2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)
June 28-30, 2023. Seattle, Washington, USA

Optimal and Adaptive Engine Switch Control for a Parallel Hybrid Electric Vehicle Using a Computationally Efficient Actor-Critic Method

Tong Liu , *Member, IEEE,* Kaige Tan , *Member, IEEE,* Wenyao Zhu , and Lei Feng , *Member, IEEE*

***Abstract*— Energy management strategies (EMSs) are crucial to the fuel economy of hybrid electric vehicles (HEVs). However, due to the lack of efficient solving approaches, most of existing EMSs mainly focus on the optimal torque split between the internal combustion engine (ICE) and the electric motor but neglect improper ICE on/off switches, and thus usually suffer degraded fuel economy and even unacceptable drivability in practice. To tackle this issue, this paper presents a novel EMS that uses an efficient actor-critic (AC) method to regulate ICE switches with limited computation resources. While common AC methods use complex neural networks (NNs) with arbitrary initialization, the proposed AC uses piecewise cubic polynomials whose parameters are initialized based on optimized solutions of dynamic programming (DP). By this means, the AC can quickly converge with high computation efficiency. The testing results from processor-in-the-loop (PIL) simulations showcase that, compared with a rule-based EMS with tabular value functions, the proposed EMS can greatly improve the equivalent fuel economy by eliminating improper ICE switches after only several iterations of adaptive learning and dramatically save the onboard memory space owing to the concise AC structure.**

***Index Terms*— Hybrid electric vehicle, Energy management strategy, Engine switch, Actor-critic method, Adaptive learning**

I. Introduction

The pursuit of sustainable development is expediting the development of electrified vehicles [1], [2]. As a major representative of electrified vehicles, hybrid electric vehicles (HEVs) are popular in the automobile market nowadays [3], [4]. By flexibly allocating the total torque demand to the internal combustion engine (ICE) and the electric motor (EM), HEVs can achieve much higher fuel economy and fewer exhaust emissions than conventional vehicles solely powered by ICEs. Moreover, they effectively avoid the common challenges for full-electric vehicles, such as range anxiety and long-time charging. Nevertheless, due to multiple onboard power sources, HEVs must rely on efficient energy management strategies (EMSs) to maximize their economic and ecological profiles [5].

The currently published EMSs can be generally classified into three types, namely the rule-based (RB-), optimization-based (OB-), and learning-based (LB-) EMSs [6], [7]. RB-EMSs can satisfy the real-time requirement in the online control with decent performances but cannot ensure near-optimal fuel economy and lack adaptivity to various driving scenarios [8]. According to whether complete prior knowledge is necessary, OB-EMSs are further divided into offline and online subgroups. The offline methods, including dynamic programming (DP) [9], particle swarm optimization [10], genetic algorithm [11], and simulated annealing [12], can obtain the non-causal optima in the premise of precise future driving information. Due to the enormous computation loads, they cannot be directly implemented as online controllers but usually serve as benchmarks for other EMSs. By contrast, the online methods [13], comprising Pontryagin's minimum principle (PMP) [14], equivalent consumption minimization strategy [15], and model predictive control, only require driving information in a short-term future. Thus, they are computationally tractable to provide real-time sub-optimal solutions. Nevertheless, their performances are very sensitive to accuracies of the system model and predicted information. Benefiting from the excellent adaptivity and the distinctive model-free property, LB-EMSs, containing supervised/unsupervised learning [16], reinforcement learning [17], and deep reinforcement learning [18], have become a research hotspot in recent years. However, some inevitable defects must be well coped with in practice, for instance, the trade-off between exploitation and exploration, the tedious training process, and the trap of local optima.

In addition to the torque split between ICE and EM, the ICE on/off switch is critical to the overall HEV performance [19], especially to the parallel HEVs, whose wheel speed is directly coupled with the ICE spinning speed. The proper switches can indeed reduce fuel consumption and extend the ICE lifetime; whereas the improper ones, reflected as frequent and/or rapid switches, will not only waste much more energy for ICE restart but also destroy the drivability and even incur severe safety hazards. An effective approach is to optimize the ICE switch and the torque split simultaneously, but the introduction of integer variables will further complicate the optimal control problem (OCP) due to the complex non-linear powertrain dynamics and time-varying driving environment. Thus, the majority of studies adopt heuristic rules to decide the ICE switches, while a small number of emerging EMSs depend on complex optimization algorithms to seek optimal solutions [20], [21]. For this reason, most researchers merely present improvements in numeric results achieved by the novel EMSs but do not analyze the actual consumption of computation resources when these EMSs are executed in real-time.

To tackle the aforementioned issues and bridge the research gap, we propose a computationally efficient EMS

T. Liu, K. Tan, and L. Feng are with the Department of Engineering Design, KTH Royal Institute of Technology, Brinellvägen 83, SE-10044, Stockholm, Sweden. L. Feng is the corresponding author. {tongliu, kaiget, lfeng}@kth.se

W. Zhu is with the Division of Electronics and Embedded Systems, KTH Royal Institute of Technology, Electrum 229, SE-16440, Kista, Sweden. wenyao@kth.se

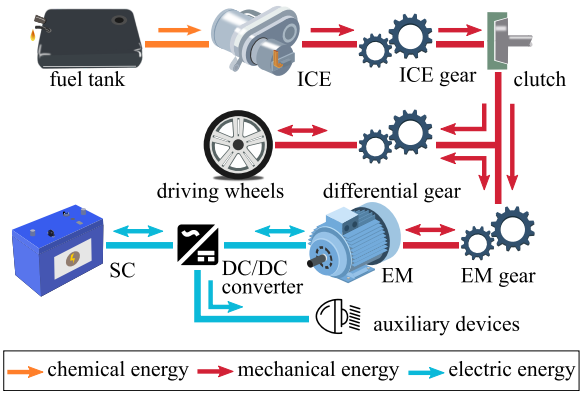


Fig. 1: Parallel HEV Powertrain Architecture

that can simultaneously optimize the ICE switch and torque split for a parallel HEV with limited computation resources. Firstly, the ICE switch is controlled by an efficient actor-critic (AC) method. Unlike the general AC methods with complex neural networks (NNs) and arbitrary initialization, both the actor and the critic in this new method are formulated as third-order piecewise polynomials, whose coefficients are initialized based on the optimal policy and value functions (VFs) by offline DP, respectively. If the ICE is switched on, a value-based PMP controller is employed to allocate the torque demand on ICE and EM. To seek optimality, the costate of PMP is derived by the approximated VF from the critic. The testing results from processor-in-the-loop (PIL) simulations verify the optimality and computation efficiency of the proposed EMS. After only several iterations of adaptive learning, the proposed EMS can effectively avoid improper ICE switches and thus improve the equivalent fuel economy to a close-to-optimal level, more than 20% higher than that by the benchmark, an RB-EMS with tabular VFs. Moreover, owing to the concise structures of piecewise polynomials in AC, the proposed EMS reduces more than 50% of the onboard memory space compared to its counterpart.

