Integrated Thermal and Energy Management of Connected Hybrid Electric Vehicles Using Deep Reinforcement Learning

Hao Zhang, Boli Chen, Nuo Lei, Bingbing Li, Rulong Li and Zhi Wang

***Abstract*—The climate-adaptive energy management system holds promising potential for harnessing the concealed energy-saving capabilities of connected plug-in hybrid electric vehicles. This research focuses on exploring the synergistic effects of artificial intelligence control and traffic preview to enhance the performance of the energy management system (EMS). A high-fidelity model of a multi-mode connected PHEV is calibrated using experimental data as a foundation. Subsequently, a model-free multistate deep reinforcement learning (DRL) algorithm is proposed to develop the integrated thermal and energy management (ITEM) system, incorporating features of engine smart warm-up and engine-assisted heating for cold climate conditions. The optimality and adaptability of the proposed system is evaluated through both offline tests and online hardware-in-the-loop tests, encompassing a homologation driving cycle and a real-world driving cycle in China with real-time traffic data. The results demonstrate that ITEM achieves a close to dynamic programming fuel economy performance with a margin of 93.7%, while reducing fuel consumption ranging from 2.2% to 9.6% as ambient temperature decreases from 15°C to -15°C in comparison to state-of-the-art DRL-based EMS solutions.**

***Index Terms*—Climate-adaptive, plug-in hybrid electric vehicles, deep reinforcement learning, integrated thermal and energy management, optimality, adaptability.**

I. Introduction

PLUG-IN hybrid electric vehicles (PHEVs) have remarkable potential to augment conventional powertrain efficiency and significantly curtail carbon emissions [1, 2]. By combining the advantages of series-parallel and power-split hybrid powertrain [3], a novel multi-mode dedicated hybrid transmission (DHT) has gained widespread attention and application, offering enhanced control flexibility and striking energy-saving benefits [4]. Though the concept of multi-mode-PHEVs have been proved promising for enhancing fuel economy, the actual performance under real driving conditions is highly dependant on the energy management system (EMS) and thermal management system (TMS) [5], since the engine efficiency and TMS-related accessory loads are greatly influenced by the coolant temperature. Current research often presupposes that the engine is pre-heated at the commencement of this mission [6], but vehicles invariably encounter cold start conditions [7]. Moreover, it is challenging to maintain coolant temperature in PHEV due to the intermittent operation of engine [8]. Given that using electric heater in cold climate can be highly energy consuming, optimizing the integrated thermal and energy management (ITEM) system could engender reliance on engine assisted heating instead of electric heating [9].

Effective control systems are required to address the inherent coupling in the EMS and TMS [10], and a handful of researches have begun to explore the synergistic optimization of ITEM [11]. Pham et al. proposed an integrated strategy for energy and thermal management applicable to parallel HEVs using the equivalent consumption minimization strategy (ECMS), where the battery state of charge (SOC) is designed to be temperature-related [12]. Zahraei et al. introduced an energy management framework incorporates engine thermal management, where they developed a cost function based on SOC and coolant temperature to obtain the optimal SOC trajectory using dynamic programming (DP) [13]. Though the above literatures have initiated the exploration of more nuanced and accurate system models to depict the reciprocal impact of EMS and TMS [14], substantial challenges still persist in this field due to the complexity of solving coupled thermal and energy management in varying climate conditions [15, 16]. To effectively address this problem, research might need to concentrate on the following areas: 1) the use of real-time traffic and terrain data provided by intelligent transportation system (ITS) and geographic information system (GIS) to achieve predictive ITEM for near-global optimality [17, 18]; 2) adopting optimization methodologies empowered by reinforcement learning (RL) to solve complex energy and thermal management problems [19].

Recently, some studies have made use of the information from ITS for model predictive control (MPC) to solve the ITEM of HEVs [17]. With reliance on traffic data, Wang et al. developed a speed prediction-based ITEM system built upon

Manuscript received XXXX, 2023; revised XXXX, 2023; accepted XXXX, 2023. Date of publication XXXX, 2023; date of current version XXXX, 2023. This work was supported in part by the State Key Laboratory of Automotive Safety and Energy under Grants ZZ2023-041 and in part by the Dongfeng Motor Corporation Ltd., China under Grants CGSQ2022111518. (Corresponding authors: Zhi Wang and Boli Chen.)

