IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE VEHICULAR TECHNOLOGY SOCIETY SECTION

Received March 2, 2021, accepted April 8, 2021, date of publication April 13, 2021, date of current version April 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3072903

Energy Management Strategy of Fuel Cell Electric Vehicles Using Model-Based Reinforcement Learning With Data-Driven Model Update

HEEYUN LEE[1], (Member, IEEE), AND SUK WON CHA[1,2], (Member, IEEE)
[1]Department of Mechanical Engineering, Seoul National University, Seoul 08826, Republic of Korea
[2]Institute of Advanced Machines and Design, Seoul National University, Seoul 08826, Republic of Korea

Corresponding author: Suk Won Cha (swcha@snu.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant by the Korean Government through Ministry of Science and ICT (MSIT) under Grant NRF-2019R1A4A1025848.

ABSTRACT Fuel cell electric vehicles use fuel cells as their main power source; the vehicle is driven by an electric motor, and have an electric battery as a secondary power source that stores regenerative braking energy and assists driving. To reduce the hydrogen fuel consumption by using these fuel cells and electric batteries efficiently, an energy management strategy is needed for the proper distribution of power among them. In this study, model-based reinforcement learning was utilized for energy management. For the optimal control of a fuel-cell electric vehicle, reinforcement learning is conducted using an internal vehicle powertrain model in the learning algorithm; initially, the model is completely unknown, but the model is learned with data from experiences as the learning process progresses. Then, reinforcement learning is conducted for the environment of the driving cycle profile to optimize the control policy. In this study, vehicle simulation was conducted using standard driving cycles, and the results showed that the learning process converged steadily and that the powertrain model was well learned. The simulated fuel consumption values show that the proposed algorithm reduces fuel consumption compared to the rule-based strategy by an average of 5.7%.

INDEX TERMS Fuel cell electric vehicles, model-based reinforcement learning, optimal control, power management, reinforcement learning.

I. INTRODUCTION

Recent developments in fuel-cell technology have attracted considerable attention owing to its advantages such as zero emission, which can augment fuel cell electric vehicle (FCEV) technology. Among the different types of fuel cells, such as polymer electrolyte membrane fuel cells (PEMFCs), direct methanol fuel cells, alkaline fuel cells, phosphoric acid fuel cells, and solid oxide fuel cells, PEMFCs are considered suitable for vehicle applications considering the operating temperatures, power density, efficiency, and cost. As they are powered by hydrogen instead of conventional fossil fuels, FCEVs do not produce harmful emissions and can be charged quickly similar to conventional internal combustion engine-based vehicles. Similar to electric vehicles (EVs), FCEVs also rely on electricity to drive vehicles using electric motors, but FCEVs generate electricity using PEMFCs. However, an electric battery is still an essential component in the FCEV powertrain to regenerate energy from braking so as to provide extra power as required and to avoid inefficient operation of the PEMFC system, such as idling and low speed. Therefore, in FCEVs, it is necessary to appropriately distribute the power required by the vehicle to two power sources, the PEMFC and electric battery, and this is called an energy management strategy (EMS) [1].

Numerous studies have been conducted on EMSs for FCEVs. Dynamic programming (DP) is a well-known strategy for finding a global optimal solution [2]–[5]. DP is based on the Bellman equation, which represents the optimization problem in a recursive form, and can solve the problem efficiently. However, the DP calculation is quite time consuming, making real-life application difficult; also, as the number of states increases, the complexity of calculation also increases rapidly owing to the so-called ''curse of dimensionality.''

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

Thus, in many studies, DP is used to determine the benchmark performance of the vehicle system, or provide a guideline for the rule-based controller. The equivalent consumption minimization strategy (ECMS) is another popular approach [6]–[9] in which the fossil fuel and electrical energy consumption is evaluated comprehensively using an equivalent factor comprising of equivalent values. In ECMS, the instantaneous cost minimization results in global optimization with a proper equivalent factor. However, ECMS requires an exact estimation of the equivalent factor for the subsequent driving cycle, and at the same time, mitigates the difficulty in prediction. Pontryagin's minimum principle (PMP)-based energy management strategy [10]–[14] can explain the theoretical background of the ECMS. In PMP, the fuel consumption of the engine and battery can be expressed by a Hamiltonian using the co-state value corresponding to the equivalent factor in the ECMS, and minimization of the Hamiltonian gives the equivalent consumption of ECMS, which should be always be minimized. However, in PMP, it is also necessary to estimate an initial co-state value, similar to ECMS, which is driving cycle dependent; thus, it also has the same difficulty. Model predictive control (MPC) has been studied in many studies [15]–[17], in which, based on optimization techniques, the MPC-based approach to various industrial problems has been successful, and accordingly, research using this approach is underway for the control problem of FCEVs.

The above approaches of DP, PMP, and MPC is mainly based on the exact modeling of the environment, which requires the modeling of the complicated powertrain system, especially in the case of DP, these difficulty of modelling of the system is called ''curse of modeling''. In addition, modeling is required not only for the vehicle system, but also for the driving cycle. The driving cycle information should be incorporated in the control strategy, but the exact forecasting of driving cycle information has its limitations. Recently, data-driven approaches based on machine learning have been developed. In particular, reinforcement learning (RL) can be applied to decision-making processes, such as the energy management of FCEVs. In RL, unlike in supervised or unsupervised learning, the agent interacts with the environment, and the control policy can be learned through this interaction; thus, the control policy is optimized based on the control action, state transition, and reward [18]. A few approaches have been examined for reinforcement learning in FCEVs. In [19], Q-learning was utilized for fuel cell/battery/ultra-capacitor hybrid electric vehicles. Here, a fuzzy filter was used to improve the performance of the RL algorithm. In [20], for FCEVs made of PEMFCs and Li-ion batteries, RL is utilized in which the battery state of charge(SOC) boundary is determined using a deep deterministic policy gradient (DDPG), and thermostat control is used within the SOC boundary. In [21], two-level RL was proposed for plug-in FCEVs, in which the global optimal solution was approximated using hierarchical RL, enabling speed predictions. These approaches are mainly based on model-free methods, and the optimal behavior of the control strategy is learned from experience. However, ensuring the convergence of the value function or control policy is a time-consuming process that requires data from many iterative simulations.

In this study, model-based reinforcement learning (MBRL) was examined for FCEVs. MBRL uses an internal model to approximate the environment and the control behavior can be learned through this model [22]–[24]. MBRL are more efficient than model-free approaches; in model-based learning, the optimal policy can be extracted from the calculation using the learned model, while in the model-free method, trial-and-error is required to determine the optimal control policy, which is a time-consuming process. In our previous studies, MBRL was utilized for the EMSs of parallel hybrid electric vehicles (HEVs) [25], [26] and in this study, MBRL was applied to the FCEVs. Unlike parallel HEVs studied previously, FCEVs have a series-type powertrain in which the fuel cell system can be operated independently regardless of the vehicle's speed, and the fuel cell system has characteristics that are completely different from the internal combustion engine. Therefore, the energy management strategy of FCEVs must also depend on these powertrain dynamics characteristics. Here, we divided the environment into the vehicle powertrain and driving cycle, and a model-based learning approach was utilized for the FCEV's powertrain using domain knowledge, while model-free learning is conducted for the environment of the driving cycle profile. The contributions of this study are as follows:

1) MBRL is applied to find the supervisory control policy of the energy management problem of FCEVs. In particular, by showing that MBRL can be successfully applied to FCEVs, we proved that MBRL can be applied to different vehicle powertrain systems.

2) In addition, we propose a data-driven model update method in MBRL, where MBRL starts the learning process with completely unknown model parameters, and the model is learned using the experience data alone. The learning performances of the model-based and model-free methods are compared, and the fuel economy performance of FCEVs using MBRL is compared with that of DP and rule-based control strategies to verify its effectiveness.

The remainder of this paper is organized as follows. In Section II, the modeling of vehicle longitudinal dynamics and the modeling of the powertrain for vehicle simulation is introduced. In Section III, an energy management strategy using RL is presented. In Section IV, the vehicle simulation results are presented, and finally, the conclusion is presented in Section V.

