controller can be adjusted according to different states of the PEMFC so that the PEMFC can meet the real-time control requirements under different operating conditions. The controller tempers sudden air flux changes, thus preventing oxygen deprivation or oversaturation in the PEMFC.

(2) A CIED-MD3 algorithm for the framework is proposed. This algorithm, which is an improvement on the DDPG, is a distributed deep reinforcement learning algorithm with great global search ability and optimization speed. The CIED-MD3 employs the collective intelligent exploration policy (i.e. multi-actor network based on different exploration principles and parameters) to enable full exploration of the environment. The classification experience replay mechanism is introduced in order to improve exploration efficiency. Various techniques are also introduced in an effort to address the Q-value overestimation problem, including clipped multiple Q-learning, delayed update of policy, and smooth regularization of target policy. Finally, the CIED-MD3 algorithm with better global search ability and optimization speed is applied to the model-free PEMFC air flux intelligent controller. The simulation results presented in this paper show that the air flux can be controlled in a timely and accurate manner. Compared to controllers based on other methods, the intelligent controller based on the CIED-MD3 offers effective and superior performance.

The remainder of this paper contains the following: in section II the PEMFC air supply system model is demonstrated; section III contains a discussion on the CIED-MD3 algorithm; section IV details the design of the intelligent controller based on CIED-MD3; in section V the simulation results are discussed and analyzed; and, the findings in his paper are summarized in section VI.

II. MODEL OF PEMFC AIR SUPPLY SYSTEM

The PEMFC air supply system consists of the PEMFC stack, humidifier, air compressor, controller, radiator, hydrogen tank, gas-liquid separator and other devices [31]. The schematic diagram of this system is as shown in Figure 3. The operating principle of system is as follows: the air compressor is used to compress air into the supply pipe, and then, the air enters cathode through humidifier; under high pressure within the tank, the hydrogen in tank enters anode through humidifier; in the catalyst layer, the hydrogen molecules are decomposed into protons and electrons. The hydrogen protons pass through the proton exchange membrane under the effect of electromigration and reach cathode. The hydrogen protons combine with oxygen molecules under the effect of the cathode catalyst to produce water and at the same time produce electricity. The air not fully reacted in cathode is vented to the atmosphere through the return pipe [14].

A. PRECONDITIONS

In this paper, the mathematical model of PEMFC air supply system is constructed using the mechanism modeling method. The preconditions include:

1)Assume all gases follow the ideal gas law.

2)Assume the air temperature in electrode equals to the stack temperature.

3)Assume when the relative humidity of gas is higher than 100%, the vapor will be condensed into the liquid form. The air transmission process consists of the cathode, supply pipe and return pipe.

4) The hydrogen is used as the anode reaction gas.

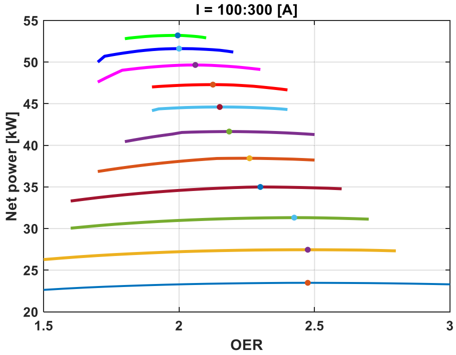


FIGURE 1. The relation curves between oxygen excess ratio and net power at different stack currents.

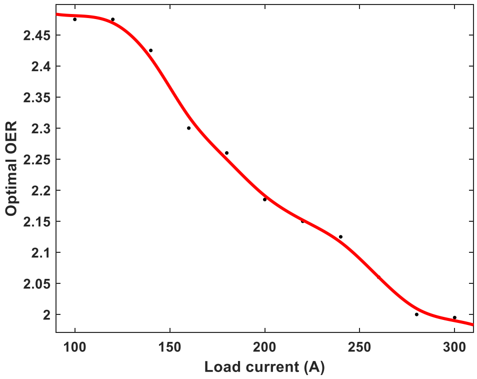


FIGURE 2. Fitting curve for the optimal OER and load currents.

B. CATHODE

Given the conservation of mass, the continuity equation of nitrogen and oxygen is obtained when the continuous equilibrium and masses of nitrogen in the cathode and oxygen's two components are considered [32]:

$$\frac{dm_{O_2}}{dt} = W_{O_2,\text{in}} - W_{O_2,\text{ out}} - W_{O_2,\text{ret}} \tag{1}$$

$$\frac{dm_{N_2}}{dt} = W_{N_2,\text{in}} - W_{N_2,\text{ out}} \tag{2}$$

where m_{O_2} and m_{N_2} are the masses of oxygen and nitrogen respectively; $W_{O_{2,in}}$ and $W_{N_{2,in}}$ represent the flows of oxygen and nitrogen entering the stack; $W_{O_{2,out}}$ and $W_{N_{2,out}}$ are the flows of oxygen and nitrogen flowing out of stack; $W_{O_{2,ret}}$ is the flow of oxygen consumed due to reaction in the stack.

Oxygen excess ratio (OER) is a key variable that significantly influences the fuel cell system's performance, expressed as

$$\lambda_{O_2} = \frac{W_{O_2,in}}{W_{O_2,\text{ret}}} \tag{3}$$

Figure 1 shows oxygen excess ratios and the net power at various stack currents. Figure 2 presents a lookup table generated by OER for the maximum net power under the load current.

C. ANODE

Hydrogen is used as the anode's reaction gas and supplied by a high-pressure tank, where a valve controls its flow. A high-energy source and the quick regulation of hydrogen flow are guaranteed. The control of the flow rate can help narrow the pressure gap between the exchange membrane's two sides, i.e., between the anode and cathode. A linear relationship exists between the control of the hydrogen in the anode and the pressure difference. The hydrogen flow is assumed to be under the direct control of the pressure gap's feedback. Yet, it is hard to directly measure the anode and cathode's pressure. Therefore, we apply a controller to the pressure of the cathode supply manifold. In terms of the anode, we assume that the supply manifold is small and inseparable from the anode volume, resulting in the same pressure. Thus, we express the anode pressure met by the controller as

$$W_{\text{an,in}} = K_1 \left(K_2 p_{\text{sm}} - p_{\text{an}}\right) \tag{4}$$

where p_{rm} is the return pipe's pressure and p_{rm} is the hydrogen pressure, $K_1 = 2.1$, $K_2 = 0.9$.

D. SUPPLY PIPE

The connection parts between the pipe and air compressor/cathode runner are included in the cathode's supply pipe. The mass inflow and outflow of the supply pipe are W_{cp} and W_{sm}, respectively. Thus, we get from the conservation of mass as well as energy [33]:

$$\frac{dm_{\text{sm}}}{dt} = W_{\text{cp}} - W_{\text{sm}} \tag{5}$$

$$\frac{dp_{\text{sm}}}{dt} = \frac{\gamma R_a}{V_{\text{sm}}} \left(W_{\text{cp}} T_{\text{cp}} - W_{\text{sm}} T_{\text{sm}}\right) \tag{6}$$

where m_{sm} and p_{sm} are the mass of air and the pressure in the supply pipe, respectively; γ is the air's specific heat ratio; R_a is the air's gas constant; V_{sm} is the supply pipe's volume; T_{cp} is the temperature of the air compressed into the compressor; T_{sm} is the air temperature in the supply pipe.

