Cloud-based health-conscious energy management of hybrid battery systems in electric vehicles with deep reinforcement learning

Weihan Li[a,b,*], Han Cui[a], Thomas Nemeth[a,b], Jonathan Jansen[a], Cem Ünlübayir[a,b], Zhongbao Wei[d], Xuning Feng[e], Xuebing Han[e], Minggao Ouyang[e], Haifeng Dai[f], Xuezhe Wei[f], Dirk Uwe Sauer[a,b,c]

[a] *Chair for Electrochemical Energy Conversion and Storage Systems, Institute for Power Electronics and Electrical Drives (ISEA), RWTH Aachen University, Jaegerstrasse 17/19, 52066, Aachen, Germany*
[b] *Juelich Aachen Research Alliance, JARA-Energy, Germany*
[c] *Helmholtz Institute Münster (HI MS), IEK-12, Forschungszentrum Jülich, Germany*
[d] *National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China*
[e] *State Key Laboratory of Automotive Safety and Energy, School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China*
[f] *National Fuel Cell Vehicle Powertrain System Research Engineering Center, School of Automotive Studies, Tongji University, Shanghai 201804, China*

Abstract

In order to fulfill the energy and power demand of battery electric vehicles, a hybrid battery system with a high-energy and a high-power battery pack can be implemented as the energy source. This paper explores a cloud-based multi-objective energy management strategy for the hybrid architecture with a deep deterministic policy gradient, which increases the electrical and thermal safety, and meanwhile minimizes the system's energy loss and aging cost. In order to simulate the electro-thermal dynamics and aging behaviors of the batteries, models are built for both high-energy and high-power cells based on the characterization and aging tests. A cloud-based training approach is proposed for energy management with real-world vehicle data collected from various road conditions. Results show the improvement of electrical and thermal safety, as well as the reduction of energy loss and aging cost of the whole system with the proposed strategy based on the collected real-world driving data. Furthermore, processor-in-the-loop tests verify that the proposed strategy can achieve a much higher convergence rate and a better performance in terms of the minimization of both energy loss and aging cost compared with state-of-the-art learning-based strategies.

Keywords: energy management, vehicle-to-cloud, reinforcement learning, battery aging, lithium-ion, battery safety

1. Introduction

The popularity of electric vehicles (EVs) provides a promising solution to the increasingly severe greenhouse effect in the world [1]. Compared with hybrid electric vehicles (HEVs) and plug-in hybrid electric

*Corresponding author. Chair for Electrochemical Energy Conversion and Storage Systems, Institute for Power Electronics and Electrical Drives (ISEA), RWTH Aachen University, Jaegerstrasse 17/19, 52066, Aachen, Germany
Email address: batteries@isea.rwth-aachen.de, weihan.li@isea.rwth-aachen.de (Weihan Li)

vehicles, battery electric vehicles (BEVs) produce zero carbon dioxide emissions through electrification with high powertrain efficiency and renewable-energy integration possibility. However, challenges are still existing regarding system cost reduction and performance increase. In order to fulfill the energy and power demand of BEVs, using a single type of battery as the energy source may lead to an oversized configuration of the battery system if the power-to-energy ratio of the cell does not match with that of the system requirement. In contrast, less system weight and volume are achievable by utilizing different energy sources via power electronics, which has been considered widely in the literature [2]. To cover the power requirement in the hybrid energy storage system, different energy storage technologies, e.g., batteries [3], fuel cells [4], and super-capacitors [5], have been used. Hybrid energy storage systems with lithium-ion batteries and super-capacitors have been developed for electric vehicles [6], electric ships [7], and electric trains [8], etc. With the technological progress in high-power lithium-ion batteries, e.g., batteries with lithium-titanate-oxide (LTO) anode [9], hybrid battery systems (HBSs) are attracting more and more attention from both industry and academia [10]. Compared with the single battery system (SBS), an HBS constructed with a high-energy (HE) pack and a high-power (HP) pack is able to balance the power and energy demand of the BEVs for the system scalability [11]. However, an efficient energy allocation between the battery packs relies on a reliable and robust energy management system (EMS).

Generally, the state-of-the-art EMSs for hybrid energy sources can be divided into three categories: rule-based, optimization-based, and learning-based EMSs [12]. Due to the simplicity and the low demand for computational cost, rule-based methods, such as thermostatic strategy [13] and fuzzy logic controller [14], have gained great success for HEVs. Nonetheless, rule-based strategies optimized the performance of each component of the system individually on the basis of the predetermined rule, which may lead to solutions far away from the optimality. Hence, they can be tuned to be only suitable for a specific driving cycle. Optimization-based methods, e.g., particle swarm optimization [15], equivalent consumption minimization strategy (ECMS) [16], dynamic programming (DP) [17], and model predictive control (MPC) [18], have shown the ability to achieve globally optimal control based on the prior knowledge of future driving conditions. The goal of these approaches is the optimization of the predefined objective values, considering system constraints. Kollmeyer et al. [19] developed a DP-based EMS for the optimal control of a hybrid energy storage system. In Ref. [20], a parallel DP-based energy management algorithm was designed for a battery and fuel cell hybrid train based on the matrix calculation. However, future scenarios are usually unavailable in real driving conditions, limiting the DP-based strategy to an offline benchmark for performance evaluation of other approaches. In Ref. [21], model predictive control was applied to explore the appropriate power-split strategy for hybrid energy sources, where the cost function is minimized through the calculation of optimal control sequence in a prediction horizon. However, the determined control action is sub-optimal as the problem was divided into several time steps.

Learning-based methods, e.g., reinforcement learning (RL), can learn from historical experiences and

Nomenclature

$C_{1,2}$	polarization capacitances
C_N	nominal capacity
C_{heat}	heat capacity
I	current
I_{DC-DC}	current in DC-DC converter
N	size of the mini-batch
N_E	size of the replay buffer
P_t	power
Q	critic network
Q'	target critic-network
Q_{DC-DC}	energy loss in DC-DC converter
R_0	ohmic resistance
$R_{1,2}$	polarization resistances
T	battery temperature
T_{amb}	ambient temperature
V_t	terminal voltage
α	reward weight
α_C	aging weight
α_R	aging weight
β	reward weight
β_C	aging weight
β_R	aging weight
$\dot{Q}$	heat generation rate
η	coulomb efficiency
γ	discount factor
γ_1	reward weight
γ_2	reward weight
γ_3	reward weight
γ_4	reward weight
γ_5	reward weight
γ_6	reward weight
μ	action network
μ'	target actor-network
ν	soft update of the target network
ψ	reward weight
τ	replacement cost
θ	network weights
a_t	action
h	heat transfer coefficient
l_a	learning rate for the actor-network
l_c	learning rate for the critic-network
p	aging weight
p_C	aging weight
p_R	aging weight
r_t	reward
s_t	state
t	time
v_t	velocity

List of abbreviations

BEV	battery electric vehicle
$DDPG$	deep deterministic policy gradient
DNN	deep neural network
DoD	depth of discharge
DP	dynamic programming
DQL	deep Q-learning
DRL	deep reinforcement learning
ECM	equivalent circuit model
$ECMS$	equivalent consumption minimization strategy
EM	electric motor
EMS	energy management system
EV	electric vehicle
HBS	hybrid battery system
HE	high energy
HEV	hybrid electric vehicle
HP	high power
LMO	lithium-manganese-oxide
LTO	lithium-titanate-oxide
ML	machine learning
MPC	model predictive control
NCA	lithium-nickel-cobalt-aluminum-oxide
OCV	open-circuit voltage
PiL	processor-in-the-loop
RL	reinforcement learning
SoC	state of charge

optimize the control scheme gradually through the interaction with the environment, which provides self-adapted energy management strategies concerning different driving conditions [22]. With the support of cloud computing and the internet of things, EV data can be measured and transmitted to the cloud seamlessly [23], where learning-based methods show significant advantages over the other methods facing the operation-related big data [24]. As a popular RL method, tabular Q-learning (QL) was first introduced to EMS to control the power-split in HEVs [25]. The implemented Q-table contains all possible Q-values of the

discretized state-action pairs, which need to be updated continuously by exploring the estimated maximum total reward before the convergence is reached. However, RL is not suitable for optimization problems with high dimensional state- and action-spaces on account of the 'curse of dimensionality.' Compared to RL, deep reinforcement learning (DRL) methods, such as deep Q-learning (DQL), use deep neural networks (DNNs) to approximate Q-values, avoiding the limitation of state-space discretization efficiently. Therefore, the DQL outperforms the QL for solving the optimization problem with multi-dimensional states [26]. In Ref. [27], DQL-based EMS was introduced to minimize the fuel consumption of an HEV. In order to further stabilize the training results of the Q network, the experience replay buffer [28] and the target network [29] were adopted. Li et al. [30] introduced the prioritized experience replay buffer to improve the quality of the sampled data for the parameter update of the Q network, whereas the computational burden was significantly increased. In Ref. [31], a DQL-based energy management strategy was introduced for a hybrid battery system in electric vehicles consisting of an HE and an HP battery pack. The energy loss was minimized and both the electrical and the thermal safety of the system was guaranteed by the EMS. Although the DQL-based EMSs have been proven to be suitable for a continuous state space, their performance can be influenced by the discretization of the action space, which constrains their applications in continuous action scenarios. In order to apply the DRL to solve the continuous control problems, the actor-critic strategy, e.g., deep deterministic policy gradient (DDPG), was developed [32], in which the actor is the decision-maker and the critic evaluates its taken action. Although in Ref. [33], the DDPG-based EMS for HEVs was developed to minimize the fuel consumption while maintaining the state of charge (SoC) of the battery at an appropriate level, they only adapted a simplified battery model concerning the battery's electrical behaviors. However, battery dynamics are affected by the temperature as well. To address this problem, Li et al. [34] used an equivalent circuit model (ECM) coupled with a thermal model to simulate the dynamics of the battery system. To further develop a health-conscious EMS, some research work [35] implemented a dynamic semi-empirical battery degradation model [36] considering the battery's capacity loss during usage. In Ref. [37], the state of health of the battery was updated utilizing a look-up table concerning temperature, cell capacity, and cycle number. However, these methods only consider the capacity loss caused during the battery operation, neglecting other aging factors such as the inspected time range and the increase of the internal resistance.

To the best knowledge of the authors, no efforts have been made to develop a learning-based health-conscious energy management strategy for HBSs in BEVs. This paper aims to bridge the aforementioned research gap and proposes a continuous DRL-based EMS for HBSs in BEVs considering battery aging. A DDPG-based energy management strategy was proposed to explore the optimal continuous power-split scheme for hybrid battery packs in a BEV for the first time. A novel reward function was designed incorporating multi-objective reward terms, including the electrical and thermal constraints, the energy loss and the aging cost of the whole system. Electro-thermal models and aging models were developed based on experiments to simulate the real dynamics of the battery packs with high accuracy. The proposed aging

models consider both the capacity fade and power fade of the battery in the calendar and cyclic aging. Based on a vehicle-to-cloud energy management framework, we carried out the training process on the cloud-server with real-world vehicle data under different dynamics collected from various road conditions. Processor-in-the-loop (PiL) tests were carried out with a machine learning (ML)-capable embedded device to validate the DDPG-based EMS and verify its low computation burden. The proposed strategy can continuously control the power-split between two battery packs, guaranteeing the safe and efficient operation and prolonged longevity of the entire system. Compared with QL and DQL-based EMSs, the proposed DDPG-based strategy achieved a much higher convergence rate as well as less energy loss and aging cost.

