Enhanced Q-learning for real-time hybrid electric vehicle energy management with deterministic rule

Yang Li1,2, Jili Tao1, Liang Xie2, Ridong Zhang3, Longhua Ma1 and Zhijun Qiao4

Abstract
Power allocation plays an important and challenging role in fuel cell and supercapacitor hybrid electric vehicle because it influences the fuel economy significantly. We present a novel Q-learning strategy with deterministic rule for real-time hybrid electric vehicle energy management between the fuel cell and the supercapacitor. The Q-learning controller (agent) observes the state of charge of the supercapacitor, provides the energy split coefficient satisfying the power demand, and obtains the corresponding rewards of these actions. By processing the accumulated experience, the agent learns an optimal energy control policy by iterative learning and maintains the best Q-table with minimal fuel consumption. To enhance the adaptability to different driving cycles, the deterministic rule is utilized as a complement to the control policy so that the hybrid electric vehicle can achieve better real-time power allocation. Simulation experiments have been carried out using MATLAB and Advanced Vehicle Simulator, and the results prove that the proposed method minimizes the fuel consumption while ensuring less and current fluctuations of the fuel cell.

Keywords
Q-learning, deterministic rule, hybrid electric vehicle, energy management controller

Date received: 22 January 2020; accepted: 4 June 2020

Introduction
Energy shortage, air pollution, and global warming have pushed the development of fuel cell (FC)-driven vehicles to replace pure fuel-driven vehicles.1–4 However, possessing quick dynamic response and load-following ability is difficult for current FC.5 Furthermore, rapid load variation has bad effects on the lifetime of FC.6 Thus, pure FC vehicles are still in their early development stages, which will probably last for the next decade. A hybrid propulsion, such as the supercapacitor (SC), with fast charge/discharge attributes, long life cycles, and high power density seems to be the most economical and feasible solution so far. Hybrid electric vehicles (HEVs) composed of the FC and the SC may be a good choice. When HEV is in braking, climbing, or acceleration condition, SC can be used as a power buffer,7–9 and the combination of FC and SC as the hybrid propulsion is an efficient way to overcome the slow dynamic response and rapid load variation while achieving braking energy recovery.10,11 How to control the energy flow between the hybrid FC/SC has been the core issue.

Literature review
The conventional energy management method can be generally classified into the following two trends: rule-based and optimization-based.12 The former strategy can be subdivided into deterministic and fuzzy rule-based methods, while the latter can be subdivided into off-line global and real-time optimization-based methods. The deterministic rule methods are the most direct

1School of Information Science and Engineering, NingboTech University, Ningbo, China
2College of Control Science and Engineering, Zhejiang University, Hangzhou, China
3The Belt and Road Information Research Institute, Automation College, Hangzhou Dianzi University, Hangzhou, China
4Ningbo CRRC New Energy Technology Co., Ltd., Ningbo, China

Corresponding authors:
Zhijun Qiao, Ningbo CRRC New Energy Technology Co., Ltd., Ningbo 315012, China.
Email: qiaozhijun0928 @163.com
and widely used strategy with easy implementation and low calculation burden. Jalil et al.\textsuperscript{13} proposed a rule-based strategy in which the power demand is allocated between the engine and the battery, by which those power sources can be used efficiently. The proposed rules ensure efficient operation of the engine and battery at any situation, but it is applicable only in series hybrid structure because of its simplicity. In the study by Phillips,\textsuperscript{14} a type of state machine was utilized to supervise the control of a more general parallel HEV; however, in terms of achieving the goals of fuel economy and emission reduction, it has not gained good performance optimization. To further realize improvement of the performances of energy management system (EMS) for HEV, literally, fuzzy logic and their modified variants, instead of using deterministic rules, seem to be the most effective way to solve the problem of robustness and adaptability,\textsuperscript{15–17} because they are not only tolerable to fuzzy measurement but also easy to adapt, if necessary. Multi-objective optimization strategies based on fuel economy, FC lifetime and so on are also researched widely and have obtained good simulation results.\textsuperscript{18–20} These rule-based control strategies are optimized by minimization of a loss function generally representing the control objectives under a fixed driving cycle, which means a prior knowledge of a predefined driving cycle is used. Obviously, they cannot be directly used in real-time energy management. Recently, several optimization-based methods, such as real-time control based on equivalent fuel consumption, were proposed to develop a loss function in instantaneous optimization.\textsuperscript{21–23} Model predictive control\textsuperscript{24} and dynamic programming\textsuperscript{25,26} have also been widely used to develop the advanced on-line EMS. Furthermore, to obtain a prior knowledge of driving cycle, we\textsuperscript{27} proposed a driving pattern recognition-based EMS using neural network, which also achieved real-time control while accomplishing less current fluctuation and fuel consumption. Intelligent algorithms have developed rapidly in recent years, and the learning-based energy management method has been considered as a viable solution to apply decision and control problems in electric power system.\textsuperscript{28} A learning-based EMS aims to take appropriate actions automatically according to the states, not relying on any manual predefined rules, and converges to an optimal policy without any optimization algorithms. In addition, the learning-based EMS has shown its self-learning ability to adapt to different driving conditions.\textsuperscript{29–31} Statistical learning method is a significant way to optimization approach, and related studies may help to improve the robustness of the EMSs.\textsuperscript{32–34}