H. Zhang, N. Lei and Z. Wang are with the State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China (e-mail: hao_thu@foxmail.com; wangzhi@tsinghua.edu.cn).

B. Chen and B. Li are with the Department of ELectronic and Electrcial engineering, University College London, London WC1E 6BT, UK (e-mail: boli.chen@ucl.ac.uk; bingbli@seu.edu.cn).

Rulong Li is with the Dongfeng Motor Corporation Ltd., Wuhan 430058, China (e-mail: lirl@dfmc.com.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

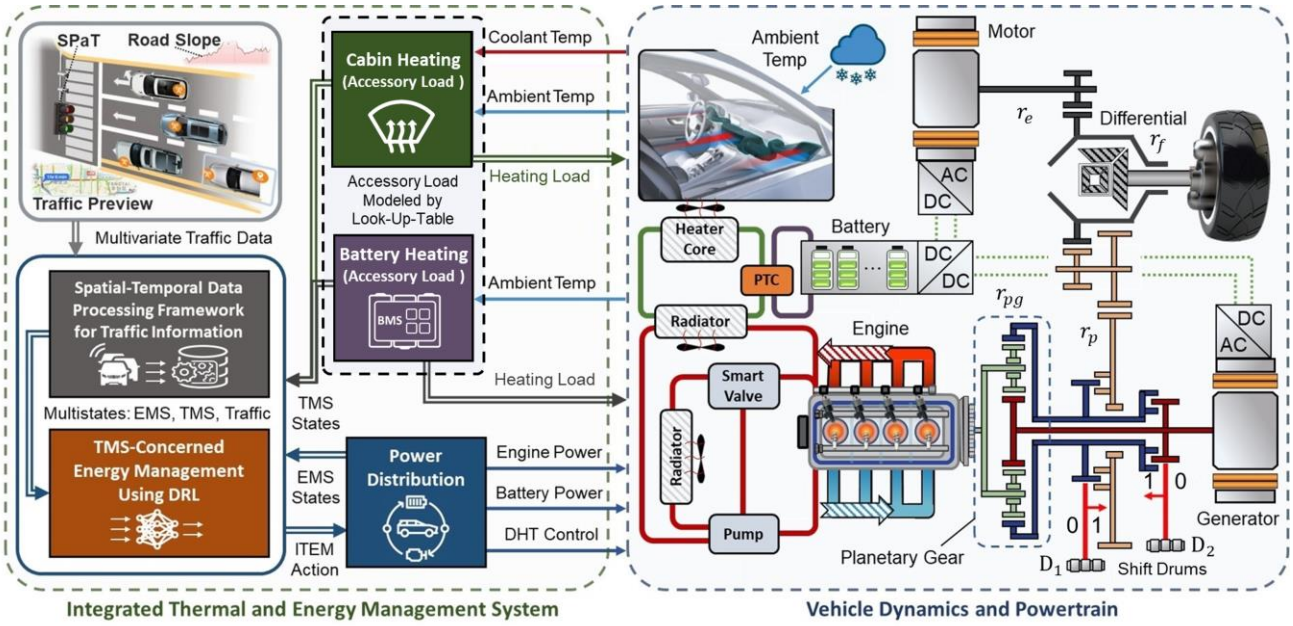


Fig. 1. The overall design of the intelligent ITEM system and dual-mode PHEV powertrain configuration.

MPC to reduce the energy consumption of cabin air conditioning by optimization compressor schedule [20]. Hu et al. proposed a multi-horizon MPC for ITEM to utilize both short and long-range speed prediction, considering that the EMS and TMS differ a lot in terms of time constants [21]. However, the shortcomings of MPC in terms of real-time performance severely impact its industrial applications [22], and its effectiveness heavily relies on the model accuracy [23].