The remainder of this paper is structured as follows. In Section 2, the topology of the HBS in the BEV is introduced, followed by the electro-thermal and aging modeling details. In Section 3, the framework of the DDPG-based EMS with collected real-world driving data is introduced. The training results in the cloud and validation results in PiL tests are shown and discussed in Section 4. Section 5 draws the conclusions.

2. Hybrid battery system model

2.1. Vehicle modeling and HBS topology

The study focuses on an HBS for a compact all-wheel-drive vehicle, where each axle is propelled by an electric motor (EM), as shown in Fig. 1. The HBS is constructed with an HE pack and an HP pack, as derived from [38], where the HE pack serves as the primary energy source and the HP pack supplies the vehicle with additional power to fulfill high power demands. The connection of each battery pack to the DC-link can be realized in either a direct or indirect way. The latter requires an additional DC-DC converter, which offers the flexibility to scale the battery pack regarding the voltage range at the expense of increased system costs and complexity. As a compromise between the flexibility and system cost, one DC-DC converter is implemented in the HBS to connect the HP pack to DC-link, whereas the HE pack is connected to DC-link directly. This HBS topology contributes to the stabilization of the DC-link voltage at highly dynamic loads [11]. The implemented DC-DC converter is Brusa BDC546, whose energy loss concerning the input current is derived from [19] and can be calculated with a second-order polynomial as follows:

$$Q_{DC-DC} = 1.56 \times 10^{-2} I_{DC-DC}^2 - 1.44 I_{DC-DC} + 388.90 \tag{1}$$

where Q_{DC-DC} and I_{DC-DC} are the energy loss and current of the DC-DC converter, respectively.

Two different battery types are utilized to construct the HE and HP pack, in which the differences between battery cells in each pack are neglected. Based on the range requirement and peak power requirement of the vehicle under standard driving cycles, different system configurations with various hybridization ratios were simulated with the simulation tool, considering the differences in the system weight. With increased

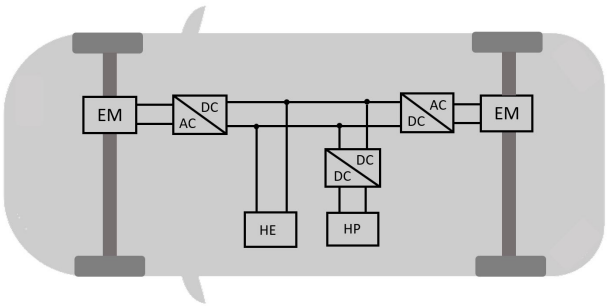


Fig. 1. The topology of the HBS in the BEV.

hybridization radio (HP pack energy / HE pack energy) and reduced total system weight, the energy consumption decreases but the system cost increases. The final optimized system configuration is determined based on the trade-off between energy consumption and system cost. The total energy capacity of the battery system is 27.3 kWh. As summarized in Table 1, the HE pack contains total energy of 25.4 kWh, which is constructed with 1440 HE cells with a nominal capacity of 4.9 Ah for each cell.

With their graphite anode and lithium-nickel-cobalt-aluminum-oxide (NCA) cathode, the HE cells are cylindrical cells of the new 21700 format similar to the cell format used in the Tesla Model 3, and have a high energy density but a limited power capacity. In contrast, the HP pack with 1.9 kWh is composed of 270 HP cells with a nominal capacity of 2.9 Ah for each cell. With lithium-titanate-oxide (LTO) as the anode material and lithium-manganese-oxide (LMO) as the cathode material, the HP cells possess a high power density despite the relatively low energy density (45 Wh/kg). Due to the high lithiation potential of 1.55 V vs. Li in LTO, safety critical lithium plating, the formation of a solid electrolyte interface, and the growth of dendritic lithium is prevented, even at low temperatures. LTO offers good thermal stability and no mechanical stress occurs in the material during the charge and discharge processes (zero-strain behavior), leading to an excellent cycle lifetime. The weight of the HE pack and the HP pack is 139.1 kg and 56.7 kg, respectively.

2.2. Battery modeling

2.2.1. Electro-thermal modeling

In order to ensure the accuracy without increasing the computational burden greatly, we use an extended Thevenin model with two RC pairs, rather than electrochemical models [39], to simulate the electrical dynamics of each battery cell, as shown in Fig. 2, where V_t is the terminal voltage, OCV is the open-circuit voltage, I is the current, R_0 is the ohmic resistance, $R_{1,2}$ and $C_{1,2}$ are the polarization resistances and polarization capacitances, respectively. The change rate of SoC is calculated by

$$\dot{SoC} = \frac{\eta}{C_N} I \tag{2}$$

Table 1

Specifications of the HBS.

	HE cell	HP cell
Cell Manufacturer	Samsung	Toshiba
Chemistry (Cathode / Anode)	NCA / C	LMO / LTO
Cell nominal capacity	4.9 Ah	2.9 Ah
Cell nominal voltage	3.6 V	2.4 V
Cell voltage limits (min / max)	2.5 V / 4.2 V	1.5 V / 2.9 V
Cell current limits (DCH / CHA)	2 C / 1 C	70 C / 70 C
Energy density	250 Wh/kg	45 Wh/kg
Power density (10 s, DCH / CHA)	1.3 kW/kg / 0.5 kW/kg	3.2 kW/kg / 3.2 kW/kg
Cell weight	69 g	150 g
Pack configuration	90s 16p	90s 3p
Pack energy	25.4 kWh	1.9 kWh
Pack weight	139.1 kg	56.7 kg

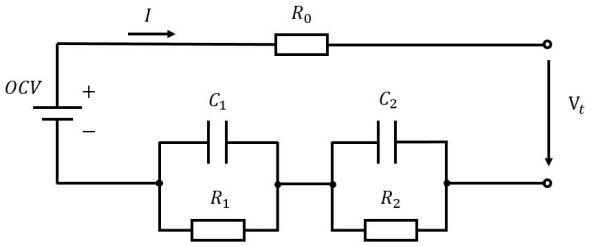


Fig. 2. An extended Thevenin model of a battery cell.

where η is the Coulomb efficiency and C_N is the nominal capacity. The change rate of the voltage over each polarization resistance is determined by

$$\dot{V}_i = -\frac{1}{R_i(SoC,T)C_i(SoC,T)}V_i + \frac{1}{C_i(SoC,T)}I \tag{3}$$

where $i = 1, 2$, the values of resistance and capacitance are dependent on SoC and temperature. The terminal voltage is then calculated as follows:

$$V_t = OCV(SoC) + V_1 + V_2 + R_0(SoC,T)I \tag{4}$$

where OCV is dependent on the SoC with a nonlinear function that can be measured experimentally. The OCV was determined by measuring the terminal voltage of the fully relaxed cell on different SoC levels in the OCV test. In order to determine the parameters in this ECM, we carried out pulse tests under different temperatures varying from - $25°C$ to $40°C$ and various SoCs between 0% and 100%. A fourth-order polynomial is used to represent each element of ECM regarding temperature and SoC based on the experiment data [40].

Considering the influence of temperature on the values ECM elements observed in the experiments, we adopt a thermal model developed in [41] in this work to characterize the dynamic of battery temperature. The energy balance of the cell is described as follows:

$$C_{heat}\frac{\partial T}{\partial t} = -h(T - T_{amb}) + \dot{Q} \tag{5}$$

where C_{heat} is the heat capacity, h is the heat transfer coefficient, T, T_{amb}, $\dot{Q}$ are the battery temperature, surrounding temperature, and heat generation rate, respectively. The heat generation rate can be further calculated by

$$\dot{Q} = I(OCV - V_t) + IT\frac{\partial V_t}{\partial T} \tag{6}$$

where the first part of the equation on the right side indicates the irreversible heat, while the second part represents the reversible entropic heat. In this work, the heat capacity, C_{heat}, for HE and HP cells are 950 J/K and 1120 J/K, respectively, and the heat transfer coefficient, h, for HE and HP cells is 12 W/K [31].

2.2.2. Aging modeling

Since the power output of the battery pack is influenced by cell degradation, we implemented aging models based on aging tests to simulate the degradation of both HE and HP cells in driving tasks. Generally, the degradation of the lithium-ion battery can be divided into calendar and cyclic aging. The calendar aging mainly comprises the formation of passivation layers at the electrode-electrolyte interfaces, such as the solid electrolyte interphase (SEI) at the anode, which is the predominant aging factor for batteries when their idle interval is longer than their operation period [42]. In contrast, cyclic aging indicates the aging during discharging and charging processes. While calendar aging is influenced by temperature and cell voltage, as well as the inspected time range, cyclic aging relies on average cell voltage and depth of discharge (DoD) [43]. Both aging processes lead to capacity loss and resistance increase.

In this work, the aging model for HE cells is derived from the previous work of our research group [44], in which an aging model concerning both capacity fade and resistance increase was developed. Based on the same modeling methodology, an aging model for HP cells [45] is developed based on experiments in this work and will be introduced as follows.