E. RETURN PIPE

When designing the return pipe, we need to take temperature changes into account. The gas temperature T_m in the return pipe is the same as that of gas leaving the cathode. According to the ideal gas law and the conservation of mass, the pressure p_{rm} of return pipe is expressed as,

$$\frac{dp_{\text{rm}}}{dt} = \frac{R_a T_{\text{rm}}}{V_{\text{m}}} \left(W_{\text{ca}} - W_{\text{rm}}\right) \tag{7}$$

where W_{ca} is the air flux in stack cathode; W_{rm} is the air flux at the outlet of return pipe; V_{rm} is the volume of the return pipe.

F. AIR COMPRESSOR

Figure 3 illustrates the system controlling PEMFC air flux. Referring to the model's dynamic characteristics from [32],

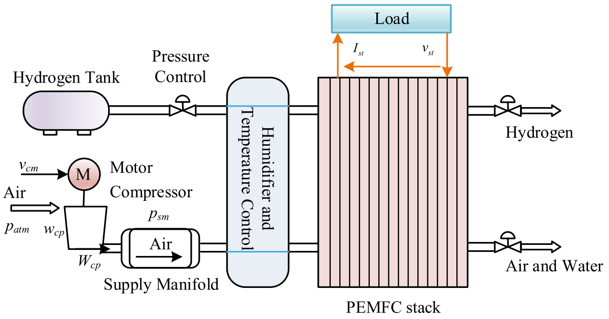


FIGURE 3. PEMFC air flux control system.

we obtain the air compressors rotational model as:

$$\begin{cases} J_{\mathrm{cp}} \dfrac{\mathrm{d}\omega_{\mathrm{cp}}}{\mathrm{d}t} = \tau_{\mathrm{cm}} - \tau_{\mathrm{cp}} \\ \tau_{\mathrm{cm}} = \eta_{\mathrm{cm}} \dfrac{k_{\mathrm{t}}}{R_{\mathrm{cm}}} \left(v_{\mathrm{cm}} - k_{\mathrm{v}}\omega\right) \\ \tau_{\mathrm{cp}} = W_{\mathrm{cp}} \dfrac{C_{\mathrm{p}} T_{\mathrm{atm}}}{\omega \eta_{\mathrm{cp}}} \left[\left(\dfrac{p_{\mathrm{sm}}}{p_{\mathrm{atm}}} \right)^{\frac{\gamma-1}{\gamma}} - 1 \right] \end{cases} \tag{8}$$

where J_{cp}, ω_{cp}, τ_{cm}, and τ_{cp} are the compressor's rotational inertia, speed, motor torque and load torque, respectively; k_t, R_{cm} and k_v are the motor constants; η_{cm} and v_{cm} are the motor's mechanical efficiency and control voltage, respectively; C_p is the air's specific heat capacity; η_{cp} is the compressor's efficiency; p_{atm} is the atmospheric pressure; T_{atm} is the temperature.

III. INTELLIGENT CONTROLLER BASED ON CIED-MD3

A. DEEP REINFORCEMENT LEARNING

The reinforcement learning (RL) [34] is one of the paradigms and methodologies of machine learning, which is used to describe and solve the problem that how the agent realizes maximal reward or achieves specific goal based on learning policy during its interaction with the environment.

The deep reinforcement learning combines the perception ability of deep learning with the decision ability of reinforcement learning, which can directly develop control strategy according to the state of environment. It is an artificial intelligence (AI) method closer to human thinking.

B. CIED-MD3

The collective intelligence exploration distributed multi-delay deep deterministic policy gradient (CIED-MD3) is a deep reinforcement learning algorithm expanded on the original DDPG [25]. To address the Q-value overestimation problem [25], the algorithm employs three techniques: the clipped multiple Q-learning, policy delayed updating, and smooth target policy regularization, so that the algorithm can achieve better stability and higher training efficiency.

During update of parameters, the traditional reinforcement learning algorithm adopts single neural network in agent for continuous update. This will lead to significant redundancy in the information utilized during update of agent, which will cause slow parameter update and tendency to fall into local optimum, particularly for the DDPG with high requirement for exploration ability. During training, if only one actor network is employed for environment exploration, the diversity of samples cannot be guaranteed. In order to solve this problem, this paper introduces the three techniques including classification experience replay mechanism, distributed training framework and collective intelligent exploration policy into the CIED-MD3, so as to enable the algorithm to improve the exploration efficiency and achieve better optimization results. The training framework and flow are shown in Figure 4 and Figure 5.

C. TRICKS

1) CLIPPED MULTIPLE Q-LEARNING

Inspired by the double-deep Q-learning (DDQN) [35] method, the current actor network is responsible to select optimal action. the policy can be evaluated by the target critic network:

$$y_t = r\left(s_t, a_t\right) + \gamma Q_{\theta'}\left(s_{t+1}, \pi_{\phi}\left(s_{t+1}\right)\right) \tag{9}$$

The target value is calculated by the clipped multiple Q-learning method in the CIED-MD3:

$$y_t^1 = r\left(s_t, a_t\right) + \gamma \min_{i=1,2,3} Q_{\theta_i'}\left(s_{t+1}, \pi_{\phi_1}\left(s_{t+1}\right)\right) \tag{10}$$

2) POLICY DELAYED UPDATING

After the critic network is updated for d times, the actor network will be updated once.

3) SMOOTH TARGET POLICY REGULARIZATION

Stochastic noise will be added to the target policy and the values of a mini-batch are averaged for the implementation of smooth regularization:

$$y_t = r\left(s_t, a_t\right) + \mathrm{E}_{\varepsilon}\left[Q_{\theta'}\left(s_{t+1}, \pi_{\phi'}\left(s_{t+1}\right) + \varepsilon\right)\right] \tag{11}$$

A stochastic noise is added to the target strategy:

$$y_t = r\left(s_t, a_t\right) + \gamma \min_{i=1,2,3} Q_{\theta_t}\left(s_{t+1}, \pi_{\phi'}\left(s_{t+1}\right) + \varepsilon\right) \tag{12}$$

$$\varepsilon \sim clip(N(0, \sigma), -c, c) \tag{13}$$

where ε is the noise added and $\min_{i=1,2} Q_{\theta_t}(s_{t+1}, \pi_{\phi'}(s_{t+1}) + \varepsilon)$ is the minimum Q value.