Comparatively, data-driven approaches hold promise in addressing the aforementioned issues. By reframing the EMS into Markov decision processes (MDPs) [24], several researches used deep Q-network (DQN) [25] and double DQN [26] to establish EMS with discretized action space, namely engine output power, and the results proved RL methods outperform conventional online optimization methods, such as ECMS, in both fuel economy optimality and computational burden [27]. Further, Actor-critic (AC) framework can be adopted to enable continuous action-based EMS [28, 29]. For example, Deep Deterministic Policy Gradient (DDPG) has been adopted for competent EMS with continuous output of engine power [30, 31]. Moreover, some improved RL algorithms capable of handling high-dimensional state variables [32], such as twin-delayed DDPG (TD3) can be combined with multivariate trip information for traffic-aware EMS, which have been applied to improve the optimality and adaptability of EMSs in complex driving cycles with less computational resource [33]. Though excellent examples of RL have already been demonstrated in the field of EMS [34], its application in ITEM remains largely untapped.

Therefore, it is of utmost significance to develop DRL-based ITEM system for multi-mode PHEVs [35]. Also, The combination of artificial intelligence control, real-time traffic and terrain data has the potential to yield synergistic effects, ultimately enhancing the overall efficiency of the powertrain [36]. To the best of the authors' understanding, there is a scarcity of research focused on the trip-oriented integrated thermal and energy management system for PHEVs using DRL. This area of study is crucial in the development of the powertrain domain control unit (DCU) for PHEVs that operate under varying climatic conditions during real-driving scenarios. To fill this gap, this research is based on the promising multi-mode PHEV that embodies series, parallel, and power-split functionalities. In this context, a powertrain model with high-fidelity thermal and energy consumption characteristics is built. Then, a DRL-based ITEM system is proposed, as depicted in Fig.1, which integrates multi-source traffic and terrain information processed by a spatio-temporal data processing (STDP) framework in real-time.

The principal features of the proposed ITEM strategy encompass the following elements: (1) a novel model-free optimization methodology is proposed for the trip-oriented ITEM of a multi-mode PHEV with engine-assisted heating for cold climate operation, where the powertrain model is modeled by experimental data; (2) multivariate states, which encompass the ambient temperature and engine coolant temperature, in addition to traffic and terrain information, are integrated into the state space of the RL agent, which promotes the optimal decision-making in real-world driving situations; (3) numerous features including a spatio-temporal data processing (STDP) framework, bounded double Q-values, and delayed policy updates are merged to design the ITEM agent for enhanced learning ability, and the results under homologation and real-world driving cycles verify the optimality and adaptability of the proposed control system.

The rest of this article is structured as follows. Section II describes the powertrain model of a connected PHEV. Section III presents the design of DRL-based ITEM strategy. In Section IV, the testing and validation setup are explained and the benchmark strategies are introduced. Section V analyses the optimality and adaptability of the proposed RL-ITEM in

different temperature conditions and under various driving scenarios. Section VI summarizes the concluding remarks.

II. Powertrain Model and Experimental Calibration

A. Multi-Mode PHEV Longitudinal Dynamic Model

The PHEV in this work adopts a planetary gear (PG)-based DHT with dual clutches, as shown in Figure 1, where the shift drums D_1 and D_2 control the system operating in series hybrid (SH), parallel hybrid (PH), and power-split hybrid (PSH) modes. When operating in the PSH mode, the PG functions as an electric continuously variable transmission (ECVT), where generator (GEN) acts as a speed regulator. The longitudinal dynamic model is explained from Eq. (1) to Eq. (3), and the powertrain parameters can be found in Table I.