**Motivation and innovation**

The main goal of this study was to propose a novel Q-learning strategy with deterministic rule (QLDR) for real-time HEV energy management that satisfies the driver’s demand for traction power while achieving decreased fuel consumption and load fluctuation. In particular, we focus on improving the Q-learning (QL)-driven agent’s adaptation to different driving cycles. The main contributions of this paper are as follows. (1) To reduce the load fluctuation of the FC, we innovatively propose two optional sets of the maximum FC power output, which will be selected by the deterministic rule. The smaller set is for the general driving conditions that are frequently used. And the alternative set is provided for extreme driving conditions like continuous high power demand, which will only be used when these extreme situations occur. (2) The deterministic rule is combined with the QL policy to further improve the adaptation to different driving conditions. (3) To realize less fuel consumption, we manage to keep the best Q-table with minimal fuel consumption and maintain the state of charge (SoC) of the SC within a safe range so that the SC is always able to recover the braking energy. This idea differs greatly from the conventional energy conservation\textsuperscript{35–37} based on the “load-leveling” concept for efficient actual operation and for achieving as close as possible to the optimal point, which is dependent deeply on the precise measurement and prior knowledge of driving conditions.

**Organization of this paper**

The structure of this paper is as follows. Section “HEV energy system” describes the power-train of the HEV system and the modeling of the FC and SC. Then the HEV energy optimization problem is formulated. Section “QL-based HEV energy optimization” describes the key concepts of QL in HEV energy management control. Then the proposed QLDR algorithm is designed and the real-time energy management strategy is provided. The evaluation results of the proposed method are shown in section “Simulation results,” and section “Conclusion” concludes the paper.

**HEV energy system**

In this section, we first describe the power-train of the HEV system. Then the modeling of the FC and SC is achieved. Finally, we present how the HEV energy optimization problem formulates.

**Power-train description**

The power-train of the HEV energy system is shown in Figure 1, where the black and red arrows represent the control signals and the direction of power flow, respectively. The power demand of the HEV can be satisfied by controlling the direction of the power flow between the FC stack and the SC storage. By means of the proposed QLDR, the power supply required is derived, which is packed as signals sent to the corresponding energy source. Among the energy components, the FC is used as the primary power source. The SC, with the characteristics of fast charge and discharge, is equipped
as a power buffer for leveling the peak power during cold start and hard acceleration and for recovering the braking energy. The SoC of the SC is fed back to the proposed QLDR for better control. To control the power flow between the FC and the SC, different energy sources are equipped with different types of converters. The unidirectional DC/DC converter is for the FC, whereas the bidirectional DC/DC converter is for the SC. The power supplied from these converters is gathered by the DC link and then flows to the motor through a DC/AC converter.

The modeling of the FC and the SC

1. FC model: To calculate the output power of the FC, we need to obtain the output voltage of the FC ($V_{out}$) first, which is given as follows

$$V_{out} = N_0E_{cell} - V_{act} - V_{ohm}$$

$$V_{act} = B_1\ln (B_2I)$$

$$V_{ohm} = IR_{ohm}$$

where $V_{act}$ and $V_{ohm}$ are the activation voltage and overall internal ohmic voltage, respectively; $N_0$ is the number of FC in series; $B_1$ and $B_2$ are constants; $I$ is the output current of the FC; $R_{ohm}$ is the internal resistance; and $E_{cell}$ is the Nernst cell voltage, which is calculated as follows

$$E_{cell} = E_i - k_F(T - 298) + \frac{R_gT}{2F}\ln \left(\frac{pH_2 \sqrt{PO_2}}{k}\right) - E_{dcell}$$

where $E_i$ is the standard reference potential per FC, $T$ is the FC stack temperature, $R_g$ is the gas constant, $F$ is the Faraday constant, and $pH_2$ and $PO_2$ are the hydrogen and oxygen partial pressure, respectively, and they are constants here for simple analysis. $E_{dcell}$ is obtained by a first-order transfer function as can be seen in equation (5)

$$E_{dcell}(s) = \frac{\lambda_e \tau_s}{\tau_s + 1} r(s)$$

where $\lambda_e$ is a constant factor and $\tau_s$ is an overall flow delay, and the total hydrogen consumption of FC can be derived as follows in equation (6)

$$m_{H_2} = \frac{M_{H_2}N_0AF_{FC}}{2F} I$$

where $M_{H_2}$ is the molecular weight of hydrogen, $AF_{FC}$ is the active area of each FC, and $F$ is the Faraday constant. The main parameters of FC are shown in Table 1. To analyze the power and current fluctuation of the FC, we provide the following definition as can be seen in equations (7) and (8), where $V_P$ and $V_I$ denote the variance ratio of the power and current of the FC, respectively.

$$V_P = \frac{\Delta(V_{out}\times I)}{\Delta I}$$

$$V_I = \frac{\Delta I}{\Delta I}$$

2. SC model: We use the resistor-capacitor circuit model to emulate the internal part of the SC, which is described in Figure 2.

