Reinforcement Learning based Energy Management for Fuel Cell Hybrid Electric Vehicles

Liang Guo, Zhongliang Li, Rachid Outbib

Reinforcement Learning based Energy Management for Fuel Cell Hybrid Electric Vehicles

Liang GUO, Zhongliang LI, Rachid OUTBIB
Aix-Marseille University, Labratery LIS (UMR CNRS 7020), Marseille, France
liang.guo@lis-lab.fr, zhongliang.li@lis-lab.fr, rachid.outbib@lis-lab.fr

***Abstract*—In the paper, a self-learning energy management strategy is proposed for fuel cell hybrid electric vehicles (FCHEV). The studied energy system for FCHEV is composed of fuel cells and lithium batteries. A reinforcement learning (RL) based energy management strategy (EMS) for FCHEV is studied to achieve the power allocation of the two energy sources. The objective is to learn a satisfactory EMS from scratch and only through the interaction of environments. Specifically, Q-Learning, one of the RL methods, is applied to minimize fuel consumption and ensure battery sustainability. Compare with Dynamic Programming (DP), which can reach the best performance of sequential decision problems theoretically, Q-Learning based EMS can achieve results close to DP based EMS. During the process, different objective functions are optimized to be suitable for Q-Learning. Finally, the simulation results with python verify the effectiveness of the method proposed in this paper.**

Keywords—energy management strategy, fuel cell, hybrid electric vehicle, reinforcement learning

I. INTRODUCTION

Fuel cell hybrid electric vehicles are attracting increasing attention. Energy management strategy (EMS), dedicated to allocating power between different energy sources, is one of the key elements to achieve high fuel efficiency [1]. Energy management strategies can be divided into three categories: rule-based EMS, optimization-based EMS, and learning-based EMS.

Rule-based strategies realize EMS goals according to the rules which are established based on the characteristics of the concerned powertrain and load. Among those proposed, fuzzy logic rule-based EMS has been demonstrated to be an effective one in a wide range of hybrid electric vehicles [2][3]. Easy implementation and reliable performance make the rule-based EMS the most widely used strategy. However, over-reliance on experience and difficulty to reach optimal results are its main limitations to higher energy economic.

An EMS problem can often be formulated as a constrained sequence optimization problem. The optimization goal is to find an optimized trajectory with respect to an objective function. Among the optimization methods, the numerical global optimal solution can be found using dynamic programming (DP) [4]. However, DP implementation needs to the entire information of external system input in advance which is hardly available in practice. In addition, DP implementation is computationally heavy which blocks its real-time implementation. Other model-based optimal control methods have been investigated for EMS problems. Among them, Pontryagin's minimum principle (PMP), equivalent consumption minimization strategy (ECMS) and stochastic dynamic programming (SDP) are three methods that can be used for real-time implementation[5][6][7].However, the performance of the three methods is highly dependent on the initial parameter settings or identifications that are related to driving conditions. Model predictive control (MPC) is also widely studied for EMS problems [8]. The main drawback of MPC is that the control performance is heavily dependent on the model prediction performance and model accuracy.

To tackle the modeling complexity and the uncertain external input information, learning-based approaches are recently receiving attention in both academic and industrial communities [9][10]. Among them, reinforcement learning (RL, which achieved remarkable advances in recent years, is considered a promising alternative for EMS. Through RL, control policy can be learned in a model-free way and only through interaction with the environment [11]. The dependency of EMS design on the precise system model could be alleviated and the uncertain information could be handled naturally.

In the paper, EMS based on RL is studied for FCHEV. More precisely, one basic RL method Q-learning is investigated to solve the EMS problem. In practice, the major challenge of using the RL-based method is to improve learning efficiency. For this, two novel objective functions are designed and tested. The performance of the proposed RL-based EMS with the novel objective functions is evaluated by comparing to the benchmarks using DP on a simulation platform.

II. SYSTEM MODELING

The studied FCHEV energy system is as follows: it consists of two energy systems, one is a fuel cells system, another is a batteries system. Each energy source will be cascaded with a DC/DC converter to control their work points and improve system reliability.

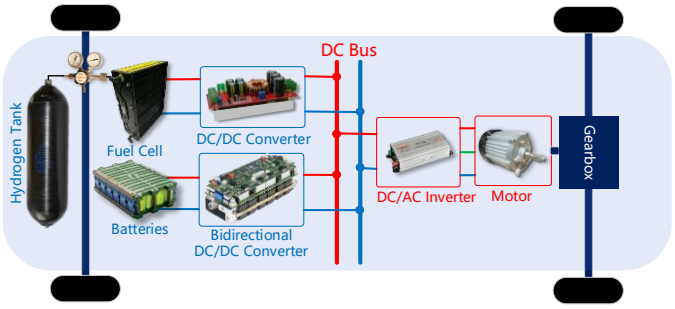


Fig.1. Energy system for fuel cell hybrid electric vehicle.