II. VEHICLE SIMULATION MODEL

In this study, the proposed algorithm was verified through a vehicle simulation. The concept of the powertrain structure for the vehicle simulation model is shown in Fig. 1. In the structure, the vehicle is driven using an electric motor, and the energy required to operate the electric motor is supplied by a PEMFC and an electric battery. In addition, there is

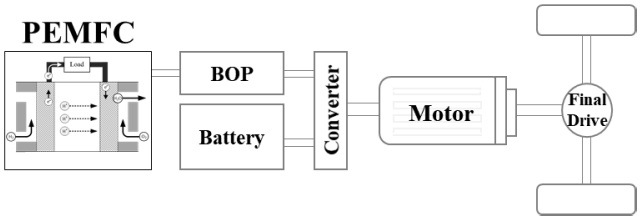

FIGURE 1. Vehicle simulation model.

the balance of plant (BOP), which is a subsystem component for operating the PEMFC, such as a compressor, heat exchanger, and humidifier, that requires additional power. In regenerative braking, the energy generated on braking is stored in the battery. Quasi-static modeling was used for the vehicle powertrain modelling, and only the longitudinal vehicle dynamics were considered. The detailed subsystem modeling is as follows.

First, the vehicle dynamics can be expressed as (1), considering only the longitudinal dynamics

$$\dot{v} = \frac{T_{whl}/R_{tire} - F_{brk} - F_{load}}{m_v}, \tag{1}$$

where v is the vehicle speed, m_v is the vehicle mass, T_{whl} is the wheel torque, R_{tire} is the tire radius, F_{brk} is the brake force, and F_{load} is the vehicle road load, which can be expressed as (2):

$$F_{load} = \frac{1}{2}\rho_a A_f C_d v^2 + C_r m_v g\cos\theta + m_v g\sin\theta, \tag{2}$$

where ρ_a is the air density, A_f is the vehicle front area, C_d is the vehicle drag coefficient, C_r is the vehicle rolling resistance coefficient, and θ is the slope. Then, the vehicle powertrain can be represented by:

$$T_{whl} = T_m \cdot \eta_{fd}^{sgn(T_m)} \cdot \gamma_{fd}, \tag{3}$$

$$\omega_m = \gamma_{fd} \cdot \omega_{whl}, \tag{4}$$

where T_m is the motor torque, η_{fd} is the efficiency of the final drive ,γ_{fd} is the gear ratio of the final drive, ω_m is the motor speed, and ω_{whl} is the wheel speed, which is $\omega_{whl} = v/R_{tire}$. The efficiency of the final drive is assumed to be constant at 0.98. The electric power demand P_{elec} can then be expressed as follows:

$$P_{elec} = \eta_{elec}^{\ -sgn(T_m)} \cdot T_m \cdot \omega_m, \tag{5}$$

where η_{elec} is the efficiency of the motor and converter, which can be expressed as a function of the motor torque and speed, as shown in Fig. 2. In the case of an electric battery, an internal resistance model is used, and the battery state of charge (SOC) is expressed as follows:

$$\dot{SOC} = -\frac{1}{2}\frac{V_{ocv} - \sqrt{V_{ocv}^2 - 4P_{bat}R_{bat}}}{Q_{bat}R_{bat}}, \tag{6}$$

where V_{ocv} is the open-circuit voltage, P_{bat} is the battery power, and R_{bat} is the internal resistance of the battery, both of which can be expressed as a function of the battery SOC, as shown in Fig. 3. Q_{bat} represents the battery capacity. For

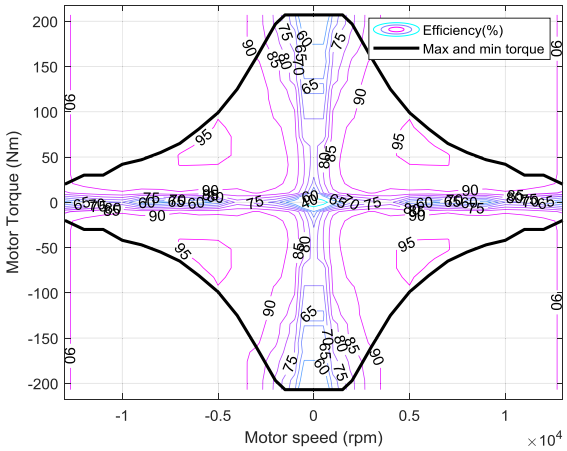

FIGURE 2. Efficiency of the electric motor including the converter.

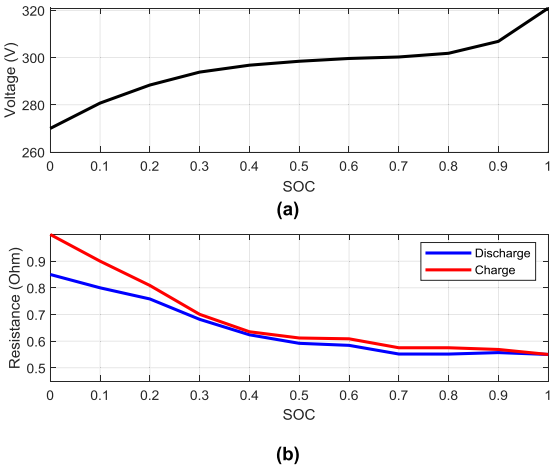

FIGURE 3. Open circuit voltage (a) and internal resistance (b) of the electric battery as a function of battery SOC.

the PEMFC, a steady-state model was utilized [27], [28]. The fuel consumption of the PEMFC can be expressed as follows:

$$\dot{m}_{h_2} = \frac{N_{cell} \cdot M_{h_2}}{n \cdot F} \cdot \lambda \cdot I_{stack}, \tag{7}$$

where $\dot{m}_{h_2}$ is the hydrogen fuel consumption rate, N_{cell} is the number of cells, M_{h_2} is the molar mass of hydrogen, n is the number of electrons acting in the reaction, F is the Faraday constant, λ is the hydrogen excess ratio, I_{stack} is the stack current, and the stack current and stack power of the fuel cell system are shown in Fig. 4. (a). For the fuel cell system, auxiliary power consumption in BOP, P_{BOP}, and the net power P_{fcs}, and the stack provided power P_{stack} have the following relationship:

$$P_{fcs} = P_{stack} - P_{BOP} \tag{8}$$

while the fuel cell system net power P_{fcs} is required to satisfy $P_{fcs} = P_{elec} - P_{bat}$. Then, Fig. 4. (b) shows the BOP power consumption for net power of the given fuel cell system, and hydrogen fuel consumption can be expressed as a function of the net power of the fuel cell system, as shown in Fig. 4. (c). A summary of the vehicle model parameters is

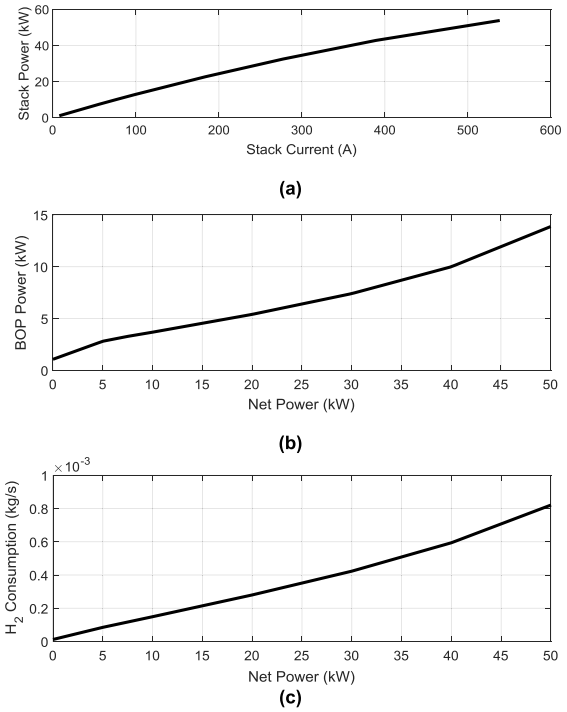

FIGURE 4. Characteristics of the fuel cell system; (a) stack power over stack current; (b) auxiliary power in BOP corresponds to net power; (c) Hydrogen fuel consumption rate corresponds to fuel cell system net power.

