Paper ID APEN-MIT-2022_4945
Applied Energy Symposium: MIT A+B
July 5-8, 2022 • Cambridge, USA

Intelligent Battery Health-Aware Energy Management Strategy for Hybrid Electric Bus: A Deep Reinforcement Learning Method

Ruchen Huang
National Engineering Laboratory for Electric Vehicles,
Beijing Institute of Technology
Beijing, China
Ruchen_Huang@163.com

Hongwen He*
National Engineering Laboratory for Electric Vehicles,
Beijing Institute of Technology
Beijing, China
hwhebit@bit.edu.cn

Abstract—This paper proposes an intelligent battery health-aware energy management strategy (EMS) for the hybrid electric bus (HEB) with a deep reinforcement learning (DRL) method. Firstly, an EMS based on twin delayed deep deterministic policy gradient (TD3) algorithm considering battery health is innovatively designed to minimize the total operating cost of the HEB. Secondly, the superiority of the proposed EMS over the state-of-the-art deep deterministic policy gradient (DDPG) based strategy is validated. Simulation results show that the proposed EMS accelerates the convergence by 24.00% and reduces the total operating cost by 9.58% compared with the EMS based on DDPG.

Keywords—hybrid electric bus, energy management, battery health, deep reinforcement learning, twin delayed deep deterministic policy gradient (TD3)

I. INTRODUCTION

Being famous for long-range and low-emission, hybrid electric buses (HEBs) equipped with appropriate energy management strategies (EMSs) provide a popular solution for the new requirements of electrification and decarbonization of the urban public transport [1]. EMSs have been studied in a large number of research works and can be generally classified into three categories including rule-based strategies [2], optimization-based strategies [3], and reinforcement learning-based strategies [4].

Rule-based strategies mainly include logic threshold strategy and fuzzy logic strategy, having low computation cost and strong practicability [5]. However, the designed control rules are extracted from engineering intuition, making it far away from satisfactory control performance. Optimization-based strategies consist of global optimization strategies such as dynamic programming (DP) and real-time optimization strategies such as model predictive control (MPC) and equivalent consumption minimization strategy (ECMS), which can obtain the global optimal or near-optimal control performance [6]. Nevertheless, the model-based attribute hinders further improvement of the optimization effect.

More recently, reinforcement learning (RL) especially deep reinforcement learning (DRL) has attracted a lot of research attention owing to its model-free attribute and great adaptability [7]. From Q-learning to deep deterministic policy gradient (DDPG), energy management for HEBs based on DRL has become a research hotspot in recent years [8]. However, there are still three major deficiencies waiting for further overcoming:

1) Most of the research on energy management for HEBs is based on standard driving cycles, which is unfavorable to improving the fuel economy of HEBs operating on fixed bus routes.

2) Most of the research works focusing on DRL-based EMSs tend to optimize the fuel economy unilaterally. Since the replacement of the battery system costs a lot, it is vital to take the battery's health into account.

3) Although DDPG is the state-of-the-art DRL algorithm for energy management, it still suffers from several inherent defects. The designing of more intelligent EMSs requires more advanced DRL algorithms.

To bridge the aforementioned research gaps, this paper proposes a DRL-based and battery health-aware EMS for an urban power-split HEB. This paper encompasses three perspectives that may contribute to relevant research:

1) The real-world velocity data are used as the training and testing datasets for the DRL agent to evaluate the practical operating costs of the HEB accurately.

2) Twin delayed deep deterministic policy gradient (TD3) is used as a more efficient DRL method to further explore the energy conservation potential of the DRL-based EMSs.

3) The degradation of the onboard battery system is especially considered with precise modeling in the form of second-order RC.

To the best of our knowledge, this is a pioneer research work to adopt the TD3 algorithm for energy management of the urban HEB with special awareness of battery health.

The remainder of this paper is organized as follows. The powertrain modeling of the power-split HEB with a detailed battery aging model is presented in Section 2. In Section 3, a TD3-based EMS to deal with the optimization problem is formulated. Simulation results are analyzed in Section 4. Finally, Section 5 draws major conclusions.

II. POWERTRAIN MODELING

A. HEB Configuration

The adopted HEB powertrain is a power-split configuration, which is shown in Fig.1. The power-split device is composed of a dual planetary gear set. The detailed description and main parameters of the configuration can refer to our previous research [9].