TABLE I
Specifications of the Vehicle

Component	Parameters	Value
Vehicle	Total mass	1830 kg
	Rolling resistance	0.0062
	Air resistance coefficient	0.325
Engine	Maximum power	120 kW
	Maximum torque	270 N·m
	Maximum speed	5200 rpm
Generator	Maximum power	50 kW
	Maximum torque	115 N·m
Motor	Maximum power	70 kW
	Maximum torque	150 N·m
Battery	Battery Capacity	85 A·h
	Nominal Voltage	345 V
	Number of cells in series	105

$$T_{dem} = r\left(m_v \frac{dv}{dt} + (f + sin\vartheta)m_v g + \frac{1}{2}\rho C_D A v^2\right) \quad (1)$$

$$\omega_{DM} = \frac{r_f r_e}{r} v \quad (2)$$

$$\begin{cases} r_f r_p T_{ICE} + r_f r_e T_{DM} = T_{dem} \\ (1 + r_{pg}) r_e \omega_{ICE} = r_p r_{pg} \omega_{DM} \end{cases}, if \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{cases} r_f r_e T_{DM} = T_{dem} \\ \omega_{ICE} = \omega_{GEN} \end{cases}, if \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\begin{cases} \frac{r_p r_{pg}}{(1 + r_{pg})} T_{ICE} + r_e T_{DM} = \frac{T_{dem}}{r_f} \\ (1 + r_{pg})\omega_{ICE} = \frac{r_p r_{pg}}{r_e}\omega_{DM} + \omega_{GEN} \end{cases}, if \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (3)$$

where T_{dem} refers to the torque demand, and a represents acceleration. Air density is denoted by ρ, while gravitational acceleration is represented by g. Additionally, m_v, v and A represent the vehicle mass, velocity, and frontal area. Also, r_{PG}, r_f, r_e and r_f refer to the gear ratio of the PG, final dive and the transmission sets shown in Fig.1. Moreover, the wheel radius is represented by r, respectively. Rolling resistance is defined as f. The air resistance coefficient and slope gradient are indicated by C_d and ϑ, respectively. $T_{ICE}, T_{GEN}, T_{DM}, \omega_{ICE}, \omega_{GEN}$, and ω_{DM} refer to the torque and angular speed of the engine, generator, and drive motor, respectively.

B. Powertrain Model Calibrated with Experimental Data

The efficiency map of the generator can be observed in Fig. 2 (a). The power of electric machines (EMs) can be calculated according to Eq. (4), where T_{EM}, ω_{EM} and η_{EM} represent the torque, angular speed and efficiency of EMs. Besides, the engine fuel consumption model considering thermal effect is corrected on the basis of the steady-state map, which is shown in Fig. 2 (b), and the dynamic model is represented by Eq. (5) and Eq. (6). In addition, T_{tar} and T_{col} are the target and actual value of the engine coolant temperature, respectively, also, the α and β are fitting parameters calibrated with experimental data. And the thermal dynamics of engine coolant T_{col} is modeled based on Eq. (7). Also, the engine speed is limited to 1200 rpm when the coolant temperature is less than 40℃.

$$P_{EM} = T_{EM}\omega_{EM}\eta_{EM}(T_{MG}, \omega_{EM})^{\mu} \quad (4)$$

$$\dot{m} = \left(1 + \alpha\left(\frac{T_{tar} - T_{col}}{T_{tat} - 20}\right)^{\beta}\right) P_{ICE} BSFC(T_{ICE}, \omega_{ICE}) \quad (5)$$

$$\omega_{ICE} \in \begin{cases} [0,1200rpm], & T_{col} < 40℃ \\ [0,4500rpm], & T_{col} \geq 40℃ \end{cases} \quad (6)$$

$$\dot{T}_{col} = \frac{(LHV\dot{m} - P_{ICE}) - (\dot{Q}_{exh} + \dot{Q}_{rad} + \dot{Q}_{cab})}{m_{ICE} C_{ICE}} \quad (7)$$

where μ determines whether the EM operates as a motor or a generator. It is assigned a value of -1 when the EM functions as a motor. Conversely, μ is set to 1 when the EM serves as a generator. Also, LHV refers to the lower heating value of gasoline, and the heat brought by exhaust gases, emitted via air convection and delivered for cabin heating through radiator are denoted as $\dot{Q}_{exh}$, $\dot{Q}_{rad}$ and $\dot{Q}_{cab}$, respectively. Besides, the m_{ICE} and C_{ICE} represent the equivalent thermal mass and specific heat capacity of the engine body and cooling system.