To determine the calendar aging, we stored multiple battery cells in two different scenarios for 700 days, considering the effects of both temperature and voltage on calendar aging. In the first scenario, the aging of battery cells at 2.57 V was tested at three different temperatures. In contrast, the batteries with five different voltages were stored at $45°C$ in the second scenario. The capacity and resistance concerning calendar aging is determined by

$$C_{calendar} = (1 - \alpha_C t^p)C_{init} \tag{7}$$

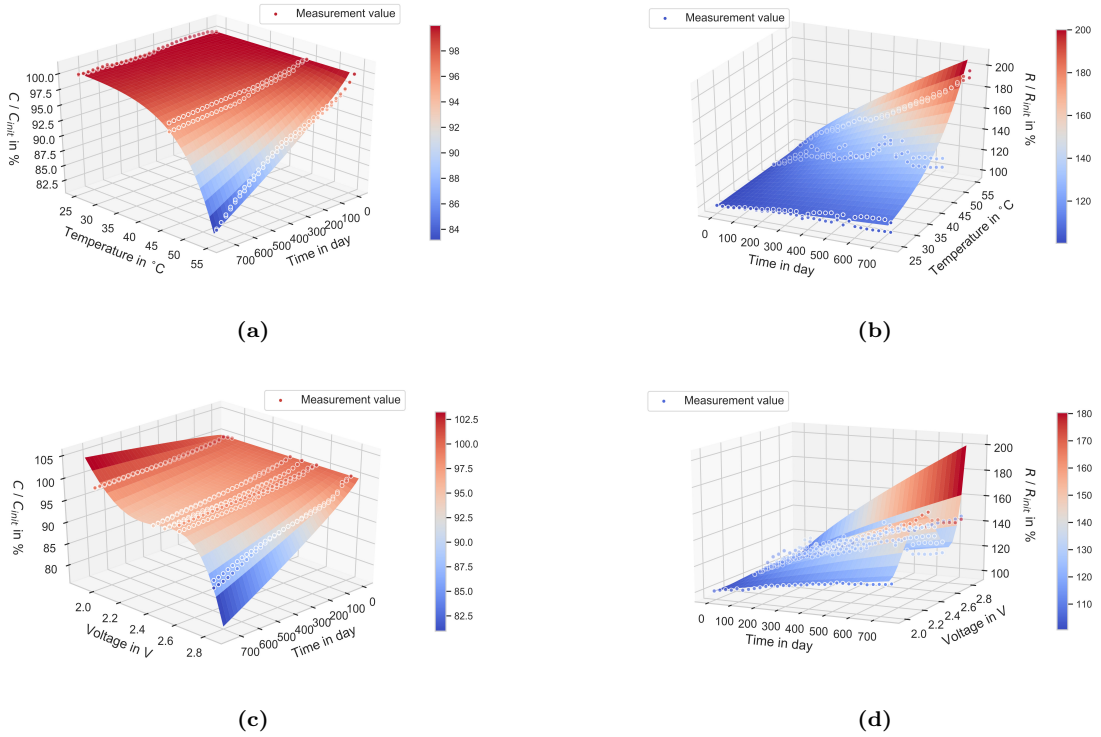


Fig. 3. The developed calendar aging model for HP cells compared with measurement data at the constant voltage of 2.57 V: (a) capacity fade, (b) resistance increase. The calendar aging model at 45°C: (c) capacity fade, (d) resistance increase.

$$R_{calendar} = (1 + \alpha_R t^p) R_{init} \tag{8}$$

where weights α_C and α_R are dependent on temperature and voltage, t is the time in day, p is the time-related weight, which is determined by fitting the curve to the measurement data. C_{init} and R_{init} are the initial capacity and resistance, respectively. The parameterization results of the calendar aging are summarized in Table 2. The whole calendar aging model of the HP cell is shown in Fig. 3. Fig. 3(a), (b) depict the capacity fade and resistance increase in different temperatures and the mean absolute errors are 1.50% and 3.38%, respectively. Fig. 3(c), (d) illustrate the capacity fade and resistance increase at different voltages, where the mean absolute errors are 0.89% and 2.57%.

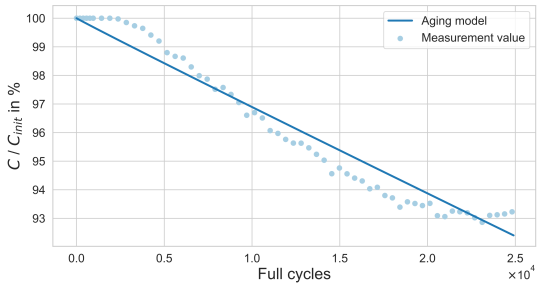
The cyclic aging experiment was carried out under 10 C at 25°C. Since calendar aging occurs during the measurements of cyclic aging, we calibrated all test data of the capacity and resistance by getting rid of the calendar aging part to acquire the authentic cycle aging. The capacity fade and resistance increase caused by cyclic aging are described as follows:

$$C_{cyclic} = (1 - \beta_C Q_c^{p_C}) C_{init} \tag{9}$$

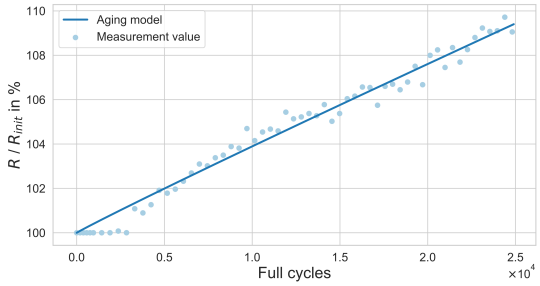
Table 2

Parameterization results of the aging model for the HP cell.

Parameter	Value
α_C	$5 \times 10^{17}(0.0058V_t^4 - 0.049V_t^3 + 0.15V_t^2 - 0.21V_t + 0.1)e^{\frac{-15.782}{T}}$
α_R	$6.92 \times 10^{7}(0.1V_t^4 - 0.93V_t^3 + 3.25V_t^2 - 4.98V_t + 2.84)e^{\frac{-7.771}{T}}$
p	0.82
β_C	6.67×10^{-7}
β_R	9.77×10^{-7}
p_C	0.98
p_R	0.97



(a)



(b)

Fig. 4. The developed cyclic aging model for HP cells compared with measurement data at $25°C$ with 10 C current: (a) capacity fade, (b) resistance increase.

$$R_{cyclic} = (1 + \beta_R Q_c^{p_R}) R_{init} \tag{10}$$

where Q_c is the charge throughput, β_C, β_R, p_C, p_R are parameters that are determined based on the test data and are summarized in Table 2. The fitting results of the cyclic aging are shown as follows and depicted in Fig. 4. The mean absolute error for capacity fade and resistance increase is 0.45% and 0.36%, respectively.

Thus, the overall aging model for the HP cells can be described by the superposition of calendar and

cycle aging as follows:

$$C = (1 - \alpha_C t^p - \beta_C Q_c^{p_C}) C_{init} \tag{11}$$

$$R = (1 + \alpha_R t^p + \beta_R Q_c^{p_R}) R_{init}. \tag{12}$$

3. Deep deterministic policy gradient-based EMS

For most of the learning-based methods, e.g., QL and DQL, the state or action space needs to be discretized, which may lead to less efficient control due to discretization errors. In order to guarantee the optimal power-split between the HE and HP pack, a DDPG-based EMS is explored in this work, which can deal with continuous multi-dimensional state and action space.

3.1. Deep deterministic policy gradient

DDPG is an off-policy, model-free DRL method, which is developed based on the actor-critic architecture [32]. Utilizing the high potential of DNNs in dealing with high-dimensional states, DDPG can explore the most favorable strategy to solve the optimization problems with continuous state and action spaces. The actor-network $\mu(s|\theta^\mu)$ with weights θ^μ behaves like the policy in conventional RL methods, which determines the action regarding the observed environmental states. The critic-network $Q(s, a|\theta^Q)$ with weights θ^Q evaluates the taken action with the estimated total reward. Specifically, the inputs of the actor are continuous multi-dimensional environmental states, and the output is a continuous action. At the same time, the critic takes both the states and actions as its inputs and outputs the Q-value.

Moreover, an experience replay buffer and two more networks are adopted in DDPG to increase the stability and speed up the convergence. The experience replay buffer breaks the temporal correlation of continuous transition pairs, which contributes greatly to the smaller variance of the Q-value by ignoring the unsatisfied prediction in a short time range. Since it is likely to cause divergence by merely using the Q-value of the single critic-network and the estimated action of the actor-network, a target critic-network $Q'(s, a|\theta^{Q'})$ and a target actor-network $\mu'(s|\theta^{\mu'})$ are applied to calculate the temporal difference error between the target and the evaluation value. The estimated optimal Q-value is determined by

$$y_t = r_t + \gamma Q'(s_{t+1}, \mu'(s_{t+1}|\theta^{\mu'})|\theta^{Q'}) \tag{13}$$

where r_t is the immediate reward, and $\gamma \in (0, 1)$ is a discount factor to assure the convergence of the estimated Q-value. The update of the critic follows the rule as follows:

$$L = \frac{1}{N} \sum_{t=1}^{N} (y_t - Q(s, a|\theta^Q))^2 \tag{14}$$

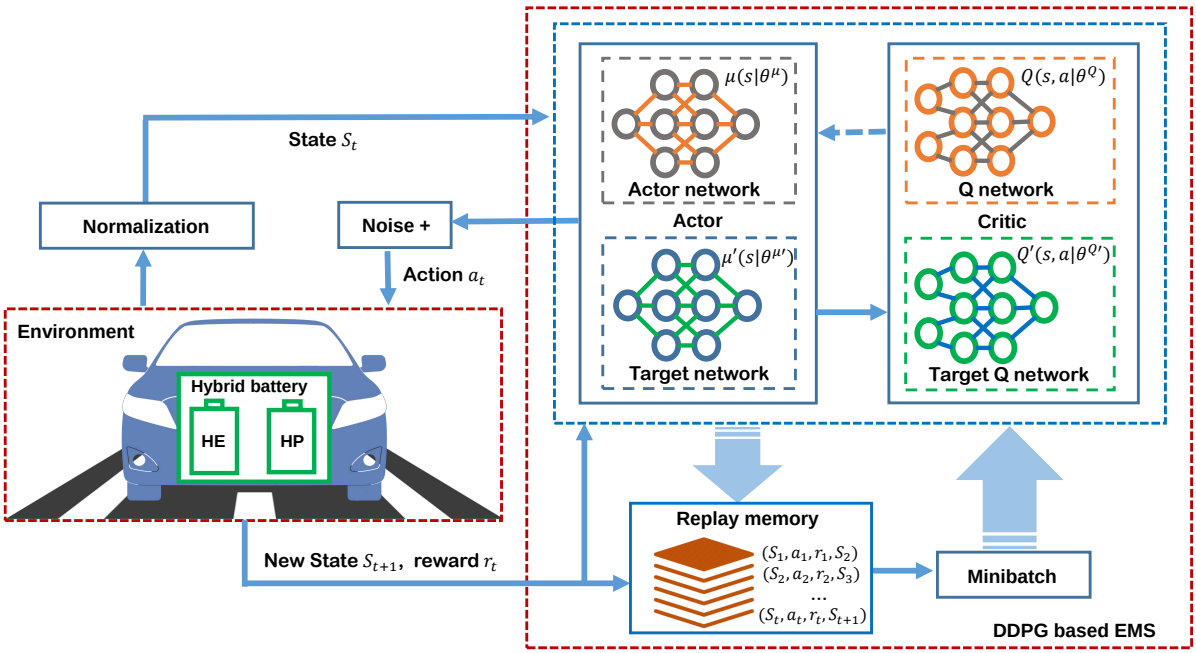


Fig. 5. The framework of the DDPG-based EMS for the hybrid battery packs in electric vehicles.

where L is the mean loss, N is the fixed size of a mini-batch, which is randomly selected from the experience replay buffer. Since the objective of the optimal policy is to maximize the expected Q-value, the actor-network $\mu(s|\theta^\mu)$ can be updated in the direction of maximum Q-value with the help of the derivative of the Q-value regarding the parameters θ^μ in the actor-network, where the chain rule is applied as follows:

$$\bigtriangledown_{\theta^\mu} J = \bigtriangledown_{\theta^\mu} E[Q(s_t, a_t)] \approx \bigtriangledown_{a_t} Q(s_t, a_t) \bigtriangledown_{\theta^\mu} \mu(s_t|\theta^\mu). \tag{15}$$

The update of parameters in the target network follows the soft update. It guarantees the stability of the trained networks by slowly tracking the weights of the evaluation network as follows:

$$\theta' \leftarrow \nu\theta + (1-\nu)\theta', \quad \nu << 1 \tag{16}$$

where ν is the soft update factor, θ' and θ are the parameter in target networks and original networks, respectively. The overall algorithm of DDPG is summarized in Algorithm 1.