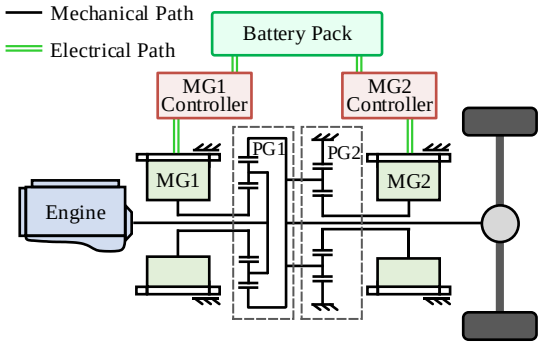


Fig. 1. Powertrain configuration.

B. Vehicle Dynamics Modeling

The backward vehicle simulation model is adopted to simplify the calculation, and the driving force demand can be calculated as [10]:

$$F_t = mgf\cos\varphi + mg\sin\varphi + \frac{C_d A}{21.15}v^2 + \delta m\frac{\mathrm{d}v}{\mathrm{d}t} \tag{1}$$

where F_t is the driving force demand, m is the vehicle mass, g is the gravity acceleration, f is the rolling resistance coefficient, φ is the angle of road slope, C_d is the drag coefficient, A is the front area, v is the velocity, δ is the rotational mass coefficient.

The coupling relationship brought by the power-split device can be formulated as:

$$\begin{cases} T_{eng} = (1+\dfrac{1}{k_1})T_{out} - \dfrac{(1+k_1)(1+k_2)}{k_1}T_{mg2} \\ \omega_{eng} = \dfrac{1}{1+k_1}\omega_{mg1} + \dfrac{k_1}{(1+k_1)(1+k_2)}\omega_{mg2} \\ T_{out} = \dfrac{F_t \cdot R_{wh}}{i_f} \end{cases} \tag{2}$$

where T_{eng}, T_{mg1}, T_{mg2} are the torque of the engine, MG1, and MG2. ω_{eng}, ω_{mg1}, ω_{mg2} are the rotational speed. T_{out} is the output torque of the power-split device. k_1 and k_2 are the gear ratio of the ring gear to the sun gear of PG1 and PG2 respectively. R_{wh} and i_f are the rolling radius and final gear ratio respectively.

C. Power Components Modeling

The diesel engine, MG1, and MG2 are modeled in the form of quasi-static models. Since the efficiency of MG1 and MG2 are quite higher than that of the diesel engine, the optimization work is mainly carried out for the engine in the HEB, only the efficiency map of the engine is displayed in this paper, which is shown in Fig. 2.

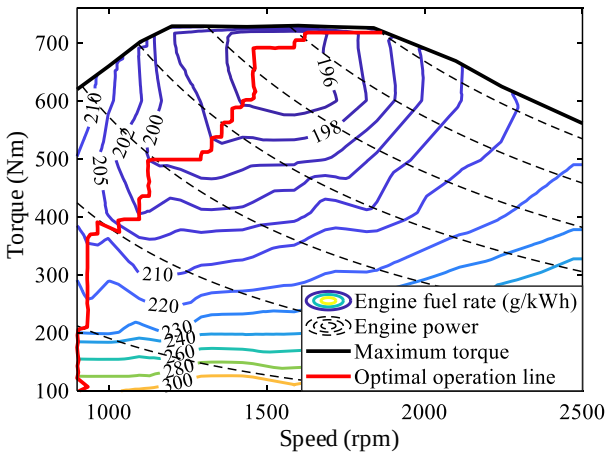


Fig. 2. Engine efficiency map.

D. Battery Aging Modeling

The second-order RC model shown in Fig. 3 is adopted in this paper. The governing equations are formulated as [11]:

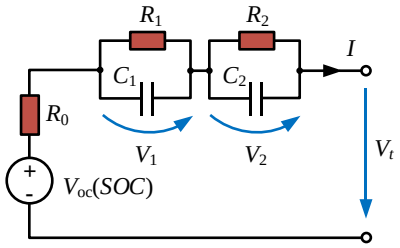


Fig. 3. Second-order RC model.