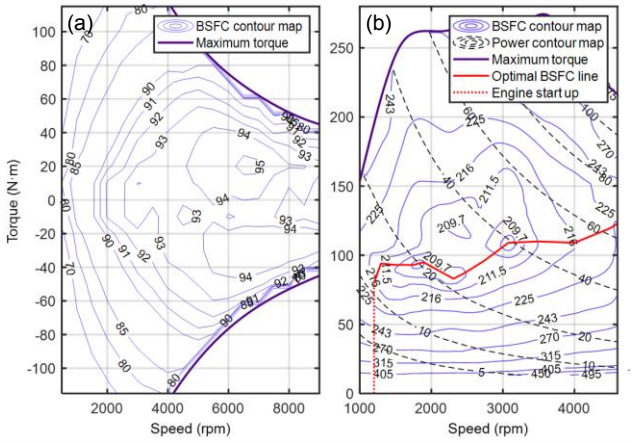


Fig. 2. Generator map (a) and engine BSFC map (b).

The battery's charge-discharge characteristics are modeled with the open circuit voltage and internal resistance, depicted in Fig. 3. By employing Eq. (8), the battery output power can be obtained. Also, the current can be calculated using Eq. (9).

$$P_{bat} = P_{trac} - P_{ICE} - P_{aux} \quad (8)$$

$$I_{bat} = \frac{U_{oc}(SOC) - \sqrt{{U_{oc}}^2(SOC) - 4R_{bat}P_{bat}}}{2R_{bat}} \quad (9)$$

where P_{trac}, P_{ICE} and P_{aux} represent the power of traction, engine and auxiliary components. This paper mainly focuses

on the ITEM operation in cold climate, with an ambient temperature range of [-15°C, 15°C], thus the accessory load of TMS-related auxiliary components can be approximated by its average value according to experiments, as shown in Fig. 4. Rather than using positive temperature coefficient (PTC) device only, the heat of coolant can be utilized for cabin heating when its temperature reaches 75 °C, the warm state of engine coolant, resulting in reduction of accessory load.

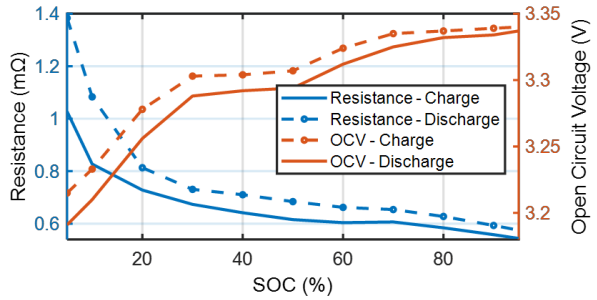


Fig. 3. Battery open circuit voltage and resistance.

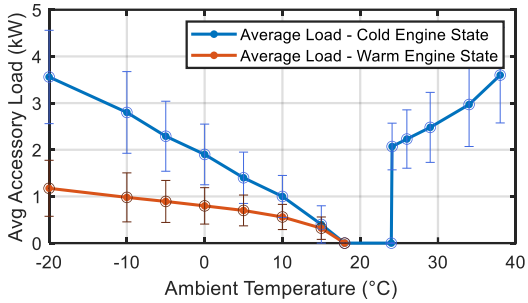


Fig. 4. Average accessory load of the TMS-related auxiliary components in case of cold and warm status of engine coolant.

The simulation results of the model after being calibrated using experimental data are presented in Fig. 5, which exhibit a high level of consistency with the bench test data. Notably, the actual energy management strategy implemented in the vehicle controller of Dongfeng Motor is designated as the benchmark (BMK) strategy, which is calibrated using finite state machine for mode selection and look-up-table for determining output power of the engine and battery, as shown in Fig. 5 (a). The simulated engine power aligns well with the actual strategy, and the trend of coolant temperature and SOC captured by the simulation accurately characterize the real system, as illustrated in Fig. 5 (b) and Fig. 5 (c).