A simulation platform is built in this study. The models composing the platform are presented in this section.

A. *Fuel cell model*

The output voltage V_{fc} of the fuel cell can be expressed as follows:

$$V_{fc} = n_{cell} \cdot (E_{nst} - V_{act} - V_{con} - V_{ohm}) \tag{1}$$

where n_{cell} is the number of single fuel cells, E_{nst} is the theoretical voltage called the Nernst electromotive force, V_{act} is the voltage drop due to the phenomenon of activated polarization, V_{con} is the voltage drop caused by concentration polarization, and V_{ohm} is the ohmic voltage loss. The specific model of each part of the fuel cell is shown in (2):

$$\begin{cases} E_{nst} = E_0 + \dfrac{\Delta TS}{nF} - \dfrac{RT}{nF}\ln\left(\dfrac{P_{H_2O}}{P_{H_2}\sqrt{P_{O_2}}}\right) \\ V_{act} = \dfrac{RT}{\alpha F}\ln\left(\dfrac{i_{fc}+i_{loss}}{i_0}\right) \\ V_{con} = \dfrac{RT}{nF}\ln\left(\dfrac{I_{lim}}{I_{lim}-i_{fc}}\right) \\ V_{ohm} = i_{fc}R_{ohm} \end{cases} \tag{2}$$

where $E_0 = 1.23V$ is the open-circuit voltage of fuel cell reaction at standard atmospheric pressure, $R = 8.3145$ is gas constant, $T = 333.15K$is the fuel cell temperature, $F = 96485$ is Faraday constant, $\alpha = 1$ is the transfer coefficient, P is the local pressure of the reactants and products at this atmospheric pressure. i_{fc} is the current density. $i_{loss} = 2mA/cm^2$ is the current loss, $i_0 = 0.003mA/cm^2$ is the exchange current density. $I_{lim} = 1.6A/cm^2$ is the limiting current density. R_{ohm} is the fuel cell resistance.

$$\dot{m}_{H_2} = M_{H_2}\frac{I_{fc}}{nF} \Rightarrow P_{fc} = \frac{2V_{fc}F}{M_{H_2}} \cdot \dot{m}_{H_2} \tag{3}$$

where $\dot{m}_{H_2}$ is the rate at which hydrogen is consumed, and M_{H_2} is the molar mass of hydrogen. P_{fc} is the output power of fuel cells. The converter model will only concern about its power characteristics. The DC/DC converter model for fuel cells is as follows:

$$P_{fc} = P_{fc}' / \eta_{dc}(P_{fc}') + P_{aux} \tag{4}$$

where P'_{fc} is the output power of the fuel cells system. It is considered that P'_{fc} is equal to the power command from the control strategy. η_{dc} is the efficiency of DC/DC converter for fuel cells. P_{aux} is the auxiliary system, and it can be considered as a constant current load $I_{aux} = 2.0A$.

The fuel cells parameters are $n_{cell} = 200$, the effective area of the electrode is $A_{fc} = 324cm^2$, the pressure of anode hydrogen is 50 kPa, and anode oxygen is obtained from the air by natural aspiration. As shown in Fig.2, when the current is: 437A, the FC power can reach the max power: 104kW, and the efficiency will be: 43.19%; When the current is: 63.2A, the FC efficiency can reach the max efficiency: 54.49%, and the power will be: 15.7kW.

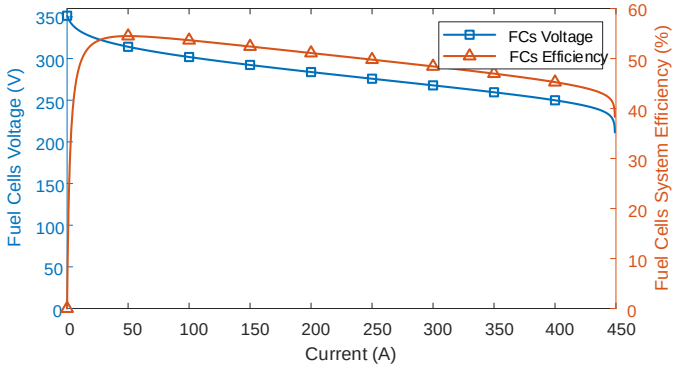


Fig.2 The output voltage and efficiency of fuel cells

B. *Battery model*

Battery output power P_{bat} is modelled as

$$P_{bat} = V_{oc}(SOC_{bat})I_{bat} - I_{bat}^2 R_{bat}(SOC_{bat}) \tag{5}$$

where V_{oc} is the battery open-circuit, R_{bat} is the internal resistance of the battery. I_{bat} is the output current of the battery. When the $I_{bat} > 0$, the battery discharge. When the $I_{bat} < 0$, the battery discharge.