$$\frac{\mathrm{d}SOC(t)}{\mathrm{d}t} = -\frac{I(t)}{3600Q_{bat}} \tag{3}$$

$$\frac{\mathrm{d}V_1(t)}{\mathrm{d}t} = -\frac{V_1(t)}{R_1(t)C_1(t)} + \frac{I(t)}{C_1(t)} \tag{4}$$

$$\frac{\mathrm{d}V_2(t)}{\mathrm{d}t} = -\frac{V_2(t)}{R_2(t)C_2(t)} + \frac{I(t)}{C_2(t)} \tag{5}$$

$$V_t(t) = V_{oc}(SOC) - V_1(t) - V_2(t) - R_0(t)I(t) \tag{6}$$

where SOC is the state of charge of the battery, I is the battery current, V_{oc} is the open-circuit voltage, V_t is the terminal voltage, Q_{bat} is the battery capacity, R_0 is the internal resistance, R_1, R_2, and C_1, C_2 are the equivalent resistance and capacitance of the two RC branches, V_1 and V_2 are the polarization voltage across the two RC branches.

The ANR26650M1 battery is adopted and its related parameters have been identified and validated scientifically in [12]. The capacity loss can be estimated as [13]:

$$\Delta Q = B(c)\cdot\exp\left(\frac{-E_a(c)}{R_g\cdot(T_a+273.15)}\right)\cdot Ah(c)^z \tag{7}$$

where ΔQ is the capacity loss, B is the pre-exponential factor which is listed in Table 2. c is the C-rate, R_g is the universal gas constant, T_a is the internal average temperature in the unit of ℃, Ah is the discharged ampere-hour throughput, z is the power-law factor, E_a is the activation energy.

The end-of-life (EOL) capacity loss of the onboard LIB is usually considered 20%, so the total discharged throughput $Ah(c,T_a)$ and the total number of cycles $N(c,T_a)$ before EOL can be calculated as:

$$Ah(c,T_a)=\left[20\Big/\left(B(c)\cdot\exp\left(\frac{-E_a(c)}{R_g\cdot(T_a+273.15)}\right)\right)\right]^{\frac{1}{z}} \tag{8}$$

$$N(c,T_a)=\frac{3600Ah(c,T_a)}{Q_{bat}} \tag{9}$$

The state-of-health (*SOH*) of the battery is defined as:

$$\frac{\mathrm{d}SOH(t)}{\mathrm{d}t}=-\frac{|I(t)|}{2N(c,T_a)Q_{bat}} \tag{10}$$

III. EMS Based on TD3

A. Formulation of TD3 Algorithm

The TD3 algorithm is derived from DDPG to effectively overcome several defects that DDPG suffers from, including overestimation, overfitting, and poor update [14]. Limited by pages, only the main differences between TD3 and DDPG are presented in this paper.

Firstly, two critic networks $Q_1(s,a|\theta^{Q_1})$ and $Q_2(s,a|\theta^{Q_2})$ are designed and each critic network corresponds to a target network $Q_1'(s,a|\theta^{Q_1'})$ and $Q_2'(s,a|\theta^{Q_2'})$ respectively, thus two different action-values of the next state can be calculated:

$$\begin{cases}Q_1'\left(s_{t+1},\hat{a}_{t+1}\middle|\theta^{Q_1'}\right)=Q_1'\left(s_{t+1},\mu'\left(s_{t+1}\middle|\theta^{\mu'}\right)\middle|\theta^{Q_1'}\right)\\ Q_2'\left(s_{t+1},\hat{a}_{t+1}\middle|\theta^{Q_2'}\right)=Q_2'\left(s_{t+1},\mu'\left(s_{t+1}\middle|\theta^{\mu'}\right)\middle|\theta^{Q_2'}\right)\end{cases} \tag{11}$$

Then clipped double Q-learning is adopted to eliminate the overestimation in DDPG:

$$y_t=r(s_t,a_t)+\gamma\min_{i=1,2}Q_i'\left(s_{t+1},\hat{a}_{t+1}\middle|\theta^{Q_i'}\right) \tag{12}$$

Secondly, the clipped normal distribution noise is added to the output actions to eliminate the overfitting:

$$\hat{a}_{t+1}=\mu'\left(s_{t+1}\middle|\theta^{\mu'}\right)+\xi,\quad \xi\sim\mathrm{clip}\left(\mathcal{N}\left(0,\sigma^2\right),-c,c\right) \tag{13}$$

Thirdly, the updating frequency of the actor and target networks is reduced and their update can be completed only when the critic networks have been updated for a fixed number of steps.