III. Reinforcement Learning-based Integrated Thermal and Energy Management Strategy

A. Problem Formulation with Reinforcement Learning

In our research, we harness the reinforcement learning for the optimization of integrated thermal and energy management. The core capability of RL lies in its competency to manage assignments that follow the MDP. This property empowers it to construct a precise correlation between environmental states and the ideal responses. Four elements form the nucleus of the RL iterative process: (a) environment

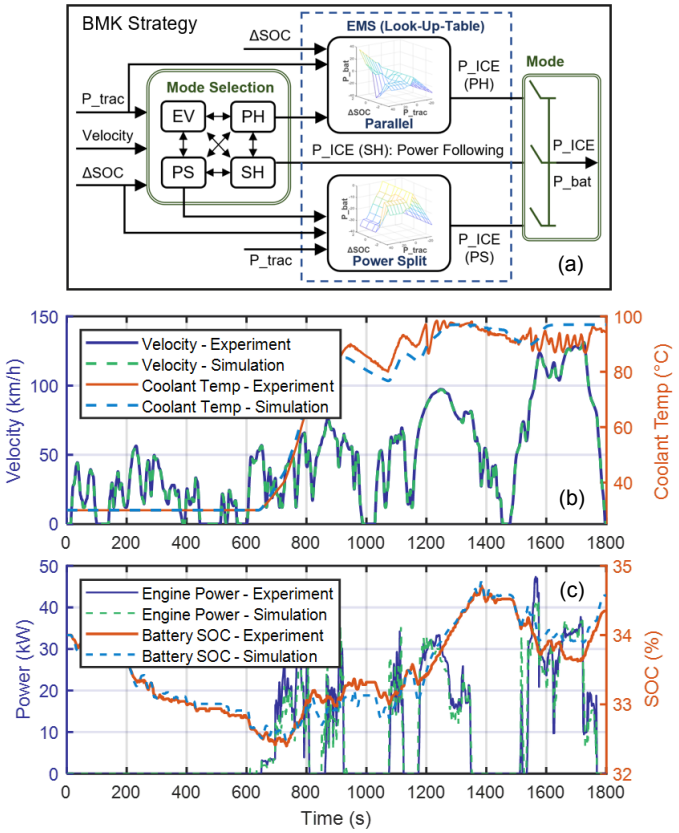


Fig. 5. Validation of model and benchmark control system.

model, (b) value function $V^{\pi}(s_t)$, (c) samples $e_t = (s_t, a_t, s_{t+1})$ and (d) policy $\pi_\theta(s_t)$ parametrized by θ. Within this context, the control policy, symbolized as π_t^*, is developed focusing on boosting the overall return in the long run rather than seeking immediate gains. The optimization is to maximize the total discounted reward, as shown below

$$\pi_t^* = \underset{\pi}{\operatorname{argmax}}\, \mathbb{E}_\pi \left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t(s_t, a_t) \right] \tag{10}$$

where $r_t(s_t, a_t)$ denotes the reward associated with the state and action at time step t. Also, γ is defined as the discount factor belongs to [0,1]. The reward function r_t is designed to include the engine fueling rate (mL/s), denoted as $\dot{m}_f(a_t)$, and the battery SOC penalty $\varsigma(s_t)$, which are both set to negative as shown in Eq. (11). The weight λ balances the engine consumption and battery charge depletion.

$$r(s_t, a_t) = -\left[\lambda \dot{m}_f(a_t) + (1-\lambda)\varsigma(s_t)\right] \tag{11}$$

$$\varsigma(s_t) = \begin{cases} 0, & \Delta SOC_t \geq 0 \\ \Delta SOC_t^2, & \Delta SOC_t < 0 \end{cases} \tag{12}$$

where the SOC difference is defined as $\Delta SOC_t = SOC_t - SOC_{SC}$, where SOC_{SC} is indicative of the SOC target value.