TABLE 1. Parameters of the vehicle model.

Parameter	Value
Fuel Cell	Max. power: 50 (kW)
Electric motor	Max. power: 60 (kW)
Electric battery	Ni-MH, Capacity: 6.5 (Ah)
Final drive gear	Final Gear ratio: 10 Efficiency : 98%
Converter	Efficiency : 95 %
Vehicle mass	1500 (kg)
Tire radius	0.305 (m)
Drag coefficient	0.29
Front area	1.8 (m^2)
Air density	1.23 (kg/m^3)
Rolling resistance coefficient	0.007

presented in Table. 1. The data used in this paper for vehicle simulation is obtained from "OC SIM," an Optimal Control Simulator developed for the control problem of electrified vehicles [29]. Based on this vehicle model, a simulation was conducted to test and verify the newly proposed algorithm. The energy management strategy using model-based reinforcement learning is described in detail in the following sections.

III. ENERGY MANAGEMENT STRATEGY

A. OPTIMAL CONTROL PROBLEM DEFINITION

In this section, the energy management strategy of FCEVs using MBRL is presented. The objective of the optimization problem is to find the control policy π to minimize cost $J_\pi(x_0)$, which is the expected cost when the vehicle system starts from an initial condition of x_0 and follows the control policy π, which can be expressed as follows:

$$\min J_\pi(x_0) = \lim_{N\to\infty} E\left\{\sum_{k=0}^{N-1} \gamma^k g(x_k, \pi(x_k))\right\}, \quad (9)$$

where x_k is the state variable at time k, which is a three – dimensional state, as shown below (10)

$$x_k = [SOC_k, P_{dem,k}, v_k], \quad (10)$$

where $P_{dem,k}$ represents the power demand. Battery SOC, SOC_k evolves with the nonlinear vehicle powertrain dynamics described in (1)–(6). The transition of vehicle speed v_k and corresponding power demand $P_{dem,k}$, can be modeled using a Markov chain as a discrete time stochastic dynamic process using random variable w_k, which is the probability distribution of driving cycle information. It can be expressed as follows.

$$\Pr\left\{w_k = [P_{dem}, v]_{k+1} \mid [P_{dem}, v]_k\right\} \quad (11)$$

Thus, the one-step transition probability in (11) can be estimated from the driving cycle as per recent driving cycle records; however, it is not required to model in MBRL, in which the Q function can be updated based on sampling. For P_{dem}, it can be calculated from the vehicle dynamics in (1) for a given driving cycle. The control variable $u = \pi(x)$ is the fuel cell net power P_{fcs}, which is equally discretized as

$$P_{fcs} \in \{P_{fcs}^l | l = 1, 2, \ldots, N_u\}, \quad (12)$$

where N_u is the number of control inputs, and P_{dem}=P_{fcs} + P_{bat}. Vehicle powertrain dynamics are constrained by equations (13)–(15).

$$\omega_{m,min} \le \omega_m(k) \le \omega_{m,max}, \quad (13)$$

$$T_{m,min}(\omega_m(k), SOC(k)) \le T_m(k) \le T_{m,max}(\omega_m(k), SOC(k)), \quad (14)$$

$$SOC_{min} \le SOC(k) \le SOC_{max}. \quad (15)$$

The instantaneous cost, also called the reward g, is composed of the hydrogen fuel consumption W_{fuel} and battery SOC sustaining term W_{soc}

$$g = W_{fuel} + W_{soc}(SOC), \quad (16)$$

where W_{fuel} is the hydrogen fuel consumption observed for one step, and W_{SOC} is a term for keeping the SOC close to the constant target SOC reference SOC_{ref}, defined as:

$$W_{soc}(SOC) = \begin{cases} C_1 \cdot (SOC - SOC_{ref})^2 & if\ SOC > SOC_{min} \\ C_2 & if\ SOC > SOC_{min} \end{cases} \quad (17)$$

where C_1 and C_2 are the weighting coefficients. Then, the optimization problem is to minimize the discounted sum of the instantaneous cost with the discount factor γ for a cycle duration of N, when $N \to \infty$. Thus, with the stochastic

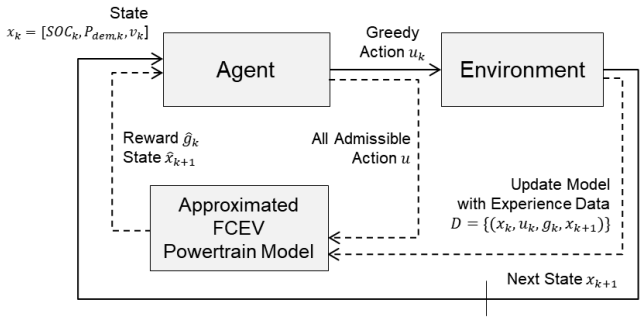

FIGURE 5. Control structure of MBRL for FCEV.

driving cycle profile, the control policy $\pi\ (x_k)$ is optimized to minimize the expected value of the hydrogen fuel consumption while the battery SOC is sustained near the target SOC reference.

B. MBRL FOR FCEVs

To solve this optimization problem, an MBRL based on Q-learning was utilized. First, the P_{dem}, v, and battery SOC are quantized as follows:

$$P_{dem} \in \{P^i_{dem} | i = 1, 2, \ldots, N_p\}, \tag{18}$$
$$v \in \{v^j | j = 1, 2, \ldots, N_v\}, \tag{19}$$
$$SOC \in \{SOC^m | m = 1, 2, \ldots, N_{SOC}\}, \tag{20}$$

where N_p, N_v, and N_{SOC} represent the number of discretized power demands, vehicle speed, and battery SOC, respectively. In Q-learning [30], the Q function can be defined using the instantaneous cost g_k and the cost of the next state $J_\pi\ (x_{k+1})$ with the control u as follows :

$$Q\left(x_k, u_k\right) = g\left(x_k, u_k\right) + \gamma \cdot J\left(x_{k+1}\right), \tag{21}$$

where $Q\left(x_k, u_k\right)$ are the action-value functions. Using Q, the optimal cost $J^*\left(x_k\right)$ can be defined by simply finding the minimum value of Q^* as below (21), and the optimal control policy $\pi^*\left(x_k\right)$ can be found as follows (22):

$$J^*\left(x_k\right) = \min_u \left(Q^*(x_k, u)\right), \tag{22}$$
$$\pi^*\left(x_k\right) = \arg\min_u \left(Q^*(x_k, u)\right). \tag{23}$$

The Q-function value can be updated iteratively as below:

$$Q\left(x_k, u_k\right) \leftarrow Q\left(x_k, u_k\right) + \alpha(g_k + \gamma \min_u Q\left(x_{k+1}, u\right) - Q\left(x_k, u_k\right)). \tag{24}$$

Therefore, in Q-learning, the Q function can be updated in a model-free and data-driven manner by calculating the reward cost g_k based on observation according to control u_k without modeling the vehicle powertrain dynamics or driving environments.