The remainder of this paper is organized as follows. Section II establishes the dynamical model of the parallel HEV under study; Section III formulates the HEV energy management problem as an OCP; Section IV elaborates the design process of the proposed EMS; Section V presents the testing results by PIL simulation; and Section VI draws the conclusion and raises the future work.

II. Parallel HEV Powertrain Model

The HEV under study is a lightweight prototype and has a parallel powertrain, which contains a gasoline-driven ICE in the fuel path and a brushless direct current EM in the electric path. Owing to the higher specific power, a supercapacitor (SC) rather than a battery pack is selected as the electric energy storage. The architecture of this HEV powertrain is depicted in Fig. 1, with essential parameters listed in TABLE I.

The total driving time t_f of a driving cycle is uniformly divided into N steps with identical interval $t_s = t_f/N$. At the k^{th} step, $k \in \{0, 1, 2, \cdots, N-1\}$, the net tractive torque

TABLE I: Essential Parameters of HEV

Parameter	Sign	Value	Unit
HEV gross mass	M	216	kg
Gravitational acceleration	g	9.81	$kg{\cdot}m{\cdot}s^{-2}$
Equivalent mass ratio	δ	1.04	/
Tire radius	r	0.26	m
Frontal area	A_f	1.05	m^2
Air drag coefficient	c_d	0.15	$kg{\cdot}m^{-3}$
Rolling resistance coefficient	c_r	0.011	/
ICE gear ratio	R_{ce}	38/31	/
EM gear ratio	R_{em}	38/36	/
Differential gear ratio	R_p	10	/
Lumped efficiency in drive shaft	η_d	0.9	/
Lumped efficiency to recharge SC	η_{rc}	0.25	/
Average SC efficiency	η_{sc}	0.98	/
SC terminal voltage	V_{sc}	40-50	V
SC Nominal capacitance	C_{sc}	107	F
SC Nominal charge capacity	Q_{sc}	5350	C
Average auxiliary power	P_{aux}	10	W
ICE maximum torque	$T_{ce}^{\max}$	3.1	Nm
ICE maximum power	$P_{ce}^{\max}$	1.5	kW
EM maximum torque	$T_{em}^{\max}$	11	Nm
EM maximum power	$P_{em}^{\max}$	3.55	kW

on driving wheels $T_{t,k}$ is calculated by,

$$T_{t,k} = r\left[\delta M a_k + A_f c_d v_k^2/2 + Mg(c_r \cos\alpha_k + \sin\alpha_k)\right], \quad (1)$$

$$a_k = (v_{k+1} - v_k)/t_s, \quad (2)$$

where a, v, and α denote the HEV acceleration, speed, and road slope angle in the longitudinal direction, respectively.

In the parallel powertrain, $T_{t,k}$ can be satisfied individually by the ICE or the EM, or jointly by both, expressed as,

$$T_{t,k} = R_p\left(T_{ce,k} R_{ce} \eta_d + T_{em,k} R_{em} \eta_d^{sign(T_{em,k})}\right), \quad (3)$$

where T_{ce} and T_{em} are the ICE and EM torque outputs.

To efficiently estimate the accumulated energy consumption on fuel and electric paths, the quasi-static modeling method is employed to analyze the energy efficiencies of ICE and EM under stable status. The details of fast dynamics concerning ICE on/off switches and clutch dis/engagement are neglected since they have a negligible effect on the total fuel consumption of long-time driving. For simplification, it is assumed that one ICE switch can be completed within one step. Denote by $s_{ce} \in \{0, 1\}$ a binary variable indicating the current ICE on/off status ("0" means off and "1" means off), and $u_{ce} \in \{0, 1\}$ another one representing the ICE switch instruction. The dynamics and energy consumption concerning ICE switches are described as,

$$s_{ce,k+1} = u_{ce,k}, \quad (4)$$

$$m_{sw,k} = \begin{cases} 0; & s_{ce,k} = u_{ce,k} \\ m^\star; & s_{ce,k} \neq u_{ce,k} \end{cases}, \quad (5)$$

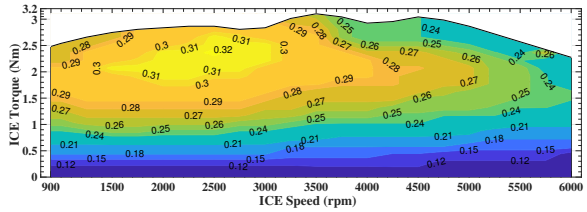
where m_{sw} is the potential equivalent fuel consumption for one switch and is equal to either $m^\star$ or 0.

After the ICE is switched on, the transient fuel consumption during one step, m_{ce}, is defined as,

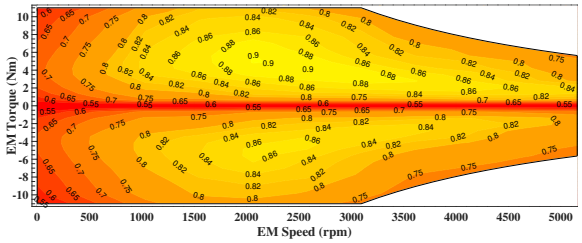
$$m_{ce,k} = t_s \frac{P_{ce,k}}{Q_f} = t_s \frac{T_{ce,k}\omega_{ce,k}}{\eta_{ce}(T_{ce,k}, \omega_{ce,k}) Q_f}, \quad (6)$$

$$\omega_{ce,k} = R_p R_{ce} v_k / r, \quad (7)$$

$$s_{ce,k} = 0 \Rightarrow T_{ce,k} = 0, \quad (8)$$



(a) ICE



(b) EM

Fig. 2: Actuator Efficiency Maps

$$s_{ce,k}=1 \Rightarrow T_{ce,k}\in\left[T_{ce}^{\min}(v_k),\ T_{ce}^{\max}(v_k)\right], \tag{9}$$

where P_{ce}, ω_{ce}, and $\eta_{ce}(\cdot)$ represent the ICE power consumption, spinning speed, and net efficiency, respectively. Q_f is the gasoline lower heating value. The admissible range of T_{ce} is subject to both s_{ce} and v. If the ICE is off, i.e., $s_{ce}=0$, T_{ce} must be 0; otherwise, its upper and lower limits, $T_{ce}^{\max}$ and $T_{ce}^{\min}$, are determined by ω_{ce}, further coupled with v.

