

An Intelligent Power and Energy Management System for Fuel Cell/Battery Hybrid Electric Vehicle using Reinforcement Learning

Namireddy Praveen Reddy, David Padeloup, Mehdi Karbalaye Zadeh, Roger Skjetne

Norwegian University of Science and Technology

namireddy.p.reddy@ntnu.no, dfpasdel@stud.ntnu.no, mehdi.zadeh@ntnu.no, roger.skjetne@ntnu.no

Abstract—Hybrid electric vehicles powered by fuel cells and batteries have attracted significant attention as they have the potential to eliminate emissions from the transport sector. However, fuel cells and batteries have several operational challenges, which require a power and energy management system (PEMS) to achieve optimal performance. Most of the existing PEMS methods are based on either predefined rules or prediction that are not adaptive to real-time driving conditions and may give solutions that are far from the actual optimal solution for a new drive cycle. Therefore, in this paper, an intelligent PEMS using reinforcement learning is presented, that can autonomously learn the optimal policy in real time through interaction with the onboard hybrid power system. This PEMS is implemented and tested on the simulation model of the onboard hybrid power system. The propulsion load is represented by the new European drive cycle. The results indicate that the PEMS algorithm is able to improve the lifetime of batteries and efficiency of the power system through minimizing the variation of the state of charge of battery.

Keywords—Fuel cell, Hybrid electric vehicle, Power and energy management system, Reinforcement learning, Intelligent systems, Onboard DC power systems

I. INTRODUCTION

Transportation of passengers and goods accounts for about a quarter of world's energy consumption and contributes to more than 20% of global emissions according to the study of the intergovernmental panel on climate change (AR5'2014, IPCC). The transport sector is the fastest growing contributor to the world's energy consumption and global emissions. The main drivers of this growth are land-based transport, mainly light-duty vehicles such as passenger cars and freight transport. Currently, conventional diesel engines are used in most of the transport applications, which produces a significant amount of emissions. This emissions lead to air pollution that causes several diseases and reduces the quality of life for residents in urban areas. Therefore, it is necessary to investigate the advanced technologies that enable optimal usage of onboard emission-free energy systems for transport applications.

In the present market, the potential options for light-duty vehicles are hybrid electric vehicles (HEVs), plugin hybrid electric vehicles (PHEVs) and electric vehicles powered by batteries (BEVs). HEVs and PHEVs powered by conventional diesel engines and batteries merely reduce the emissions, but cannot eliminate completely. Moreover, BEVs have limitations such as long charging times, limited driving distance per charge. FCHEVs powered by fuel cells and batteries offer emission-free operation while overcoming the limitations on driving distance per charge and long

charging times. Therefore, FCHEVs have gained significant attention in recent years.

Fuel cells and batteries have several operational challenges, which directly impacts their lifetime and reliability. Therefore, for onboard hybrid power system powered by fuel cells and batteries, a power and energy management system (PEMS) is required to achieve the optimal performance. The existing research studied and developed several types of PEMS for transport applications. Sulaiman et al. [1] present a critical review of different energy management strategies for FCHEVs. Li et al. [2] present a review of EMS objectives and strategies for FCHEVs. There are several types of energy management strategies exist such as the state-based EMS [3] [4] [5], rule-based fuzzy logic strategy [6], charge depleting and charge sustaining (CDCS) strategy [7], wavelet transform based strategy [8], variable frequency control techniques [9], classical PI and PID strategies [17], equivalent consumption minimization strategy (ECMS) [10] [11] [12] [13], Potryagin's minimum principle (PMP) [14] [15], model predictive control (MPC) [16] [17], stochastic dynamic programming [18], and adaptive optimal control [19].

Most of the existing PEMS methods are based on either predefined rules or known drive cycle or predicted driving conditions. These methods are not adaptive to real-time system conditions and give solutions far away from the actual optimal solution for a new driving condition. Therefore, it is necessary to implement a PEMS that can learn in real time. For this reason, learning-based PEMS methods are gaining interest among researchers. Within machine learning, reinforcement learning (RL) has attracted a lot of attention in recent years by beating world champion human players in Chess and Go games. Some advantages of RL over classical optimal control techniques are; it is model-free, autonomously learns the optimal policy in real time, digital twins can be used to train the algorithm. Q learning is a powerful reinforcement learning approach and particularly suitable for applications where the system model is unknown or not accurate or continuously changing. For example, the onboard battery pack represents such a system [20]. The characteristics of battery systems keep changing as batteries go through numerous charging/discharging iterations. Q-learning can continuously learn from the observed experiences and automatically adjust to the evolving dynamics. Hsu et al. [21] applied Q-learning on the fuel cell hybrid electric vehicles to minimize fuel cell consumption. Li et al. [22] implemented multi-agent reinforcement learning for optimal control of microgrid in grid-connected

mode. Some more implementations of reinforcement learning for energy management could be found in [23], [24], [25], [26], [27], [28].