However, as mentioned in the Introduction, since the powertrain of FCEVs can be modeled, the use of this domain knowledge is beneficial for the efficient learning process of RL. The main idea of MBRL is shown in Fig. 5. Here, MBRL utilizes the internal approximation model of the FCEV powertrain and can determine the optimal action in association with learning process. MBRL can be divided into two main approaches. The first is the internal loop with the agent and the approximation model created by domain knowledge, such as dynamics equations, and the second is the outer loop to construct a model through learning based on the experience obtained from the interaction with the environment. More specifically, in MBRL for FCEVs, we can estimate the fuel consumption $\hat{W}_{fuel,k}$ for a given $u_k = P_{fcs,k}$, and the battery SOC change $\widehat{\Delta soc_k}$ can be predicted for the given $P_{dem,k}$, v_k, and $P_{fcs,k}$ values from the FCEV powertrain model, as follows:

$$\hat{W}_{fuel,k} = \hat{f}_{fuel}\left(P_{fcs,k}\right), \tag{25}$$
$$\widehat{\Delta soc_k} = \hat{f}_{soc}(P_{dem,k}, v_k, P_{fcs,k}), \tag{26}$$

where $\hat{f}_{fuel}$ and $\hat{f}_{soc}$ are estimated models for fuel consumption and battery SOC use, and $P_{fcs,k}$, v_k, and $P_{dem,k}$ are discretized using the nearest neighbor quantization. In addition, instead of using the known model, these estimation models can also be learned from the experience data of $\mathcal{D} = \{(x_k, u_k, g_k, x_{k+1})\}$ and observation of $W_{fuel,k}$ for a given input u_k, as follows:

$$\hat{f}_{fuel} \leftarrow \hat{f}_{fuel}\left(P^l_{fcs,k}\right) + \alpha_m\left(W_{fuel,k} - \hat{f}_{fuel}\left(P_{fcs,k}\right)\right), \tag{27}$$

$$\hat{f}_{soc} \leftarrow \hat{f}_{soc}(P_{dem,k}, v_k, P_{fcs,k}) + \alpha_m(\widehat{\Delta soc_k} - \hat{f}_{soc}(P_{dem,k}, v_k, P_{fcs,k})), \tag{28}$$

where α_m is the learning rate of the model learning process. Then, the MBRL runs the vehicle in a virtual driving cycle using the experience replay of $P_{dem,k} \rightarrow P_{dem,k+1}$, and $v_k \rightarrow v_{k+1}$, but for all admissible control inputs $u^l = P^l_{fcs}$ with $l = 1, 2, 3, \ldots, N_u$, and for different battery SOC, SOC^m with $m = 1, 2, 3, \ldots, N_{SOC}$, as follows:

$$Q \leftarrow (1-\alpha) \cdot Q\left(x_k = [SOC^m, P_{dem,k}, v_k], u^l\right) + \alpha \cdot \left(\hat{g}_k + \gamma \min Q\left(\hat{x}_{k+1} = \left[\widehat{SOC}^m_{k+1}, P_{dem,k+1}, v_{k+1}\right], u\right)\right), \tag{29}$$

where $\hat{g}_k$ is estimated cost (reward) that can be calculated as

$$\hat{g}_k = \hat{W}_{fuel,k} + W_{soc}(SOC_k), \tag{30}$$

and the next state of battery SOC $\widehat{SOC}^j_{k+1}$, for the previous battery SOC, SOC^j can be estimated using $\hat{f}_{soc}$ as follows.

$$\widehat{SOC}^j_{k+1} = \hat{f}_{soc}\left(x_k, u^l\right) + SOC^j \tag{31}$$

The overall MBRL algorithm for FCEVs is given in Algorithm 1. Similar to [25], in FCEVs, the driving cycle information, which is difficult to expect, remains completely model-free, to be learned in the Q-function with control policy through the data-driven learning process, while the vehicle powertrain dynamics can be modeled using domain knowledge; however, in this study, instead the model is

Algorithm 1 MBRL algorithm for FCEVs' Energy Management Strategy

input: data $\boldsymbol{x_k}$, size $\boldsymbol{N}$
repeat
Observe x_k
Choose the greedy action
$u_k = \underset{u}{\text{argmin}}\,(Q(x_k, u))$
Observe reward g_k, and next state x_{k+1}
Update model using D $=\{(x_k, u_k, g_k, x_{k+1})\}$

$$\hat{f}_{fuel} \leftarrow \hat{f}_{fuel}\left(P_{fcs,k}\right) + \alpha_m\left(W_{fuel,k} - \hat{f}_{fuel}\left(P_{fcs,k}\right)\right)$$

$$\hat{f}_{soc} \leftarrow \hat{f}_{soc}\left(P_{dem,k}, v_k, P_{fcs,k}\right) + \alpha_m\left(\widehat{\Delta soc}_k - \hat{f}_{soc}\left(P_{dem,k}, v_k, P_{fcs,k}\right)\right)$$

Update Q using the approximated vehicle model for all u
for $l = 1$ **to** N_u **do**
for $m = 1$ **to** N_{soc} **do**

$$Q \leftarrow (1-\alpha)\cdot Q\left(x_k = [SOC^m, P_{dem,k}, v_k], u^l\right) + \alpha \cdot \left(\hat{g}_k + \gamma \min Q\left(\hat{x}_{k+1} = \left[\widehat{SOC}^m_{k+1}, P_{dem,k+1}, v_{k+1}\right], u\right)\right)$$

end for
end for
$k \leftarrow k+1$
until Simulation Stop

given, it is designed such that the model can be learned from the data-driven update. Thus, reinforcement learning is conducted based on a framework using a model, but the model is updated with the experience data, without a model of the dynamic equation. Compared to the HEV case study in our previous research, in the present study, FCEVs have a series-type powertrain, where hydrogen fuel consumption is only dependent on the control input of the fuel cell power and not on the constrained vehicle speed; thus, the estimation model of hydrogen fuel consumption can be expressed as (25). For the model of battery SOC change, $\hat{f}_{soc}$ is a function of P_{dem}, v, and P_{fcs}, and is not a function of battery SOC; in reality, it is also a function of SOC that Δsoc is affected by the current battery SOC, but it is assumed that the difference in Δsoc due to the current SOC value is not large, and the SOC dynamics of the battery in the decision-making process of the RL are relatively slow, such that a sudden change in the SOC of the battery does not occur within a short time; thus, it can be ignored in the estimation model, while it is included in the Q-function value update in (29). Therefore, using estimation models, the Q-function value can be updated quickly and stably.

Considering the practical applications of this algorithm, the learning process requires considerable time and computational load. In this study, the model update and Q function update using the vehicle approximation model in Algorithm 1 are carried out at every step. However, the update does not necessarily have to be performed at every step, but it can be performed asynchronously considering the computational load. In addition, as with other Q-learning, once the Q table learns a general control policy for different driving cycles, it is possible to use it as a base policy and to adaptively change the control policy by learning according to the characteristics of the current driving cycle. To adapt the control policy more quickly, the 'for loop' in Algorithm 1 can be replaced with an update method using parameterization methods such as neural networks, or the update method itself can be changed using methods such as prioritized sweeping [31], which is beyond the scope of this paper. In addition, the vehicle approximation model does not change significantly once it is trained, so the model need not be updated regularly. The proposed energy management strategy is tested on the FCEV model in Section II, and the vehicle simulation results are presented in Section IV.

TABLE 2. Discretization of state and control variables.

Parameter	Discretization level	Range
Fuel cell power (W)	$N_u = 51$	[0, 50000]
Power demand (W)	$N_{P_{dem}} = 43$	[-34000, 50000]
Vehicle speed (m/s)	$N_v = 29$	[0,28]
Battery SOC	$N_{SOC} = 3000$	[0.45 0.75]

IV. VEHICLE SIMULATION

MATLAB was used for the vehicle simulation and to write and implement the algorithm based on MBRL. The vehicle simulation, including the training, was executed on a workstation with an i7-8700k CPU @ 3.70GHz, with a 32.0 GB RAM. The discretization levels of the state and control variables are listed in Table 2. Note that, generally, the better the discretization proceeds, the better the performance that can be expected. However, this should be considered in terms of learning and determined appropriately by taking into account the state variable that needs to be measured with the disturbance. In addition, for a Q function configured in the form of a lookup table, the size of the Q function increases rapidly depending on the number of state variables and the level of discretization; therefore, it is necessary to take this into account. The parameters for RL are given in Table 3, where the parameters C_1 were defined to evaluate the hydrogen fuel consumption and battery SOC deviation comprehensively, while C_2 has a relatively large value to assign a penalty for SOC less than the minimum value. The learning rate α and model learning rate α_m are determined by trial and error. The discounted factor is assumed to be close to 1, which means that the total expected value of the future cost is considered, instead of conducting instantaneous minimization.