4) DISTRIBUTED TRAINING FRAMEWORK

The CIED-MD3 adopts the distributed reinforcement learning training framework, which involves multiple explorers, one learner and two experience buffer pools. The learner includes three critic networks and one actor network. Each explorer includes one actor network and has its own network model as well as environment. Multiple explorers explore the environment in a parallel way. First of all, the explorer generates experience based on its own environment, and adds the experience to the two experience buffer pools according to the standard. Then, the learner samples the experience from the experience buffer pools based on the standard and keeps

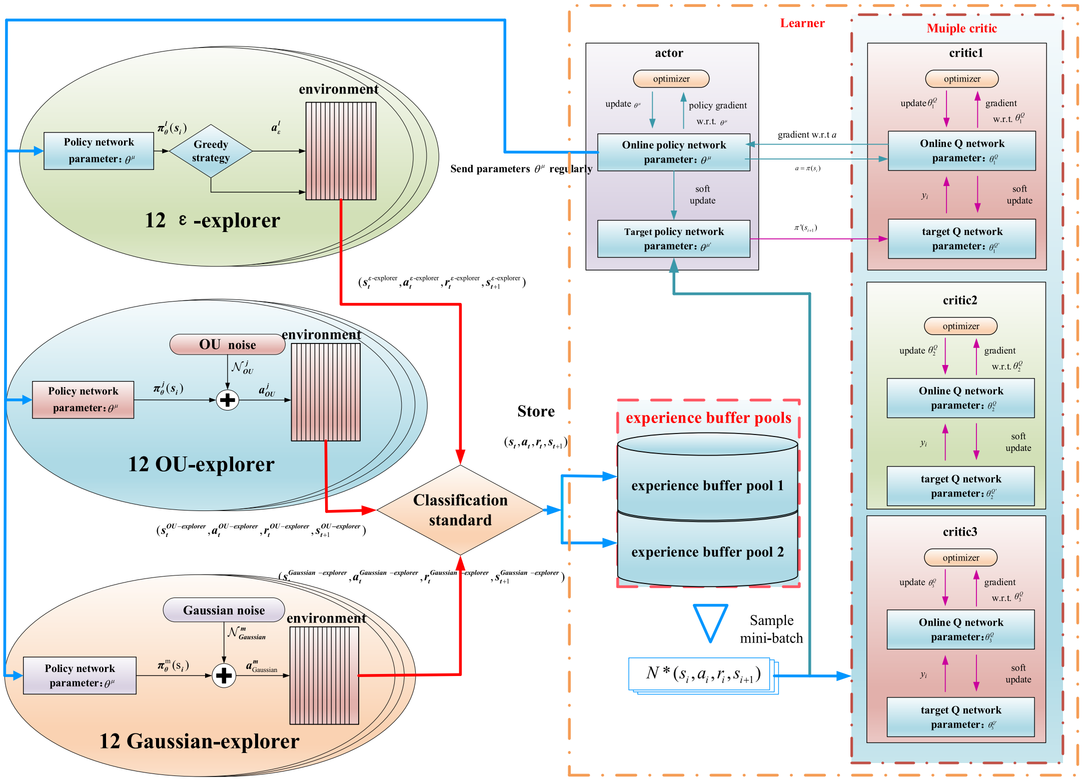


FIGURE 4. Distributed training framework of PEMFC intelligent controller based on the CIED-MD3.

learning. Finally, the actor network among explorers updates its network parameters regularly based on the latest actor network of learner.

5) COLLECTIVE INTELLIGENT EXPLORATION POLICY

This paper adopts three different exploration principles: the greedy strategy, Gaussian noise and Ornstein-Uhlenbeck (OU) noise [25]. The actor network in different explorer employs different exploration policy.

In Q-learning, the ε-greedy strategy means selecting an action within the action space that has some probability. Hence, referring to Q learning's exploration policy, we set that policy of the actor network in 12 explorers as the greedy strategy, and call it the ε-explorer, the action of which is expressed as:

$$a_\varepsilon^l = \begin{cases} \pi_\theta^l(s) & \text{With } \varepsilon \text{ probability} \\ a_{\text{rand}}^l & \text{With } 1-\varepsilon \text{ probability} \end{cases} \tag{14}$$

where $\pi_\theta^l(s)$ is the actor network policy of lth ε-explorer and a_{rand} is the action within the total action space.

Furthermore, we take the actor network's exploration policy in 12 explorers as the OU noise, called the OU-explorer. Random OU noise that has different variances is applied, leading to different noises in different explorers. This method effectively reduces sample repetition.

The action of OU-explorer is shown as follows:

$$a_{OU}^j = \pi_\theta^j(s) + \mathcal{N}_{OU}^j \tag{15}$$

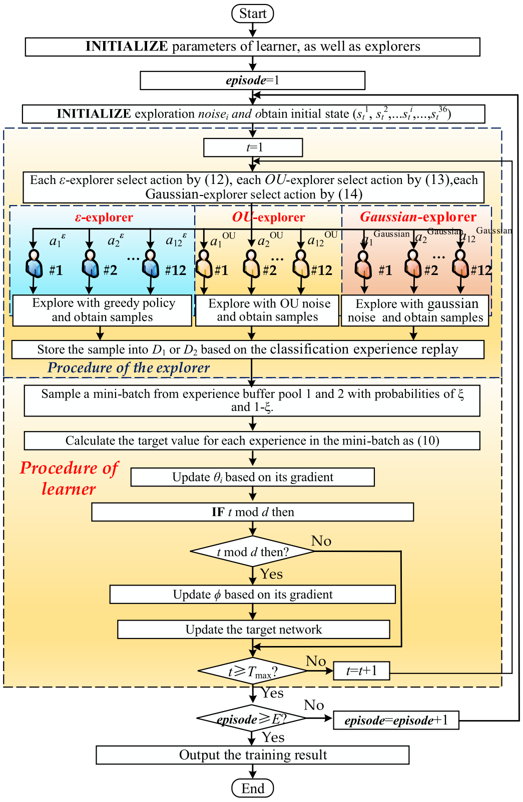

Start
INITIALIZE parameters of learner, as well as explorers
episode=1
INITIALIZE exploration noise$_i$ and obtain initial state $(s_t^1, s_t^2, \ldots s_t^i, \ldots, s_t^{36})$
t=1
Each ε-explorer select action by (12), each OU-explorer select action by (13),each Gaussian-explorer select action by (14)
ε-explorer
a_1^ε a_2^ε a_{12}^ε
#1 #2 ... #12
OU-explorer
a_1^{OU} a_2^{OU} a_{12}^{OU}
#1 #2 ... #12
Gaussian-explorer
$a_1^{Gaussian}$ $a_2^{Gaussian}$ $a_{12}^{Gaussian}$
#1 #2 ... #12
Explore with greedy policy and obtain samples
Explore with OU noise and obtain samples
Explore with gaussian noise and obtain samples
Store the sample into D_1 or D_2 based on the classification experience replay
Procedure of the explorer
Sample a mini-batch from experience buffer pool 1 and 2 with probabilities of ξ and 1-ξ.
Calculate the target value for each experience in the mini-batch as (10)
Procedure of learner
Update θ_i based on its gradient
IF t mod d then
t mod d then?
No
Yes
Update ϕ based on its gradient
Update the target network
$t \geqslant T_{\max}$?
No
$t=t+1$
Yes
episode$\geqslant E$?
No
episode=episode+1
Yes
Output the training result
End

FIGURE 5. Flow of CIED-MD3.

where $\pi_\theta^j(s)$ is the actor network policy of jth OU -explorer and N_{OU} is the OU noise.

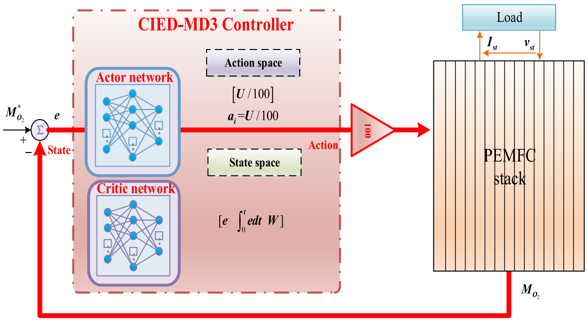


FIGURE 6. PEMFC air flux control based on the CIED-MD3.