In Figure 2, $v_c$ is the SC internal capacitance voltage; $R$ and $i$ are the internal resistor and current, respectively; and $v$ and $P$ are the terminal voltage and power, respectively. When $P > 0$, the SC is discharged, whereas when $P < 0$, it is charged. The relationship between $v$ and $i$ is derived as follows.

Figure 1. The power-train of the HEV energy system.

Table 1. The main parameters of the FC.

| Parameters   | Value | Unit | Parameters   | Value | Unit |
|--------------|-------|------|--------------|-------|------|
| $R_{ohm}$    | 0.004 | $\Omega$ | $E_i$       | 0.9   | V    |
| $AF_{FC}$    | 204   | cm$^2$ | $N_0$       | 750   | –    |
| $T$          | 368   | K     | $R_g$       | 8314.47 | – |
| $F$          | 9.65e+4 | C/mol  | $B_1$       | 0.0478 | – |
| $PO_2$       | 2.2   | atm   | $B_2$       | 0.0136 | – |
| $PH_2$       | 2.0   | atm   | $\lambda_e$ | 0.00333 | – |
| Rated power  | 40    | kW    |              |       |      |

FC: fuel cell.
Provided that the impedance matching is satisfied in this model, the SC is capable of supplying maximum power, which is derived by:

\[
P_{\text{max}} = \frac{v^2}{4R} \text{ or } P_{\text{max}} = \frac{v^2}{R}.
\]

Thus, the terminal voltage \( v \) can be obtained by solving equation (10)

\[
\frac{dv}{dt} + RPv^{-1} + \frac{P}{C}v^{-1} = 0 \tag{10}
\]

Ignoring the negligible impact of the internal resistor, the SoC of the SC is defined in equation (11)

\[
\text{SoC} = \frac{v}{v_{\text{max}}} \times 100(\%)
\]

where \( v_{\text{max}} \) is the maximum voltage of SC. The main parameters of the SC per cell are listed in Table 2.

### Table 2. The main parameters of the SC per cell.

| Parameters          | Value | Unit | Parameters          | Value | Unit |
|---------------------|-------|------|---------------------|-------|------|
| \( C \)             | 1700  | F    | \( R \)             | 0.083 | \( \Omega \) |
| Storage capacity    | 160   | Wh   | Rated voltage       | 2.5   | V    |
| Number of cells     | 84    | –    |                     |       |      |

SC: supercapacitor.

QL-based HEV energy optimization

Reinforcement learning is introduced as the theoretical foundation of the proposed QLDR strategy. We first describe the key concepts of QL in HEV energy control. Then, a deterministic rule is designed and combined in the QL controller. Finally, the proposed QLDR real-time energy management strategy is given in this section.

**QL in HEV energy control**

Given an episode under the defined driving cycle, that is, the time-continuous sequence of the power demand from the HEV, the goal of the proposed algorithm is to complete the power allocation by satisfying the power demand while keeping the SoC of SC in the safe range and saving the H\(_2\) fuel consumption. To accomplish this, QL is introduced as the baseline to carry out energy management. QL belongs to the reinforcement learning family that learns by interacting with the environment. During the learning process, the QL-driven controller observes the state of the power system, such as the power demand and SoC of SC; then performs the action of power split between FC and SC; and calculates the reward value by assessing the safety range of SoC of SC. Finally, the value function that accumulates the total rewards over time is updated. When the value function converges, the learning process ends and the control policy is obtained. Using this connection can produce a lot of information about causality, behavioral consequences, and what should be done for higher rewards and achieving goals. To further explain the QL in HEV energy control, the key concepts applied in the proposed QLDR are formulated.

**Policy.** A policy assigns how learned agents behave in a given state. In other words, the state of the environment is perceived first, and then the strategy is mapped to the actions to be taken in these states. The policy is
generated from the Q-table, that is, a lookup table filled with value functions. Specifically, the Q-table is represented by a simple two-dimensional array that contains the SoC of SC and split coefficient of power demand to FC. When the agent is in one of these states, by checking the maximum value function of the state in the Q-table, the corresponding action is selected and performed.