A. LEARNING CURVE

First, vehicle simulation was conducted using an urban dynamometer driving schedule (UDDS). The MBRL

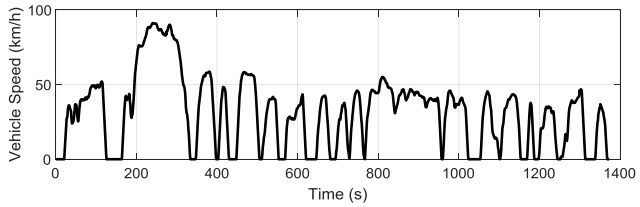

FIGURE 6. UDDS driving cycle.

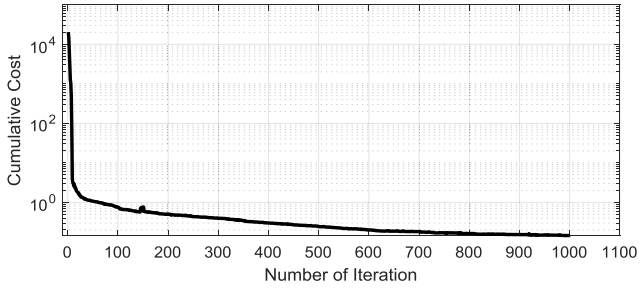

FIGURE 7. Learning curve on the log scale for MBRL for driving cycle of UDDS.

algorithm is run iteratively to learn the model and the corresponding control policy. As mentioned earlier, models of $\hat{f}_{fuel}$ and $\hat{f}_{soc}$ are completely unknown initially, and are updated based on the data from these iterations. The iteration was repeated 1000 times. The learning curve is shown in Fig. 7. Here, the cumulative cost, which is the accumulated value of the instantaneous cost g in (15), when the vehicle runs through the UDDS cycle is presented in the log scale, and it can be seen that the learning is conducted effectively; the cost decreases progressively with each iteration, and the decrease is sharp initially. The simulated motor torque, fuel cell power, battery power, hydrogen fuel consumption, and battery SOC are shown in Fig. 8. The battery mainly operates around the target SOC of 0.6, and the SOC performs well. Therefore, the MBRL successfully learns the optimal control policy for the UDDS cycle and drives the vehicle while maintaining the battery SOC.

On the other hand, the fuel consumption and battery SOC estimation models $\hat{f}_{fuel}$ and $\hat{f}_{soc}$ are also updated successfully. In Fig. 9 and Fig. 10, the plant model used as the environment in the vehicle model simulation and the estimation model used inside the MBRL are presented. After iterative learning, the estimation models tend to resemble the plant model. In the case of the fuel consumption model, which is relatively simply expressed as a function of the net power of the fuel cell system, it can be confirmed that the model is estimated to be remarkably close to the actual model in all areas, as shown in Fig. 9. The SOC can be predicted when the current battery SOC, power demand, vehicle speed, and control input of the fuel cell power are given. The SOC predictions made using the known powertrain and power electric dynamics are remarkably accurate, as shown in Fig. 10. Therefore, as in model-free learning, even without information or initialization of the vehicle model, with only information about which state determines the future states and constitutes the

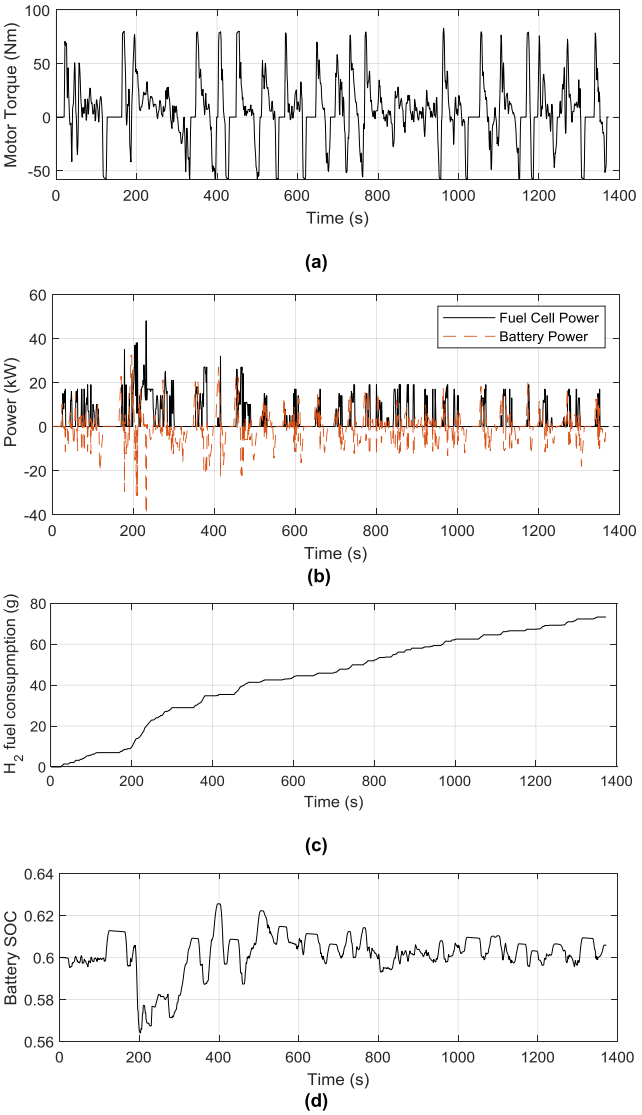

FIGURE 8. Simulation results of MBRL on the UDDS driving cycle; (a) motor torque, (b) fuel cell and battery power, (c) hydrogen fuel consumption, (d) battery SOC trajectory.

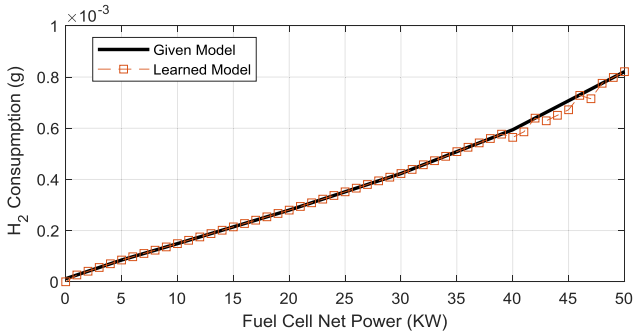

FIGURE 9. Hydrogen fuel consumption model.

instantaneous cost, it is possible to proceed with the learning through the interaction between the agent and the driving cycle environment. The model can be learned based on data-driven updates with experience while the control policy is simultaneously optimized based on the learned model.

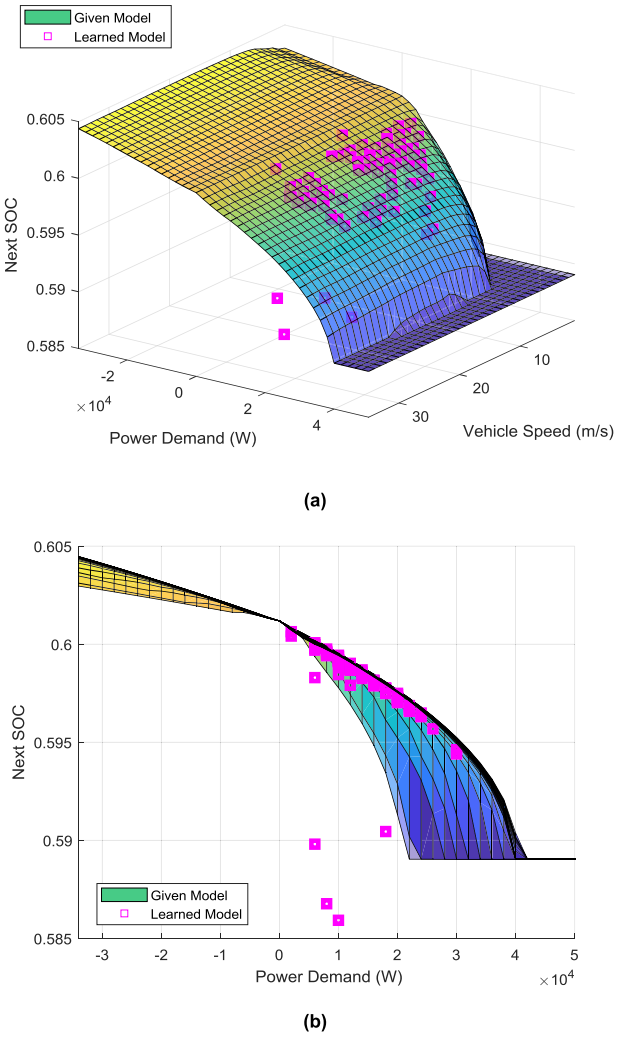

FIGURE 10. Estimation of next battery SOC when Battery SOC is 0.6 and fuel cell output power is 10 kW; (a) Overall view, (b) Side view.