In addition, the Gaussian noise is obtained by the actor network's optimization policy in 12 explorers, called the Gaussian-explorer. Similarly, in different explorers, we apply a random Gaussian noise that has different variances.

$$a^{m}_{\text{Gaussian}} = \pi^{\text{m}}_{\theta}(s) + \mathcal{N}^{m}_{Gaussian} \tag{16}$$

where $\pi^{m}_{\theta}(s)$ is the actor network policy of *mth* Gaussian - explorer and $N_{Gaussian}$ is the Gaussian noise.

As a result, by employing the above exploration policy based on different principles, the randomness and diversity of samples explored by explorers can be enhanced.

6) CLASSIFICATION EXPERIENCE REPLAY

Given the application of the experience replay mechanism, we further employ the immediate reward standard and the classification experience replay method. Experience samples are stored by two CIED-MD3 experience buffer pools that are independent. When the network model is initialized, for all samples in the two pools, we set the average immediate reward value as 0. Then we compare the immediate reward value with the value of one sample. When the latter is larger than the average value r_{a} for all experience samples, we will store this sample in pool 1; otherwise, we store it in pool 2. Then the average immediate reward value will be updated.

IV. DESIGN OF INTELLIGENT CONTROLLER BASED ON THE CIED-MD3

For the intelligent controller based on the CIED-MD3, the specific control strategy is shown in the Figure.6. Because the CIED-MD3 is adopted as the control algorithm of controller, the control interval is set as 0.01s according to the system response speed. The control objective is to make the PEMFC air flux strictly and accurately follow the reference air flux and make the oxygen excess ratio precisely tracking the reference value of oxygen excess ratio.

A. ACTION SPACE

The objective of air flux control is to control the motor voltage of air compressor, so as to control the air compressor and further control the air flux. In order to achieve the above objective and facilitate optimization at the same time, the action a is set as $U/100$, so that the action space will be within the range of [0,10]. The actual voltage obtained by the motor of air compressor is U, as shown in (17).

$$\begin{cases} a = [U/100] \\ 0 \leq U \leq U_{\max} \end{cases} \tag{17}$$

where U is the motor voltage of air compressor, and U_{max} is the upper limit of motor voltage of air compressor.

B. STATE SPACE

The states include the error $e(t)$ (air flux error) between the real air flux and reference air flux, its integral to t, and the air flux $W(t)$, it is shown in (18).

$$[e(t) \quad \int_0^t e(t)dt \quad W(t)] \tag{18}$$

C. REWARD FUNCTION

the comprehensive reward function i is represented as:

$$\text{r}(t) = -\left[\mu_1 e^2(t) + \mu_2 a^2(t-1)\right] + \beta \tag{19}$$

$$\beta = \begin{cases} 1 & e^2(t) \leq 0.01 \\ 0 & e^2(t) > 0.01 \end{cases} \tag{20}$$

where t is the discrete moment, $e(t)$ is the air flux error at moment t; $a(t-1)$ is the action of agent at moment $t-1$; β is the control reward term and when the control error $e(t)$ is no bigger than 0.01, a positive reward will be granted to the agent.

V. SIMULATION

In order to achieve better control of the air flux in the offline training process, the control interval is set to 0.01s. Moreover, to ensure randomness and diversity of samples, a step load current with varying amplitude (from 100 to 200A) is introduced into PEMFC for training, and the training interval of each episode is set to 5s. The parameters used in the model are given in Table AI. The fuel cell stack employs 75kW stacks used in the FORD P2000 fuel cell prototype vehicle [36]. The active area of the fuel cell is calculated from the peak power of the stack. The compressor model is based on the Allied Signal compressor detailed in [37]. The membrane properties of Nafion 117 membrane are obtained from [38]. The values of volumes are approximated from the dimensions of the P2000 fuel cell system. The training graph is shown in Figure 5 and the relevant parameters are listed in Table 1. Both the simulation model and programs described in this paper have been developed using a server consisting of 48 CPUs. The single CPU is a 2.10GHz Intel Xeon Platinum processor, and the RAM of the server is 192GB. The simulation software package used is MATALB/Simulink version 9.8.0 (R2020a). The operating parameters are as follows: anode and cathode gases of the fuel cell are fully humidified, the working pressure ranges from 0.14MPa to 0.22MPa, and the working temperature is 353K.

TABLE 1. Parameter settings.

Parameter	Value
Learning rate of critic	0.002
Learning rate of actor	0.002
Discount factor	0.9
Number of explorers	36
Noise variance of *m*th Gaussian-explorer	0.06+0.006**m*
Noise variance of *j*th OU-explorer	0.05+0.003**j*
Probability ε of *l*th ε-explorer	0.9
Probability of choosing experience pool 1	0.85
Interval of policy network update	2
Sizes of experience pools 1 and 2	900000
Target action noise variance	0.01

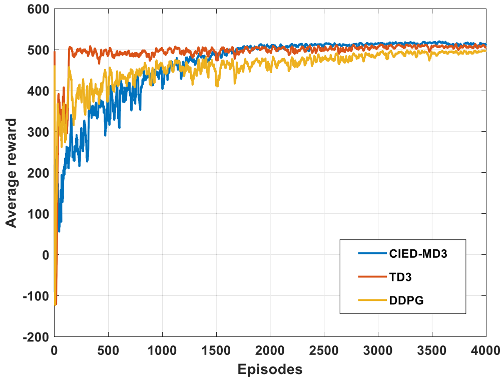


FIGURE 7. Training chart.

A. PARAMETER SELECTION

The weight coefficient used in the reward function and the hyperparameters in the offline training process are designed, as shown in Table 1.

B. PRE-LEARNING

The training graph is shown in Figure 7 and the relevant parameters are listed in Table I.

In Figure 7, the curves represent the average of corresponding episode rewards for various algorithms. Among them, both TD3 and DDPG algorithms show a slow pace of learning and exhibit visible oscillation during the learning process. In comparison, CIED-MD3 algorithm undergoes a relatively steady learning process, and its final average reward value is high, suggesting that the CIED-MD3 algorithm based on collective intelligent exploration policy is effective in working out an optimal solution with a higher value. In the meantime, applying the distributed optimization method requires many different types of explorers to perform optimization simultaneously for the CIED-MD3 algorithm to converge to the optimal solution sooner, which implies that the distributed training method is effective in improving the quality of solution as obtained.