Given P_{bat}, I_{bat} can be calculated according to (5), as .

$$I_{bat} = \frac{V_{oc}(SOC_{bat}) - \sqrt{V_{oc}^2(SOC_{bat}) - 4R_{bat}(SOC_{bat})P_{bat}}}{2R_{bat}(SOC_{bat})} \tag{6}$$

The battery state of charge $SOC_{bat}(t)$ can be obtained by ampere time integration:

$$SOC_{bat}(t) = SOC_{bat}(0) - \int_0^t I_{bat}(t)/Q_{bat} \tag{7}$$

where Q_{bat} is the battery capacity.

The efficiency of the batteries η_{bat} is:

$$\eta_{bat} = \begin{cases} \dfrac{V_{oc}(SOC_{bat}) - I_{bat}R_{bat}(SOC_{bat})}{V_{oc}(SOC_{bat})} & (I_{bat} > 0) \\ \dfrac{V_{oc}(SOC_{bat})}{V_{oc}(SOC_{bat}) - I_{bat}R_{bat}(SOC_{bat})} & (I_{bat} < 0) \end{cases} \tag{8}$$

V_{oc} and R_{bat} are two empirical functions of SOC shown in Fig. 3, and formed in looking-up table form. Considering the power loss of the battery-side DC/DC converter, the battery output power is expressed as

$$P_{bat} = \begin{cases} P_{bat}' / \eta_{bdc} & (P_{bat}' > 0) \\ P_{bat}' \cdot \eta_{bdc} & (P_{bat}' < 0) \end{cases} \tag{9}$$

where P'_{bat} is the output power of the power converter whose efficiency is η_{abc}.

We choose the capacity of batteries is: 6.6 Ah, the serial number and parallel number of batteries are 68 and10. Therefore, the standard voltage will be 244.8V. Fig.3 shows the characteristics of the batteries, including the open circuit voltage and the internal resistance of the batteries.

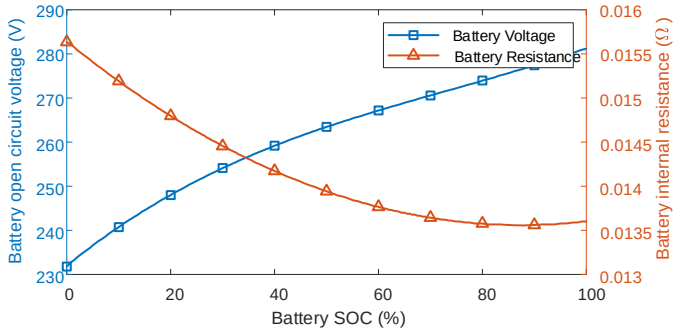


Fig.3. The characteristics of the batteries

C. *Vehicle dynamics model*

Suppose a vehicle is moving forward at velocity v on a road with gradient θ, its dynamic equation is:

$$\begin{aligned} F_m &= F_{air} + F_f + F_s + F_a \\ &= \frac{1}{2} C_D A \rho v^2 + Gf\cos\theta + G\sin\theta + m\frac{dv}{dt} \end{aligned} \tag{10}$$

Where F_m represents the driving force provided by the motor, F_{air} represents air resistance, F_f represents rolling resistance, and F_s shows slope resistance and F_a represents acceleration resistance. ρ and C_D represent air density and air resistance coefficient respectively. A represents the windward surface volume of the car body, and v represents the vehicle velocity. m represents the vehicle mass. $G = mg$ represents the gravity of the vehicle, and f represents the sliding resistance coefficient.

The required power for the vehicle is:

$$P_{veh} = F_m \cdot v / \eta_m \tag{11}$$

where, P_{veh} represents the required power of the motor, η_m represents the transmission efficiency of the electric machine. According to the power balance, the required power of the motor is provided by fuel cell and battery:

$$P_{veh} = P'_{fc} + P'_{bat} \tag{12}$$

In the instantiation of the vehicle, we set the weight of the vehicle to be 2500kg, the windward area is 1.8m2, air density is 1.25 kg/m2, the air resistance coefficient is 0.3, the rolling friction coefficient is 0.01, and the total mechanical transmission efficiency is 90%, the gravity acceleration is 9.8m/s2.