B. MODEL-BASED VS MODEL-FREE

In addition, the MBRL was compared with the existing model-free RL method based on simulations. Tabular Q-learning was used as a model-free method and all other conditions were defined in the same way as MBRL. For learning, the Q-function value was updated as in (24), while the ε-greedy method with $\varepsilon = 0.1$ is used to balance exploration and exploitation, where ε represents the probability of choosing an action randomly for exploration and exploitation dilemma [32]. The simulation results are shown in Fig. 11. This shows that the cumulative cost of MBRL and Q-learning for the UDDS driving cycle decreases as the iterations proceed. MBRL takes more time to perform one iteration than Q-learning. However, after several iterations, the cumulative cost decreases more quickly in MBRL. Q-learning takes considerably more time to learn the control policy reflecting both vehicle powertrain dynamics and driving cycle information, especially as learning is performed through trial and error. Thus, the cumulative cost decreases slowly; whereas in MBRL, learning can be performed efficiently using the

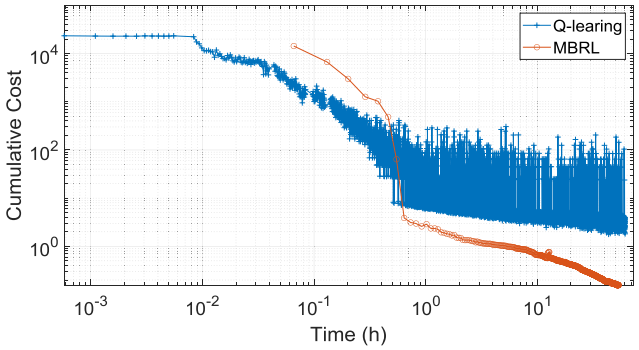

FIGURE 11. Learning curve on the log scale for Q-learning, and MBRL for driving cycle of UDDS.

powertrain dynamics model. As Q-learning is a completely model-free method, it is not necessary to model the environment, and so, it is free from problems caused by modeling difficulties or inaccurate modeling. However, the proposed MBRL also has the same advantage as Q-learning because it is not necessary to know model parameters or detailed dynamics equations to implement it, and only the framework needs to be constructed based on the domain knowledge of model dynamics.

C. COMPARISON WITH ENERGY MANAGEMENT STRATEGY USING DP, AND RULE-BASED CONTROL

To verify the fuel efficiency performance of the proposed MBRL algorithm, the DP and rule-based strategies were used. As mentioned in the introduction, DP is the algorithm that shows the maximum performance that the system can achieve when the driving cycle information is given in advance, and so, the performance of the MBRL can be compared with the best solution. On the other hand, in the case of the rule-based strategy, thermostat control [33] is used that, according to the series powertrain structure, enables the fuel cell to operate in the most efficient area, as shown in (32).

$$P_{fc} = \begin{cases} P_{opt} & if\ S_{fc}(k) = 1 \\ 0 & if\ S_{fc}(k) = 0, \end{cases} \tag{32}$$

where P_{opt} is the fuel cell power at the most efficient area, and S_{fc} is the on/off signal for the fuel cell system, which is determined to maintain the battery SOC near the target SOC, as follows:

$$S_{fc}(k) = \begin{cases} 0 & if\ SOC(k) > SOC_{upper} \\ S_{fc}(k-1) & if\ SOC_{lower} \le SOC(k) \le SOC_{upper} \\ 1 & if\ SOC(k) < SOC_{lower}, \end{cases} \tag{33}$$

where SOC_{upper}, and SOC_{lower} are the upper boundary and lower boundary of SOC, separately.

The simulation results for UDDS and highway fuel economy test driving cycle (HWFET) are shown in Fig. 12 and Fig. 13, respectively. In Fig. 12 (a) and Fig. 13 (a), because

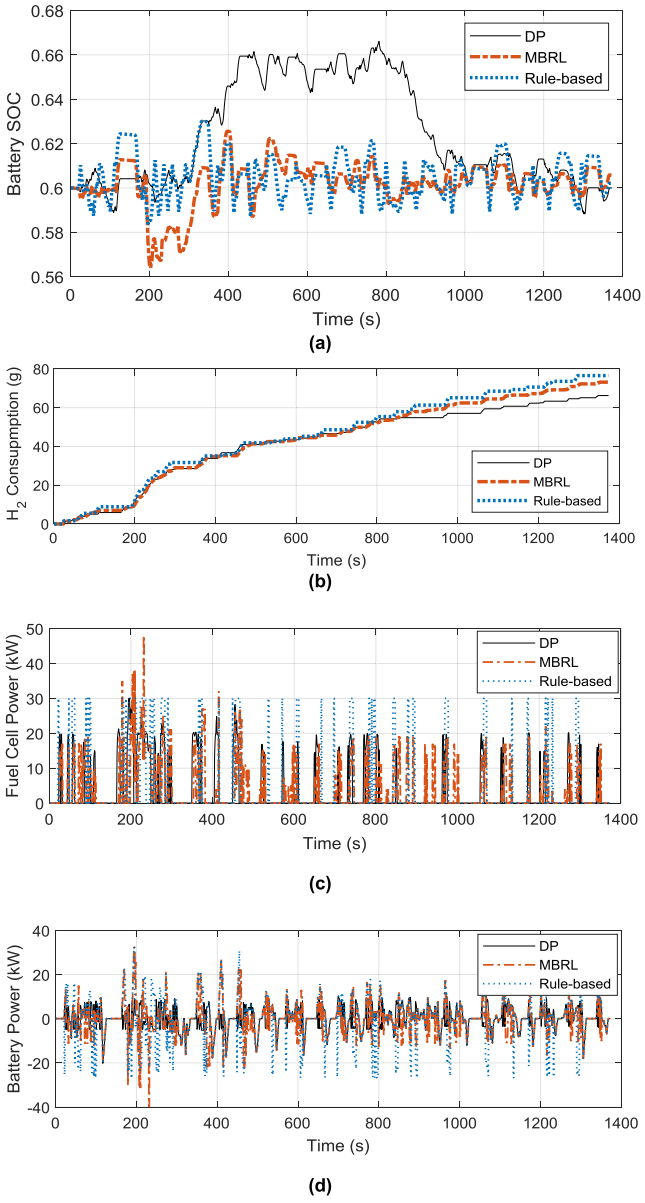

FIGURE 12. Simulation results on UDDS driving cycle for DP, MBRL, and rule-based strategy; (a) battery SOC, (b) hydrogen fuel consumption, (C) fuel cell net power, (d) battery power.