Similarly, the transient net power across the SC, P_{sc}, is described as,

$$P_{sc,k}=\frac{P_{em,k}+P_{aux}}{\eta_{sc}^{sign(P_{em,k}+P_{aux})}}, \tag{10}$$

$$P_{em,k}=\frac{T_{em,k}\omega_{em,k}}{\eta_{em}(T_{em,k},\omega_{em,k})^{sign(T_{em,k})}}, \tag{11}$$

$$\omega_{em}=R_pR_{em}v_k/r, \tag{12}$$

where P_{em}, ω_{em}, and $\eta_{em}(\cdot)$ indicate the EM power consumption, spinning speed, and net efficiency, respectively.

Therefore, the SC dynamics can be derived as,

$$V_{sc,k+1}=V_{sc,k}-t_sP_{sc,k}/(C_{sc}V_{sc,k}), \tag{13}$$

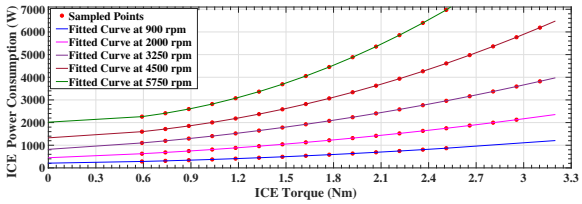
$$SOC_k=C_{sc}V_{sc,k}/Q_{sc}. \tag{14}$$

In general, the state of charge (SOC) of an SC cannot be directly measured during HEV driving. Thanks to the linear relationship between the SOC and V_{sc}, V_{sc} serves as the indicator of SOC in the following EMS design and test.

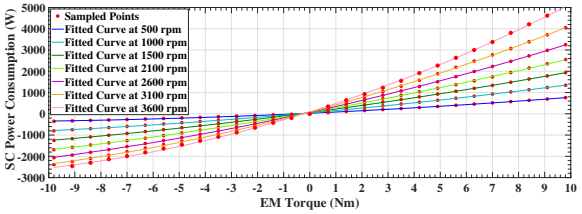
It is noteworthy that both $\eta_{ce}(\cdot)$ and $\eta_{em}(\cdot)$ are originally modeled as explicit 2D maps, with their corresponding torque outputs and spinning speeds, (T_{ce},ω_{ce}) and (T_{em},ω_{em}), as inputs, shown in Fig. 2. However, this method will result in enormous computation overhead and memory occupation in online calculation. To tackle this issue, P_{ce} is approximated as a second-order polynomial of T_{ce}, and similarly, P_{sc} by T_{em}, exhibited by Fig. 3 and expressed as,

$$P_{ce,k}=p_2(\omega_{ce,k})T_{ce,k}^2+p_1(\omega_{ce,k})T_{ce,k}+p_0(\omega_{ce,k}), \tag{15}$$

$$P_{sc,k}=q_2(\omega_{em,k})T_{em,k}^2+q_1(\omega_{em,k})T_{em,k}+q_0(\omega_{em,k}), \tag{16}$$



(a) ICE



(b) SC

Fig. 3: Power Consumption Approximation

where p_2, p_1, and p_0 are coefficients associated to ω_{ce}, while q_2, q_1, and q_0 are coupled to ω_{em}.

III. Optimal Control Problem Formulation

The EMS for this HEV aims to minimize fuel consumption with proper ICE switches over a driving route. More specially, the ICE is expected to continuously work for a relatively long period after being switched on instead of being frequently switched in a short period. Additionally, the SC charge should be well preserved during driving; otherwise, it should be recharged by the ICE after driving.

If the HEV can strictly follow the speed trajectory of a specific driving cycle, $v(k), k\in\{0,1,\cdots,N\text{–}1\}$, then a_k and $T_{t,k}$ at each time step can be calculated by (1) and (2). Thus, the formulated OCP concerning HEV energy management can be derived as below.

$$\underset{\mathbf{u}_k}{\operatorname{argmin}}\sum_{k=0}^{N-1}\left[m_{ce}(\mathbf{x}_k,\mathbf{u}_k)+m_{sw,k}(\mathbf{x}_k,\mathbf{u}_k)\right]+m_{rc}(\mathbf{x}_N) \tag{17}$$

subject to (3)-(9), (12), (13), (15), (16) and the following,

$$\mathbf{x}_k=[V_{sc,k},\ s_{ce,k}]^T, \tag{17a}$$

$$\mathbf{u}_k=[T_{ce,k},\ u_{ce,k}]^T, \tag{17b}$$

$$\mathbf{x}_0=[V_{sc,0},\ 0]^T, \tag{17c}$$

$$V_{sc,k}\in\left[V_{sc}^{\min},\ V_{sc}^{\max}\right]^T, \tag{17d}$$

$$T_{em,k}\in\left[T_{em}^{\min}(v_k),\ T_{em}^{\max}(v_k)\right]^T, \tag{17e}$$

$$m_{rc}(\mathbf{x}_N)=C_{sc}\left(V_{sc,0}^2-V_{sc,N}^2\right)/\left(2\eta_{rc}Q_f\right), \tag{17f}$$

where $\mathbf{x}$ and $\mathbf{u}$ are the state and control vectors, and m_{rc} is the extra fuel consumption to recharge the SC after driving. Due to the mutual restraint in (3), T_{em} is not an independent control variable anymore, but its upper and lower limits, $T_{em}^{\min}$ and $T_{em}^{\max}$, should be considered.

The OCP (17) is a mixed integer nonlinear programming problem and is firstly solved offline by DP. The optimized solutions, including the optimal control policy on ICE switch $u_{ce,k}^*=\pi(\mathbf{x}_k,t_k)$, the tabular VF $Y(\mathbf{x},t)$, and the optimal ICE on/off status profile $s_{ce}^\circ(t_k)$, are employed to design the online EMS.

IV. Online EMS Design

To satisfy the real-time requirement and attain near-optimal fuel economy in online applications, a computationally efficient EMS of hierarchical structure and adaptive learning is designed to successively determine the optimal ICE on/off status and calculate the torque split solution. The ICE switch control is implemented by an efficient AC method, wherein the actor and critic are designed based on the policy $\pi(\cdot)$ and the VF $Y(\cdot)$ from offline DP. The torque split control is realized by a value-based PMP method, which also relies on $Y(\cdot)$ to estimate its optimal costate.