This paper presents a PEMS based on classical Q learning for FCHEVs. This research work includes 1) Building the simulation model of the onboard hybrid power system in MATLAB/Simulink, 2) Developing a classical Q learning based PEMS, which can autonomously learn the optimal policy in real time through interaction with simulation model of the onboard hybrid power system, 3) Training the Q learning algorithm. The simulation results indicate that the learning process is capable of converging to the optimal control behavior. Furthermore, the Q learning based PEMS is able to improve the lifetime of batteries and minimize the power losses through minimizing the variation of the state of charge of battery.

The rest of the paper is organized as; Section-I presents the introduction, motivation, literature review, contribution, and structure of the paper. Section-II presents a description of the onboard hybrid power system. Section-III presents the objectives considered for this work. Section-IV presents the implementation of Q learning based PEMS including discretized state space, discretized action space, reward function, and controller, etc., Section-V presents the simulation results and section-VI presents the conclusion.

II. ONBOARD HYBRID POWER SYSTEM

The schematic of the onboard hybrid power system is shown in Fig. 1. It is a DC power system powered by fuel cell and battery as power sources. Fuel cell system is connected to a DC bus through unidirectional DC/DC boost converter. The battery pack is connected to DC bus through a bidirectional DC/DC converter. The electric motor load is connected to the DC bus through load converter. The propulsion load is represented by the New European Drive Cycle (NEDC). A simple model for all the components is used for this work, as the focus is on verifying the effectiveness of Q learning based PEMS.

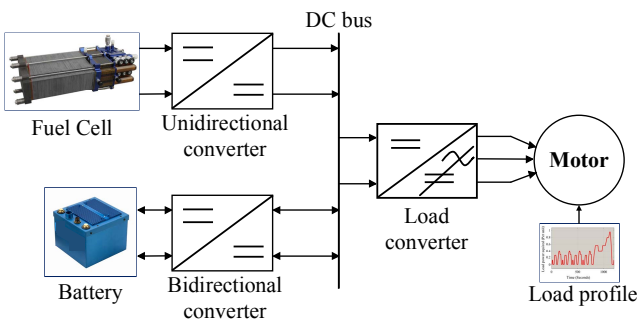


Fig. 1. Schematic of onboard hybrid power system

In this work, Proton Exchange Membrane Fuel Cell (PEMFC) is used, as it is the most suitable fuel cell type for transport applications. High price and low lifetime are the main challenges that are hindering the widespread commercialization of fuel cells. The main cause of aging and degradation in fuel cells is fuel starvation. During fast load changes, the risk of fuel starvation is high due to the inertia of actuators and valves in the fuel delivery system. Thus, fuel starvation limits the dynamics of fuel cell stack. Therefore, batteries are usually combined with fuel cells to

compensate for the slow dynamics of the fuel cell. PEMS decides the power to be produced by fuel cell, which sets the operating conditions for the components of fuel cell, thus affecting the aging and degradation of components [29] [30] [31].

For this work, a simple model of the fuel cell shown in Fig. 2 is used. Rate limiter represents the slow dynamics by limiting the rate of change in power and saturation sets the limits for minimum and maximum power.

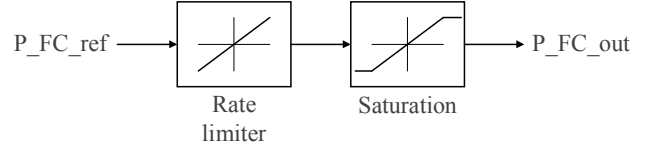


Fig. 2. Fuel cell model

The coulomb counting or current integration method is used to represent the lithium-ion battery model,

$$SOC = SOC_{initial} + \frac{1}{C_{rated}} \int I_{Battery} dt \quad (1)$$

Where SOC is state of charge, C_{rated} is rated capacity, $I_{Battery}$ is battery current.

All the three power electronic converters are represented as ideal and lossless i.e. 100% efficient with respect to power. DC bus is represented using the following equation,

$$P_{Load} = P_{FC} + P_{Battery} \quad (2)$$

III. LEARNING BASED POWER AND ENERGY MANAGEMENT SYSTEM

There are two alternative ways to supply load power as shown in Fig. 3. One way is to supply load power directly from fuel cell. The other is to supply through discharging the battery. When battery is discharged to supply the load power, it has to be charged by fuel cell. In the second case, as shown in Fig. 3, there are more components involved compared to first option, which implies more component power losses. Therefore, the component power losses can be minimized by reducing the charging/discharging cycles of the battery. This could be achieved through minimum variation of SOC. Furthermore, the lifetime of batteries is improved through minimum variation of SOC. In Q learning based PEMS, the reward function is formulated to achieve maximum reward for minimum variation of SOC.