III. PROBLEM FORMULATION AND RL-BASED EMS

The EMS for FCHEV can be considered as a constrained sequence optimization problem. The optimization goal is to find an optimized trajectory to minimize an objective function

$$\begin{cases} X_{\min} \le x_i \le X_{\max} \\ U_{\min} \le u_i \le U_{\max} \end{cases} \tag{13}$$

subject to the constraints of state and control variables. It is found that the performance of RL-based EMS is dependent on the property of the objective function. In the sequel, several objective functions are formulated to investigate.

A. *Objective function design*

Objective function formulation is a key element for optimal control. In the paper, minimizing hydrogen consumption and maintaining battery SOC are two general objectives of EMS. Intuitively, the objective function can be formulated as

$$\min J_0 = \min \sum_{i=0}^{T-1} \{\dot{m}_{H_2}(i) + \alpha \cdot [SOC_{bat}(i) - SOC_{ref}(i)]^2\} \tag{14}$$

where m_{H_2} is the cumulative mass of hydrogen consumed. α is a positive real coefficient.$SOC_{ref}(i)$ is the tracking reference of batteries' SOC.

To enable the SOC to track the predefine trajectory faster, the redesigned objective function for fuel consumption minimization and battery SOC tracking would be:

$$\min J_1 = \min \sum_{i=0}^{T-1} \{\dot{m}_{H_2}(i) + \alpha \cdot S_i \cdot \Delta S_i\} \tag{15}$$

To avoid the occurrence of too large a positive value or too small negative value ΔS_i, a nonlinear mapping is performed on it based on the hyperbolic tangent function $Tanh(x)$ to constrain the size of ΔS_i. As a result, the *improved objective function* (16) is designed as follows:

$$\min J_2 = \min \sum_{i=0}^{T-1} \{\dot{m}_{H_2}(i) + \alpha \cdot S_i \cdot \mathrm{Tanh}(\Delta S_i / \delta)\} \tag{16}$$

where δ is a coefficient greater than zero, which means that when the absolute value of ΔS_i is less than δ, $Tanh(\Delta S_i/\delta)$ is close to$\Delta S_i/\delta$. When $|\Delta S_i/\delta| \ge 10$, $|Tanh(\Delta S_i/\delta)|$ will be close to 1.

To evaluate the performance of different objective functions, a unified evaluation function is needed. Since the change of SOC in the process is dynamic and the battery will store or release energy, the concept of equivalent fuel consumption is introduced and used as the indicator for evaluating fuel consumption optimization. we can convert the energy charge or discharge in the batteries into the corresponding equivalent hydrogen mass. By combining with (15), a unified evaluation function can be obtained：

$$\begin{aligned} E &= \lambda[SOC_{bat}(T) - SOC_{bat}(0)] \\ &\quad + \sum_{i=0}^{T-1} \{\dot{m}_{H_2}(i) + \alpha[SOC_{bat}(i) - SOC_{ref}(i)]^2\} \\ \lambda &= \frac{Q_{bat} \cdot \bar{V}_{bat}}{\bar{Eff}_{bat} \cdot \bar{Eff}_{fc}} \cdot \frac{M_{H_2}}{\Delta H} \end{aligned} \tag{17}$$

where λ is the equivalent fuel factor, Q_{bat} is the capacity of batteries, V_{bat} is the voltage of Batteries, Eff_{fc} is the average efficiency of the fuel cell system, Eff_{bat} is the average efficiency of batteries system. $M_{H_2} = 2g/mol$ is the molar mass of hydrogen. $\Delta H = 284kJ/mol$ is the high calorific value of hydrogen.

With the evaluation function, we can test the performance of different objective functions under different optimization methods.

B. *Constraints*

Considering the practical application of energy management optimization problems in FCHEV, constraints are required for both the system state variables $\boldsymbol{x}(t)$ and action variable $\boldsymbol{u}(t)$. In

the paper, the states of the studied system are chosen as the SOC of batteries $SOC_{bat}(t)$ and the required power of the vehicle $P_{veh}(t)$.Then, the safety working range of the batteries is set as $SOC_{bat}(t) \in [20\%, 90\%]$. The allowed power demand of the vehicle is set as $P_{veh}(t) \in [-100kW, 100kW]$. In addition, the action variable is chosen as the output power of the fuel cell system $P_{fc}(t)$.The designed maximum power of the fuel cell is 117kW. Taking into account the efficiency loss of the auxiliary system and the DC/DC converter of fuel cells system, the range of the control variable is set to $P_{fc}(t) \in [0,100kW]$. The optimal control of the EMS problem will be carried out under these constraints.