A. ICE Switch Rules Derived by DP

Although DP cannot directly provide real-time solutions, it is generally treated as the most effective approach to the global optimum and widely utilized to design or refine other online EMSs. Thus, effective ICE switch rules $u^*_{ce,k}=\pi(\mathbf{x}_k,t_k)$ can be derived from DP solutions. Examining the values of $\pi(\mathbf{x}_k,t_k)$ at any fixed pair of $s_{ce,k}$ and t_k, one can find that $u^*_{ce,k}$ is determined by a threshold value of $V_{sc,k}$ corresponding to $s_{ce,k}$. If $s_{ce}=1$ or 0, then the threshold value is $V^1_{sc}(t_k)$ or $V^0_{sc}(t_k)$. Consequently, $u^*_{ce,k}$ is determined by the following inequality rules:
when $s_{ce,k}=0$,

$$u^*_{ce,k}=\begin{cases}0; & V_{sc,k}\geq V^0_{sc}(t_k)\\ 1; & V_{sc,k}<V^0_{sc}(t_k)\end{cases}; \tag{18}$$

when $s_{ce,k}=1$,

$$u^*_{ce,k}=\begin{cases}0; & V_{sc,k}>V^1_{sc}(t_k)\\ 1; & V_{sc,k}\leq V^1_{sc}(t_k)\end{cases}. \tag{19}$$

The rationale of the rules is that at each time step k when the actual SC voltage is larger than its threshold value, the ICE should be off, i.e., $u^*_{ce,k}=0$; otherwise, the ICE should be on, i.e., $u^*_{ce,k}=1$. For online usage, the two threshold vectors, $V^0_{sc}(t_k)$ and $V^1_{sc}(t_k)$, should be pre-stored into the onboard processor.

These ICE switch rules enjoy the advantage of rapid execution but suffer from two deficiencies, namely potential risks to huge memory overhead and unstable control performance. Since $V^1_{sc}(t_k)$ and $V^0_{sc}(t_k)$ are two time-based vectors, their sizes will linearly go up as the driving time increases. More importantly, DP solutions are non-causal and thus cannot ensure a robust performance when the prior knowledge cannot match the real driving scenario. As a consequence, an adaptive method for ICE switches is designed in Subsection IV-B to overcome the two issues.

B. AC Method for Adaptive ICE Switch Control

Iteratively upgrading the control policies through the information interaction with the environment, LB-EMSs can effectively eliminate the adverse impact from the deviation between prior knowledge and real driving scenarios, and finally attain close-to-optimal performances. As a typical LB-EMS, the AC method employs two NNs. The actor NN generates the control actions at each step, while the critic NN estimates the state-action value (Q value) for the coefficient update of two NNs. Compared to other prevalent LB-EMSs, such as Q-learning, deep Q-network, and deep deterministic policy gradient, the AC method can provide continuous control actions with only two NNs, inferring its superior computation power with a limited memory occupation.

However, general AC methods have two drawbacks caused by the usage of NNs. The first one is the lengthy training time and the possible divergence due to the "cold start" at the initial training stage. The second one is the enormous computation overheads when NNs have complex structures. To solve these issues, this proposed AC method employs piecewise polynomials of simple structures to replace NNs and exploits DP solutions to initialize their parameters before online learning.

Among DP solutions from Section III, $\pi(\cdot)$ and $Y(\cdot)$ are used to formulate and initialize the actor and critic, and $s^\circ_{ce}(t_k)$ serves as the basis to divide the entire driving cycle into several segments since s_{ce} has the most critical effect on the evolution of $\mathbf{x}$ and $Y(\cdot)$. Our investigation shows that the functions of $\pi(\cdot)$ and $Y(\cdot)$ are too complex to be approximated by simple polynomials, but are more suitable to piecewise polynomials. The segments along the timeline are correlated with $s^\circ_{ce}(t_k)$. Note that although $s^\circ_{ce}(t_k)$ is subject to the initial value of V_{sc}, similar segments of $Y(\cdot)$ can be defined and the fitting errors are very small [22]. Since s_{ce} is a binary variable, $Y(\cdot)$ can be divided into 2 parts $Y^1(V_{sc},t)$ and $Y^0(V_{sc},t)$ concerning ICE on and off for easy usage. Denote by N_{sw} the number of ICE switches in $s^\circ_{ce}(t_k)$ and $\boldsymbol{W}=[\boldsymbol{w}^1;\boldsymbol{w}^2;\cdots;\boldsymbol{w}^{N_{md}+1}]$ the coefficient matrix. The approximated VFs, parameterized by $\boldsymbol{W}$, $\tilde{Y}^1_{\boldsymbol{W}}(\cdot)$ and $\tilde{Y}^0_{\boldsymbol{W}}(\cdot)$, can be expressed as piecewise cubic polynomials,

$$\begin{aligned}\tilde{Y}^1_{\boldsymbol{W}}(V_{sc},t)=&w^n_1V^3_{sc}+w^n_2V^2_{sc}t+w^n_3V_{sc}t^2+w^n_4t^3+\\ &w^n_5V^2_{sc}+w^n_6V_{sc}t+w^n_7t^2+w^n_8V_{sc}+w^n_9t+w^n_{10},\end{aligned} \tag{20}$$

$$\begin{aligned}\tilde{Y}^0_{\boldsymbol{W}}(V_{sc},t)=&w^n_{11}V^3_{sc}+w^n_{12}V^2_{sc}t+w^n_{13}V_{sc}t^2+w^n_{14}t^3+\\ &w^n_{15}V^2_{sc}+w^n_{16}V_{sc}t+w^n_{17}t^2+w^n_{18}V_{sc}+w^n_{19}t+w^n_{20},\end{aligned} \tag{21}$$

where $\boldsymbol{w}^n=[w^n_1,w^n_2,\cdots,w^n_{20}]$ is the coefficient vector for the n^{th} section. Hence, the explicit VF $Y(\cdot)$ in the critic can be replaced by its approximations $\tilde{Y}^1_{\boldsymbol{W}}(\cdot)$ and $\tilde{Y}^0_{\boldsymbol{W}}(\cdot)$, which only require 20 $(N_{md}+1)$ coefficients.

Since the output of the actor should be continuous and differentiable but u_{ce} is binary, the AC method cannot directly implement $\pi(\cdot)$ illustrated by (18) and (19). To solve this issue, we reformulate $\pi(\cdot)\in(0,\ 1)$ as a logistic function,

$$\pi(\mathbf{x}_k,t_k)=\begin{cases}1-\dfrac{1}{1+e^{-\left(V_{sc,k}-V^0_{sc}(t_k)\right)}}; & s_{ce,k}=0\\ 1-\dfrac{1}{1+e^{-\left(V_{sc,k}-V^1_{sc}(t_k)\right)}}; & s_{ce,k}=1\end{cases}. \tag{22}$$

The value of $\pi(\cdot)$ depends on the difference between V_{sc} and its reference $V^1_{sc}(t_k)$ or $V^0_{sc}(t_k)$. If V_{sc} is larger than its reference, $\pi(\cdot)$ is smaller than 0.5; if V_{sc} is smaller, $\pi(\cdot)$ is larger than 0.5; otherwise, if they are identical, $\pi(\cdot)$ is exactly

equal to 0.5. The actual control decision u^*_{ce} is determined by the value of $\pi(\cdot)$, written as,

$$u^*_{ce,k} = \begin{cases} 1; & \pi(\mathbf{x}_k, t_k) > 0.5 \\ 0; & \pi(\mathbf{x}_k, t_k) < 0.5 \\ s_{ce,k}; & otherwise \end{cases}. \quad (23)$$