3.2. The architecture of the DDPG-based EMS

The objective of the energy management for the HBS is to find the optimal power-split scheme between the HE and HP pack. After each decision-making of the power allocation, the driving condition and vehicle states will change, and the allocation method should be updated consequently. Thus, this energy management problem can be formulated as the Markov decision process problem, which can be solved by DDPG.

Algorithm 1 DDPG-based EMS

Initialization of the experience replay buffer E

Initialization of the action network $\mu(s|\theta^{\mu})$ with random weights θ^{μ}

Initialization of the weights of the target actor-network $\mu'(s|\theta^{\mu'})$ with $\theta^{\mu'} = \theta^{\mu}$

Initialization of the critic-network $Q(s, a|\theta^{Q})$ with random weights θ^{Q}

Initialization of the weights of the target critic-network $Q'(s, a|\theta^{Q'})$ with $\theta^{Q'} = \theta^{Q}$

for epoch = 1 to M **do**

  Reset the environment state to s_0

  for time = 1 to N **do**

    Select an action $a_t = \mu(s_t|\theta^{\mu}) + \epsilon N(0, 1)$

    Clip the action to the range of [0,1]

    Get the reward r_t and new state s_{t+1} from the environment

    Store transition(s_t, a_t, r_t, s_{t+1}) in memory E

    Select a fixed-size mini-batch from E if E is full

    $y_t = r_t$ if $time = N$, otherwise

    $y_t = r_t + \gamma Q'(s_{t+1}, \mu'(s_{t+1}|\theta^{\mu'})|\theta^{Q'})$

    Update the weights of actor-network using sampled policy gradient

    $\bigtriangledown_{\theta^{\mu}} J \approx \frac{1}{N} \sum_{t=1}^{N} \bigtriangledown_a Q(s_t, a_t|\theta^{Q}) \bigtriangledown_{\theta^{\mu}} \mu(s_t|\theta^{\mu})$

    Update the weights of the critic-network by minimizing the loss

    $L = \frac{1}{N} \sum_{t=1}^{N} (y_t - Q(s, a|\theta^{Q}))^2$

    Update the weights of the target actor-network

    $\theta^{\mu'} \leftarrow \nu\theta^{\mu} + (1 - \nu)\theta^{\mu'}$

    Update the weights of the target critic-network

    $\theta^{Q'} \leftarrow \nu\theta^{Q} + (1 - \nu)\theta^{Q'}$

  end for

end for

The overall architecture of the proposed DDPG-based EMS is illustrated in Fig. 5, where the vehicle and the surrounding driving conditions act as the environment and the DDPG-based EMS serves as the agent. The essential elements of DDPG, i.e., states, action, reward, need to be carefully designed to address the energy management problem for the HBS.

The states indicate the status of the environment, from which the agent learns the surroundings and makes a proper decision. In this work, we consider not only the electrical and thermal safety of the whole system but also the energy loss and balance of the aging effect between the HE pack and the HP pack. SoC, temperature as well as the relative capacity of HE and HP cells considering aging are selected as the states

of the HBS. The vehicle's speed and total power demand of the HBS are chosen to represent the dynamics of the vehicle. The total state space S is shown as follows:

$$S = [SoC_{t,HE}, SoC_{t,HP}, T_{t,HE}, T_{t,HP}, C_{t,HE}, C_{t,HP}, v_t, P_t] \tag{17}$$

where $SoC_{t,HE}$, $SoC_{t,HP}$, $T_{t,HE}$, $T_{t,HP}$ are SoC and temperature of each battery pack at time t. $C_{t,HE}$ and $C_{t,HP}$ represent the actual capacity of the HE and HP pack, respectively, v_t is vehicle speed, and P_t is the total power demand.

Since the objective of the proposed DDPG-based EMS is to split the power between the HE and HP pack optimally, the power ratio of the HE pack in the total power is adopted as the action, which varies within the range of [0, 1]. The power extracted from the HP pack can be determined under the consideration of the DC-DC converter's efficiency.

Through the interaction with the environment, the agent can improve its power-split scheme by maximizing the long-term accumulative reward. In this work, the reward function is defined as follows:

$$\begin{aligned} r =& \alpha(Q_{HE} + Q_{HP} + Q_{DC-DC}) + \beta P_{\mathrm{HP}}^2 sgn(P_{total}) \\ &+ \gamma_1 max(I_{HE,min} - I_{HE}, 0) \\ &+ \gamma_2 min(I_{HE,max} - I_{HE}, 0) \\ &+ \gamma_3 min(T_{HE,max} - T_{HE}, 0) \\ &+ \gamma_4 max(I_{HP,min} - I_{HP}, 0) \\ &+ \gamma_5 min(I_{HP,max} - I_{HP}, 0) \\ &+ \gamma_6 min(T_{HP,max} - T_{HP}, 0) \\ &+ \psi G \end{aligned} \tag{18}$$

where Q_{HE}, Q_{HP}, and Q_{DC-DC} are the energy loss of HE pack, HP pack, and DC-DC converter, respectively. Compared with the HE pack, the HP pack contains a relatively smaller internal resistance. Therefore, the first reward term will encourage the use of the HP pack while reducing the total energy loss. In order to explore the optimal operating range of the HP battery pack, we introduce the second reward term, $P_{\mathrm{HP}}^2 sgn(P_{\mathrm{total}})$, to balance the discharging trend caused by the first reward term. The positive P_{total} indicates the recuperation phase and its negative value stands for the power consumption phase of the BEV. The physical constraints of the cells are also considered in the reward function to ensure both the electrical and thermal safety of the battery systems, where $I_{HE,min}$, $I_{HE,max}$ and $T_{HE,max}$ are the maximum discharging and charging currents and maximum temperature limit of the HE cells. Similarly, $I_{HP,min}$, $I_{HP,max}$, and $T_{HP,max}$ are the maximum discharging and charging currents and maximum temperature limit of the HP cells. The idea of these reward terms is to penalize the agent when the current or the temperature is beyond the safety range. When the current and the temperature are within the safety range, no penalization will

Table 3

Hyperparameters of DDPG-based EMS.

Hyperparameters	Description	Value
ν	Soft update of the target network	0.0214
l_a	Learning rate for the actor-network	10^{-3}
l_c	Learning rate for the critic-network	10^{-4}
γ	Discount factor	0.99
N_E	Size of the replay buffer	10^5
N	Size of the mini-batch	64

be given to the agent. We name the reward term regarding the current and temperature as electrical safety term and thermal safety term for simplification. Furthermore, G is the reward term regarding equivalent aging cost, which represents the immediate replacement cost caused by capacity fade, calculated by

$$G = (\frac{\tau_{HE}(C_{HE,t-1} - C_{HE,t})}{20\%C_{HE,init}} + \frac{\tau_{HP}(C_{HP,t-1} - C_{HP,t})}{20\%C_{HP,init}}) \tag{19}$$

where weights τ_{HE} and τ_{HP} represent the replacement cost of each battery pack. Utilizing the BacPac tool [46], τ_{HE}, which is calculated to be 5715 \$ with 225 $\$/kWh$ and τ_{HP} is calculated to be 2850 \$ with 1500 $\$/kWh$. $C_{HE,init}$ and $C_{HP,init}$ are the initial capacity of HE pack and HP pack, respectively. $C_{HE,t-1}$, $C_{HP,t-1}$ are the capacity of two packs at time $t-1$. The EoL of each pack is defined if 20% of capacity degradation is reached. The replacement cost regarding battery aging is distributed into each time interval of 1 s. Different weights, i.e., α, β, γ_1, γ_2, γ_3, γ_4, γ_5, γ_6, ψ are implemented to balance the influence of each reward term.

To speed up the convergence during the training process, we normalize the environmental states in the range of [0, 1] before feeding them into the actor-network and the critic-network at each time point. Through trial and error, the hyperparameters of the proposed system are determined and summarized in Table 3. The actor-network, $\mu(s|\theta^\mu)$, contains two hidden layers with 400 and 300 neurons, where rectifier (Relu) is used as the activation function. The sigmoid function is implemented in the output layer to limit the action in the range of [0, 1]. The actor-network's outputs and the normalized states are inputs of the critic-network and they are combined in the critic-network's second layer with 300 neurons. The construction of the actor and critic-network are shown in Fig. 6. The target actor-network $\mu'(s|\theta^{\mu'})$ and the target critic-network $Q'(s, a|\theta^{Q'})$ have the same architecture as $\mu(s|\theta^\mu)$ and $Q(s, a|\theta^Q)$, respectively. Adam optimizer [47] is chosen to train the networks. At the beginning of the training process, a random action chosen from the uniform distribution is implemented to the environment before the buffer is full. To balance the exploration and exploitation, we add additional noise to the output of the actor-network. In this work, the additive Gaussian noise is applied.

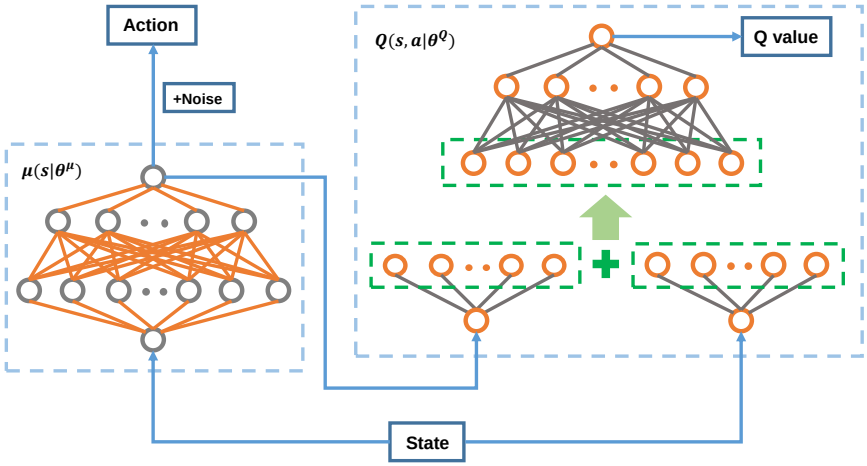


Fig. 6. The construction of the actor- and the critic-network.