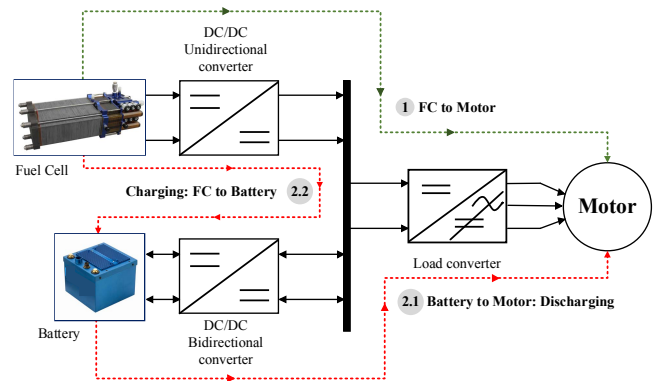


Fig. 3. Alternative ways of supplying the load power

The Q learning based PEMS algorithm is coded as a MATLAB script, autonomously learns online through

interaction with the onboard hybrid power system modeled in Simulink. The control framework for the onboard hybrid power system is shown in Fig. 4. Where, P_{FC} is power delivered by fuel cell, $P_{Battery}$ is the power delivered or absorbed by battery, $SOC_{Battery}$ is state of charge of battery, V_{FC} is fuel cell voltage, $V_{Battery}$ is battery voltage, V_{DC} is DC bus voltage, i_{FC} is fuel cell current, $i_{Battery}$ is battery current, P_{FC-ref} is fuel cell power reference, $P_{Battery-ref}$ is battery power reference, i_{FC-ref} is fuel cell current reference, $i_{Battery-ref}$ is battery current reference, u_{FC} is control reference to DC/DC unidirectional converter, $u_{Battery}$ is control reference to DC/DC bidirectional converter, t is the current time step, $t+1$ is the next time step, ϵ is the greedy policy, α is learning rate, γ is discount factor.

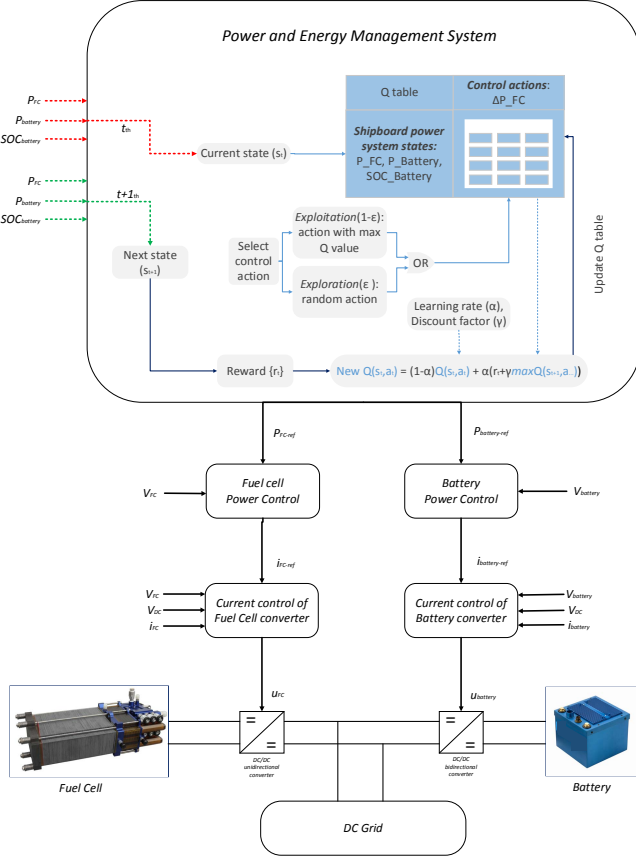


Fig. 4. The control framework for FCHEV

The Q learning based PEMS is the highest level of control, provides the power reference to main power source and energy storage, in this case, fuel cells and battery respectively. Fuel cell as the main source of power responsible for supplying the load power and charging the battery. The battery power is used during fast load changes to compensate for the slow dynamics of fuel cell. Generally, the source converter with faster dynamics is used for the DC bus voltage regulation; in this case, it is battery converter. Then, the fuel cell converter is used for maintaining the state of charge in the battery. As a result, there are two controlled variables namely DC bus voltage V_{DC} , battery state of charge $SOC_{Battery}$ or battery voltage $V_{battery}$. These are controlled by control variables namely battery current reference $i_{battery-ref}$ and fuel cell current reference i_{FC-ref} respectively. Hereafter, this work focuses on the highest level control i.e., Q learning based PEMS.

The key elements of Q learning based PEMS including discretized state space, discretized action space, reward function, and controller are presented below.