$V^1_{sc}(t_k)$ and $V^0_{sc}(t_k)$ are approximated by piecewise cubic polynomials as well. Their approximations parameterized by the coefficient matrix $\boldsymbol{F} = [\boldsymbol{f}^1; \boldsymbol{f}^2; \cdots; \boldsymbol{f}^{N_{md}+1}]$, $\tilde{V}^1_{sc,\boldsymbol{F}}(t_k)$ and $\tilde{V}^0_{sc,\boldsymbol{F}}(t_k)$, are written as,

$$\tilde{V}^1_{sc,\boldsymbol{F}}(t_k) = f^n_1 t^3_k + f^n_2 t^2_k + f^n_3 t_k + f^n_4 \quad (24)$$

$$\tilde{V}^0_{sc,\boldsymbol{F}}(t_k) = f^n_5 t^3_k + f^n_6 t^2_k + f^n_7 t_k + f^n_8 \quad (25)$$

where $\boldsymbol{f}^n = [f^n_1, f^n_2, \cdots, f^n_8]$ is the coefficient vector for the n^{th} section. Since $\tilde{V}^1_{sc,\boldsymbol{F}}(t_k)$ and $\tilde{V}^0_{sc,\boldsymbol{F}}(t_k)$ have only 1 free variable t_k, $\boldsymbol{f}^n$ is composed of 8 coefficients, less than $\boldsymbol{w}^n$, and $\boldsymbol{F}$ contains 8 $(N_{md}+1)$ coefficients totally.

The policy used in the actor is denoted by $\pi_{\boldsymbol{F}}(\cdot)$. It follows the principle by (22) but replace $V^1_{sc}(t_k)$ and $V^0_{sc}(t_k)$ by their approximations $\tilde{V}^1_{sc,\boldsymbol{F}}(t_k)$ and $\tilde{V}^0_{sc,\boldsymbol{F}}(t_k)$.

During online control, the temporal difference (TD) learning [23] is employed to update the critic coefficient matrix $\boldsymbol{W}$. More specially, the TD error $e_{\boldsymbol{W}}$ is defined as,

$$e_{\boldsymbol{W},k} = m_{ce,k} + m_{sw,k} + \tilde{Y}_{\boldsymbol{W}}(\mathbf{x}_{k+1}, t_{k+1}) - \tilde{Y}_{\boldsymbol{W}}(\mathbf{x}_k, t_k), \quad (26)$$

where $\tilde{Y}_{\boldsymbol{W}}(\cdot)$ is the approximation of $Y(\cdot)$.

The loss function l based on an individual sample and its gradient $\boldsymbol{\Delta}_c \in \mathbb{R}^{20}$ are written as,

$$l_k = e^2_{\boldsymbol{W},k}/2, \quad (27)$$

$$\boldsymbol{\Delta}_{c,k} = \frac{\partial l_k}{\partial \boldsymbol{w}^n}. \quad (28)$$

Accordingly, the update of actor coefficient matrix $\boldsymbol{F}$ aims to minimize the cumulative fuel consumption. Hence, gradient descent is performed to update $\boldsymbol{F}$ in the direction of reducing the Q value, $Q(\mathbf{x}_k, \mathbf{u}_k, t_k)$, written as,

$$\boldsymbol{\Delta}_{a,k} = \frac{\partial Q(\mathbf{x}_k, \mathbf{u}_k, t_k)}{\partial u^*_{ce,k}} \cdot \frac{\partial \pi_{\boldsymbol{F}}(\mathbf{x}_k, t_k)}{\partial \boldsymbol{f}^n} \quad (29)$$

where $\boldsymbol{\Delta}_a \in \mathbb{R}^8$ is the gradient of $Q(\cdot)$ in terms of $\boldsymbol{f}^n$.

Because u^*_{ce} is binary and the critic can only provide $\tilde{Y}(\cdot)$ rather than $Q(\cdot)$, $\boldsymbol{\Delta}_a$ cannot be directly calculated by (29). To tackle this issue, we make use of the system dynamics to estimate $Q(\cdot)$ and adopt difference operation referring to different values of u_{ce} to replace the first partial differential operation on the right-hand side of (29). Denote by $\breve{u}^*_{ce}$ the opposite value of u^*_{ce}, and then this partial differential operation can be rewritten as (30).

To ensure the training robustness with limited onboard computation resources, batch gradient descent [24] is employed to update $\boldsymbol{W}$ and $\boldsymbol{F}$. A batch of maximum size $\mathcal{K}$ samples $s_k = \{\mathbf{x}_k, \mathbf{u}_k, m_{ce,k}, m_{sw,k}\}$ is initialized as empty before an episode starts. At each step, if the batch is not full, the new sample will be appended to the batch, but the coefficient update will not be performed; otherwise, $\boldsymbol{W}$ and $\boldsymbol{F}$ will be updated immediately based on the existing samples in the batch, and then the batch will be reset to empty for accommodating new samples. The expressions for updating $\boldsymbol{W}$ and $\boldsymbol{F}$ are given below.

$$\boldsymbol{W}' \leftarrow \boldsymbol{W} - \beta_c \sum_{k \in \mathcal{K}} \boldsymbol{\Delta}_{c,k}, \quad (31)$$

$$\boldsymbol{F}' \leftarrow \boldsymbol{F} - \beta_a \sum_{k \in \mathcal{K}} \boldsymbol{\Delta}_{a,k}, \quad (32)$$

where β_c and β_a are the learning rates of the critic and actor, respectively; and for distinction, $\boldsymbol{W}'$ and $\boldsymbol{F}'$ are denoted to indicate the newly updated coefficient matrices.

C. Torque Split Control

When the ICE is switched on, i.e., $s_{ce}=1$, the torque split control is responsible for calculating T^*_{ce} and T^*_{em} to satisfy T_t. Since the main contribution of this paper is the efficient ICE switch control, we directly employ a value-based PMP controller from our previous work to perform this task [22]. Above all, the Hamiltonian $\mathcal{H}$ is defined as,

$$\mathcal{H}_k = P_{ce,k}/Q_f - \lambda_k P_{sc,k}/(C_{sc} \cdot V_{sc,k}), \quad (33)$$

$$T^*_{ce,k} = \underset{T_{ce,k}}{\operatorname{argmin}} \mathcal{H}_k \quad (34)$$

where λ is the costate.