C. *Reinforcement Learning (RL) based EMS*

An RL-based EMS is studied in the paper. As shown in Fig. 4, a general RL controller observes the state s_t and the reward r_t from the environment, then chooses an action a_t with the learned policy at moment *t*. As a result, the environment will give feedbacks on rewards r_{t+1} and the next state s_{t+1} information. In the study, state variables are composed by the vehicle driving power and the battery SOC, as $[P_{veh}, SOC_{bat}]$. The power command for fuel cells system $[P_{fc}]$ is the action variable. The instantaneous reward r_t is defined as the opposite instantaneous cost in objective functions J_0, J_1 and J_2.

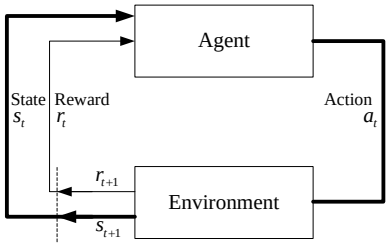


Fig.4. Reinforcement Learning principle

RL seeks the optimal policy to maximize the expected discounted cumulated reward, as

$$V^*(s) = \max_{\pi} E\left[\sum_{0}^{\infty} \gamma^t r_t\right] \tag{18}$$

where $\gamma \in (0,1)$ is the discount coefficient factor. For this, Q-value $Q(s_t, a_t)$ is introduced to express the expected total reward from state s_t and taking action a_t under one policy. The optimal Q-value $Q^*(s_t, a_t)$ represents the maximum total reward when the action a_t is taken on state s_t. $V^*(s)$ can be obtained from $Q^*(s_t, a_t)$ by taking the optimal action a_t, as:

$$V^*(s_t) = \max_{a_t} Q^*(s_t, a_t) \tag{19}$$

Thus, once $Q^*(s_t, a_t)$ is obtained, the optimal policy can be determined as:

$$\pi^*(s_t) = \arg\max_{a_t} Q^*(s_t, a_t) \tag{20}$$

$Q^*(s_t, a_t)$ can be expressed using Bellman optimality equation:

$$Q^*(s_t, a_t) = E[r_t + \gamma \max Q(s_{t+1}, a_{t+1}) \mid s_t, a_t] \tag{21}$$

According to the Bellman optimality equation, Q-Learning proposes an approach to calculate $Q^*(s_t, a_t)$ iteratively [Sutton book]. Q-learning algorithm used in the study is summarized in Table I. More details about Q-learning implementation can be found in [12].

TABLE I. THE PROCEDURE OF THE Q-LEARNING ALGORITHM

Q-Learning
Initialize $Q(s,a), \forall s \in \mathcal{S}, a \in \mathcal{A}(s)$, arbitrarily, and $Q(terminal,\cdot) = 0$
Repeat for each episode:
Reset the FCHEV environment with the initialize states S
Repeat for each step of the episode:
Choose A from S using policy derived from Q (e.g., $\varepsilon - greedy$)
Observe the reward R and next state S'
$Q(S,A) \leftarrow Q(S,A) + \alpha[R + \gamma max_a Q(S',a) - Q(S,A)]$
Update state $S \leftarrow S'$
Until S is terminal

IV. RESULTS AND DISCUSSION

A Python-based training and testing platform have been established for the proposed RL-based EMS. In this section, the results of the proposed EMS and the performance of different objective function settings are analyzed and discussed.

A. *Test driving cycles*

The proposed EMS will be tested using the driving cycle named *Urban Dynamometer Driving Schedule* (UDDS). The velocity and power of the specific FCHEV under the driving cycle are shown in Fig. 5.

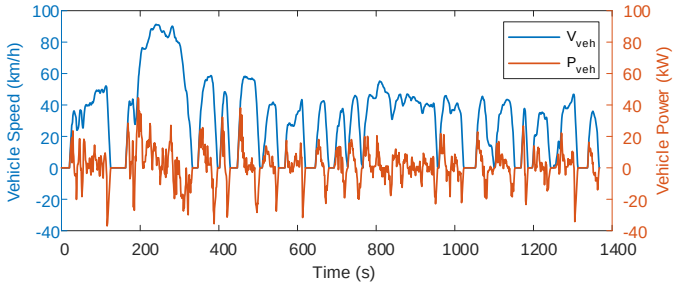


Fig.5. the velocity and power of the FCHEV under the driving cycle UDDS

B. *Dynamic programming results*

DP is deployed for the different objective functions to obtain EMS benchmarks. For DP implementation, SOC_{bat} and P_{fc} are considered respectively as state control variable. The discretization steps for SOC_{bat} and P_{fc} are respectively 0.1% and 1 kW. EMS time step is set to 1s. The details of DP implementation can be found in [1]. In the test, the values of factor α in the objective functions J_0, J_1, J_2 are set as 100, 1.44 and 2.5 separately, and factor δ in function J_2 is 0.001. The initial value of the battery's SOC and the final state value are both set to 50%.