**State space definition.** The instantaneous SoC of the SC, represented by SoC, is selected to represent the system state, which is a continuous variable. To discretize the state variables, SoC should be discretized by equations (12) and (13)

\[
\text{SoC} \rightarrow \left[ \frac{\text{SoC} - \text{SoC}_{\text{min}}}{d_1} + 1 \right] \quad (12)
\]

\[
d_1 = \frac{\text{SoC}_{\text{max}} - \text{SoC}_{\text{min}}}{\text{num}_{s}} - 1 \quad (13)
\]

where \(d_1\) represents the discretization degree; SoC\(_{\text{max}}\) and SoC\(_{\text{min}}\) represent the maximum and minimum SoC, respectively; and num\(_s\) represents the number of states. It is worth mentioning that after discretization, SoC no longer represents the SoC value, but a one-based searching index in the state space to the corresponding SoC value. By this transform technique, the state is not only discretized but also better indexed in the Q-table.

**Action space definition.** We choose the output power of the FC, \(a_i\), as the control action in this study. The same discretization technique is applied to the action dimension as can be seen in equations (14) and (15)

\[
a_i = k \times d_2 \quad (14)
\]

\[
d_2 = \frac{p_{\text{max}}}{\text{num}_{a}} - 1 \quad (15)
\]

where \(d_2\) is the discretization degree, \(k\) is the one-based index to the action, \(p_{\text{max}}\) represents the maximum power output of the FCs, and num\(_a\) represents the number of actions. The output power of the SC (\(p_t\)) can be calculated by subtracting \(a_i\) from the power demand.

**Reward definition.** Immediate reward evaluates the effect of the action at the current state. The control objectives of the HEV are to satisfy the power demand and minimize the fuel consumption, which can be summarized as maintaining the SoC of the SC in the safe range, because the SC can not only provide the FC with sufficient power but also recover the braking energy only when the SoC is within the normal range. Moreover, the Q-table with minimal fuel consumption is regarded as the best Q-table and kept in the training epochs. Keeping the objectives in mind, the reward function is defined as can be seen in equation (16), where \(r_t\) is the immediate reward at time \(t\). This definition can guarantee all the objectives mentioned above

\[
r_t = \begin{cases} 
0, & \text{SoC}_{\text{min}} \leq \text{SoC}_{t+1} \leq \text{SoC}_{\text{max}} \\
-1000, & \text{SoC}_{t+1} \leq \text{SoC}_{\text{min}} \text{ or } \text{SoC}_{t+1} \leq \text{SoC}_{\text{max}} 
\end{cases} 
\]

\[s.t. \quad p_t + p_{f} = p_d \quad (16)\]

where \(p_d\) is the power demand, \(p_t\) is the output power provided by the SC, and \(p_{f}\) is the output power of the FC.

**Value function.** Value function is an estimation of future total rewards at state \(s\) and action \(a\). It is calculated by updating the value of the two-dimensional Q-table according to the definition of one-dimensional state and action given in policy definition. Mathematically, it is formulated as the sum of future immediate rewards as can be seen in equation (17)

\[
Q(s_t, a_t) = E[r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots | s_t, a_t] 
\]

where \(Q(s_t, a_t)\) represents the value function obtained by taking action \(a_t\) at state \(s_t\). \(\gamma\) is the discount factor that attributes to the convergence of the infinite sum of rewards. \(Q^*\) represents the optimal value function, that is, the maximum accumulative reward; it is easy to prove that \(Q^*\) can be expressed by the Bellman equation which is decomposed into two parts as shown in equation (18)

\[
Q^*(s_t, a_t) = E[r_{t+1} + \gamma \max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1}) | s_t, a_t] 
\]

where the first part is the immediate reward \(r_{t+1}\), and the second part is the discounted value of successor state \(\gamma Q(s_{t+1}, a_{t+1})\). To obtain \(Q^*\), the Bellman equation is applied to iterate the value function as can be seen in equation (19), where \(\eta \in (0,1]\) is the learning rate. The value thus obtained gradually converges to the optimal action value function with the iteration of the algorithm.

\[
Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \eta \left( r_{t+1} + \gamma \max_{a_{t+1}} Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t) \right) 
\]

**Algorithm design.** The proposed QLDR for HEV energy management is presented in Algorithm 1. The Q-table is initialized by 0, which means the power demand is provided by the FC in default, and the learning process maximizes the reward function by tuning the action of SC. The outer loop represents the number of training epochs to update the value function, while the inner loop describes energy management policy at each step within the episode duration. The maximum number of the
When the epoch cycle begins. If the reward is repeated until the episode ends, that is, if the function is updated by the Bellman equation. This step after performing action, then switches back to the default set. Finally, the value function is updated by the Bellman equation. This step is repeated until the episode ends, that is, if the SoC is within the safe range; if not, that is, the reward is −1000, then the episode stops immediately and a new epoch cycle begins.