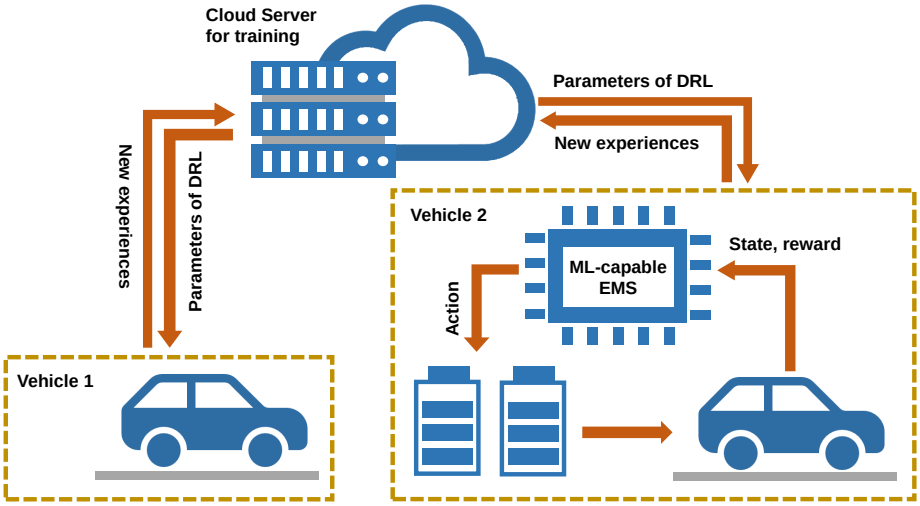
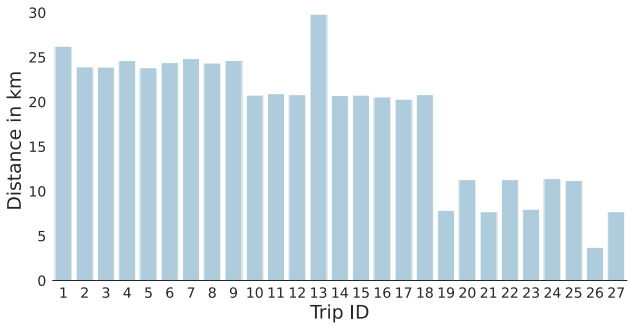


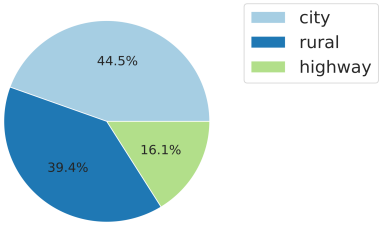
Fig. 7. The framework of the cloud-based training and employment of the EMS.

3.3. Cloud-based energy management based on real-world driving data

Considering the high computational demand in the training process, we propose a vehicle-to-cloud framework for the training and implementation of learning-based energy management strategies, as shown in Fig. 7. The final performance of the DDPG-based EMS relies significantly on the quality of training data. In order to explore the real dynamics of the HBS, such as temperature and aging development under real-world operation, we collected a large amount of real-world driving data to generate the application-oriented load profiles for the HBS. To accelerate the training process, we implement suitable hardware with high computational power is in the cloud platform to train the DRL-based strategies based on the collected data.



(a)



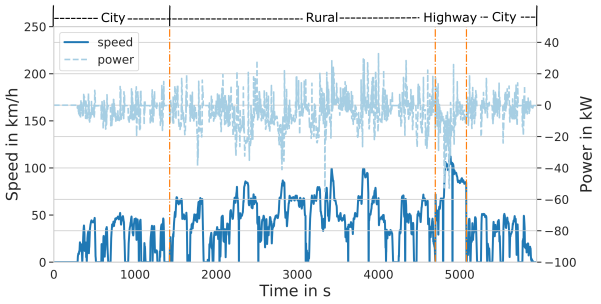
(b)

Fig. 8. (a) Collected real-world data, (b) Distribution of different road conditions.

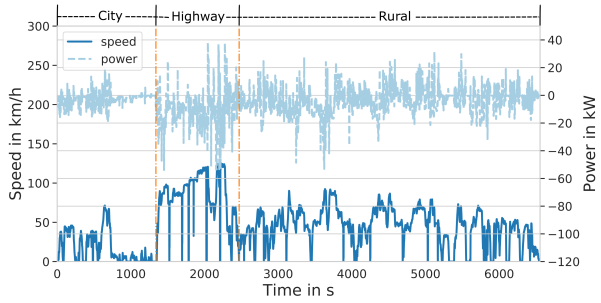
Finally, the trained DRL model will be implemented and validated on the local embedded device in BEVs.

In this work, various real-world data were collected from 27 trips on various road conditions in Aachen and nearby cities in Germany. To further increase the diversity of training data for the proposed EMS, the logged data is classified into three categories based on GPS information concerning the road conditions, i.e., urban, rural, highway. The driving lengths and the distribution of road conditions within the logged data are shown in Fig. 8.

The gathered data was uploaded to the cloud server to train the EMS, which is equipped with two Tesla V100 GPUs. In each training epoch, the driving data concerning different road conditions were randomly extracted from the trips and combined to train the proposed EMS. One of the combinations of the real-world driving data in one training epoch is shown in Fig. 9(a). Once the training of the DRL-based strategy is accomplished, parameters of the trained EMS can be transmitted to the ML-capable embedded device in BEVs to perform the power-split between the HE and HP pack. In this work, a PiL test with a relatively low-cost embedded device manufactured by Nvidia Corporation, Jetson Nano, is carried out to verify the performance and the real-time feature of the proposed strategy under new load profiles, as shown in Fig. 9(b).



(a)



(b)

Fig. 9. The combined real-world driving data for training and validation: (a) A combination of the driving data for the training, (b) Driving data for the validation.

4. Results and discussion

As introduced in Section 3.3, the real-world data regarding various road conditions were randomly combined in each epoch to train the EMS in the cloud server. In each training epoch, the initial SoCs of the HE pack and HP pack are set to be 90% and 60%, respectively, enabling the absorption of recuperation energy even at the beginning of the driving profile. The ambient temperature and the initial temperature of the battery packs are assumed to remain constant at 35°C to simulate the environment in summer. According to the cell's specifications, the maximum safe operating temperatures, $45°C$ for the HE cell and $55°C$ for the HP cell, are selected as the temperature limit. In the validation process within the PiL test, the trained EMS is tested with the same initial environment states.

4.1. Results of the training process

After 120 training epochs, the DDPG-based EMS converged and the training results are illustrated in Fig. 10 and Fig. 11. Fig. 10(a) depicts the total power and the power of the HP pack, where the positive value represents the energy recuperation and the negative value represents the power demand for propelling the vehicle. A large amount of recuperation power and high power consumption are supported by the HP pack. Fig. 10(b) illustrates the tendencies of each pack's SoC. The SoC of the HE pack decreases continuously,

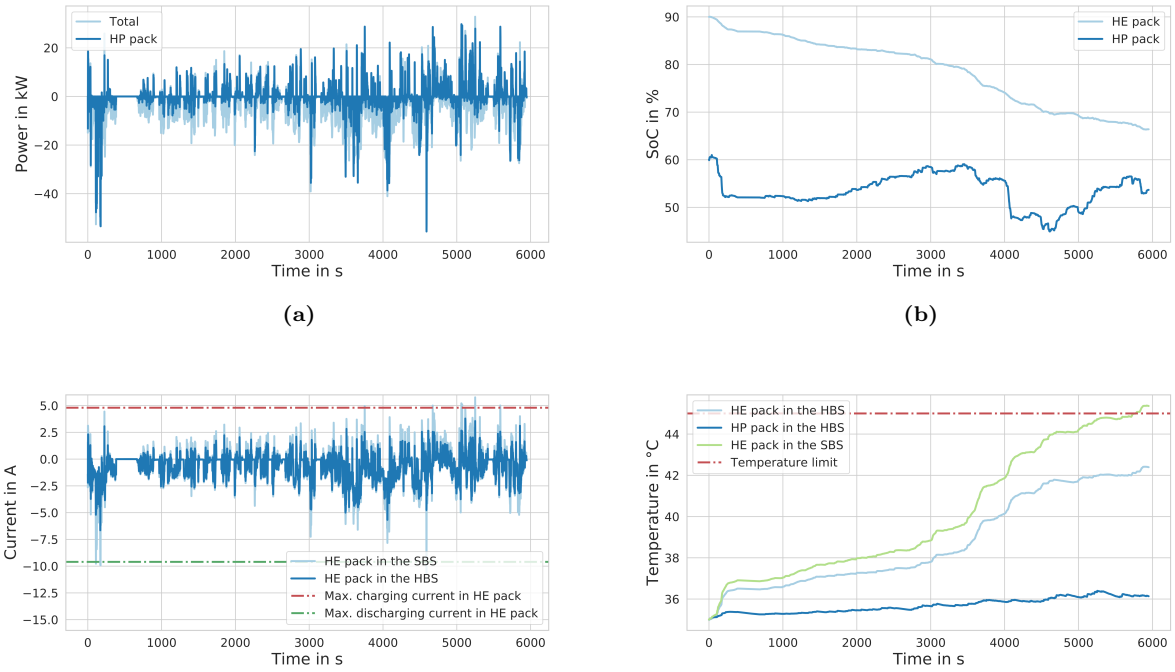


Fig. 10. Training results of the DDPG-based EMS. (a) Total power of the HBS and power of the HP pack. (b) The SoC tendencies of HE and HP packs within the last training epoch. (c) Comparison between the SBS and HBS in terms of the HE cell's current. (d) Temperature profiles of each pack in SBS and HBS.

which accords with the fact that the HE pack serves as the primary energy source. In contrast, the HP pack's SoC varies around 55% with a limited range, demonstrating the effectiveness of the balancing between the first two reward terms in Eq. (18) and avoiding the overcharging and overdischarging of the HP pack.

To verify the improvement of the electrical and thermal safety of the HBS compared with the single battery system (SBS), we implement an SBS with an increased number of the same HE cell to have the same total energy capacity as that of the HBS under identical load profile as a benchmark. Fig. 10(c) compares the HE cell's current in SBS with that in HBS. While the HE cell's current in SBS exceeds the electrical safety limit, the HE cell in HBS works in the safe area over the entire driving distance, which demonstrates the improvement of the electrical safety of the system with the proposed EMS. Fig. 10(d) illustrates the temperature profiles of the SBS and HBS. While the temperature of the HE pack in SBS exceeds the temperature limit, utilizing the HBS with the proposed EMS guarantees the thermal safety of the HE pack. The difference between the HE and HP pack's temperature lies in a higher heat generation within the HE pack because of its larger internal resistance and the larger energy supply for the vehicle.