After implementing DP, Fig.7 shows the specific values of the output power of fuel cells, batteries, and the required power of the vehicle for the concerned driving cycle. During the test process with 3 different objective functions, the SOC termination value is close to the initial SOC value in each test. The quantitative results are summarized in Table II. It is seen that using J_0, DP-based EMS can obtain the smallest fuel loss. Further, the optimal control sequences generated using J_0, J_1, J_2

are tested using evaluation function defined in (17). The evaluation values of J_0, J_1, J_2 results are 33.49, 38.53 and 35.91. The objective function J_0 still have the smallest evaluation value.

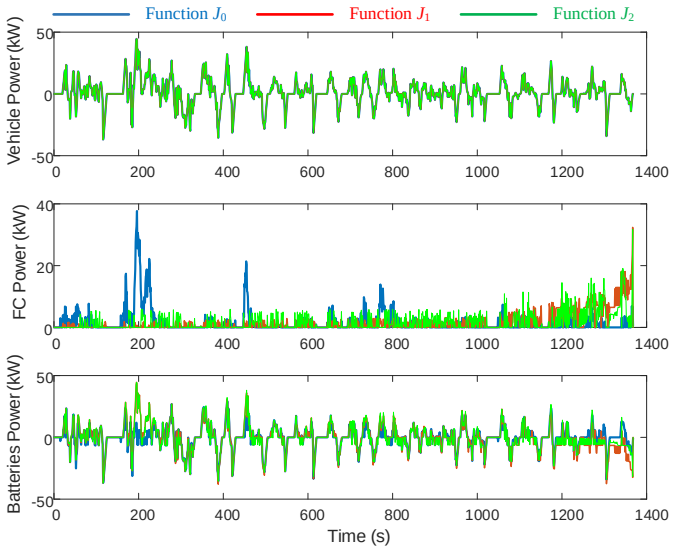


Fig.6. Power allocation of the FCHEV with DP based EMS

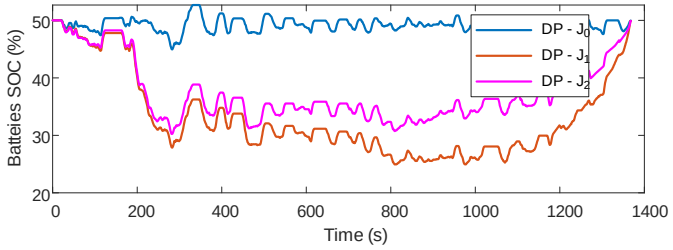


Fig.7. The SOC trajectory of batteries with DP based EMS

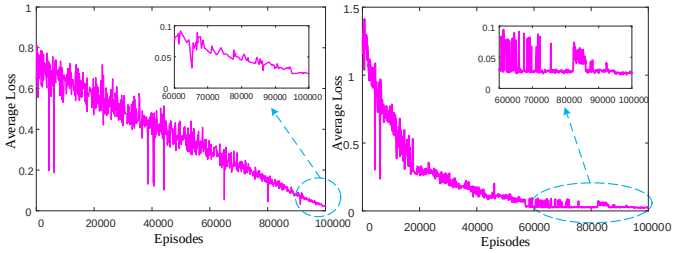
TABLE II. DYNAMIC PROGRAMMING TEST RESULTS

Objective Functions	*Hydrogen Consumption (g)*	*ΔSOC (%)*	*Cumulative Loss*	*Evaluation Value*
Function J_0	33.49	-0.05%	56.86	33.49
Function J_1	35.15	-0.10%	35.16	38.53
Function J_2	34.14	-0.10%	2.87	35.91

C. *Q-Learning based EMS Test*

In this section, the implementation of RL-based EMS is talked about. In the Q-Learning setting, the declining exploration rate ϵ from 1.0 to 0.001 is used. The learning rate α of Q-Learning is set as 0.01, and the decay rate γ of Q-Learning is set as 0.99. State and control variables are discretized as follows. the step sizes of P_{veh}, SOC_{bat} and $P_{fc}(t)$ are respectively 1 kW, 10% and 1 kW. Under the setting, 100,000 episodes of Q-Learning are carried out. In the test, the parameters of the objective functions are consistent with those in the DP test.