### Real-time energy management

The proposed QLDR for HEV energy management in section “Algorithm design” is implemented off-line, which means the agent is trained under the specific driving cycle. However, because of applying the deterministic rule, the converged agent is adaptive under different driving cycles for real-time EMS. For example, the agent is trained under the urban dynamometer driving schedule (UDDS) episode and then applied directly in the highway fuel economy certification test (HWFET) episode. Unlike the traditional driving cycle recognition–based algorithms, the proposed HEV energy management algorithm is a lightweight and high real-time method that does not depend on the driving pattern recognition. We aim at controlling the SoC of the SC in the safe range, and the agent is capable of carrying out the dynamic power management under other driving cycles. The framework for power system decision and control is described in Figure 3.

As we can see, there is a little difference between the learning module and the execution module. In the simulation environment, the agent tries to explore more information by action with the $\epsilon$–greedy algorithm. In this way, the agent enlarges the scope of cognition about the environment by going through the Q-table as complete as possible. However, in the real system, the agent no longer takes risks to obtain more information by $\epsilon$–greedy, but still receives reward from the environment to help adapt to different driving conditions.

#### Simulation results

**Off-line training**

To verify and evaluate the effectiveness of the proposed QLDR algorithm, this study uses the joint simulation environment of MATLAB and Advanced Vehicle Simulator (ADVISOR) to carry out simulation experiments. The main hyper-parameters of the QLDR algorithm involved in the simulation are summarized in Table 3. The specific values of the hyper-parameters are obtained by trial and error—especially the number of states and actions, which should be set carefully by trial and error because if the number is too small, the controller accuracy is too low; on the contrary, the calculation complexity is too high. In addition, it is worth mentioning that the value of $p^{\max}$ is an optional set determined by the rule, which is described in Algorithm 1. The QL-driven agent is trained under the driving cycle of HWFET. The power demand and the power allocation results in this driving condition are shown in Figure 4. We can see from the blue line that the SC

---

**Algorithm 1**: The proposed QLDR algorithm design

Initialize value function $Q(s, a) \leftarrow 0$
Initialize $p^{\text{max}} \leftarrow P_{\text{low}}$

for epoch = 1 to N do
    Reset environment: $s_0 = SoC$
    for $t = 1$ to $T$ do
        With probability $\epsilon$ select a random action $a_t$
        otherwise select $a_t = \arg \max Q(s_t, a_t)$
        Take action $a_t$, calculate $SoC_{t+1}$, and observe the reward $r_{t+1}$
        if $SoC_{t+1} < 0.6$ then
            $p^{\text{max}} = P_{\text{high}}$
        else
            $p^{\text{max}} = P_{\text{low}}$
        end
        if terminal $s$ then
            $Q(s_t, a_t) := Q(s_t, a_t) + \eta(r_{t+1} - Q(s_t, a_t))$
        else
            $Q(s_t, a_t) := Q(s_t, a_t) + \eta(r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))$
        end
        if $r_{t+1} = -1000$ then
            break
        end
    end
    Calculate $m_{\text{H2}}$
    Record $Q$-table with minimal $m_{\text{H2}}$
    if minimal $m_{\text{H2}}$ kept unchanged for 300 consecutive times
        break
    end
end
functions well as a power buffer. Here, the negative output power denotes that the SC has recovered the braking energy, and the positive power represents the auxiliary power to compensate for the instantaneous power demand, which assists the FC with less output burden.

The error distribution between the power demand and the hybrid FC/SC output power is shown in Figure 5. It is obvious that the QL-based controller acts well to satisfy the power demand with slight deviation. The SoC of the SC is maintained in the predefined range \([0.45, 0.95]\), which is given in Figure 6.

### Real-time application

In the real-time application, the trained agent under HWFET was directly tested under the four typical driving cycles mentioned by the Environmental Protection Agency (EPA). They are congested urban roads, flowing urban roads, and subway and highway, which are represented by Manhattan bus drive cycle (MBDC), EPA urban dynamometer driving schedule (UDDS), West Virginia suburban driving schedule (WVUSUB), and HWFET, correspondingly. The characteristics of each driving cycle are described in Table 4.

The simulation results under the four combined driving cycles are provided in the following figures. Figure 7 gives the real-time power allocation of the power demand. The error distribution between the power demand and the hybrid FC/SC output power is shown in Figure 8. It can be seen that the error is within \([-4 \times 10^{-12}, 4 \times 10^{-12}]\), which means that the proposed QL-based controller performs well on-line to satisfy the power demand with the slight deviation.