B. TD3-based EMS

To comprehensively reflect the operating characteristic of the urban HEB, the state space is set to be composed of velocity, acceleration, battery *SOC*, and battery *SOH*:

$$S=[v,acc,SOC,SOH] \tag{14}$$

The action variable can be solely set as the engine output power:

$$A=\left[P_e\middle|P_e\in[0,140\mathrm{kW}]\right] \tag{15}$$

The optimization objective is to reduce the total operating cost including fuel consumption and battery aging, as well as to sustain the *SOC* within a certain optimal fluctuating range, hence the reward function can be designed as:

$$R=w_1\cdot\dot{m}_{fuel}(t)+w_2\cdot\Delta SOH(t)+\varepsilon\cdot\left[SOC(t)-SOC_{tar}\right]^2 \tag{16}$$

where $\dot{m}_{fuel}$ is the fuel consumption rate, ΔSOH is the battery degradation rate, SOC_{tar} is the *SOC* target value which is set to 0.5. w_1 and w_2 are the unit price of the diesel oil and the LIB replacement cost, which are set to 6.7 CNY/L and 1500 CNY/kWh (Note: CNY means Chinese Yuan) respectively. ε is the weight factor of *SOC* sustaining.

The optimization is subjective to the physical constraints of the powertrain system:

$$\begin{cases}\omega_{eng\_min}\le\omega_{eng}\le\omega_{eng\_max},T_{eng\_min}\le T_{eng}\le T_{eng\_max}\\ \omega_{mg1\_min}\le\omega_{mg1}\le\omega_{mg1\_max},T_{mg1\_min}\le T_{mg1}\le T_{mg1\_max}\\ \omega_{mg2\_min}\le\omega_{mg2}\le\omega_{mg2\_max},T_{mg2\_min}\le T_{mg2}\le T_{mg2\_max}\\ I_{cell\_min}\le I_{cell}\le I_{cell\_max}\end{cases} \tag{17}$$

where the subscripts max and min represent the upper and lower bound of each physical quantity.

Accordingly, the control framework of the TD3-based EMS is shown in Fig. 4.

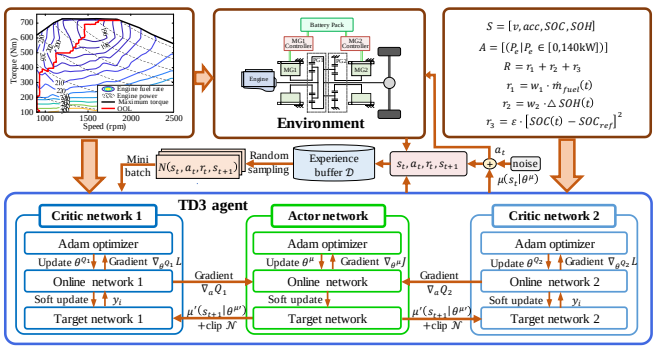
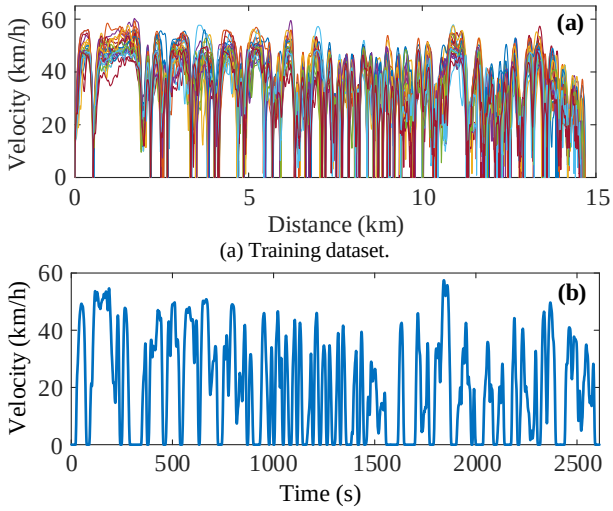


Fig. 4. Control framework of the TD3-based EMS.

IV. Results and Discussion

A. Conditions of Validation

In this paper, the real-world velocity data collected from a fixed bus route that the urban HEB operating on in Zhengzhou, China, is adopted as the training dataset for the proposed DRL-based EMS [15]. Moreover, a reconstructed driving cycle that reflects the driving characteristics and traffic scenarios of the test bus route is used as the testing dataset. The training dataset and testing dataset are shown in Fig. 5.