The average step losses of J_1 and J_2 setting during training processes are shown in fig.8. It can be seen that the average loss using J_1 demonstrates converging trend, while the loss for J_2 tend to be converged after 72,000 episodes. However, the classic objective function J0 still fails to converge after 100,000 episodes of training. This shows that using modified objective functions J_1 and J_2 can effectively improve the training efficiency of Q-Learning and increase the convergence rate.



(1) Objective function J_1 (2) Objective function J_2

Fig.8. the average loss of Q-learning based EMS for FCHEV

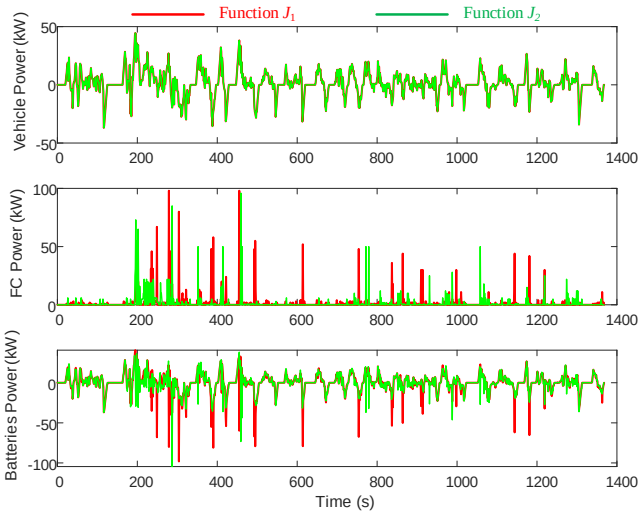


Fig.9.The test result of Q-Learning based EMS for FCHEV

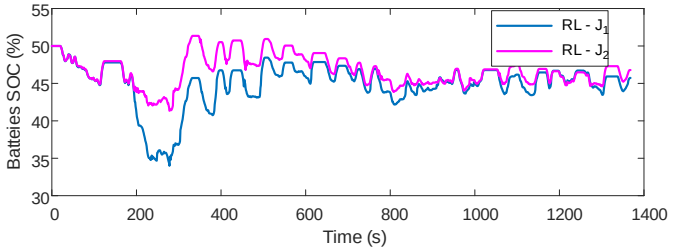


Fig.10. The SOC trajectory of batteries with RL based EMS

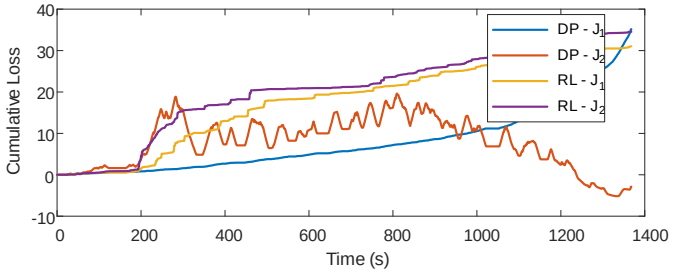


Fig.11. The cumulative loss graphs of EMS test processes

After the Q-learning process, EMSs based on the learned Q tables are tested. During the test process, the agent will choose the action at each step according to (20). The specific values of the output power of fuel cells, batteries and the required power

of the FCHEV are shown in fig.9. And the SOC trajectory of batteries with RL-based EMS is shown in fig.10. The cumulative loss graphs of EMS test processes based on DP and Q-learning under different objective functions are shown in fig.11. It can be seen when the SOC is far from the preset reference trajectory, more powerful actions will be taken to make the SOC return to the normal trajectory as soon as possible.

As shown in Table III, the values using J_1 and J_2 are respectively 31.2 and 33.2, which are both smaller than DP results. The hydrogen consumptions are 31.0 g and 29.4 g with function J_1 and J_2 which are also smaller than DP results. The reason is that the final state constrain is released for RL-based EMS. The evaluation function values using J_1 and J_2 are respectively 31.34 and 29.54 which are also better than DP results. The evaluation value of the function J_2 is 5.74% lower than that of function J_1, which has better performance. Therefore, by using the RL-based EMS and the proposed objective functions, satisfactory EMS results, in terms of consumption reduction and battery charge maintenance, can be achieved.