C. ONLINE TEST

1) LOAD ADDING/SHEDDING CONDITION

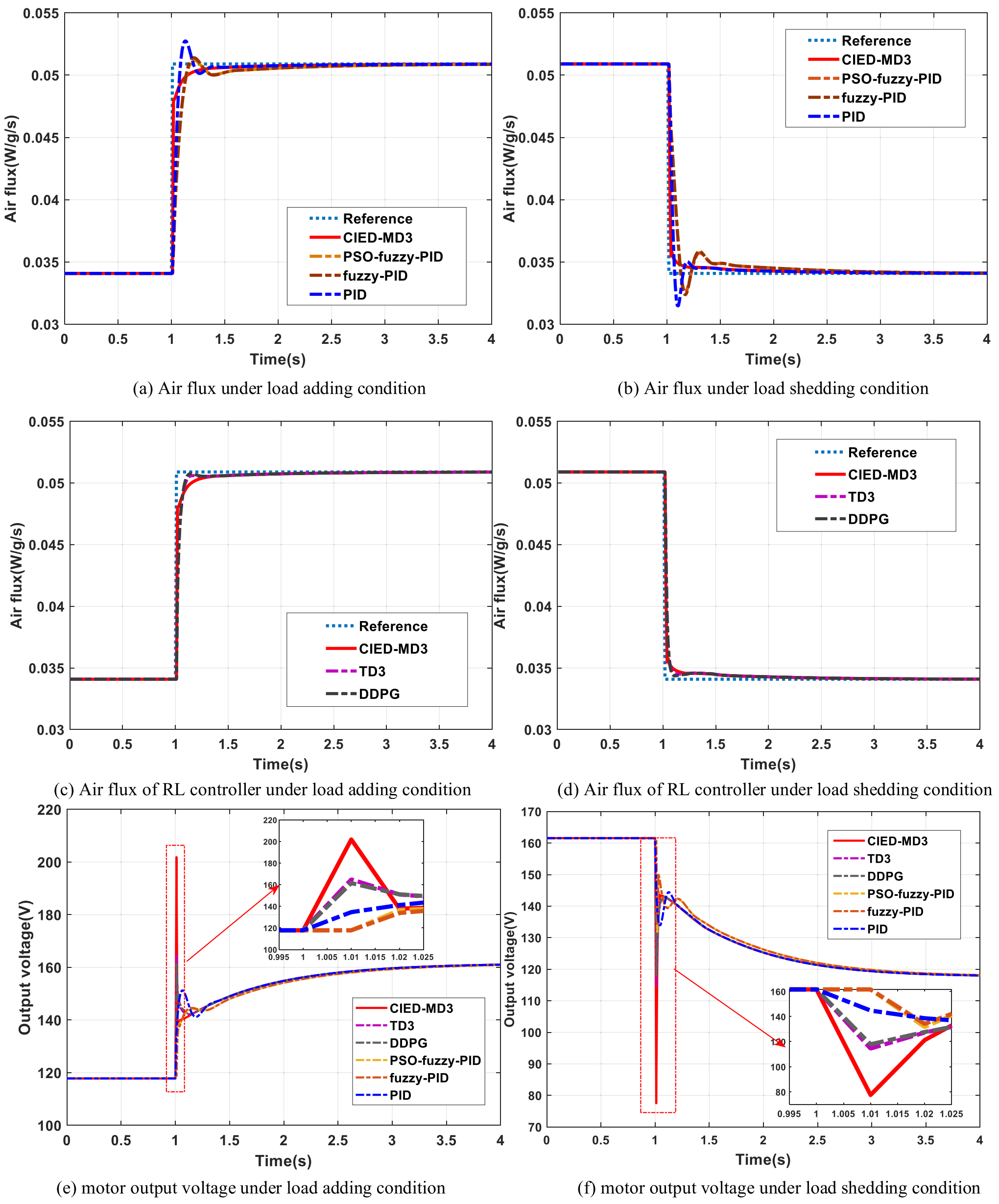
During the simulation process, it can be found out that the time it takes to shift from Unstable state to steady operation is far less than 10s. Hence, the operating time of working condition is set to 4s, which is a precondition for analysis. Under the load adding condition, at 1s, the load current instantly increases from 100A to 160A. While under the load shedding condition, at 1s, the load current drops from 100A to 160A momentarily. The result of simulation is shown in Figure 8(a)∼(d) and Table 2. According to Table 2, the rise time represents the time taken to reach the median between two reference values, and the stable time refers to the time taken to stabilize within the range of 0.01% of *e* reference value. In order to prove the availability of the CIED-MD3 controller, the TD3 controller, DDPG controller [25], PSO-fuzzy-PID controller [16], Fuzzy-PID controller [19] and PID controller are used as the examples. The PEMFC parameters are shown in Table 3.

As shown in Figures 8(a)∼(d) and Table 2, under the load adding/shedding condition, the CIED-MD3 controller can adapt to the fast change in load and the change in air flux caused by load change. The PEMFC controller therefore possesses greater adaptability and is an improvement on other controllers due to small overshoot and short response time. During the air transmission system's response, the small overshoot prevents the occurrence of oxygen over-saturation or deprivation in the PEMFC that can result from the significant fluctuations during transfer of air from the air compressor to the PEMFC. The short time of the dynamic response is conducive to ensuring sufficient PEMFC response speed, which improves the operating efficiency of the cell. Hence, other control algorithms are outperformed by CIED-MD3. Even though the PID controller is capable of making a fast dynamic response, it still requires big overshoots. The fuzzy-PID and PSO-fuzzy-PID controllers are prone to overshoot and oscillation. The CIED-MD3 controller proposed in this paper can exhibit better control performance and achieve a higher level of stability under the load adding/shedding condition.

Within the motor control system, the short response time and small overshoot are achieved via a compromise on the stability of the motor. In order to reach a high jump speed within a short space of time, the motor requires a high jump driving acceleration as the motor is required to reach a wider jump amplitude of output voltage quickly. To verify this statement, the output voltage of the motor is analyzed in this paper. According to the output voltage of the motor, as shown in Figures 8(e) and 8(f), when step change occurs with the load current, the output voltage of the instantaneous controller is changed abruptly to control the air flux. Under the load adding condition, at 1s, the output voltage of the CIED-MD3 controller will increase from 140V to 200V instantaneously. While under the load shedding condition, the output voltage drops sharply from 150V to 75V, and the voltage of air compressor exhibits significant fluctuations.

TABLE 2. Response parameters of PEMFC air transmission system.

Load change	Parameter	CIED-MD3	TD3	DDPG	PSO-fuzzy-PID	Fuzzy-PID	PID
adding	Rise time $T_r/10^{-3}$s	**1.16**	2.25	2.43	4.97	5.23	3.76
	Stable time T_s/s	**3.5**	3.54	3.59	3.90	3.97	3.60
	Overshoot σ/%	**0**	0	0	0.57	0.98	3.65
shedding	Rise time $T_r/10^{-3}$s	**2.82**	2.40	3.20	6.25	6.33	4.3
	Stable time T_s/s	**3.4**	3.45	3.48	3.80	3.85	3.59
	Overshoot σ/%	**0**	0	0	6.13	5.05	7.69



(a) Air flux under load adding condition

(b) Air flux under load shedding condition

(c) Air flux of RL controller under load adding condition

(d) Air flux of RL controller under load shedding condition

(e) motor output voltage under load adding condition

(f) motor output voltage under load shedding condition

FIGURE 8. Simulation results of PEMFC under load adding/shedding conditions.