B. Design of the State and Action Space

Multivariate state space: the state space of ITEM agent $\boldsymbol{S}_t$ is expected to consider key parameters related to three parts: (a) thermal management variables $\boldsymbol{S}_t^{TMS}$, (b) energy management variables $\boldsymbol{S}_t^{EMS}$ and (c) real-time traffic information $\boldsymbol{S}_t^{ITS}$.

$$\boldsymbol{S}_t = [\boldsymbol{S}_t^{TMS}, \boldsymbol{S}_t^{EMS}, \boldsymbol{S}_t^{ITS}] \tag{13}$$

Hybrid action space: the action $\boldsymbol{A}_t$ include continuous engine power $P_t^{ICE} \in [0kW, 60kW]$ and the priority of each drive mode with a value belongs to [0,1], namely V_t^{SH}, V_t^{PH} and V_t^{PSH}. The ultimate drive mode M_t^{DHT}, either series hybrid (SH), parallel hybrid (PH) or PSH, is chosen with the highest value, a principle analogous to DQL action selection mechanism, namely $M_t^{DHT} = \underset{M_{DHT}}{\text{argmax}}(V_t^{SH}, V_t^{PH}, V_t^{PSH})$.

$$\boldsymbol{A}_t = [P_t^{ICE}, M_t^{DHT}] \tag{14}$$

The thermal states of the powertrain and cabin are taken into account when optimizing power distribution, thus the state space in this work includes the power demand of thermal management system P_t^{Aux}, engine coolant temperature T_t^e and ambient temperature T_t^a, which can be found in Eq. (15).

$$\boldsymbol{S}_t^{TMS} = [P_t^{Aux}, T_t^e, T_t^a] \tag{15}$$

The energy management related variables, shown in Eq. (16), include the vehicle's velocity v_t, acceleration χ_t, and the battery SOC denoted as SOC_t.

$$\boldsymbol{S}_t^{EMS} = [v_t, \chi_t, SOC_t] \tag{16}$$

Multi-source traffic and terrain data processed by an STDP framework is incorporated as shown in Eq. (17), where d_t^{rem} refers to the remaining mileage of the journey, and TL_t, defined in Eq. (18), represents the traffic light state which is determined based on the signal phasing and timing (SPaT) data $cd_t^{G/R}$ and SPaT type S_{TL}. Here, cd_t^G and cd_t^R are defined as the countdown of green and red traffic light, respectively. Besides, T_G and T_R respectively refer to the total time length of the green or red phase. The definition of the SPaT type is as follows: S_{TL}=1 is defined as when the vehicle does not enter the traffic light waiting area; S_{TL}=0 means when the vehicle appears in the traffic light waiting area while the traffic signal is green; and S_{TL}=-1 means when the vehicle appears in the traffic light waiting area and the traffic signal is red.

$$\boldsymbol{S}_t^{STDP} = [d_t^{rem}, TL_t, \boldsymbol{v}_t, \boldsymbol{\vartheta}_t] \tag{17}$$

$$TL_t = \begin{cases} 1, & if\ S_{TL} = 1 \\ \dfrac{cd_t^G}{T_G}, & if\ S_{TL} = 0 \\ \dfrac{T_R - cd_t^R}{T_R}, & if\ S_{TL} = -1 \end{cases} \tag{18}$$

where $\boldsymbol{v}_t = \boldsymbol{v}(t - T_b : t + T_f)$ and $\boldsymbol{\vartheta}_t = \boldsymbol{\vartheta}(t - T_b : t + T_f)$ represent the vector of velocity and road slope for the past and future horizon, provided by the digital map service from time step $t - T_b$ to $t + T_f$. Here, T_b is symbolic of backward indexed time steps, while T_f signifies forward indexed time steps, i.e., the length of the future horizon. The future horizon's vehicle velocity at the initial time step and final time step T are both set to zero, implying a stationary state. The recorded historical velocity and road slopes are denoted as $\boldsymbol{v}(t - T_b : t - 1)$ and $\boldsymbol{\vartheta}(t - T_b : t - 1)$. Correspondingly, the future horizon's $\boldsymbol{v}(t : t + T_f)$ and $\boldsymbol{\vartheta}(t : t + T_f)$ can be obtained by transforming the map-forecasted future velocity $\boldsymbol{v}(t_i : t_{i+n})$ and GIS-based road slopes $\boldsymbol{\vartheta}(t_i : t_{i+n})$ from the spatial domain to the desired time span. Here, d_i, t_{d_i} and v_{d_i} symbolize the driving distance, time, and velocity at a specific location.