(a) Training dataset.

(b) Testing dataset.

Fig. 5. Training dataset and Testing dataset.

In addition, the EMSs based on TD3 with and without the consideration of battery health are represented as TD3@FH and TD3@F respectively in the following. Similarly, two EMSs based on DDPG are represented as DDPG@FH and DDPG@F, respectively.

B. *Validation of Convergence Speed*

The mean reward of each training episode is a representative indicator to reflect the convergence process. The mean rewards of EMSs based on TD3 and DDPG are shown in Fig. 6. Since the reward function is set as the positive value, the mean reward close to zero means satisfactory convergence performance. It is shown in Fig. 6 that it takes 50 episodes for the DDPG-based EMS to get the convergence, while it only takes 38 episodes for the EMS based on TD3. Benefiting from the delay updating mechanism for the actor-network and target-networks in the TD3 algorithm, the convergence speed of the TD3-based EMS is improved by 24.00% to that of the DDPG-based EMS.

Besides, it also can be found in Fig. 6 that the mature mean reward of the EMS based on TD3 is obviously less than that of the DDPG-based EMS, demonstrating a better learning ability, which attributes to the fact that the usage of clipped double Q-learning and the adding of noise to target actions in the TD3 algorithm have eliminated the overestimating and overfitting in DDPG, thus achieving a quite satisfactory learning ability.

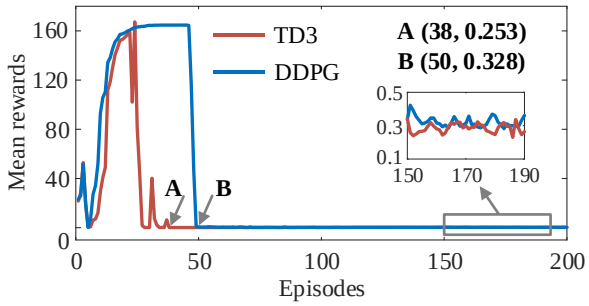


Fig. 6. Comparison of mean rewards.

C. *Validation of Degradation Control*

The *SOC* of the battery system regarding different EMSs is shown in Fig. 7. It is shown that all *SOC* trajectories are maintained effectively within the range of 0.45 to 0.55. Besides, the *SOC* trajectories of EMSs considering battery health fluctuate within a comparatively narrower range than that of the EMSs neglect battery health, which is favorable for decreasing the degradation of the battery.

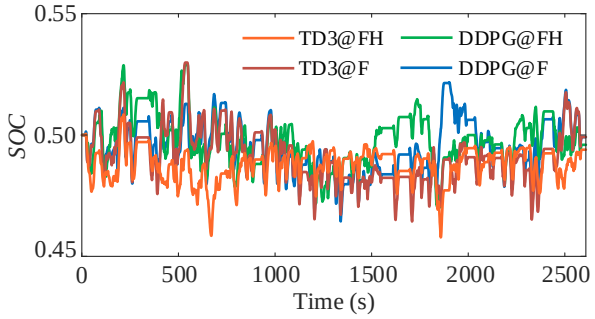


Fig. 7. *SOC* trajectories of different EMSs.

The *SOH* of the battery system regarding different EMSs is shown in Fig. 8. It is shown that the neglecting of battery health aggravates battery aging obviously, which can be explained that the current will be out of control to some extent when the battery health is overlooked. The proposed EMS achieves the best degradation control and reduces the degradation by 57.54% when battery health is considered. Moreover, although DDPG@FH considers battery health, its degradation control performance is 34.24% inferior to that of TD3@FH, due to the superior learning ability of TD3.