A. Discretized state space

In Q learning, the control action is determined directly by the state of onboard power system. For this work, the power delivered by fuel cell (P_{FC}), the power delivered or absorbed by battery ($P_{Battery}$) and state of charge of the battery ($SOC_{Battery}$) are selected to form the three-dimensional state space. All the values in this work are in per unit for the reusability of the algorithm. To apply classical Q learning, the parameters representing the onboard hybrid power system state need to be discretized into a finite number of levels. P_{FC} is discretized as [0.7 0.9]. $P_{Battery}$ is discretized as [-1 -0.8 -0.55 -0.25 0.25 0.55 0.8 1]. The negative sign indicates that the battery is charging and the positive sign indicates that the battery is discharging. $SOC_{Battery}$ is discretized as [0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1].

Table 1: Discretized state space

System states	P_{FC}	$P_{Battery}$	$SOC_{Battery}$
1	0.7	-1	0.4
.	.	.	.
.	.	.	.
.	.	.	.
104	0.9	1	1

B. Discretized action space

The main task of Q learning based PEMS is to distribute the load power required among fuel cells and batteries while maximizing the reward. Therefore, in this work, the change in fuel cell power (ΔP_{FC}) is selected to form the one-dimensional action space. In classical Q learning, similar to state space, action space also needs to be discretized. ΔP_{FC} is discretized as [0, +0.15, -0.15], where 0 represents remains same, +0.15 represents increase by 0.15 per unit and -0.15 represents decrease by 0.15 per unit. However, the final control signals given by Q learning based PEMS are fuel cell power reference (P_{FC-ref}) and battery power reference ($P_{Battery-ref}$). They are calculated as,

$$P_{FC-ref} = P_{FC}(t) + \Delta P_{FC} \text{ and } P_{Battery-ref} = P_{Load} - P_{FC-ref}$$

C. Reward function

For this work, the main objective is minimizing the variation of $SOC_{Battery}$ to improve battery lifetime and minimize component power losses. $SOC_{Battery}$ varies from 0.4 per unit to 1 per unit. Therefore, in this work, the reward is maximized when $SOC_{Battery}$ is closer to the mean value of 0.7 per unit. The reward function is,

$$\text{Reward}(t) = r_t = \frac{1}{2} + \frac{m}{2} \left\{ 0.5 - \left| \frac{SOC(t+1) - 0.7}{0.3} \right| \right\}$$

Where SOC (t+1) is state of charge of the battery during next time step, m is a multiplier that depends on the SOC (t+1) value and the control action a_t ,

$$-1 \leq \text{Reward}(t) \leq 1$$

If $[0.65 \leq \text{SOC}(t+1) \leq 0.75]$ or
 $[\text{SOC}(t+1) \leq 0.55 \ \& \ \Delta P_{FC} \leq 0]$ or
 $[\text{SOC}(t+1) \geq 0.85 \ \& \ \Delta P_{FC} \geq 0]$
m=2
else **m=1** end

The value of reward belongs to $[-1 \ 1]$ to comply with per unit system, a positive value indicates good reward and a negative value indicates bad reward. When SOC (t+1) belongs to $[0.65 \ 0.75]$, the value of m is 2 to maximize the positive reward, as SOC (t+1) is very close to the desired value of 0.7 per unit. When SOC (t+1) is lower than 0.55 and fuel cell power decreases, which is not a desirable scenario, the value of m is 2 to maximize the negative reward. Similarly, when SOC (t+1) is higher than 0.85 and fuel cell power increases, the value of m is 2 to maximize the negative reward. In all other cases, the value of m is 1.

D. Controller

In Q learning, the controller is essentially Q table, which has Q values for the given state of system and control action represented as $Q(s_t, a_t)$. $Q(s_t, a_t)$ is the expected cumulative discounted reward when the action a_t is performed by the controller on onboard hybrid power system in state s_t .

There are two ways to select the control action; exploitation and exploration, which depends on greedy policy $\epsilon \in [0 \ 1]$. During exploitation, for the given state, the control action with maximum Q value is chosen. Whereas, during exploration, a random control action is chosen. The probability of exploitation is $1-\epsilon$ and the probability of exploration is ϵ . It is desirable to do more exploration during the initial stages of training and more exploitation once the training is sufficient. Therefore, the value of ϵ is kept high in the beginning and decayed as the controller gains experience.

The controller performs an action a_t on onboard hybrid power system in state s_t . As a result of action a_t , the onboard hybrid power system transitions from current state s_t to next state s_{t+1} . Based on the SOC (t+1), the controller receives a reward (r_t), as explained in the previous section. Then, the Q value is updated complying with Bellman's equation,

$$Q^{\text{new}}(s_t, a_t) = (1 - \alpha)Q(s_t, a_t) + \alpha\{r_t + \gamma \max_a Q(s_{t+1}, a)\}$$

Where, $Q^{\text{new}}(s_t, a_t)$ is the new value of Q, $Q(s_t, a_t)$ is old value of Q, α is learning rate, γ is discount factor, r_t is reward, $\max Q(s_{t+1}, a)$ is the estimate of optimal future value. Since Q-learning is an iterative algorithm, it requires an initial condition. All the Q values are initialized to zeros in order to predict systems behavior better compared to the arbitrary initial condition [32].