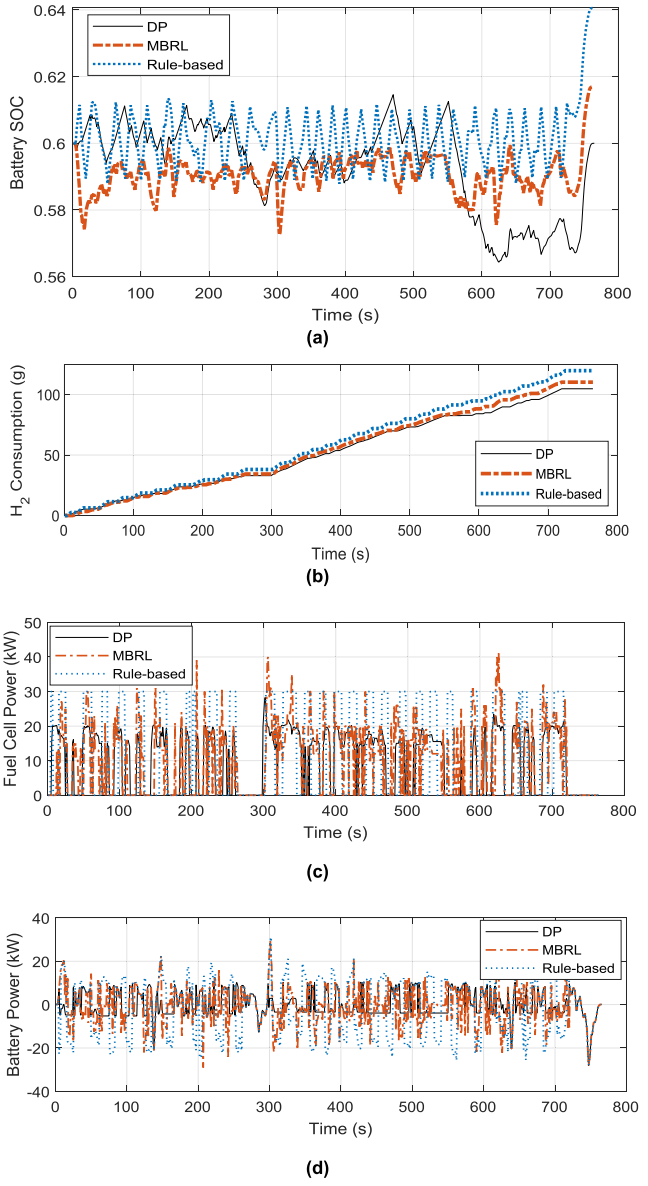

FIGURE 13. Simulation results on HWFET driving cycle for DP, MBRL, and rule-based strategy; (a) battery SOC, (b) hydrogen fuel consumption, (C) fuel cell net power, (d) battery power.

all the driving cycle information is known in advance for DP, it can be seen that there is a tendency to increase the battery SOC early and then decrease it slowly. However, in the case of MBRL and rule-based algorithms, the battery SOC is controlled to be maintained near the target SOC, and in the case of MBRL, optimization was performed to minimize fuel consumption while maintaining the battery SOC. On the other hand, in Fig. 12 (c) and Fig. 13 (c), for the rule-based strategy, the fuel cell system is operated in the most efficient area of approximately 30 kW, while the fuel cell system is turned on and off according to the battery SOC. In the case of MBRL, compared with the rule-based strategy, MBRL shows that the control is closer to the DP result. The fuel consumption for each case is listed in Table 4. Because the final SOC is different for each simulation, the equivalent fuel consumption representing the fuel consumption when the final SOC is the same as the initial SOC was calculated as in [34]. The equivalent fuel consumption shows that DP performs the best, while the fuel consumption associated with MBRL is inevitably larger than the optimal result of DP; however, the fuel consumption is reduced by nearly 5.7% compared to the rule-based fuel consumption. Note that in DP, information regarding the entire driving cycle and powertrain is given, and for the rule-based strategy, the control policy is based on the information about the powertrain (efficiency of the fuel cell system). However, for MBRL, there was no initial information, but the control policy could be extracted from the interaction between the controller and the environment; using the internal model of the powertrain learned from experience,

TABLE 3. Parameter values of RL.

Parameter	Value
C_1	0.5
C_2	10
α	0.0015
γ	0.99995
α_m	0.95
SOC_{ref}	0.60
SOC_{min}	0.45

TABLE 4. Comparison of the equivalent fuel consumption(g) results for DP, MBRL, and Rule-based strategy (percentage relative to the Rule-based results).

Algorithm	Driving Cycle	
	UDDS	HWFET
DP	66.3 (87.1%)	104.7 (91.8%)
MBRL	72.6 (95.4%)	108.6 (95.3%)
Rule-based	76.1 (100%)	114.0 (100%)

the MBRL strategy could also be used to identify and analyze the characteristics of powertrain systems.

D. OFFLINE CONTROL STRATEGY

MBRL has the advantage of being able to extract the optimal control policy through learning but formulating the control policy suitable for the current driving cycle is time-consuming. Instead, MBRL can be used for extracting a general control policy in advance before driving using offline processing with historic driving cycle data. Thus, the extracted general control policy can be used when driving, while the control policy can be updated simultaneously to reflect the characteristics of the current driving cycle. In Table 5, the offline control policy using MBRL was tested in the UDDS and HWFET driving cycles. The control policy is trained using different driving cycles of UDDS, such as the worldwide harmonized light vehicles test cycles (WLTC), supplemental federal test procedure (SC03), new European driving cycle (NEDC), and HWFET, in which each learning proceeded in the corresponding order. The simulation results are presented in Table 5. In the case of the UDDS driving cycle, the fuel consumption is 75.5 (g), which increases slightly compared with the UDDS result of 72.6 (g)in Table 4 as the control policy is generalized for different driving cycles; thus, the fuel economy performance is reduced for the specific driving cycle of the UDDS. However, it shows that fuel consumption is still less than the 76.1 (g) associated with the rule-based strategy in Table 4, and the extracted general control policy still performs better than the rule-based strategy. The HWFET shows an improved fuel consumption performance of 106.4 (g), compared to the 108.6 (g) of MBRL for HWFET, as shown in Table 4. This is because, during offline training, the HWFET cycle is the last data used for learning, and combined with the general control policy previously formulated from the other driving cycles, it shows improved performance than when only the HWFET is used, as shown in Table 4.

TABLE 5. Fuel consumption (g) of Offline control policy using MBRL: (percentage relative to the Rule-based results).

Algorithm	Driving Cycle	
	UDDS	HWFET
MBRL offline *	75.5 (99.3%)	106.4 (93.3%)

* Control policy was learned for different driving cycles in the order of UDDS, WLTC, SC03, NEDC, and HWFET.

V. CONCLUSION

In this study, MBRL was used as the energy management strategy in FCEVs. Unlike existing model-free RL methods, MBRL allows learning based on a model. In this study, the model is used for the learning process of a vehicle powertrain in which a control policy according to the driving cycle part, which has a stochastic characteristic, remains for model-free approaches. In addition, for model-based learning, instead of the vehicle model represented by dynamic equations, a powertrain model is learned with a data-driven model update. The model parameters, however, were unknown initially. The simulation results show that the learning process is very stable, and it was confirmed that the model was also well-learned. The performance of the algorithm was compared with DP and the rule-based strategy; for MBRL, the fuel consumption decreased by 5.7% on average compared to the rule-based strategy, while MBRL consumes more fuel than DP.

In the future, we plan to perform various simulations before conducting experimental validations to check and improve the performance of MBRL algorithms for FCEVs. In addition, we will consider the implementation aspect as well, because the current computational requirements of MBRL cannot be handled by the electric control unit in an actual vehicle powertrain for usage as an online controller. However, considering the expandability of MBRL and its applicability to the system, it is thought that it can be applied to the complex control problem of an FCEV system.

REFERENCES

[1] N. Sulaiman, M. A. Hannan, A. Mohamed, E. H. Majlan, and W. R. Wan Daud, "A review on energy management system for fuel cell hybrid electric vehicle: Issues and challenges," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 802–814, Dec. 2015.

[2] D. Fares, R. Chedid, F. Panik, S. Karaki, and R. Jabr, "Dynamic programming technique for optimizing fuel cell hybrid vehicles," *Int. J. Hydrogen Energy*, vol. 40, no. 24, pp. 7777–7790, Jun. 2015.

[3] M. Ansarey, M. S. Panahi, H. Ziarati, and M. Mahjoob, "Optimal energy management in a dual-storage fuel-cell hybrid vehicle using multi-dimensional dynamic programming," *J. Power Sources*, vol. 250, pp. 359–371, Mar. 2014.