Similarly, the other two RL controllers – the TD3 controller and DDPG controller – in order to obtain a faster response to the air flow, also accommodate different degrees of changes to motor voltage. Under the load adding condition, the peaks

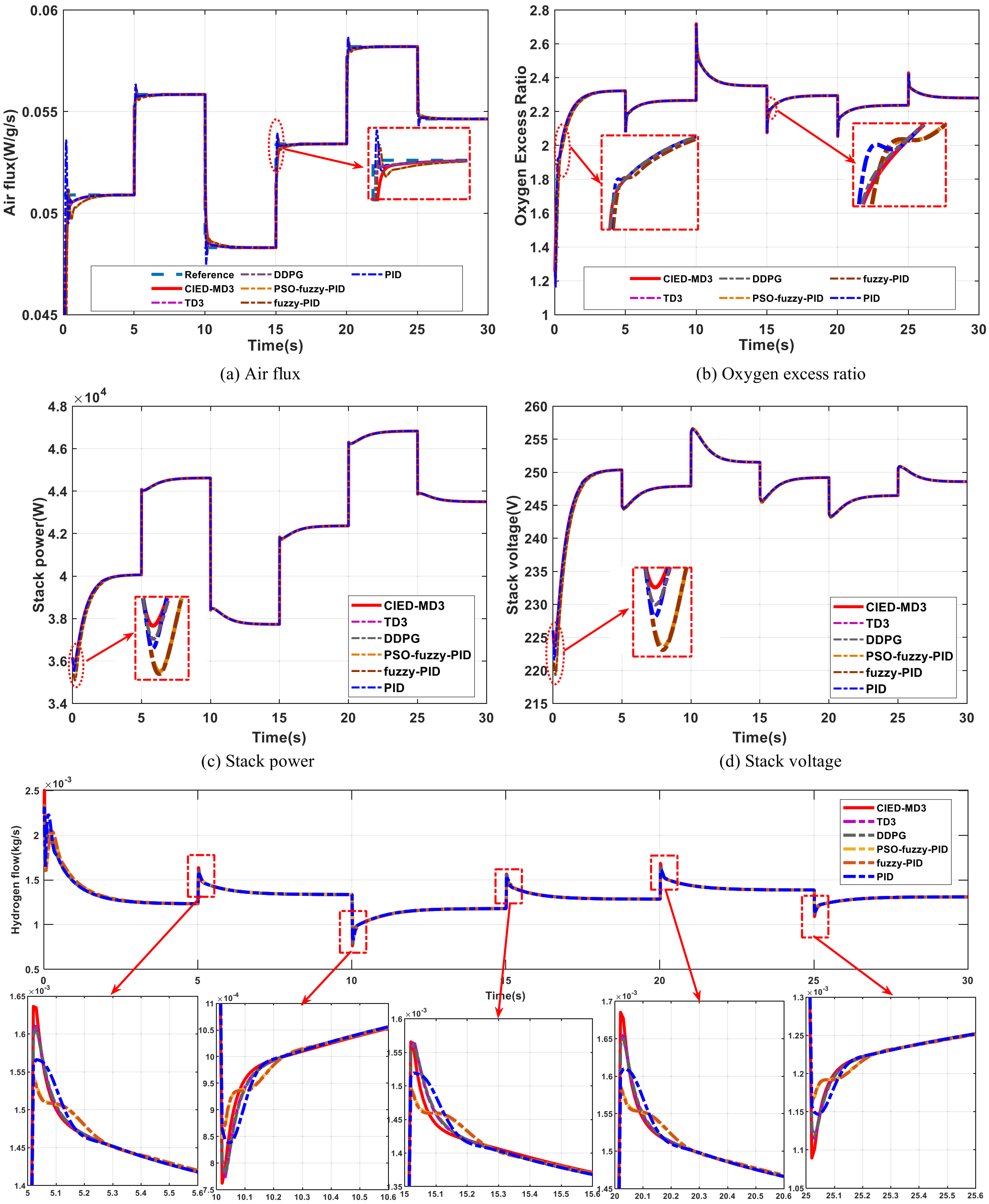
TABLE 3. PEMFC parameters.

symbol	parameter	value
$\rho_{m,dry}$	Number of cells in fuel-cell stack	0.002kg/cm^3
$M_{m,dry}$	membrane dry equivalent weight	1.1kg/mol
t_m	membrane thickness	0.01275cm
n	number of cell in fuel cell stack	381
A_{fc}	fuel cell active area	280cm^2
d_c	compressor diameter	0.2286m
J_{cp}	compressor and motor inertia	5*10^{-5}kg*m^2
V_{an}	anode volume	0.005m^3
V_{ca}	cathode volume	0.01m^3
V_{sm}	supply manifold volume	0.02m^3
V_{rm}	return manifold volume	0.005m^3
$C_{D,\,rm}$	return manifold throttle discharge coefficient	0.0124
$A_{T,\,rm}$	return manifold throttle discharge coefficient	0.002m^2
$k_{rm,out}$	supply manifold outlet orifice constant	0.3629*10^{-5} kg/(s*Pa)
$k_{ca,out}$	cathode outlet orifice constant	0.2177*10^{-5} kg/(s*Pa)
M_a	Air molar mass	29*10^{-3} kg*mol^{-1}
M_{O_2}	Oxygen molar mass	32*10^{-3} kg*mol^{-1}
M_{N_2}	Nitrogen molar mass	28*10^{-3} kg*mol^{-1}
M_v	Oxygen molar mass	18.02 kg*mol^{-1}
k_v	Compressor motor constant	0.0153V/(rad/sec)
k_t	Compressor motor constant	0.0153N-m/Amp
R_{cm}	Compressor motor constant	0.82Ω
η_{cm}	Compressor motor mechanical efficiency	98%
p_{atm}	Atmospheric pressure	101.325 kPa
T_{atm}	Atmospheric temperature	298.15 K
γ	Ratio of specific heat of air	1.4
C_p	Constant pressure specific heat of air	1004 J/(mol*K)
ρ_a	Air density	1.23 kg/m^3
R	Universal gas constant	8.3145 J/(mol*K)
R_a	Air gas constant	286.9 J/(mol*K)
R_{O_2}	Oxygen gas constant	259.8 J/(mol*K)
R_{N_2}	Nitrogen gas constant	296.8 J/(mol*K)
R_v	Vapor gas constant	461.5 J/(mol*K)
R_{H_2}	Hydrogen gas constant	4124.3 J/(mol*K)

of the controllers are 165V and 162V respectively. Under shedding load conditions, their minimum values are 114.5V and 117.8V. However, because their dynamic response speed is not as fast as that of the CIED-MD3 controller, the instantaneous motor output voltage change is not as large as that of the proposed controller. Conventional control controllers such as PSO-fuzzy-PID, fuzzy-PID and PID have a slow response speed; as a result, there is no sudden change in the motor voltage, which slowly increases and oscillates and finally stabilizes to the reference value.

2) RANDOM DISTURBANCE CONDITION

In order to verify the robustness and control performance of the controller proposed on the basis of CIED-MD3, the random load is adopted in the system, as shown in Appendix, where the disturbance lasts 5s each time, and the overall



(a) Air flux

(b) Oxygen excess ratio

(c) Stack power

(d) Stack voltage

(e) Hydrogen flow

FIGURE 9. **Simulation results of PEMFC under random load condition.**

simulation time is 30s. The simulation results are shown in Figure 9(a)∼(d).

As shown in Figure 9(a), since the proposed method has already learned many samples under different load conditions during offline training, it shows high adaptivity and robustness, which makes it capable of arriving at an optimal decision automatically under current state conditions. Therefore, even in the case of varying load disturbance, the proposed method is still capable of applying a steady control on air flux, thus attaining optimal control performance. During varying load conditions, the controller disciplines the actual air flux with the aid of a reference air flux value. Furthermore, the better control performance and high robustness of the CIED-MD3 controller enable the PEMFC to enforce a steady regulation of the oxygen excess ratio during load change disturbances, as shown in Figure 9(b). In addition, it can apply changes to net power and render the stack voltage steadier, as shown in Figures 9(c)∼(d).