Table 2: Initial Q table

Q Table		Control actions (ΔP_{FC})		
		Unchanged	Increase	Decrease
System states	1	0	0	0

	104	0	0	0

IV. SIMULATION RESULTS

The simulations in this work are focused on tuning different parameters of Q learning algorithm to achieve the convergence towards the objective of minimizing the variation of $\text{SOC}_{\text{Battery}}$. For this purpose, the training of algorithm is divided into 1000 episodes, where each training episode has a duration of 2000 seconds. Furthermore, each episode is divided into 500 iterations, where each iteration has a time step of 4 seconds. The simulation parameters used in the Q learning based PEMS algorithm are tabulated in Table 3. Simulation results obtained during the beginning of training and towards the end of training are presented below. The results presented are for three different initial values of $\text{SOC}_{\text{Battery}}$ including a minimum value of 0.3 per unit, the desired value of 0.7 per unit and the maximum possible value of 1 per unit. The simulation results presented for each episode are divided into three parts namely (a), (b), and (c) that show different parameters of the onboard hybrid power system and the controller. Part (a) of the episodic parameters shows NEDC load profile, the power delivered by fuel cell (P_{FC}) and power delivered or absorbed by battery (P_{Battery}). Part (b) of the episodic parameters shows the state of charge of battery (SOC) and whether the nature of controller action is exploitation or exploration. The presence of orange bar indicates that the nature of controller action is exploitation and the absence of orange bar shows that the nature of controller action is exploration. In addition, part (b) also shows two lines corresponding to the SOC value of 0.6 per unit and 0.8 per unit, which are closer to the desired value of 0.7 per unit. These lines are included in the plot to enhance the visibility of convergence. Part (c) shows the per unit value of reward obtained after each iteration of the episode.

Table 3: Parameters used in the Q learning algorithm

Parameters	Value
Epsilon (ϵ)	1
Epsilon decay ($\Delta\epsilon$)	0.99999
Discount factor (γ)	0.9
Learning rate (α)	1

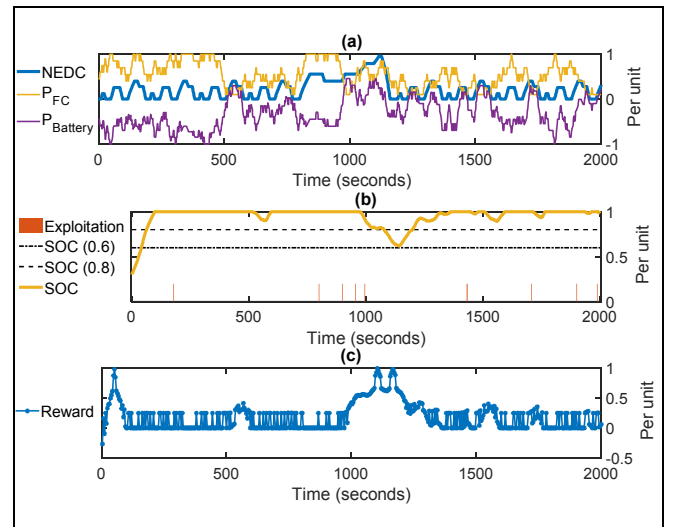


Fig. 5. Simulation parameters during episode-1 for $\text{SOC}_{\text{initial}} = 0.3$ per unit

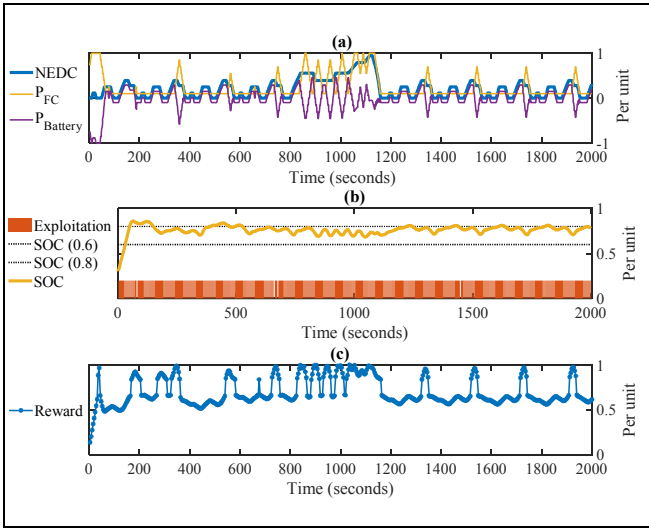


Fig. 6. Simulation parameters during episode-998 for $SOC_{initial} = 0.3$ per unit