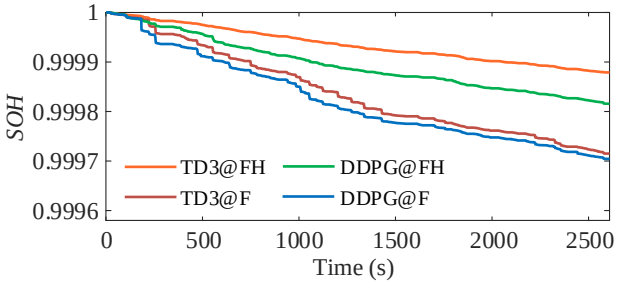
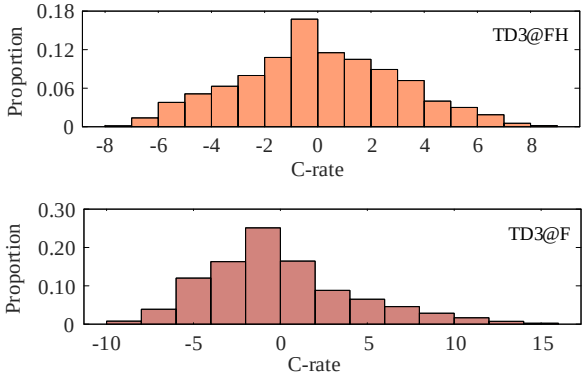


Fig. 8. *SOH* trajectories of different EMSs.

In this paper, the battery internal temperature is assumed to be constant under the assumption that the battery thermal management is well addressed. Therefore, the task of the degradation management is to control the C-rate. It is worth noting that the severity factor smaller than 4 is regarded as the healthy condition of the battery used in this paper, corresponding to the C-rate of 8.65. The distribution of the C-rate when it is not zero regarding different EMSs is shown in Fig. 9. It can be seen that the proposed EMS can ensure that the battery system works healthily. DDPG@FH could not make the battery completely works within the healthy region even though battery health is considered.



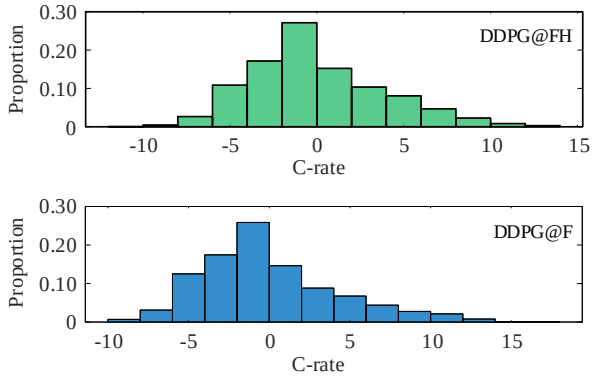


Fig. 9. Distribution of C-rate regarding different EMSs.

D. Validation of Cost Optimization

The fuel economy of different EMSs is compared in Fig. 10. It can be seen that EMSs based on TD3 achieve better fuel economy than DDPG-based EMSs. More importantly, considering battery health causes extra fuel consumption. It can be explained that the battery degradation constraint item in the reward function attempts to protect the battery from being used abusively, therefore the fuel consumption will be increased accordingly to satisfy the driving power demand.

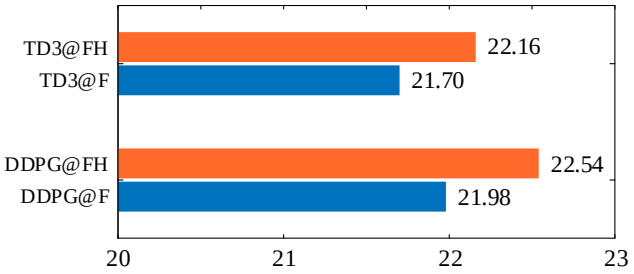


Fig. 10. Comparison of fuel economy regarding different EMSs.

The total operating costs of all EMSs under the testing dataset are listed in Table 1. The proposed EMS owns the minimum total cost among all EMSs. When considering battery health, the total costs have been reduced obviously with a slight sacrifice of fuel economy. The proposed EMS reduces the total cost by 8.06% in comparison with TD3@F, and the reduction proportion even reaches 9.58% to DDPG@F which is regarded as the most representative state-of-the-art DRL-based EMS.