In order to further look into the performance of the EMS in terms of temperature control, an additional

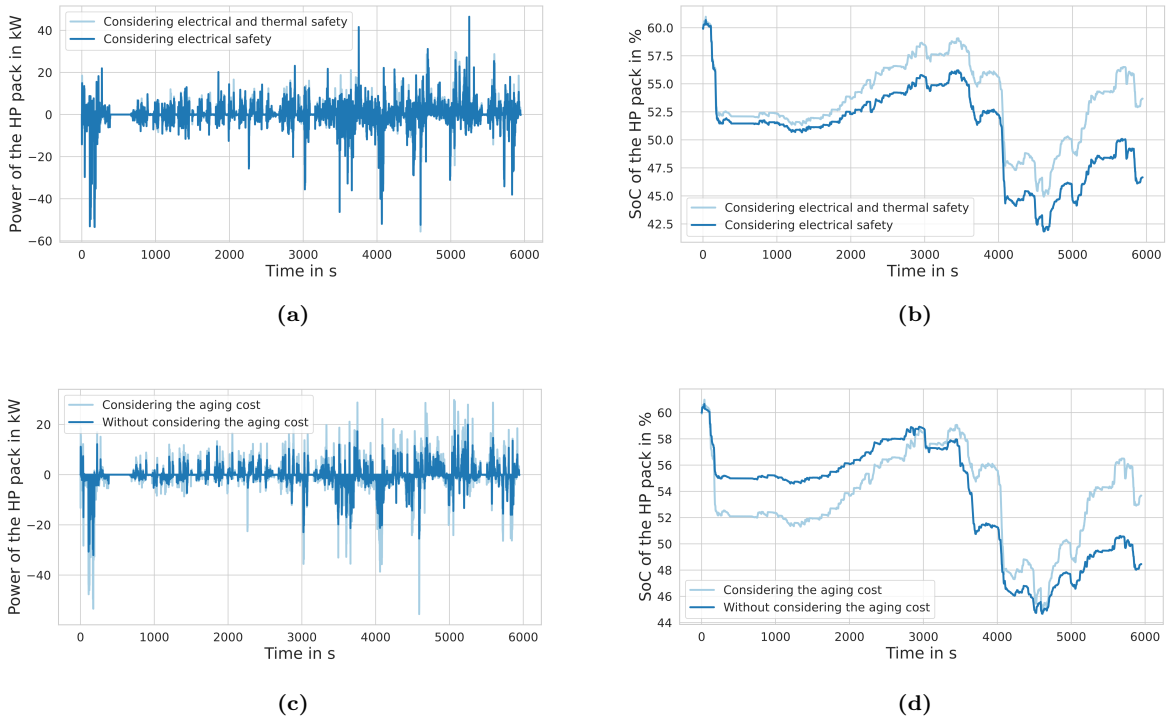


Fig. 11. Training results of the DDPG-based EMS. (a) Comparison of HP pack's power using strategies with and without thermal safety reward term. (b) Comparison of HP pack's SoC using strategies with and without thermal safety reward term. (c) Comparison of HP pack's power using strategies with and without the aging reward term. (d) Comparison of HP pack's SoC using strategies with and without the aging reward term.

EMS without considering thermal safety is adopted as a benchmark and trained with the same data set. Since thermal safety is not critical with the HP pack, the HP pack was used more frequently when thermal safety is taken into consideration, especially in the recuperation phase, as shown in Fig. 11(a). As more regenerated power is absorbed by the HP pack, the SoC of the HP pack is higher than that of the HP pack without thermal safety term in the reward function, as shown in Fig. 11(b). To verify the effectiveness of the minimization of the aging cost, we further design a new DDPG-based EMS without considering the aging cost. Using the same data set to train the new EMS for 120 epochs, the comparison results regarding the HP pack's power and SoC are presented in Fig. 11(c) and (d). Because of the slower aging process of the HP pack, the HP pack can be operated more actively, absorbing and delivering more energy when the aging cost is taken into account. Accordingly, its SoC varies more dynamically compared with that without considering the aging cost. In order to further explore the reduction of the aging cost, the sum of immediate aging cost G, as defined in Eq. (19), over the whole driving distance is evaluated as the total aging cost. As summarized in Table 4, the aging cost per 10000 km of the EMS, considering aging reduction is 47.7 $

Table 4

Aging cost per 10000 km by using different strategies in training results.

With consideration of aging cost ($)	Without consideration of aging cost ($)
1440.5	1488.2

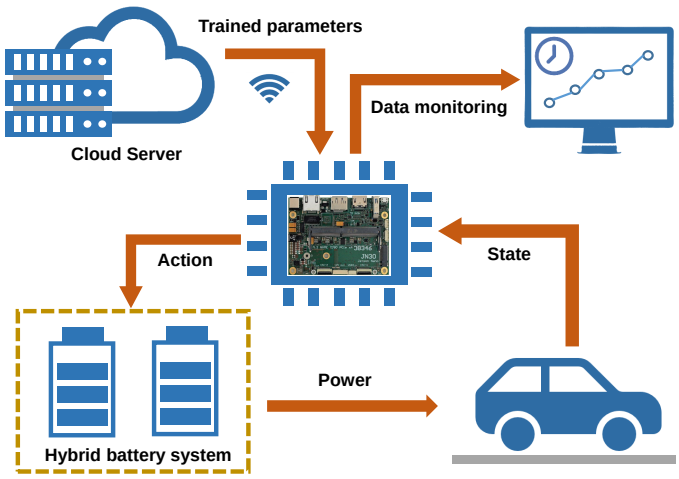


Fig. 12. Schematic of the processor-in-the-loop test with Jetson Nano.

lower than that without aging consideration in the reward function. The longevity of the entire system is prolonged since the aging cost is reduced. The energy loss due to the heat dissipation of the packs and DC-DC converter decreases by 6.015 kJ. This can be explained by a more active operation of the HP pack. Consequently, the energy supplied by the HE pack decreases, which contributes to the lower heat generation of the system.

4.2. Results of the processor-in-the-loop test

After the training process in the cloud server, the trained DDPG-based EMS is further validated in the PiL test with a low-cost embedded device, Nvidia Jetson Nano, as shown in Fig. 12, which demonstrates not only the effectiveness and reliability of the power-split scheme but also the onboard performance, computational burden, and real-time feature. The trained parameters are transmitted to the local embedded device through wireless transmission. The onboard EMS performs the decision-making of the power-split for the HBS under new load profiles, as shown in Fig. 9(b). The validation process and results can be monitored through an additional screen.

The PiL test results are illustrated in Fig. 13 and Fig. 14. Fig. 13(a) shows that the HP pack absorbs a large amount of the regenerated power and delivers sufficient power back to the vehicle under high power

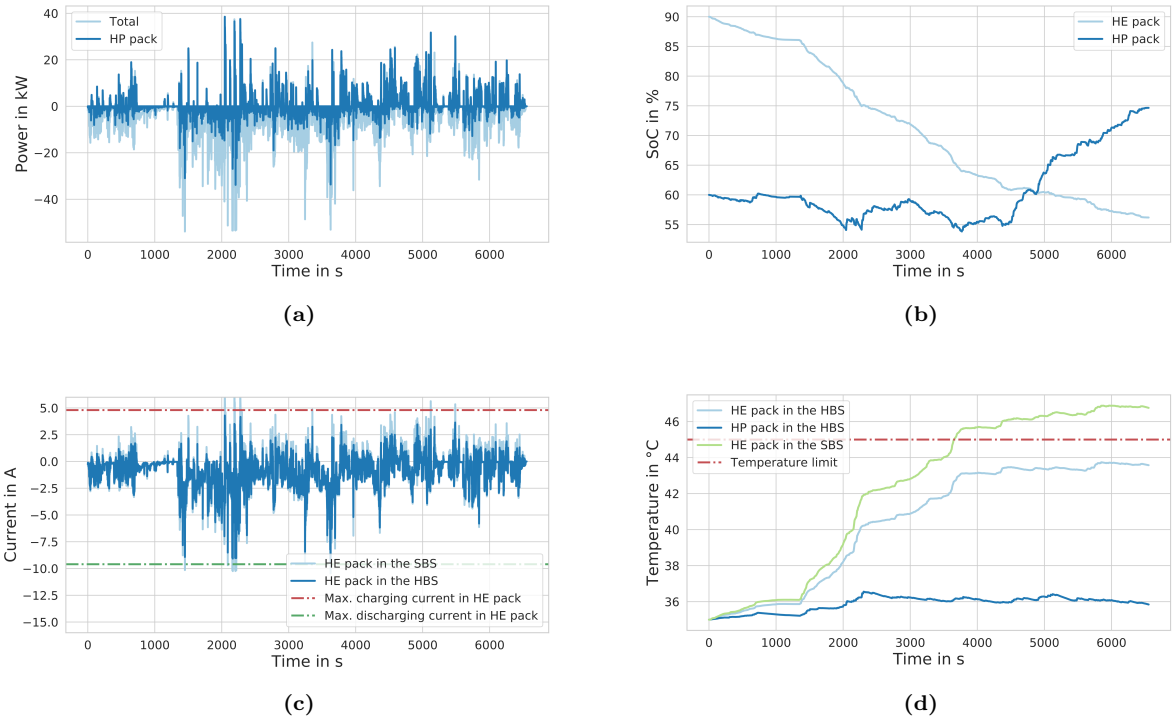


Fig. 13. Validation results of the DDPG-based EMS with the PiL test. (a) Total power of the HBS and power of the HP pack. (b) The SoC tendencies of HE and HP packs within the last training epoch. (c) Comparison between the SBS and HBS in terms of the HE cell's current. (d) Temperature profiles of each pack in SBS and HBS.

Table 5

Aging cost per 10000 km by using different strategies in validation results.

Considering the aging cost ($)	Without considering the aging cost ($)
1441.6	1523.1

demand. From Fig. 13(b), we can see the tendency of each pack's SoC over the entire driving range. In contrast to the continuous decrease of the HE pack's SoC, the SoC of the HP pack fluctuates around 65%, which accords with the fact that the HP pack works as the secondary energy source, even under new load conditions. As shown in Fig. 13(c), unlike the HE cell's overlarge current at some peaks in SBS, the safe operation of the HE pack in HBS is guaranteed by the DDPG-based EMS. Fig. 13(d) depicts the temperature profiles of each battery pack in both SBS and HBS. The HE pack's temperature in HBS is lower than that in SBS, demonstrating that the proposed EMS can constrain the temperature of the HE pack within the safe area. On account of the lower resistance and smaller energy throughput over the validated driving distance, the temperature of the HP pack varies slightly.