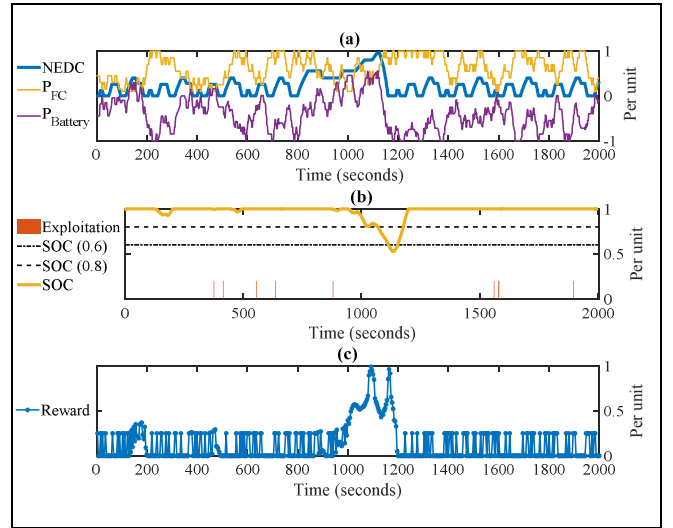


Fig. 9. Simulation parameters during episode-3 for $SOC_{initial} = 1$ per unit

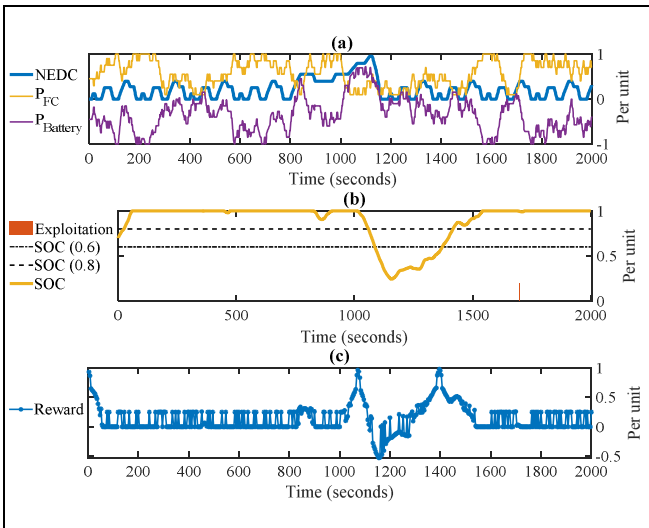


Fig. 7. Simulation parameters during episode-2 for $SOC_{initial} = 0.7$ per unit

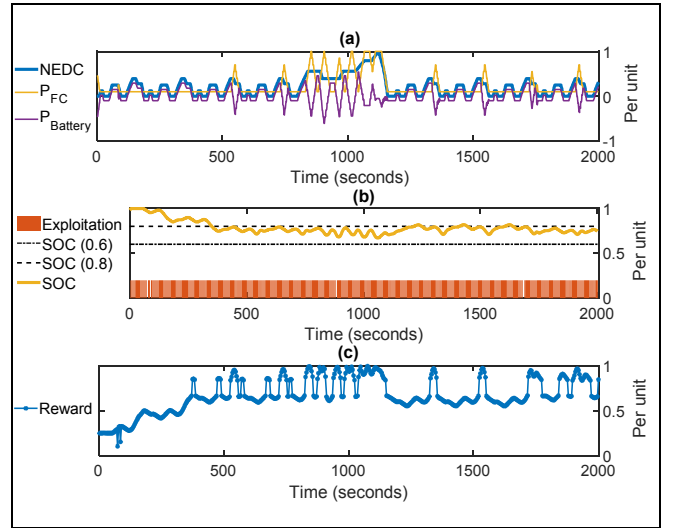


Fig. 10. Simulation parameters during episode-1000 for $SOC_{initial} = 1$ per unit

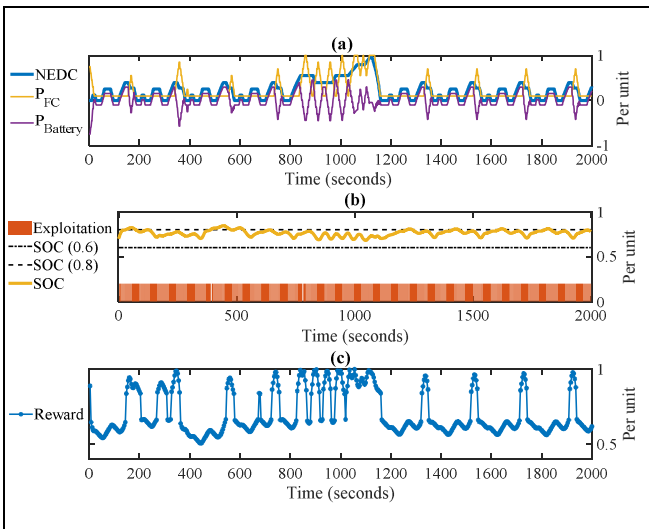


Fig. 8. Simulation parameters during episode-999 for $SOC_{initial} = 0.7$ per unit

From the simulation results, it can be observed that, during the beginning of training, the controller actions are dominated by exploration and randomly selected. Therefore, the $SOC_{Battery}$ and reward are not converging towards the objective of algorithm. Whereas, towards the end of the training, the controller actions are dominated by exploitation and selected to maximize the reward. Therefore, the $SOC_{Battery}$ and reward are converging towards the objective of algorithm. These observations are made irrespective of the initial value of $SOC_{Battery}$ for episode. However, the value of $SOC_{Battery}$ is mostly above the desired value of 0.7 per unit. This is due to the fact that the charging of the battery can be controlled by the algorithm, whereas the discharging of battery is based on the load profile.