The PMP performance is dominated by the optimality of λ. Thanks to the essential equivalence between DP and PMP, the optimal costate λ^* can be quickly derived from $Y^1(\cdot)$,

$$\lambda^*_k = \frac{\partial Y^1(V_{sc,k}, t_k)}{\partial V_{sc,k}}. \quad (35)$$

Since $Y^1(\cdot)$ can be obtained from offline DP solutions in the format of a 2-dimension lookup table, it is directly utilized to design the benchmark EMS without adaptivity. By contrast, the proposed EMS makes use of a costate estimator to always receive the up-to-date $\tilde{Y}^1_{\boldsymbol{W}}(\cdot)$ from the critic for estimating λ^* in online usage, expressed as,

$$\lambda^*_k \approx \frac{\partial \tilde{Y}^1_{\boldsymbol{W}}(V_{sc,k}, t_k)}{\partial V_{sc,k}}. \quad (36)$$

Next, T^*_{ce} can be solved by (34) and T^*_{em} by (3) thereafter.

To summarize, the complete framework of the proposed EMS containing two modules is shown in Fig. 4.

V. PIL Simulation Results

To demonstrate the advantages in computational efficiency and adaptivity of the proposed EMS, PIL simulations are conducted based on a resource-constrained microprocessor to compare the proposed LB-EMS with a benchmark EMS that employs the same ICE switch rules defined in Subsection IV-A but neither approximates the tabular VF by piecewise polynomials nor updates the VF online. The comparison includes not only the control performances, such as SC charge sustain, ICE switches, and equivalent fuel efficiency, but also the computation efficiencies, represented by CPU utilization and memory occupation.

$$\frac{\partial Q(\mathbf{x}_k,\mathbf{u}_k,t_k)}{\partial u^*_{ce,k}} \approx \frac{m_{ce,k|u^*_{ce,k}} + m_{sw,k|u^*_{ce,k}} + \tilde{Y}\left(\mathbf{x}_{k+1}|u^*_{ce,k},t_{k+1}\right) - m_{ce,k|\check{u}^*_{ce,k}} - m_{sw,k|\check{u}^*_{ce,k}} - \tilde{Y}\left(\mathbf{x}_{k+1}|\check{u}^*_{ce,k},t_{k+1}\right)}{u^*_{ce,k} - \check{u}^*_{ce,k}} \quad (30)$$

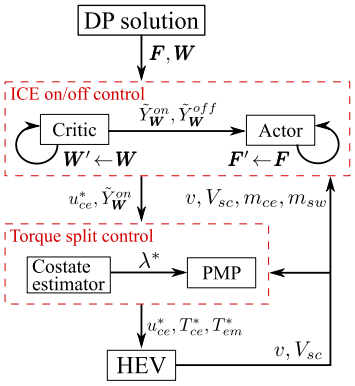


Fig. 4: EMS Architecture

A. Driving Route

To obtain the reliable estimation of control performances in realistic operating conditions, an actual driving cycle containing several steep uphills and downhills instead of the standard ones is selected to design and test the proposed and benchmark EMSs. Shown in Fig 5(a), this cycle is extracted from the university campus of KTH Royal Institute of Technology and provided by the project AD-EYE[1]. The overall length of this cycle is roughly 2300 m, and the total driving time is 345 s. The profiles of road slope angle and HEV speed&acceleration are illustrated in Figs 5(b) and (c). It is noteworthy that there are two slope angle curves in Fig. 5(b). The solid purple one is milder and represents a low-fidelity estimation of reality, depicted by the dashed green one. The former is used to design EMSs, whereas the latter is to test EMSs.

B. PIL Simulation Platform

As exhibited in Fig. 6, the PIL simulation platform consists of three modules, namely an onboard processor for executing the online EMS, a host computer for running the HEV model, and a USB cable for real-time communication. To indicate the outstanding computation efficiency of the proposed EMS, a microprocessor of very limited computation resources, *STM32L476RGT6*[2] (up to only $80\,MHz$ frequency with maximal $1\,Mbyte$ flash memory and $128\,Kbyte$ SRAM), is selected to store and run the EMS, which is converted into C code by *MATLAB/Simulink*.

To balance the control performance and computation overhead, the sampling periods for the ICE switch control and toque split control are set as $1\ s$ and $0.1\ s$, respectively. Additionally, $V_{sc,0}$ is set as $48\ V$, a value close to but lower than the upper limit for flexible operation at the start.

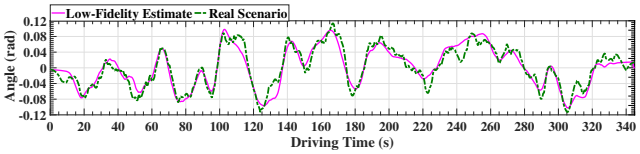
C. Initialization of Actor and Critic

Figs. 7 - 9 present the approximation results by converting the explicit DP results into piecewise polynomials. The red

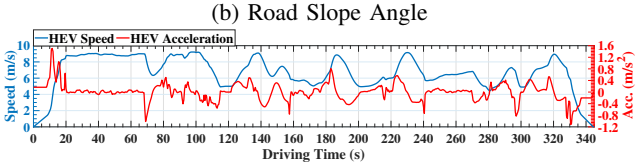
[1]https://www.adeye.se/open-kth

[2]https://www.st.com/en/microcontrollers-microprocessors/stm32l476rg.html

(a) Road Map



(b) Road Slope Angle



(c) Speed and Acceleration Profiles

Fig. 5: Driving Route Information

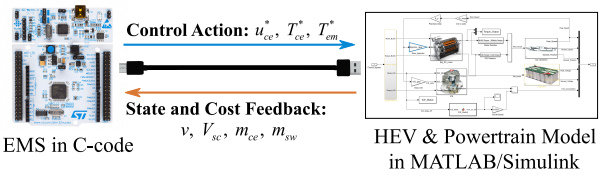
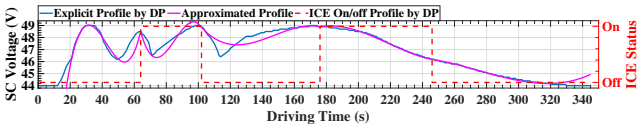


Fig. 6: PIL Simulation Platform

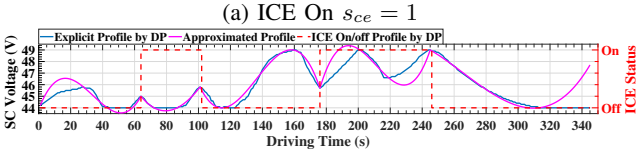
dashed line in Fig. 7 implies that the ICE is switched 4 times by offline DP; thus, the complete driving cycle is divided into 5 sections. As a result, The approximations of $V^1_{sc}(t_k)$ and $V^0_{sc}(t_k)$, $\tilde{V}^1_{sc,\boldsymbol{F}}(t_k)$ and $\tilde{V}^0_{sc,\boldsymbol{F}}(t_k)$, are represented by 5 independent polynomials for actor initialization and illustrated by the purple curves in Fig. 7(a) and (b). It can be seen that the approximations can capture the decisive characteristics of the original curves, with the normalized root mean square error (NRMSE) smaller than 4%. Similarly, two tabular VFs, $Y^1_{sc}(\cdot)$ and $Y^0_{sc}(\cdot)$ shown in Figs. 8(a) and 9(a), are also approximated by piecewise polynomials at corresponding sections for critic initialization. The fitting errors are shown in Figs. 8(b) and 9(b), with NRMSE smaller than 2.5%.