TABLE III. Q-LEARNING TEST RESULTS

Objective Functions	***Hydrogen Consumption (g)***	**ΔSOC (%)**	***Cumulative Loss***	***Evaluation Value***
Function J_1	31.0	-3.21%	31.2	31.34
Function J_2	29.4	-3.21%	33.2	29.54

V. CONCLUSION

In the paper, a reinforcement learning (RL) based energy management strategy is studied for fuel cell hybrid electric vehicles. In the strategy, several objective functions are formulated aiming at reducing hydrogen consumption and maintaining battery SOC. The proposed RL-based EMS has been tested and compared with the benchmarks provided by DP. The results show that using the proposed objective functions for RL-based EMS, the learning efficiency can be increased significantly. The quasi-optimal EMS performance can also be achieved. Ongoing work is focused on the theoretical investigation of the effects of the objective function on the learning process.

Acknowledgements

This work has been partially funded by the National Research Agency distinguished young scholar program (ANR JCJC) of France through the project DEAL (ANR-20-CE05-0016-01), and the CNRS Energy unit (Cellule Energie) through the project GIALE.

REFERENCES

[1] S. Onori, L. Serrao, and G. Rizzoni, Hybrid electric vehicles: Energy management strategies, no. 9781447167792. 2016.

[2] Y. Shen, P. Cui, X. Wang, X. Han, and Y. X. Wang, "Variable structure battery-based fuel cell hybrid power system and its incremental fuzzy logic energy management strategy," Int. J. Hydrogen Energy, vol. 45, no. 21, pp. 12130–12142, 2020, doi: 10.1016/j.ijhydene.2020.02.083.

[3] D. Phan, A. Bab-Hadiashar, M. Fayyazi, R. Hoseinnezhad, R. N. Jazar, and H. Khayyam, "Interval Type 2 Fuzzy Logic Control for Energy Management of Hybrid Electric Autonomous Vehicles," IEEE Trans. Intell. Veh., vol. 6, no. 2, pp. 210–220, 2021, doi: 10.1109/TIV.2020.3011954.

[4] C. Liu, Y. Wang, L. Wang, and Z. Chen, "Load-adaptive real-time energy management strategy for battery/ultracapacitor hybrid energy storage system using dynamic programming optimization," *J. Power Sources*, vol. 438, no. August, p. 227024, 2019, doi: 10.1016/j.jpowsour.2019.227024.

[5] X. Li, Y. Wang, D. Yang, and Z. Chen, "Adaptive energy management strategy for fuel cell/battery hybrid vehicles using Pontryagin's Minimal Principle," *J. Power Sources*, vol. 440, no. September, p. 227105, 2019, doi: 10.1016/j.jpowsour.2019.227105.

[6] Y. Li and X. Jiao, "Energy management strategy for hybrid electric vehicles based on adaptive equivalent consumption minimization strategy and mode switching with variable thresholds," *Sci. Prog.*, vol. 103, no. 1, pp. 1–20, 2020, doi: 10.1177/0036850419874992.

[7] H. Marefat, M. Jalalmaab, and N. L. Azad, "Energy management of battery electric vehicles hybridized with supercapacitor using stochastic dynamic programming," *SICE ISCS 2018 - 2018 SICE Int. Symp. Control Syst.*, vol. 2018-January, no. c, pp. 199–205, 2018, doi: 10.23919/SICEISCS.2018.8330176.

[8] H. He, S. Quan, F. Sun, Y. X. Wang, and Y. X. Wang, "Model predictive control with lifetime constraints based energy management strategy for proton exchange membrane fuel cell hybrid power systems," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 9012–9023, 2020, doi: 10.1109/TIE.2020.2977574.

[9] R. Zhang, J. Tao, and H. Zhou, "Fuzzy optimal energy management for fuel cell and supercapacitor systems using neural network based driving pattern recognition," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 1, pp. 45–57, 2019, doi: 10.1109/TFUZZ.2018.2856086.

[10] G. Zhang *et al.*, "Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach," *Energy Convers. Manag.*, vol. 227, no. November 2020, p. 113608, 2021, doi: 10.1016/j.enconman.2020.113608.

[11] R. C. Hsu, S. M. Chen, W. Y. Chen, and C. T. Liu, "A Reinforcement Learning Based Dynamic Power Management for Fuel Cell Hybrid Electric Vehicle," *Proc. - 2016 Jt. 8th Int. Conf. Soft Comput. Intell. Syst. 2016 17th Int. Symp. Adv. Intell. Syst. SCIS-ISIS 2016*, pp. 460–464, 2016, doi: 10.1109/SCIS-ISIS.2016.0104.

[12] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction second edition*, vol. 9, no. 5. 2018.