V. CONCLUSION

In this paper, an intelligent PEMS based on classical reinforcement learning (Q-learning) has been developed for FCHEV powered by fuel cells and batteries. The Q learning based PEMS algorithm is coded as a MATLAB script and

tested on the simulation model of the onboard hybrid power system. The main objective of the developed PEMS is to minimize the variation of $SOC_{Battery}$ in order to improve the lifetime of battery and reduce the component power losses. The Q learning algorithm is able to autonomously learn in real time through interaction with the simulation model of the onboard hybrid power system. Simulation results indicate that the algorithm can converge towards the main objective of minimizing the variation of $SOC_{Battery}$ around 0.7 per unit. Therefore, the algorithm can be further modified to achieve different objectives such as minimizing fuel consumption and improving the lifetime of components.

REFERENCES

- [1] N. Sulaiman, M. A. Hannan, A. Mohamed, P. J. Ker, E. H. Majlan, and W. R. Wan Daud, "Optimization of energy management system for fuel-cell hybrid electric vehicles: Issues and recommendations," *Appl. Energy*, vol. 228, pp. 2061–2079, Oct. 2018.
- [2] H. Li, A. Ravey, A. N'Diaye, and A. Djerdir, "A Review of Energy Management Strategy for Fuel Cell Hybrid Electric Vehicle," in *2017 IEEE Vehicle Power and Propulsion Conference (VPPC)*, 2017, pp. 1–6.
- [3] J. Han, J.-F. Charpentier, and T. Tang, "An Energy Management System of a Fuel Cell/Battery Hybrid Boat," *Energies*, vol. 7, no. 5, pp. 2799–2820, May 2014.
- [4] M. K. Zadeh, L. Saublet, R. Gavagsaz-Ghoachani, B. Nahid-Mobarakeh, S. Pierfederici, and M. Molinas, "Energy management and stabilization of a hybrid DC microgrid for transportation applications," in *2016 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2016, pp. 3397–3402.
- [5] M. K. Zadeh, *Stability Analysis Methods and Tools for Power Electronics- Based DC Distribution Systems, Applicable to On-Board Electric Power Systems and Smart Microgrids*. NTNU, 2016.
- [6] S. N. Motapon, L. Dessaint, and K. Al-Haddad, "A Comparative Study of Energy Management Schemes for a Fuel-Cell Hybrid Emergency Power System of More-Electric Aircraft," *IEEE Trans. Ind. Electron.*, vol. 61, no. 3, pp. 1320–1334, Mar. 2014.
- [7] S. J. Moura, D. S. Callaway, H. K. Fathy, and J. L. Stein, "Tradeoffs between battery energy capacity and stochastic optimal power management in plug-in hybrid electric vehicles," *J. Power Sources*, vol. 195, no. 9, pp. 2979–2988, May 2010.
- [8] X. Zhang, C. C. Mi, A. Masrur, and D. Daniszewski, "Wavelet-transform-based power management of hybrid vehicles with multiple on-board energy sources including fuel cell, battery and ultracapacitor," *J. Power Sources*, vol. 185, no. 2, pp. 1533–1543, Dec. 2008.
- [9] F. Segura and J. M. Andújar, "Power management based on sliding control applied to fuel cell systems: A further step towards the hybrid control concept," *Appl. Energy*, vol. 99, pp. 213–225, Nov. 2012.
- [10] G. Paganelli, S. Delprat, T. M. Guerra, J. Rimaux, and J. J. Santin, "Equivalent consumption minimization strategy for parallel hybrid powertrains," in *Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No.02CH37367)*, 2002, vol. 4, pp. 2076–2081 vol.4.
- [11] J. Han, Y. Park, and Y. Park, "A novel updating method of equivalent factor in ECMS for prolonging the lifetime of battery in fuel cell hybrid electric vehicle," *IFAC Proc. Vol.*, vol. 45, no. 30, pp. 227–232, Jan. 2012.
- [12] J. Han, Y. Park, and D. Kum, "Optimal adaptation of equivalent factor of equivalent consumption minimization strategy for fuel cell hybrid electric vehicles under active state inequality constraints," *J. Power Sources*, vol. 267, pp. 491–502, Dec. 2014.
- [13] P. García, J. P. Torreglosa, L. M. Fernández, and F. Jurado, "Viability study of a FC-battery-SC tramway controlled by equivalent consumption minimization strategy," *Int. J. Hydrog. Energy*, vol. 37, no. 11, pp. 9368–9382, Jun. 2012.