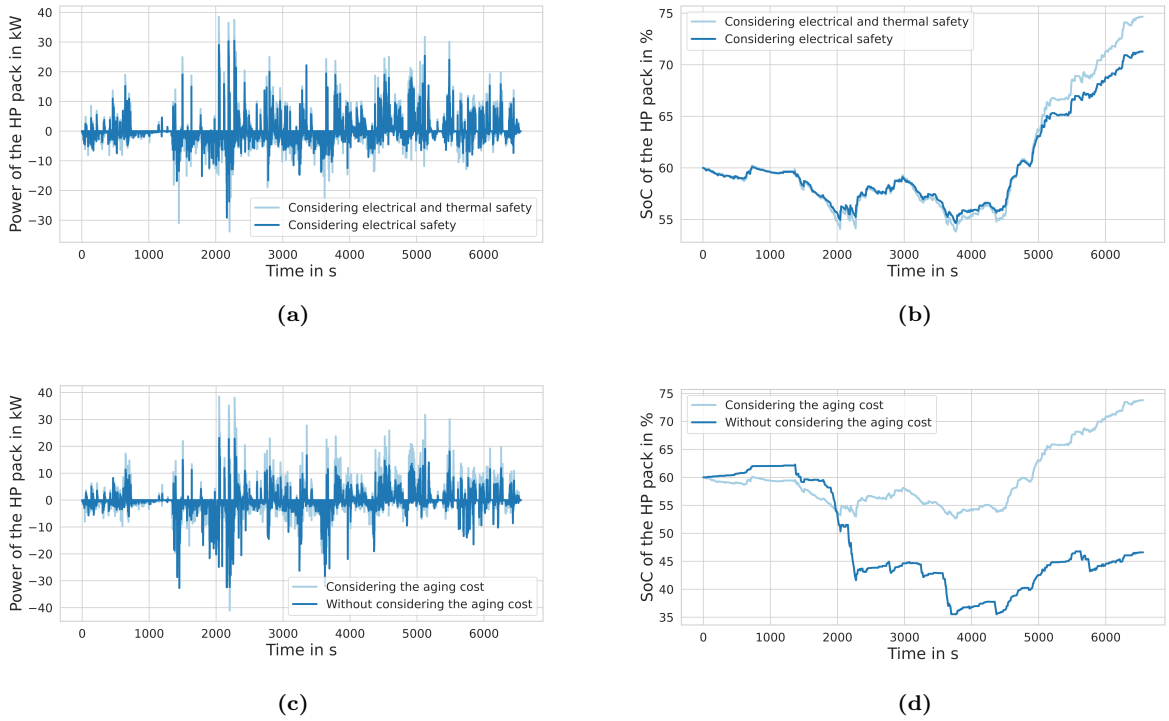


Fig. 14. Validation results of the DDPG-based EMS with the PiL test. (a) Comparison of HP pack's power using strategies with and without thermal safety reward term. (b) Comparison of HP pack's SoC using strategies with and without thermal safety reward term. (c) Comparison of HP pack's power using strategies with and without the aging reward term. (d) Comparison of HP pack's SoC using strategies with and without the aging reward term.

Fig. 14(a) and (b) compare the proposed DDPG-based strategy with its thermal safety-ignored benchmark. With the consideration of thermal safety, the HP pack delivers larger power to the BEV and absorbs more regenerated power, which generated a significant variance of its SoC. With a more active operation of the HP pack, the temperature rise of the HE pack slows down. Fig. 14(c) and (d) show the results with the EMS without consideration of the aging cost. When the equivalent aging cost is taken into account, the HP pack is operated more actively through the absorption and release of the larger amount of the power, contributing to the larger variance of its SoC. Because of the intrinsic stability and longer lifetime of the HP pack, the load imposed on the HE pack is mitigated, contributing to the extended longevity of the entire system. Table 5 summarizes the improvement of the aging cost with the designed reward term. In contrast to the EMS without considering the aging cost of the battery pack, the proposed EMS achieves a reduction of the aging cost of 81.5 $ per 10000 km before the HBS reaches its EoL, representing the prolonged lifetime of the HBS.

Furthermore, the computation burden of the proposed DRL-based strategy was also investigated in the

Table 6
Differences in different learning-based EMSs.

Strategy	Epochs before convergence	Energy loss (kJ)
QL	8000	3479.3
DQL	150	3478.0
DDPG	120	3428.5

Table 7
Aging cost of the DDPG-based EMS per 10000 km compared with that of QL and DQL-based EMSs.

QL ($)	DQL ($)	DDPG ($)
1579.1	1546.2	1441.6

PiL test. The total elapsed time with the embedded device is 46.93 s for the driving task of length 6555 s with 1 s resolution. On average, 7.16 ms is needed for each decision-making with the EMS, which is 0.12% of the time resolution, verifying the real-time feature of the proposed strategy. It can be concluded that the proposed strategy offers the possibility to be applied to the energy management of an electric vehicle in real-time.

4.3. A comparative study with QL and DQL-based EMSs

To further validate the effectiveness of the DDPG-based EMS, QL and DQL-based EMSs were additionally developed. Both of them share the same discretized action space, from 0% to 100%, with 10% intervals. For QL-based EMS, only the total power and the power of the HP pack were adopted as the states to speed up its convergence. The DQL-based EMS utilizes the same state space as the DDPG-based strategy. The same training data is applied to train these strategies. The results concerning convergence rate and energy loss are summarized in Table 6. The DDPG-based EMS achieves not only the fastest convergence rate but also the minimum energy loss, which is 49.5 kJ and 50.8 kJ less than that of DQL-based EMS and QL-based EMS, respectively.

On account of the accumulated aging cost over the driving distance in the validation process, as summarized in Table 7, the DDPG-based EMS achieves the minimum aging cost among three EMSs with the most active operation of the HP pack so that the prolonged longevity of the entire battery system is guaranteed. The aging cost reductions achieved by the DDPG-based EMS per 1000 km are 137.5 $ and 104.6 $ compared with QL-based and DQL-based EMSs, respectively.

4.4. Outlook and applications

In future work, the proposed DDPG-based energy management framework can also be applied to other hybrid energy sources in BEVs and HEVs. Efforts will be put to further increase the algorithm performance by implementing new continuous DRL algorithms. Different configurations of the hybrid battery systems will be considered to optimize the entire system cost considering electrical, thermal, and aging effects. The performance of the learning-based methods will also be compared with the optimization-based methods with experimental tests considering performance and computation efficiency. Another key expansion to the model is the investigation of the energy management strategy by implementing high-fidelity physics-based electrochemical-thermal models [48].

The proposed vehicle-to-cloud energy management framework provides excellent opportunities for improving the energy efficiency and reducing the total life-cycle system cost based on the data collected from the transportation infrastructure, e.g., traffic information, weather information, and charging station information. Future trip information can be forecasted from the historical data for the real-time adaption of the strategy according to the dynamic driving conditions [49].

5. Conclusions

This paper proposed a deep deterministic policy gradient-based health-conscious energy management strategy to achieve the continuous control of the power-split for a hybrid battery system in an electric vehicle. Apart from the electro-thermal model, the experiment-based aging model of the cells considering both capacity fade and power fade in the calendar and cyclic aging was implemented, aiming at simulating the real dynamics of the battery accurately. A new reward function was designed, which comprises the reward terms to increase electrical and thermal safety on the one hand and decrease the energy loss and equivalent aging cost of the whole system on the other hand. The proposed health-conscious strategy is trained under the vehicle-to-cloud framework and validated on a low-cost embedded system with the real-world driving data, which contains different dynamics from various road conditions. Comparative studies with state-of-the-art learning-based energy management strategies have been performed in terms of training efficiency, energy loss reduction and degradation attenuation. The major conclusions are drawn as follows.

- The proposed vehicle-to-cloud energy management strategy can continuously control the power-split between two battery packs, guaranteeing the safe and efficient operation and prolonged longevity of the entire system.
- Compared with the energy management strategy without the consideration of thermal safety, the high-power pack absorbs more regenerated power and the temperature rise of the high-energy pack is slower.

- Compared with the energy management strategy without the consideration of battery aging, the overall aging cost related to the battery replacement of the proposed strategy is lower.
- The proposed strategy reduces the training epochs by 98.5% and 20.0% compared to Q-learning and deep Q-learning-based strategies in processor-in-the-loop tests, indicating a highly improved convergence property.
- The proposed strategy also achieves 50.8 kJ and 49.5 kJ less energy loss, respectively, compared to the Q-learning and deep Q-learning-based strategies. Furthermore, the aging cost of the proposed strategy is also lower than that of Q-learning and deep Q-learning-based strategies.

Acknowledgment

This work is funded by the German Federal Ministry for Transport and Digital Infrastructure (BMVi) with the funding numbers of 03B10502B and 03B10502B2. The authors gratefully acknowledge the support by RWTH Aachen University with the computing resources under the project thes0600.

References

[1] A. Mahmoudzadeh Andwari, A. Pesiridis, S. Rajoo, R. Martinez-Botas, V. Esfahanian, A review of battery electric vehicle technology and readiness levels, Renewable and Sustainable Energy Reviews 78 (2017) 414–430. doi:10.1016/j.rser.2017.03.138.

[2] N. Mebarki, T. Rekioua, Z. Mokrani, D. Rekioua, Supervisor control for stand-alone photovoltaic/hydrogen/ battery bank system to supply energy to an electric vehicle, International Journal of Hydrogen Energy 40 (39) (2015) 13777–13788. doi:10.1016/j.ijhydene.2015.03.024.

[3] S. Ould Amrouche, D. Rekioua, T. Rekioua, S. Bacha, Overview of energy storage in renewable energy systems, International Journal of Hydrogen Energy 41 (45) (2016) 20914–20927. doi:10.1016/j.ijhydene.2016.06.243.

[4] H. F. Gharibeh, A. S. Yazdankhah, M. R. Azizian, Energy management of fuel cell electric vehicles based on working condition identification of energy storage systems, vehicle driving performance, and dynamic power factor, Journal of Energy Storage 31 (2020) 101760. doi:10.1016/j.est.2020.101760.

[5] O. Bohlen, J. Kowal, D. U. Sauer, Ageing behaviour of electrochemical double layer capacitors, Journal of Power Sources 172 (1) (2007) 468–475. doi:10.1016/j.jpowsour.2007.07.021.

[6] O. Veneri, C. Capasso, S. Patalano, Experimental investigation into the effectiveness of a super-capacitor based hybrid energy storage system for urban commercial vehicles, Applied Energy 227 (2018) 312–323. doi:10.1016/j.apenergy.2017.08.086.

[7] F. Balsamo, C. Capasso, G. Miccione, O. Veneri, Hybrid storage system control strategy for all-electric powered ships, Energy Procedia 126 (2017) 1083–1090. doi:10.1016/j.egypro.2017.08.242.

[8] F. Odeim, J. Roes, A. Heinzel, Power management optimization of an experimental fuel cell/battery/supercapacitor hybrid system, Energies 8 (7) (2015) 6302–6327. doi:10.3390/en8076302.

[9] S. Liu, M. Winter, M. Lewerenz, J. Becker, D. U. Sauer, Z. Ma, J. Jiang, Analysis of cyclic aging performance of commercial li4ti5o12-based batteries at room temperature, Energy 173 (2019) 1041–1053. doi:10.1016/j.energy.2019.02.150.

[10] J. Becker, T. Nemeth, R. Wegmann, D. Sauer, Dimensioning and optimization of hybrid li-ion battery systems for evs, World Electric Vehicle Journal 9 (2) (2018) 19. doi:10.3390/wevj9020019.