TABLE I. TOTAL OPERATING COSTS OF DIFFERENT EMSS

EMSs	Fuel cost (CNY)	Aging cost (CNY)	Total cost (CNY)	Performance
TD3@FH	21.67	1.82	23.49	100%
TD3@F	21.23	4.32	25.55	91.94%
DDPG@FH	22.04	2.77	24.81	94.68%
DDPG@F	21.51	4.47	25.98	90.42%

V. CONCLUSION

This paper proposes a DRL-based and battery health-aware energy management strategy for an urban HEB. TD3 is adopted as a more advanced DRL method to further explore the energy conservation potential of the HEB. Real-world velocity data collected from a fixed bus route is adopted as the training dataset for the DRL agents. The battery health is especially considered along with the fuel consumption. Simulation results show that the TD3-based strategy can accelerate the convergence speed by 24.00% in comparison with the EMS based on DDPG. Besides, when battery health is considered, the proposed strategy decreases the battery degradation by 57.54% compared with the battery health-neglecting strategy based on TD3 with a 2.12% sacrifice in fuel economy. Moreover, the proposed strategy reduces the total operating cost by 9.58% in comparison with the existing state-of-the-art strategy based on DDPG.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 52172377), and in part by the National Natural Science Foundation of China (Grant No. U1864205).

VI. REFERENCES

[1] Guo, H., Wang, X., & Li, L. (2019). State-of-charge-constraint-based energy management strategy of plug-in hybrid electric vehicle with bus route. Energy Conversion and Management, 199, 111972.

[2] Zhou, Y., Ravey, A., & Péra, M. C. (2020). Multi-mode predictive energy management for fuel cell hybrid electric vehicles using Markov driving pattern recognizer. Applied Energy, 258, 114057.

[3] Zhang, Z., He, H., Guo, J., & Han, R. (2020). Velocity prediction and profile optimization based real-time energy management strategy for Plug-in hybrid electric buses. Applied Energy, 280, 116001.

[4] Zhang, X., Guo, L., Guo, N., Zou, Y., & Du, G. (2021). Bi-level energy management of plug-in hybrid electric vehicles for fuel economy and battery lifetime with intelligent state-of-charge reference. Journal of Power Sources, 481, 228798.

[5] Maino, C., Misul, D., Musa, A., & Spessa, E. (2021). Optimal mesh discretization of the dynamic programming for hybrid electric vehicles. Applied Energy, 292, 116920.

[6] Zhang, F., Yang, F., Xue, D., & Cai, Y. (2019). Optimization of compound power split configurations in PHEV bus for fuel consumption and battery degradation decreasing. Energy, 169, 937-957.

[7] Li, Y., He, H., Khajepour, A., Wang, H., & Peng, J. (2019). Energy management for a power-split hybrid electric bus via deep reinforcement learning with terrain information. Applied Energy, 255, 113762.

[8] Ganesh, A. H., & Xu, B. (2022). A review of reinforcement learning based energy management systems for electrified powertrains: Progress, challenge, and potential solution. Renewable and Sustainable Energy Reviews, 154, 111833.

[9] Li, M., He, H., Feng, L., Chen, Y., & Yan, M. (2020). Hierarchical predictive energy management of hybrid electric buses based on driver information. Journal of Cleaner Production, 269, 122374.

[10] Huang, R., He, H., Meng, X., Wang, Y., Lian, R., & Wei, Y. (2021, October). Energy Management Strategy for Plug-in Hybrid Electric Bus based on Improved Deep Deterministic Policy Gradient Algorithm with Prioritized Replay. In 2021 IEEE Vehicle Power and Propulsion Conference (VPPC) (pp. 1-6). IEEE.

[11] Wei, Z., Zhao, D., He, H., Cao, W., & Dong, G. (2020). A noise-tolerant model parameterization method for lithium-ion battery management system. Applied Energy, 268, 114932.

[12] Lin, X., Perez, H. E., Mohan, S., Siegel, J. B., Stefanopoulou, A. G., Ding, Y., & Castanier, M. P. (2014). A lumped-parameter electro-thermal model for cylindrical batteries. Journal of Power Sources, 257, 1-11.

[13] Wu, J., Wei, Z., Li, W., Wang, Y., Li, Y., & Sauer, D. U. (2020). Battery thermal-and health-constrained energy management for hybrid electric bus based on soft actor-critic DRL algorithm. IEEE Transactions on Industrial Informatics, 17(6), 3751-3761.

[14] Fujimoto, S., Hoof, H., & Meger, D. (2018, July). Addressing function approximation error in actor-critic methods. In International conference on machine learning (pp. 1587-1596). PMLR.

[15] Huang, R., He, H., Meng, X., & Li, M. (2021, October). A Novel Hierarchical Predictive Energy Management Strategy for Plug-in Hybrid Electric Bus Combined with Deep Reinforcement Learning. In 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-5). IEEE.