---

**Table 3. The hyper-parameters of the proposed QLDR.**

| Hyper-parameters | Value | Hyper-parameters | Value |
|------------------|-------|------------------|-------|
| \(\eta\)         | 0.0001| \(\gamma\)       | 0.99  |
| \(\epsilon\)     | 0.1   | \(N\)            | 2000  |
| \(num_s\)        | 61    | \(num_a\)        | 31    |
| \(SoC_{\text{min}}\) | 0.45  | \(SoC_{\text{max}}\) | 0.95  |
| \(P_{\text{low}}\) | 17 kW | \(P_{\text{high}}\) | 24 kW |

QLDR: Q-learning strategy with deterministic rule.

---

**Table 4. Four typical driving cycles.**

| Driving cycles | Characteristics |
|----------------|-----------------|
| MBDC           | Low speed go-and-stop traffic driving condition |
| UDDS           | City driving condition |
| WVUSUB         | Suburban driving condition |
| HWFET          | Highway driving condition under 60 mph |

MBDC: Manhattan bus drive cycle; UDDS: urban dynamometer driving schedule; WVUSUB: West Virginia suburban driving schedule; HWFET: highway fuel economy certification test.
Figure 9 shows the real-time SoC of the SC. These results verify the ability of the proposed QL-driven agent to perform power allocation between the hybrid energy sources satisfactorily and keep the SoC in the acceptable range on-line, which is within [0.58, 0.91].

Performance comparison

To evaluate the superiority of the proposed QLDR, the adaptive fuzzy-based algorithm in EMS and the pure QL-based control strategy have been selected for comparison. For effective comparison, the maximum power output of the FC in the pure QL algorithm is set as either 17 kW, that is, 17 kW QL or 24 W, that is, 24 kW QL, as the deterministic rule defines. Two aspects of the simulation are taken into consideration, including the adaptation to the complex driving cycles and the optimization of the FC’s load and current fluctuation. First, as we can see in Figure 10, the SoC in the 17 kW QL is out of safe range in the driving duration, which means the SC has run out of energy and is no longer able to support the FC with auxiliary power. However, the proposed algorithm, fuzzy EMS of the 24 kW QL, has successfully achieved energy management under the continuous high-power output conditions with SoC kept in the safe range, which means they are more adaptive to the extreme driving conditions.

Figure 8. The power error distribution on-line.

Figure 9. The SoC of the SC on-line.

Figure 10. SoC of the SC.

Figure 11. Load of the FC.

Figure 12. Current of the FC.

Second, by applying the deterministic rule, the proposed algorithm has achieved less load and current fluctuation. As we can see in Figures 11 and 12, there are more load and current fluctuations in the 24 kW QL and the adaptive fuzzy EMS than the proposed algorithm. In Figures 13 and 14, the FC’s load and current variance rate over time are drawn, which arrives at the same conclusion in a more straightforward way. The root mean square (RMS) of the variance rate in the FC’s load and current is listed in Table 5, where we can see that the proposed QLDR achieves the least RMS of the variance rate in the FC’s load and current. The simulation results have proved the effectiveness of the proposed QLDR.
In addition, comparisons have also been performed by the three methods by taking into consideration the fuel economy. The contrastive experiments are carried out on the share dataset, that is, HWFET driving cycle for the training and combined driving cycles for the testing. The results are described in Table 6, where we can find that a small progress in the fuel consumption has been achieved by the proposed method compared to the 24 kW QL, but 9.03% of the fuel consumption is saved when compared to the fuzzy EMS, which further proves the effectiveness of the deterministic rule to QL and the superiority of the proposed method to the fuzzy EMS.

Conclusion

In this work, a novel QL method with deterministic rule is proposed for the real-time HEV energy management. To enhance the adaptation to different driving cycles, especially the extreme conditions, the deterministic rule is applied to the conventional QL algorithm as a complement to the policy. What’s more, by employing two optional sets of the maximum FC output power, which is determined by the deterministic rule, less load and current fluctuation of the FC are achieved because the smaller set of maximum output power will limit the fluctuation. The proposed algorithm is trained under the HWFET driving cycle and tested under four typical combined driving cycles. Simulation results illustrate that compared with the driving pattern–based fuzzy logic controller, the proposed QL-driven controller is more lightweight and effective. More importantly, 9.03% reduction of the fuel consumption and less load and current fluctuation of the FC have been achieved, which help to improve the fuel economy and prolong the lifetime of the FC. In addition, to prove the superiority of the proposed QLDR to the conventional QL algorithm, simulation has been conducted for performance comparison. The results show that there is small difference in the fuel consumption, but the load and current fluctuation of the FC have been greatly reduced by applying the deterministic rule. In the future, to face the open environment, instead of taking the power demand of the HEV as the only factor, information on more driving conditions including driver’s behavior will be taken into consideration to make the EMS perform well under more complex driving conditions. Moreover, the uncertainties in the system, such as the difference between the training model and the real model, will be researched in depth.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the National Natural Science Foundation of China under Grant 61603337 and the Zhejiang Province Natural Science Fund under Grant LY19F030009, Ningbo Science, and Technology Innovation 2025 Major Projects (2019B10109, 2019B10116).