- [14] C. Zheng, S. W. Cha, Y. Park, W. S. Lim, and G. Xu, "PMP-based power management strategy of fuel cell hybrid vehicles considering multi-objective optimization," *Int. J. Precis. Eng. Manuf.*, vol. 14, no. 5, pp. 845–853, May 2013.
- [15] C. H. Zheng, G. Q. Xu, Y. I. Park, W. S. Lim, and S. W. Cha, "Prolonging fuel cell stack lifetime based on Pontryagin's Minimum Principle in fuel cell hybrid vehicles and its economic influence evaluation," *J. Power Sources*, vol. 248, pp. 533–544, Feb. 2014.
- [16] T. Li, H. Liu, and D. Ding, "Predictive energy management of fuel cell supercapacitor hybrid construction equipment," *Energy*, vol. 149, pp. 718–729, Apr. 2018.
- [17] J. P. Torreglosa, P. García, L. M. Fernández, and F. Jurado, "Predictive Control for the Energy Management of a Fuel-Cell-Battery-Supercapacitor Tramway," *IEEE Trans. Ind. Inform.*, vol. 10, no. 1, pp. 276–285, Feb. 2014.
- [18] M.-J. Kim and H. Peng, "Power management and design optimization of fuel cell/battery hybrid vehicles," *J. Power Sources*, vol. 165, no. 2, pp. 819–832, Mar. 2007.
- [19] W.-S. Lin and C.-H. Zheng, "Energy management of a fuel cell/ultracapacitor hybrid power system using an adaptive optimal-control method," *J. Power Sources*, vol. 196, no. 6, pp. 3280–3289, Mar. 2011.
- [20] S. Mohan, Y. Kim, and A. G. Stefanopoulou, "Estimating the Power Capability of Li-ion Batteries Using Informationally Partitioned Estimators," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 5, pp. 1643–1654, Sep. 2016.
- [21] R. C. Hsu, S. Chen, W. Chen, and C. Liu, "A Reinforcement Learning Based Dynamic Power Management for Fuel Cell Hybrid Electric Vehicle," in *2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS)*, 2016, pp. 460–464.
- [22] F.-D. Li, M. Wu, Y. He, and X. Chen, "Optimal control in microgrid using multi-agent reinforcement learning," *ISA Trans.*, vol. 51, no. 6, pp. 743–751, Nov. 2012.
- [23] P. Kofinas, S. Doltsinis, A. I. Dounis, and G. A. Vouros, "A reinforcement learning approach for MPPT control method of photovoltaic sources," *Renew. Energy*, vol. 108, pp. 461–473, Aug. 2017.
- [24] R. Liessner, C. Schroer, A. Dietermann, and B. Bäker, "Deep Reinforcement Learning for Advanced Energy Management of Hybrid Electric Vehicles," presented at the International Conference on Agents and Artificial Intelligence, 2018, pp. 61–72.
- [25] M. Prauzek, N. R. A. Mourcet, J. Hlavica, and P. Musilek, "Q-Learning Algorithm for Energy Management in Solar Powered Embedded Monitoring Systems," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, 2018, pp. 1–7.
- [26] S. Kim and H. Lim, "Reinforcement Learning Based Energy Management Algorithm for Smart Energy Buildings," *Energies*, vol. 11, no. 8, p. 2010, Aug. 2018.
- [27] S. Etemad and N. Mozayani, "Using reinforcement learning to make smart energy storage sources in microgrid," in *2015 30th International Power System Conference (PSC)*, 2015, pp. 345–350.
- [28] C. Liu, "Optimal Power Management Based on Q-Learning and Neuro-Dynamic Programming for Plug-in Hybrid Electric Vehicles," PhD Dissertation, University of Michigan, Dearborn, Michigan, USA, 2018.
- [29] P. Thounthong and P. Sethakul, "Analysis of a Fuel Starvation Phenomenon of a PEM Fuel Cell," in *2007 Power Conversion Conference - Nagoya*, 2007, pp. 731–738.
- [30] M. A. Danzer, S. J. Wittmann, and E. P. Hofer, "Prevention of fuel cell starvation by model predictive control of pressure, excess ratio, and current," *J. Power Sources*, vol. 190, no. 1, pp. 86–91, May 2009.
- [31] F. A. de Bruijn, V. a. T. Dam, and G. J. M. Janssen, "Review: Durability and Degradation Issues of PEM Fuel Cell Components," *Fuel Cells*, vol. 8, no. 1, pp. 3–22, Feb. 2008.
- [32] H. Shteingart, T. Neiman, and Y. Loewenstein, "The role of first impression in operant learning," *J. Exp. Psychol. Gen.*, vol. 142, no. 2, pp. 476–488, May 2013.