In Figure 9(e) it can be seen that the peak hydrogen flow of the CIED-MD3 controller is the highest during the entire simulation process. This is because during earlier stages of load change, such as 5s, 10s, 15s, 20s and 25s, the CIED-MD3

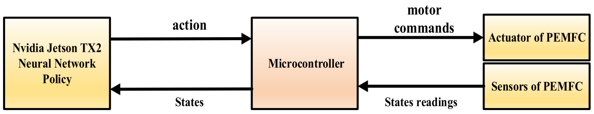


FIGURE 10. Application of CIED-MD3 air flow controller device.

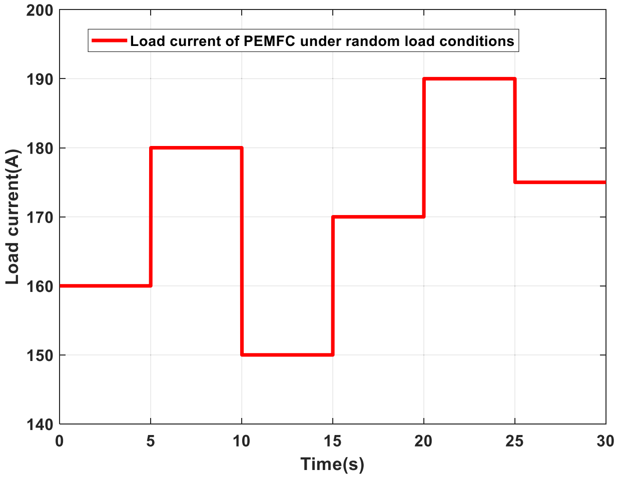


FIGURE 11. Load current of PEMFC under random load conditions.

controller has the fastest response speed (in relation to air flow), which produces a rapid rise in air pressure. According to Eq. (4), as p_{rm} rises, the flow rate of hydrogen also rises rapidly. Over time, the pressure on both sides of the exchange membrane gradually reach a state of parity, and the hydrogen flow rate becomes stabilized. The same phenomenon also occurs in other RL controllers. In addition, with a conventional controller, such as the PSO-fuzzy-PID, fuzzy-PID and PID controllers, due to their slow response rate, the hydrogen can be adjusted smoothly in the early stage. However, if there occurs a swing of air flow in the later stage, the hydrogen flow of PSO-fuzzy-PID and fuzzy-PID start to swing too, which is not conducive to the stable operation of the PEMFC.

In summary, the CIED-MD3 controller is judged as applicable to an actual air supply system due to its advantages of short response time, fast response speed, and its excellent dynamic and static performances.

D. OUTLOOK OF THE CIED-MD3 IN REALITY

In order to prove that the proposed algorithm works in reality, this paper proposes a conceived CIED-MD3 air flow controller device. The device consists of four components: microcontroller, PEMFC sensors, PEMFC actuator, and Nvidia Jetson TX2 module.

The PEMFC is equipped with sensors that can measure instantaneous air flow and other related variables. In online applications, the microcontroller can send commands to the DC motor actuator of the air compressor, receive sensor readings and perform simple calculations. However, the function of this microcontroller cannot adequately execute the neural network policy learned from deep RL. For this reason, this device requires installation of a Nvidia Jetson TX2 module in order to perform neural network inference. The TX2 module communicates with the microcontroller via the universal asynchronous transmitter (UART). In each control interval, the sensor's measurement data is collected on the microcontroller and sent back to TX2 module, where they are fed into a neural network policy to determine the action to be taken. These actions are then transferred to the microcontroller and executed by the actuator. This process is shown in Figure 10.

VI. CONCLUSION

In summary, the main contributions of this work to the field are as follows:

1) This paper has proposed an intelligent controller for the air supply system of the PEMFC, which can be achieved with the assistance of a CIED-MD3 algorithm. This algorithm has been developed from the original DDPG and it accompanies a collective intelligent exploration policy, which in turn employs a multi-actor network of different parameters and exploration principles in order to conduct distributed exploration in the environment. In addition, various other techniques are used to overcome the Q-value overestimation problem of DDPG. Consequently, this adaptive reinforcement learning control algorithm with great global search ability and optimization speed can implement fast control in accordance with a small range of PEMFC states.

2) The simulation of the PEMFC under different load conditions, and the comparative analysis of controllers operating on different principles, have shown that the CIED-MD3 controller can meet the real-time control requirements of an air supply system under different operating conditions during load current changes.

3) The CIED-MD3 algorithm proposed in this paper is a model-free control algorithm, which has better control performance, and which speeds up convergence, reduces training costs, and increases the generalization and adaptability of the controller. In view of these properties, there is considerable scope for its application in real-world PEMFCs in the foreseeable future. In future studies we aim to apply the algorithm to an actual system and verify the effectiveness of the method through a series of experiments. In addition, we aim to develop an improved algorithm by proposing more tricks with the aim of improving the generalization of the algorithm and enabling online learning.

APPENDIX

See Table 3.

REFERENCES

[1] M. Chen, Z. Cheng, Y. Liu, Y. Cheng, and Z. Tian, "Multitime-scale optimal dispatch of railway FTPSS based on model predictive control," *IEEE Trans. Transport. Electrific.*, vol. 6, no. 2, pp. 808–820, Jun. 2020.

[2] D. J. Friedman, A. Eggert, P. Badrinarayanan, and J. Cunningham, "Balancing stack, air supply, and water/thermal management demands for an indirect methanol PEM fuel cell system," in *Proc. SAE World Congr.*, Detroit, MI, USA, Mar. 2001, p. 12.

[3] F. Mitlitsky and B. Myers, "Regenerative fuel cell systems," *Energy Fuels*, vol. 12, no. 1, pp. 56–71, Jan. 1998.

[4] R. M. Aslam, D. B. Ingham, M. S. Ismail, K. J. Hughes, L. Ma, and M. Pourkashanian, ''Simultaneous direct visualisation of liquid water in the cathode and anode serpentine flow channels of proton exchange membrane (PEM) fuel cells,'' *J. Energy Inst.*, vol. 91, no. 6, pp. 1057–1070, Dec. 2018.

[5] N. Chatrattanawet, T. Hakhen, S. Kheawhom, and A. Arpornwichanop, ''Control structure design and robust model predictive control for controlling a proton exchange membrane fuel cell,'' *J. Cleaner Prod.*, vol. 148, pp. 934–947, Apr. 2017.

[6] H. Beirami, A. Z. Shabestari, and M. M. Zerafat, ''Optimal PID plus fuzzy controller design for a PEM fuel cell air feed system using the self-adaptive differential evolution algorithm,'' *Int. J. Hydrogen Energy*, vol. 40, no. 30, pp. 9422–9434, Aug. 2015.

[7] J. K. Gruber, C. Bordons, and A. Oliva, ''Nonlinear MPC for the airflow in a PEM fuel cell using a volterra series model,'' *Control Eng. Pract.*, vol. 20, no. 2, pp. 205–217, Feb. 2012.

[8] J. Han, S. Yu, and S. Yi, ''Adaptive control for robust air flow management in an automotive fuel cell system,'' *Appl. Energy*, vol. 190, pp. 73–83, Mar. 2017.

[9] Y.-X. Wang and Y.-B. Kim, ''Real-time control for air excess ratio of a PEM fuel cell system,'' *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 3, pp. 852–861, Jun. 2014.

[10] S. Laghrouche, M. Harmouche, F. S. Ahmed, and Y. Chitour, ''Control of PEMFC air-feed system using Lyapunov-based robust and adaptive higher order sliding mode control,'' *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 4, pp. 1594–1601, Jul. 2015.