ORCID iDs

Jili Tao https://orcid.org/0000-0001-7095-8968
Ridong Zhang https://orcid.org/0000-0003-4184-3955

Li et al.

1501

Table 6. The fuel consumption for 3 methods.

| Driving cycles | Proposed QLDR (g) | 24 kW QL (g) | Fuzzy EMS (g) |
|----------------|------------------|--------------|---------------|
| Training       | 0.2128           | 0.2195       | 0.2232        |
| Testing        | 0.5340           | 0.5441       | 0.5870        |

QLDR: Q-learning strategy with deterministic rule; QL: Q-learning; EMS: energy management system.

Table 5. The RMS of the variance rate in the FC’s load and current for three methods.

| RMS   | Proposed QLDR (kW/s) | 24 kW QL (kW/s) | Fuzzy EMS (kW/s) |
|-------|----------------------|-----------------|------------------|
| Vp    | 3.68                 | 7.62            | 3.99             |
| Vl    | 11.83                | 24.41           | 14.34            |

RMS: root mean square; FC: fuel cell; QLDR: Q-learning strategy with deterministic rule; QL: Q-learning; EMS: energy management system.

Figure 13. Variance rate of the FC’s load.

Figure 14. Variance rate of the FC’s current.

Table 6. The fuel consumption for 3 methods.

| Driving cycles | Proposed QLDR (g) | 24 kW QL (g) | Fuzzy EMS (g) |
|----------------|------------------|--------------|---------------|
| Training       | 0.2128           | 0.2195       | 0.2232        |
| Testing        | 0.5340           | 0.5441       | 0.5870        |

QLDR: Q-learning strategy with deterministic rule; QL: Q-learning; EMS: energy management system.
References

1. Xiong R, Sun F, Chen Z, et al. A data-driven multi-scale extended Kalman filtering based parameter and state estimation approach of lithium-ion polymer battery in electric vehicles. *Appl Energ* 2014; 113(1): 463–476.

2. Al-Alawi BM and Bradley TH. Total cost of ownership, payback, and consumer preference modeling of plug-in hybrid electric vehicles. *Appl Energ* 2013; 103(1): 488–506.

3. Lv C, Liu Y, Hu X, et al. Simultaneous observation of hybrid states for cyber-physical systems: a case study of electric vehicle powertrain. *IEEE T Cybernetics* 2018; 48(8): 2357–2367.

4. Wegmann R, Döge V and Sauer DU. Assessing the potential of a hybrid battery system to reduce battery aging in an electric vehicle by studying the cycle life of a graphite/NCA high energy and a LTO/metal oxide high power battery cell considering realistic test profiles. *Appl Energ* 2018; 226: 197–212.

5. Meacham JR, Jabbari F, Brouwer J, et al. Analysis of stationary vehicle dynamic ramping capabilities and ultra-capacitor energy storage using high resolution demand data. *J Power Sources* 2006; 156(2): 472–479.

6. Xiong R, Zhang Y, He H, et al. A double-scale, particle-filtering state prediction algorithm for lithium-ion batteries. *IEEE T Ind Electron* 2017; 65(2): 1526–1538.

7. Amjadi Z and Williamson SS. Prototype design and controller implementation for a battery-ultracapacitor hybrid electric vehicle energy storage system. *IEEE T Smart Grid* 2012; 3(1): 332–340.

8. Camara MB, Gualous H, Gustin F, et al. Design and new control of DC/DC converters to share energy between supercapacitors and batteries in hybrid vehicles. *IEEE T Veh Technol* 2008; 57(5): 2721–2735.

9. Lukic SM, Cao J, Bansal RC, et al. Energy storage systems for automotive applications. *IEEE T Ind Electron* 2008; 55(6): 2258–2267.

10. Emadi A, Rajashekar K, Williamson SS, et al. Topological overview of hybrid electric and fuel cell vehicular power system architectures and configurations. *IEEE T Veh Technol* 2005; 54(3): 763–770.

11. Khaligh A and Li Z. Battery, ultracapacitor, fuel cell, and hybrid energy storage systems for electric, hybrid electric, fuel cell, and plug-in hybrid electric vehicles: state of the art. *IEEE T Veh Technol* 2010; 59(6): 2806–2814.

12. Salmasi FR. Control strategies for hybrid electric vehicles: evolution, classification, comparison, and future trends. *IEEE T Veh Technol* 2007; 56(5): 2393–2404.

13. Jalil N, Kheir NA and Salman M. A rule-based energy management strategy for a series hybrid vehicle. In: *Proceedings of the 1997 American control conference*, Albuquerque, NM, 6 June 1997.

14. Phillips AM, Jankovic M and Bailey KE. Vehicle system controller design for a hybrid electric vehicle. In: *Proceedings of the IEEE international conference on control applications*, Anchorage, AK, 27 September 2000, pp. 297–302. New York: IEEE.