D. Control Performances

The PIL simulation results from the proposed and benchmark EMSs are summarized in Fig. 10 and TABLE II. As displayed in Fig. 10(a), the equivalent fuel efficiency by the proposed EMS is merely 161.5 km/L before online learning, indeed lower than that by the benchmark EMS of

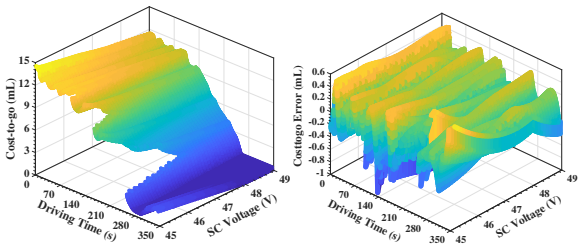


(a) ICE On $s_{ce} = 1$



(b) ICE Off $s_{ce} = 0$

Fig. 7: Threshold Vectors Approximation

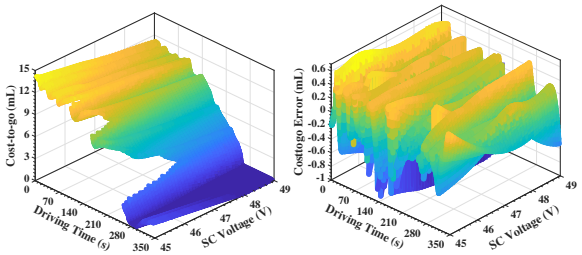


(a) Tabular Value (b) Fitting Error

Fig. 8: Value Function Concerning ICE On $s_{ce} = 1$

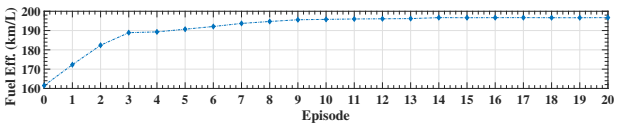
162.9 km/L; however, it surges to 196.7 km/L after less than 10 iterations of adaptive learning, reaching 94.5% of the DP optimum of 208 km/L and thus exceeding 20% higher than that by the benchmark EMS. Moreover, it can be seen from Fig. 10(b) that the final SC voltage variation by the proposed EMS is only -0.39 V, implying a much better charge sustain during driving in contrast to the one by the benchmark EMS of nearly -2 V.

The essential reasons for the better performances of the proposed EMS are twofold. First and foremost, the low-fidelity estimate of road slope angle deceives the offline DP, which provides suboptimal coefficients to the benchmark EMS and the proposed EMS at the initializing stage. However, through adaptive learning, the AC method can efficiently upgrade their coefficients and effectively eliminate those improper ICE switches thereafter. This argument can be adequately verified by the results in Fig. 10(c). Although the ICE driven by the proposed EMS works roughly 20 s longer than that driven by the benchmark EMS in total, it is only ignited once, meaning that it continuously works for a long time and provides power to propel the HEV or recharge the SC. On the contrary, the benchmark EMS

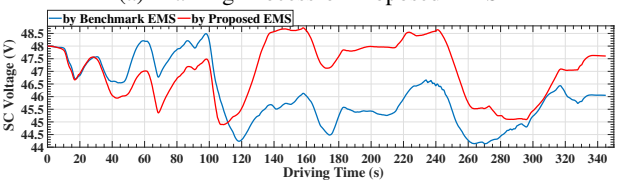


(a) Tabular Value (b) Fitting Error

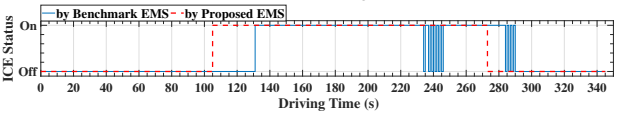
Fig. 9: Value Function Concerning ICE Off $s_{ce} = 0$



(a) Training Process of Proposed EMS



(b) SC Voltage



(c) ICE On/off Status

Fig. 10: PIL Simulation Results

TABLE II: Results Comparison

Control Strategy	Benchmark EMS	Proposed EMS
Final Voltage Variation (V)	-1.95	-0.39
Total Fuel Consumption (mL)	14.08	11.68
ICE On/off Switch (pair)	10	1
Average ICE Efficiency (%)	29.8	30.6
Equivalent Fuel Efficiency (km/L)	162.9	196.7
Flash Memory Occupation (Kbyte)	173.25	84.19
RAM Occupation (Kbyte)	42.17	46.88
Max. CPU Utilization (%)	5.02	24.70
Avg. CPU Utilization (%)	2.41	3.69

switches the ICE 20 times within 3 min, incurring a great deal amount of energy loss for multiple ICE restarts and clutch dis/engagement. In the second place, the performance of the PMP controller highly relies on the optimality of the costate, and the optimal costate is estimated by the VFs in the critic. After adaptive learning, the refine VFs enable close-to-optimal costate and thus improved PMP performance. The average ICE efficiency of the proposed EMS is 30.6%, higher than that of the benchmark EMS, 29.8%.

E. Computation Efficiency

The computation efficiency refers to the usage of onboard computation resources for running tested EMSs in a real-time context, and relevant results are listed in TABLE II. Mainly because the AC method has an adaptive learning mechanism and is more complex than the RB method, the RAM occupation and CPU utilization of the proposed EMS are higher than that of the benchmark EMS. However, the increment in average CPU utilization is only about 1.3%, and the extra RAM demand is less than 5 $Kyte$, inferring the excellent computation efficiency of the AC method in both decision-making and coefficient updates. However, by virtue of the utilization of the piecewise polynomials for approximating the threshold vectors and VFs, the proposed EMS only needs to save several tens of parameters rather than the long vectors or large lookup tables containing several tens of thousands of explicit values. Therefore, it requires only 84.19 $Kbyte$ flash memory space and saves more than 50% compared with its counterpart of 173.25 $Kbyte$.