[11] M. A. Danzer, J. Wilhelm, H. Aschemann, and E. P. Hofer, ''Model-based control of cathode pressure and oxygen excess ratio of a PEM fuel cell system,'' *J. Power Sources*, vol. 176, no. 2, pp. 515–522, Feb. 2008.

[12] Y.-B. Kim, ''Improving dynamic performance of proton-exchange membrane fuel cell system using time delay control,'' *J. Power Sources*, vol. 195, no. 19, pp. 6329–6341, Oct. 2010.

[13] J. Liu, W. Luo, X. Yang, and L. Wu, ''Robust model-based fault diagnosis for PEM fuel cell air-feed system,'' *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3261–3270, May 2016.

[14] R. J. Talj, D. Hissel, R. Ortega, M. Becherif, and M. Hilairet, ''Experimental validation of a PEM fuel-cell reduced-order model and a moto-compressor higher order sliding-mode control,'' *IEEE Trans. Ind. Electron.*, vol. 57, no. 6, pp. 1906–1913, Jun. 2010.

[15] G. Park and Z. Gajic, ''A simple sliding mode controller of a fifth-order nonlinear PEM fuel cell model,'' *IEEE Trans. Energy Convers.*, vol. 29, no. 1, pp. 65–71, Mar. 2014.

[16] P. Farhadi and T. Sojoudi, ''PEMFC voltage control using PSO-tunned-PID controller,'' in *Proc. IEEE NW Russia Young Researchers Electr. Electron. Eng. Conf. (ElConRusNW)*, Feb. 2014, pp. 32–35.

[17] C. Damour, M. Benne, C. Lebreton, J. Deseure, and B. Grondin-Perez, ''Real-time implementation of a neural model-based self-tuning PID strategy for oxygen stoichiometry control in PEM fuel cell,'' *Int. J. Hydrogen Energy*, vol. 39, no. 24, pp. 12819–12825, Aug. 2014.

[18] K. Ou, Y.-X. Wang, Z.-Z. Li, Y.-D. Shen, and D.-J. Xuan, ''Feedforward fuzzy-PID control for air flow regulation of PEM fuel cell system,'' *Int. J. Hydrogen Energy*, vol. 40, no. 35, pp. 11686–11695, Sep. 2015.

[19] W. Xia and Z. Qi, ''Dynamic modeling and fuzzy PID control study on proton exchange membrane fuel cell,'' in *Proc. 2nd Conf. Environ. Sci. Inf. Appl. Technol.*, Jul. 2010, pp. 116–119.

[20] J. Chen, Z. Liu, F. Wang, Q. Ouyang, and H. Su, ''Optimal oxygen excess ratio control for PEM fuel cells,'' *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 5, pp. 1711–1721, Sep. 2018.

[21] D. Zhao, F. Li, R. Ma, G. Zhao, and Y. Huangfu, ''An unknown input nonlinear observer based fractional order PID control of fuel cell air supply system,'' *IEEE Trans. Ind. Appl.*, vol. 56, no. 5, pp. 5523–5532, Sep. 2020.

[22] W. Guoai, Q. Shuhai, and Q. Yingchuan, ''Fuzzy logic control of air supply system in PEMFC for electric vehicles,'' in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, Dec. 2008, pp. 180–183.

[23] X. Qi, ''Rotor resistance and excitation inductance estimation of an induction motor using deep-Q-learning algorithm,'' *Eng. Appl. Artif. Intell.*, vol. 72, pp. 67–79, Jun. 2018.

[24] Y. Hu, W. Li, K. Xu, T. Zahid, F. Qin, and C. Li, ''Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning,'' *Appl. Sci.*, vol. 8, no. 2, p. 187, Jan. 2018.

[25] T. P. Lillicrap, J. J. Hunt, A. Pritzel, and N. Heess, ''Continuous control with deep reinforcement learning,'' *Comput. Sci.*, vol. 8, no. 6, p. A187, Jul. 2015.

[26] M. Zhu, X. Wang, and Y. Wang, ''Human-like autonomous car-following model with deep reinforcement learning,'' *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 348–368, Dec. 2018.

[27] P. Chen, Z. He, C. Chen, and J. Xu, ''Control strategy of speed servo systems based on deep reinforcement learning,'' *Algorithms*, vol. 11, no. 5, p. 65, May 2018.

[28] L. Xi, J. Wu, Y. Xu, and H. Sun, ''Automatic generation control based on multiple neural networks with actor-critic strategy,'' *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 14, 2020, doi: 10.1109/TNNLS.2020.3006080.

[29] J. Duan, D. Shi, R. Diao, H. Li, Z. Wang, B. Zhang, D. Bian, and Z. Yi, ''Deep-reinforcement-learning-based autonomous voltage control for power grid operations,'' *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 814–817, Jan. 2020.

[30] H. Shi, Y. Sun, G. Li, F. Wang, D. Wang, and J. Li, ''Hierarchical intermittent motor control with deterministic policy gradient,'' *IEEE Access*, vol. 7, pp. 41799–41810, 2019.

[31] J. T. Pukrushpan, *Modeling and Control of Fuel Cell Systems and Fuel Processors*. Ann Arbor, MI, USA: Univ. Michigan, 2003.

[32] R. Tirnovan and S. Giurgea, ''Efficiency improvement of a PEMFC power source by optimization of the air management,'' *Int. J. Hydrogen Energy*, vol. 37, no. 9, pp. 7745–7756, May 2012.

[33] Y. Ma, F. Zhang, J. Gao, H. Chen, and T. Shen, ''Oxygen excess ratio control of PEM fuel cells using observer-based nonlinear triple-step controller,'' *Int. J. Hydrogen Energy*, vol. 45, no. 54, pp. 29705–29717, Nov. 2020.

[34] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, ''Human-level control through deep reinforcement learning,'' *Nature*, vol. 518, no. 7540, p. 529, Feb. 2015.

[35] H. Van Hasselt, A. Guez, and D. Silver, ''Deep reinforcement learning with double q-learning,'' in *Proc. 30th AAAI Conf. Artif. Intell.*, Feb. 2016, pp. 2094–2100.

[36] J. A. Adams, W.-C. Yang, K. A. Oglesby, and K. D. Osborne, ''The development of Ford's P2000 fuel cell vehicle,'' SAE Tech. Paper 2000-01-1061, 2000.

[37] J. M. Cunningham, M. A. Hoffman, R. M. Moore, and D. J. Friedman, ''Requirements for a flexible and realistic air supply model for incorporation into a fuel cell vehicle (FCV) system simulation,'' SAE Tech. Paper 1999-01-2912, 1999.

[38] T. V. Nguyen and R. E. White, ''A water and heat management model for proton-exchange membrane fuel cells,'' *J. Electrochem. Soc.*, vol. 140, no. 8, pp. 2178–2186, 1993.

JIAWEN LI received the M.S. degree in electrical engineering from Northeast Electric Power University, Jilin, China, in 2016. He is currently pursuing the D.Eng. degree in electrical engineering with the School of Electric Power, South China University of Technology. His research interest includes automatic generation control.

TAO YU is currently a Professor of power system with the School of Electric Power, South China University of Technology (SCUT), Guangzhou, China. His research interests include nonlinear and coordinated control theory.