15. Hemi H, Ghouili J and Cheriti A. A real time fuzzy logic power management strategy for a fuel cell vehicle. *Energ Convers Manage* 2014; 80(4): 63–70.

16. Kısacıkoglu MC, Uzunoglu M and Alam MS. Load sharing using fuzzy logic control in a fuel cell/ultracapacitor hybrid vehicle. *Int J Hydrogen Energ* 2009; 34(3): 1497–1507.

17. Erdinc O, Vural B and Uzunoglu M. A wavelet-fuzzy logic based energy management strategy for a fuel cell/battery/ultra-capacitor hybrid vehicular power system. *J Power Sources* 2009; 194(1): 369–380.

18. Wu L, Wang Y, Yuan X, et al. Multiobjective optimization of HEV fuel economy and emissions using the self-adaptive differential evolution algorithm. *IEEE T Veh Technol* 2011; 60(6): 2458–2470.

19. Opila DF, Wang X, Mcgee R, et al. An energy management controller to optimally trade off fuel economy and drivability for hybrid vehicles. *IEEE T Contr Syst T* 2012; 20(6): 1490–1505.

20. Fletcher T, Thiring R and Watkinson M. An energy management strategy to concurrently optimise fuel consumption & PEM fuel cell lifetime in a hybrid vehicle. *Int J Hydrogen Energ* 2016; 41(46): 21503–21515.

21. Paganeli G, Ercole G, Brahma A, et al. General supervisory control policy for the energy optimization of charge-sustaining hybrid electric vehicles. *JSAE Rev* 2001; 22(4): 511–518.

22. Paganeli G, Delprat S, Guerra TM, et al. Equivalent consumption minimization strategy for parallel hybrid powertrains. In: *Proceedings of the vehicular technology conference (VTC spring 2002)*, vol. 4, Birmingham, AL, 6–9 May 2002, pp. 2076–2081. New York: IEEE.

23. Fang L, Xu G, Li T, et al. Optimal control of parallel hybrid electric vehicles. *IEEE T Contr Syst T* 2004; 12(3): 352–363.

24. Borhan H, Vahidi A, Phillips AM, et al. MPC-based energy management of a power-split hybrid electric vehicle. *IEEE T Contr Syst T* 2012; 20(3): 593–603.

25. Chen BC, Wu YY and Tsai HC. Design and analysis of power management strategy for range extended electric vehicle using dynamic programming. *Appl Energ* 2014; 113(1): 1764–1774.

26. Larsson V, Johannesson L and Egardt B. Analytic solutions to the dynamic programming sub-problem in hybrid vehicle energy management. *IEEE T Veh Technol* 2014; 64(4): 1458–1467.

27. Zhang R, Tao J and Zhou H. Fuzzy optimal energy management for fuel cell and supercapacitor systems using neural network based driving pattern recognition. *IEEE T Fuzzy Syst* 2018; 27(1): 45–57.

28. Glavic M, Fonteneau R and Ernst D. Reinforcement learning for electric power system decision and control: past considerations and perspectives. *IFAC-PapersOnLine* 2017; 50(1): 6918–6927.

29. Hu Y, Li W, Xu K, et al. Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning. *Appl Sci* 2018; 8(2): 187.

30. Liu T, Zou Y, Liu D, et al. Reinforcement learning–based energy management strategy for a hybrid electric tracked vehicle. *Energies* 2015; 8(7): 7243–7260.

31. Qi X, Wu G, Boriboonsomsin K, et al. Data-driven reinforcement learning–based real-time energy management system for plug-in hybrid electric vehicles. *Transp Res Record* 2016; 2572: 1–8.

32. Yang X, Liu X and Li Z. Multimodel approach to robust identification of multiple-input single-output nonlinear time-delay systems. *IEEE T Ind Inform* 2019; 16: 2413–2422.
33. Yang X, Yin S and Kaynak O. Robust identification of LPV time-delay system with randomly missing measurements. *IEEE T Syst Man Cy: S* 2017; 48(12): 2198–2208.
34. Yang X and Liu S. Variational Bayesian inference for FIR models with randomly missing measurements. *IEEE T Ind Electron* 2016; 64(5): 4217–4225.
35. Baumann BM, Washington G, Glenn BC, et al. Mechatronic design and control of hybrid electric vehicles. *IEEE/ASME T Mech* 2000; 5(1): 58–72.
36. Guzzella L, Amstutz A and Grob F. Optimal operation strategies for hybrid power-trains. *IFAC Proc Vol* 1998; 31(1): 93–98.
37. Hochgraf CG, Ryan MJ and Wiegman HL. Engine control strategy for a series hybrid electric vehicle incorporating load-leveling and computer controlled energy management. SAE technical paper 960230, 1996.