VI. Conclusion and Future Work

This paper presents a novel EMS that contains an efficient AC method to optimize ICE switches and improve the fuel economy for a parallel HEV. On the one hand, both the actor and critic are formulated as piecewise cubic polynomials to reduce the computation and memory overheads in online control; on the other hand, the coefficients of these polynomials are initialized based on the offline DP solutions for a rapid convergence. The superiority of this proposed EMS over the RB-EMS with tabular VFs is verified through the PIL simulations. The testing results reveal that only after several iterations of adaptive learning, the proposed EMS can eliminate improper ICE switches and obviously improve the fuel economy, 20% better than that of its counterpart. In addition, owing to the concise structures of the polynomials in AC, this new EMS saves more than 50% of the onboard memory space.

Due to the limitation of the research scope, the proposed EMS is only designed for parallel HEVs and tested on a specific driving route. In future studies, this method will be applied to various types of HEVs on diverse driving cycles. Moreover, apart from solely pursuing the compelling fuel economy, computationally efficient EMSs for multiple-objective optimization, including exhaust emission and component aging, ought to be explored for improving the HEV comprehensive performance.

ACKNOWLEDGMENT

This project is supported by Chinese Scholarship Council (CSC), KTH Excellence in Production Research (XPRES), and Trustworthy Edge Computing Systems and Applications (TECoSA).

References

[1] M. A. Hannan, F. A. Azidin, and A. Mohamed, "Hybrid electric vehicles and their challenges: A review," *Renew. Sustain. Energy Rev.*, vol. 29, pp. 135–150, 2014.

[2] C. M. Martinez, X. Hu, D. Cao, E. Velenis, B. Gao, and M. Wellers, "Energy Management in Plug-in Hybrid Electric Vehicles : Recent Progress and a Connected Vehicles Perspective," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4534–4549, 2017.

[3] S. F. Tie and C. W. Tan, "A review of energy sources and energy management system in electric vehicles," *Renew. Sustain. Energy Rev.*, vol. 20, pp. 82–102, 2013.

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

[5] T. Liu, W. Tan, X. Tang, J. Zhang, Y. Xing, and D. Cao, "Driving conditions-driven energy management strategies for hybrid electric vehicles: A review," *Renew. Sustain. Energy Rev.*, vol. 151, no. 111521, pp. 1–16, 2021.

[6] X. Hu, J. Han, X. Tang, and X. Lin, "Powertrain Design and Control in Electrified Vehicles: A Critical Review," *IEEE Trans. Transp. Electrif.*, vol. 7, no. 3, pp. 1990–2009, 2021.

[7] Z. Chen, Y. Liu, M. Ye, Y. Zhang, and G. Li, "A survey on key techniques and development perspectives of equivalent consumption minimisation strategy for hybrid electric vehicles," *Renew. Sustain. Energy Rev.*, vol. 151, no. 111607, pp. 1–19, 2021.

[8] S. Zhang and R. Xiong, "Adaptive energy management of a plug-in hybrid electric vehicle based on driving pattern recognition and dynamic programming," *Appl. Energy*, vol. 155, pp. 68–78, 2015.

[9] C. Maino, D. Misul, A. Musa, and E. Spessa, "Optimal mesh discretization of the dynamic programming for hybrid electric vehicles," *Appl. Energy*, vol. 292, no. 116920, pp. 1–15, 2021.

[10] S. Y. Chen, Y. H. Hung, C. H. Wu, and S. T. Huang, "Optimal energy management of a hybrid electric powertrain system using improved particle swarm optimization," *Appl. Energy*, vol. 160, pp. 132–145, 2015.

[11] X. Lü, Y. Wu, J. Lian, Y. Zhang, C. Chen, P. Wang, and L. Meng, "Energy management of hybrid electric vehicles: A review of energy optimization of fuel cell hybrid power system based on genetic algorithm," *Energy Convers. Manag.*, vol. 205, no. 112474, pp. 1–26, 2020.

[12] Z. Chen, C. C. Mi, B. Xia, and C. You, "Energy management of power-split plug-in hybrid electric vehicles based on simulated annealing and Pontryagin's minimum principle," *J. Power Sources*, vol. 272, pp. 160–168, 2014.

[13] S. Xie, X. Hu, S. Qi, X. Tang, K. Lang, Z. Xin, and J. Brighton, "Model predictive energy management for plug-in hybrid electric vehicles considering optimal battery depth of discharge," *Energy*, vol. 173, pp. 667–678, 2019.

[14] T. Liu, L. Feng, and W. Zhu, "Fuel Minimization of a Hybrid Electric Racing Car by Quasi-Pontryagin's Minimum Principle," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5551–5564, 2021.

[15] M. Razi, N. Murgovski, T. McKelvey, and T. Wik, "Design and Comparative Analyses of Optimal Feedback Controllers for Hybrid Electric Vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 2979–2993, 2021.

[16] Y. Liu, Y. Zhang, H. Yu, Z. Nie, Y. Liu, and Z. Chen, "A novel data-driven controller for plug-in hybrid electric vehicles with improved adaptabilities to driving environment," *J. Clean. Prod.*, vol. 334, no. 130250, pp. 1–14, 2022.

[17] Y. Zou, T. Liu, D. Liu, and F. Sun, "Reinforcement learning-based real-time energy management for a hybrid tracked vehicle," *Appl. Energy*, vol. 171, pp. 372–382, 2016.

[18] H. Wang, H. He, Y. Bai, and H. Yue, "Parameterized deep Q-network based energy management with balanced energy economy and battery life for hybrid electric vehicles," *Appl. Energy*, vol. 320, no. 119270, pp. 1–12, 2022.

[19] F. Ju, N. Murgovski, W. Zhuang, X. Hu, Z. Song, and L. Wang, "Predictive energy management with engine switching control for hybrid electric vehicle via ADMM," *Energy*, vol. 263 Part E, no. 125971, pp. 1–13, 2023.

[20] S. Onori and L. Tribioli, "Adaptive Pontryagin's Minimum Principle supervisory controller design for the plug-in hybrid GM Chevrolet Volt," *Appl. Energy*, vol. 147, pp. 224–234, 2015.

[21] N. Guo, J. Shen, R. Xiao, W. Yan, and Z. Chen, "Energy management for plug-in hybrid electric vehicles considering optimal engine ON/OFF control and fast state-of-charge trajectory planning," *Energy*, vol. 163, pp. 457–474, 2018.

[22] T. Liu, W. Zhu, K. Tan, M. Liu, and L. Feng, "A Low-Complexity and High-Performance Energy Management Strategy of a Hybrid Electric Vehicle by Model Approximation," in *Proc. IEEE 18th Int. Conf. Autom. Sci. Eng. (CASE)*, (Mexico City, Mexico), pp. 455–462, IEEE, 2022.

[23] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32 – 50, 2009.

[24] P. Baldi, "Gradient Descent Learning Algorithm Overview: A General Dynamical Systems Perspective," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 182 – 195, 1995.