[11] T. Nemeth, P. J. Kollmeyer, A. Emadi, D. U. Sauer, Optimized operation of a hybrid energy storage system with lto batteries for high power electrified vehicles, in: 2019 IEEE Transportation Electrification Conference and Expo (ITEC), IEEE, 19.06.2019 - 21.06.2019, pp. 1–6. doi:10.1109/ITEC.2019.8790613.

[12] C. M. Martinez, X. Hu, D. Cao, E. Velenis, B. Gao, M. Wellers, Energy management in plug-in hybrid electric vehicles: Recent progress and a connected vehicles perspective, IEEE Transactions on Vehicular Technology 66 (6) (2017) 4534–4549. doi:10.1109/TVT.2016.2582721.

[13] M. Sorrentino, C. Pianese, M. Maiorino, An integrated mathematical tool aimed at developing highly performing and cost-effective fuel cell hybrid vehicles, Journal of Power Sources 221 (2013) 308–317. doi:10.1016/j.jpowsour.2012.08.001.

[14] M. Spaner, A. Rojko, K. Jezernik, Design and realization of hybrid drive with supercapacitor and power flow control, in: 2012 12th IEEE International Workshop on Advanced Motion Control (AMC), IEEE, 25.03.2012 - 27.03.2012, pp. 1–6. doi:10.1109/AMC.2012.6197131.

[15] S. Wang, D. Guo, X. Han, L. Lu, K. Sun, W. Li, D. U. Sauer, M. Ouyang, Impact of battery degradation models on energy management of a grid-connected dc microgrid, Energy 207 (2020) 118228. doi:10.1016/j.energy.2020.118228.

[16] C. Yang, S. Du, L. Li, S. You, Y. Yang, Y. Zhao, Adaptive real-time optimal energy management strategy based on equivalent factors optimization for plug-in hybrid electric vehicle, Applied Energy 203 (2017) 883–896. doi:10.1016/j.apenergy.2017.06.106.

[17] R. Wegmann, V. Döge, J. Becker, D. U. Sauer, Optimized operation of hybrid battery systems for electric vehicles using deterministic and stochastic dynamic programming, Journal of Energy Storage 14 (2017) 22–38. doi:10.1016/j.est.2017.09.008.

[18] X. Hu, C. Zou, X. Tang, T. Liu, L. Hu, Cost-optimal energy management of hybrid electric vehicles using fuel cell/battery health-aware predictive control, IEEE Transactions on Power Electronics 35 (1) (2020) 382–392. doi:10.1109/TPEL.2019.2915675.

[19] P. Kollmeyer, M. Wootton, J. Reimers, T. Stiene, E. Chemali, M. Wood, A. Emadi, Optimal performance of a full scale li-ion battery and li-ion capacitor hybrid energy storage system for a plug-in hybrid vehicle, in: 2017 IEEE Energy Conversion Congress and Exposition (ECCE), IEEE, 01.10.2017 - 05.10.2017, pp. 572–577. doi:10.1109/ECCE.2017.8095834.

[20] H. Peng, J. Li, K. Deng, A. Thul, W. Li, L. Lowenstein, D. U. Sauer, K. Hameyer, An efficient optimum energy management strategy using parallel dynamic programming for a hybrid train powered by fuel-cells and batteries, in: 2019 IEEE Vehicle Power and Propulsion Conference (VPPC), IEEE, Piscataway, NJ, 2019, pp. 1–7. doi:10.1109/VPPC46532.2019.8952323.

[21] Amin, R. T. Bambang, A. S. Rohman, C. J. Dronkers, R. Ortega, A. Sasongko, Energy management of fuel cell/battery/supercapacitor hybrid power sources using model predictive control, IEEE Transactions on Industrial Informatics 10 (4) (2014) 1992–2002. doi:10.1109/TII.2014.2333873.

[22] J. Wu, Z. Wei, W. Li, Y. Wang, Y. Li, D. Sauer, Battery thermal- and health-constrained energy management for hybrid electric bus based on soft actor-critic drl algorithm, IEEE Transactions on Industrial Informatics (2020) 1doi:10.1109/TII.2020.3014599.

[23] W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte, D. U. Sauer, Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation, Journal of Energy Storage 30 (2020) 101557. doi:10.1016/j.est.2020.101557.

[24] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, D. U. Sauer, Online capacity estimation of lithium-ion batteries with deep long short-term memory networks, Journal of Power Sources 482 (2021) 228863. doi:10.1016/j.jpowsour.2020.228863.

[25] A. Biswas, P. G. Anselma, A. Emadi, Real-time optimal energy management of electrified powertrains with reinforcement

learning, in: 2019 IEEE Transportation Electrification Conference and Expo (ITEC), IEEE, 19.06.2019 - 21.06.2019, pp. 1–6. doi:10.1109/ITEC.2019.8790482.

[26] J. Wu, H. He, J. Peng, Y. Li, Z. Li, Continuous reinforcement learning of energy management with deep q network for a power split hybrid electric bus, Applied Energy 222 (2018) 799–811. doi:10.1016/j.apenergy.2018.03.104.

[27] Y. Hu, W. Li, K. Xu, T. Zahid, F. Qin, C. Li, Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning, Applied Sciences 8 (2) (2018) 187. doi:10.3390/app8020187.

[28] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing atari with deep reinforcement learning.
URL http://arxiv.org/pdf/1312.5602v1

[29] H. van Hasselt, A. Guez, D. Silver, Deep reinforcement learning with double q-learning.
URL http://arxiv.org/pdf/1509.06461v3

[30] Y. Li, H. He, J. Peng, J. Wu, Energy management strategy for a series hybrid electric vehicle using improved deep q-network learning algorithm with prioritized replay, DEStech Transactions on Environment, Energy and Earth Sciences (iceee). doi:10.12783/dteees/iceee2018/27794.

[31] W. Li, H. Cui, T. Nemeth, J. Jansen, C. Ünlübayir, Z. Wei, L. Zhang, Z. Wang, J. Ruan, H. Dai, X. Wei, D. U. Sauer, Deep reinforcement learning-based energy management of hybrid battery systems in electric vehicles, Journal of Energy Storage 36 (1) (2021) 102355. doi:10.1016/j.est.2021.102355.

[32] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning.
URL http://arxiv.org/pdf/1509.02971v6

[33] Y. Li, H. He, A. Khajepour, H. Wang, J. Peng, Energy management for a power-split hybrid electric bus via deep reinforcement learning with terrain information, Applied Energy 255 (2019) 113762. doi:10.1016/j.apenergy.2019.113762.

[34] G. Li, W. Zhuang, G. Yin, Y. Ren, Y. Ding, Energy management strategy and size optimization of a lfp/lto hybrid battery system for electric vehicle, in: SAE Technical Paper Series, SAE Technical Paper Series, SAE International400 Commonwealth Drive, Warrendale, PA, United States, 2019. doi:10.4271/2019-01-1003.

[35] J. Shen, A. Khaligh, Design and real-time controller implementation for a battery-ultracapacitor hybrid energy storage system, IEEE Transactions on Industrial Informatics 12 (5) (2016) 1910–1918. doi:10.1109/TII.2016.2575798.

[36] J. Wang, P. Liu, J. Hicks-Garner, E. Sherman, S. Soukiazian, M. Verbrugge, H. Tataria, J. Musser, P. Finamore, Cycle-life model for graphite-lifepo4 cells, Journal of Power Sources 196 (8) (2011) 3942–3948. doi:10.1016/j.jpowsour.2010.11.134.

[37] R. Xiong, J. Cao, Q. Yu, Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle, Applied Energy 211 (2018) 538–548. doi:10.1016/j.apenergy.2017.11.072.

[38] T. Nemeth, A. Bubert, J. N. Becker, R. W. de Doncker, D. U. Sauer, A simulation platform for optimization of electric vehicles with modular drivetrain topologies, IEEE Transactions on Transportation Electrification 4 (4) (2018) 888–900. doi:10.1109/TTE.2018.2869371.

[39] W. Li, D. Cao, D. Jöst, F. Ringbeck, M. Kuipers, F. Frie, D. U. Sauer, Parameter sensitivity analysis of electrochemical model-based battery management systems for lithium-ion batteries, Applied Energy 269 (2020) 115104. doi:10.1016/j.apenergy.2020.115104.

[40] P. Schröer, H. van Faassen, T. Nemeth, M. Kuipers, D. U. Sauer, Challenges in modeling high power lithium titanate oxide cells in battery management systems, Journal of Energy Storage 28 (2020) 101189. doi:10.1016/j.est.2019.101189.

[41] C. Forgez, D. Vinh Do, G. Friedrich, M. Morcrette, C. Delacourt, Thermal modeling of a cylindrical lifepo4/graphite lithium-ion battery, Journal of Power Sources 195 (9) (2010) 2961–2968. doi:10.1016/j.jpowsour.2009.10.105.

[42] P. Keil, S. F. Schuster, J. Wilhelm, J. Travi, A. Hauser, R. C. Karl, A. Jossen, Calendar aging of lithium-ion batteries, Journal of The Electrochemical Society 163 (9) (2016) A1872–A1880. doi:10.1149/2.0411609jes.

[43] T. Bank, J. Feldmann, S. Klamor, S. Bihn, D. U. Sauer, Extensive aging analysis of high-power lithium titanate oxide batteries: Impact of the passive electrode effect, Journal of Power Sources 473 (2020) 228566. doi:10.1016/j.jpowsour.2020.228566.

[44] J. Schmalstieg, S. Käbitz, M. Ecker, D. U. Sauer, A holistic aging model for li(nimnco)o2 based 18650 lithium-ion batteries, Journal of Power Sources 257 (2014) 325–334. doi:10.1016/j.jpowsour.2014.02.012.

[45] T. Nemeth, P. Schröer, M. Kuipers, D. U. Sauer, Lithium titanate oxide battery cells for high-power automotive applications – electro-thermal properties, aging behavior and cost considerations, Journal of Energy Storage 31 (2020) 101656. doi:10.1016/j.est.2020.101656.

[46] P. A. Nelson, K. G. Gallagher, I. D. Bloom, D. W. Dees, Modeling the performance and cost of lithium-ion batteries for electric-drive vehicles - second edition. doi:10.2172/1209682UR.

[47] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization.
URL http://arxiv.org/pdf/1412.6980v9

[48] W. Li, Y. Fan, F. Ringbeck, D. Jöst, X. Han, M. Ouyang, D. U. Sauer, Electrochemical model-based state estimation for lithium-ion batteries with adaptive unscented kalman filter, Journal of Power Sourcesdoi:10.1016/j.jpowsour.2020.228534.

[49] X. Hu, T. Liu, X. Qi, M. Barth, Reinforcement learning for hybrid and plug-in hybrid electric vehicle energy management: Recent advances and prospects, IEEE Industrial Electronics Magazine 13 (3) (2019) 16–25. doi:10.1109/MIE.2019.2913015.