[4] W. Zhou, L. Yang, Y. Cai, and T. Ying, "Dynamic programming for new energy vehicles based on their work modes part II: Fuel cell electric vehicles," *J. Power Sources*, vol. 407, pp. 92–104, Dec. 2018.
[5] C.-C. Lin, M.-J. Kim, H. Peng, and J. W. Grizzle, "System-level model and stochastic optimal control for a PEM fuel cell hybrid vehicle," *J. Dyn. Syst., Meas., Control*, vol. 128, no. 4, pp. 878–890, Dec. 2006.
[6] Z. Fu, Z. Li, P. Si, and F. Tao, "A hierarchical energy management strategy for fuel cell/battery/supercapacitor hybrid electric vehicles," *Int. J. Hydrogen Energy*, vol. 44, no. 39, pp. 22146–22159, Aug. 2019.
[7] P. Rodatz, G. Paganelli, A. Sciarretta, and L. Guzzella, "Optimal power management of an experimental fuel cell/supercapacitor-powered hybrid vehicle," *Control Eng. Pract.*, vol. 13, no. 1, pp. 41–53, Jan. 2005.
[8] 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, Nov. 2019, Art. no. 227105.
[9] H. Li, A. Ravey, A. N'Diaye, and A. Djerdir, "A novel equivalent consumption minimization strategy for hybrid electric vehicle powered by fuel cell, battery and supercapacitor," *J. Power Sources*, vol. 395, pp. 262–270, Aug. 2018.
[10] 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.
[11] C. Zheng and S. W. Cha, "Real-time application of Pontryagin's minimum principle to fuel cell hybrid buses based on driving characteristics of buses," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 4, no. 2, pp. 199–209, Apr. 2017.
[12] L. Xu, J. Li, J. Hua, X. Li, and M. Ouyang, "Adaptive supervisory control strategy of a fuel cell/battery-powered city bus," *J. Power Sources*, vol. 194, no. 1, pp. 360–368, Oct. 2009.
[13] N. Kim, S. Ha, J. Jeong, and S. W. Cha, "Sufficient conditions for optimal energy management strategies of fuel cell hybrid electric vehicles based on Pontryagin's minimum principle," *Proc. Inst. Mech. Eng. D, J. Automobile Eng.*, vol. 230, no. 2, pp. 202–214, Feb. 2016.
[14] L. Xu, M. Ouyang, J. Li, F. Yang, L. Lu, and J. Hua, "Application of Pontryagin's minimal principle to the energy management strategy of plugin fuel cell electric vehicles," *Int. J. Hydrogen Energy*, vol. 38, no. 24, pp. 10104–10115, Aug. 2013.
[15] D. Shen, C.-C. Lim, and P. Shi, "Robust fuzzy model predictive control for energy management systems in fuel cell vehicles," *Control Eng. Pract.*, vol. 98, May 2020, Art. no. 104364.
[16] D. F. Pereira, F. D. C. Lopes, and E. H. Watanabe, "Nonlinear model predictive control for the energy management of fuel cell hybrid electric vehicles in real time," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3213–3223, Apr. 2021.
[17] C. Bordons, M. A. Ridao, A. Perez, A. Arce, and D. Marcos, "Model predictive control for power management in hybrid fuel cell vehicles," in *Proc. IEEE Vehicle Power Propuls. Conf.*, Sep. 2010, pp. 1–6.
[18] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32–50, Aug. 2009.
[19] H. Sun, Z. Fu, F. Tao, L. Zhu, and P. Si, "Data-driven reinforcement-learning-based hierarchical energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles," *J. Power Sources*, vol. 455, Apr. 2020, Art. no. 227964.
[20] Y. F. Zhou, L. J. Huang, X. X. Sun, L. H. Li, and J. Lian, "A long-term energy management strategy for fuel cell electric vehicles using reinforcement learning," *Fuel Cells*, vol. 20, no. 6, pp. 753–761, Dec. 2020.
[21] J. Yuan, L. Yang, and Q. Chen, "Intelligent energy management strategy based on hierarchical approximate global optimization for plug-in fuel cell hybrid electric vehicles," *Int. J. Hydrogen Energy*, vol. 43, no. 16, pp. 8063–8078, Apr. 2018.
[22] S. Racanière, T. Weber, D. P. Reichert, L. Buesing, A. Guez, D. J. Rezende, A. P. Badia, O. Vinyals, N. Heess, Y. Li, R. Pascanu, P. W. Battaglia, D. Hassabis, D. Silver, and D. Wierstra, "Imagination augmented agents for deep reinforcement learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5690–5701.
[23] F. Berkenkamp, A. P. Schoellig, M. Turchetta, and A. Krause, "Safe model-based reinforcement learning with stability gurantees," in *Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS)*, 2014, vol. 47, no. 12, pp. 737–742.
[24] S. Bansal, R. Calandra, K. Chua, S. Levine, and C. Tomlin, "MBMF: Model-based priors for model-free reinforcement learning," 2017, *arXiv:1709.03153*. [Online]. Available: http://arxiv.org/abs/1709.03153
[25] H. Lee, C. Kang, Y.-I. Park, N. Kim, and S. W. Cha, "Online data-driven energy management of a hybrid electric vehicle using model-based Q-learning," *IEEE Access*, vol. 8, pp. 84444–84454, 2020.
[26] H. Lee and S. W. Cha, "Reinforcement learning based on equivalent consumption minimization strategy for optimal control of hybrid electric vehicles," *IEEE Access*, vol. 9, pp. 860–871, 2021.
[27] J. T. Pukrushpan, H. Peng, and A. G. Stefanopoulou, "Control-oriented modeling and analysis for automotive fuel cell systems," *J. Dyn. Syst., Meas., Control*, vol. 126, no. 1, pp. 14–25, Mar. 2004.
[28] C. H. Zheng, N. W. Kim, and S. W. Cha, "Optimal control in the power management of fuel cell hybrid vehicles," *Int. J. Hydrogen Energy*, vol. 37, no. 1, pp. 655–663, Jan. 2012.
[29] N. Kim. *OC_SIM*. [Online]. Available: https://sites.google.com/view/mdlhyu/research/oc_sim
[30] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.
[31] A. W. Moore and C. G. Atkeson, "Prioritized sweeping: Reinforcement learning with less data and less time," *Mach. Learn.*, vol. 13, no. 1, pp. 103–130, Oct. 1993.
[32] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: Bradford, 2018.
[33] Y. J. Kim and Z. Filipi, "Series hydraulic hybrid propulsion for a light truck–optimizing the thermostatic power management," *SAE Trans.*, vol. 116, pp. 1597–1609, Feb. 2007.
[34] C. H. Zheng, C. E. Oh, Y. I. Park, and S. W. Cha, "Fuel economy evaluation of fuel cell hybrid vehicles based on equivalent fuel consumption," *Int. J. Hydrogen Energy*, vol. 37, no. 2, pp. 1790–1796, Jan. 2012.

HEEYUN LEE (Member, IEEE) received the B.S. degree in mechanical engineering from Sungkyunkwan University, South Korea, in 2013, and the Ph.D. degree in mechanical and aerospace engineering from Seoul National University, South Korea, in 2018.

He is currently with the Research and Development Division, Hyundai Motor Company, South Korea. His research interests include optimal control, reinforcement learning, and modeling of electrified vehicles.

SUK WON CHA (Member, IEEE) received the B.S. degree in naval architecture and ocean engineering from Seoul National University, in 1994, and the M.S. and Ph.D. degrees in mechanical engineering from Stanford University, in 1999 and 2004, respectively.

He is currently a Professor with the Department of Mechanical Engineering, Seoul National University. His current research interests include modeling of electric vehicle modules and performance analysis of powertrains. He is also a Senior Editor of the *International Journal of Precision Engineering and Manufacturing—Green Technology*. He also serves as an Editor for the *International Journal of Automotive Technology*.