$$\boldsymbol{v}(t - T_b : t + T_f) = [\boldsymbol{v}(t - T_b : t - 1), \boldsymbol{v}(t : t + T_f)] \tag{19}$$

$$\boldsymbol{\vartheta}(t - T_b : t + T_f) = [\boldsymbol{\vartheta}(t - T_b : t - 1), \boldsymbol{\vartheta}(t : t + T_f)] \tag{20}$$

$$\begin{cases} \boldsymbol{v}(t : t + T_f) \leftarrow \boldsymbol{v}(t_{d_i} : t_{d_{i+n}}) \\ \boldsymbol{\vartheta}(t : t + T_f) \leftarrow \boldsymbol{\vartheta}(t_{d_i} : t_{d_{i+n}}) \\ t_{d_{i+n}} = t_{d_i} + \dfrac{2(d_{i+1} - d_i)}{(v_{d_{i+1}} - v_{d_i})} \end{cases} \tag{21}$$

C. Training of Integrated Thermal and Energy Management

In contrast to the notably DQN and DDPG, the state-of-the-art twin delayed deep deterministic policy gradient is favored for its faster convergence speed and stability during the training process. Its improvements can primarily be ascribed to three aspects: adoption of the target policy smoothing technique, which imparts a level of randomness to the policy for the establishment of a non-deterministic policy; delayed policy updates mechanism, where TD3 exhibits slower policy updates than those of the Q-function; and utilization of the clipped DQL mechanism, which allows TD3 to learn dual Q-functions instead of a singular one. Given these features, we leverage TD3 to train the agent. Moreover, the state space of the TD3 agent is expanded to encompass EMS and TMS and traffic/terrain information, as shown in Eq. (13). Following completion of the training, the trained network can be downloaded to powertrain DCU for online implementation.

The structure of the proposed learning algorithm is illustrated in Fig. 6. The Actor network is trained to output engine power P_{ICE} and drive mode M_{DHT} by $a_t = \pi_\theta(s_t) + N(0, \delta^2)$, where the mean-zero Gaussian noise $N(0, \delta^2)$ is added to deterministic policy for the balance of exploration and exploitation. It commences with the initialization of the replay memory D with a capacity of M before the offline training begins. The transition tuples harvested from the interaction process are preserved in the replay buffer for experience replay, while the agent continues this iterative process logging new transitions in the replay memory. If the replay buffer reaches its limit, the oldest data will be replaced with new ones. In each cycle, a mini-batch of M transition tuples are randomly selected from the replay memory for parameter update. The Evaluate Actor network is responsible for generating optimal commands, while a pair of Evaluate Critic networks calculate the Q values based on state and action inputs. The Q values parallelly computed by double Evaluate Critic networks are aimed to mitigate overestimation, with the lower of the two Q values, $Q_{w_1}(s_t, a_t)$ and $Q_{w_2}(s_t, a_t)$. They are utilized for the update of the Evaluate Critic networks by minimizing the TD error-based loss functions as shown in Eq. (22) and Eq. (23).

$$J(w_i) = \frac{1}{M}\sum_{t=1}^{M}\left[y_t - Q_{w_i}(s_t, a_t)\right]^2, \qquad i = 1, 2 \tag{22}$$

$$y_t = r_t + \gamma\, min\left[Q_{w_1'}(s_{t+1}, \tilde{a}_{t+1}), Q_{w_2'}(s_{t+1}, \tilde{a}_{t+1})\right] \tag{23}$$

where $Q_{w_1'}(s_{t+1}, \tilde{a}_{t+1})$ and $Q_{w_2'}(s_{t+1}, \tilde{a}_{t+1})$ refer to the Q-values ascertained by the pair of Target Critic networks. Moreover, $\tilde{a}_t$ symbolizes the action outputted by the Target

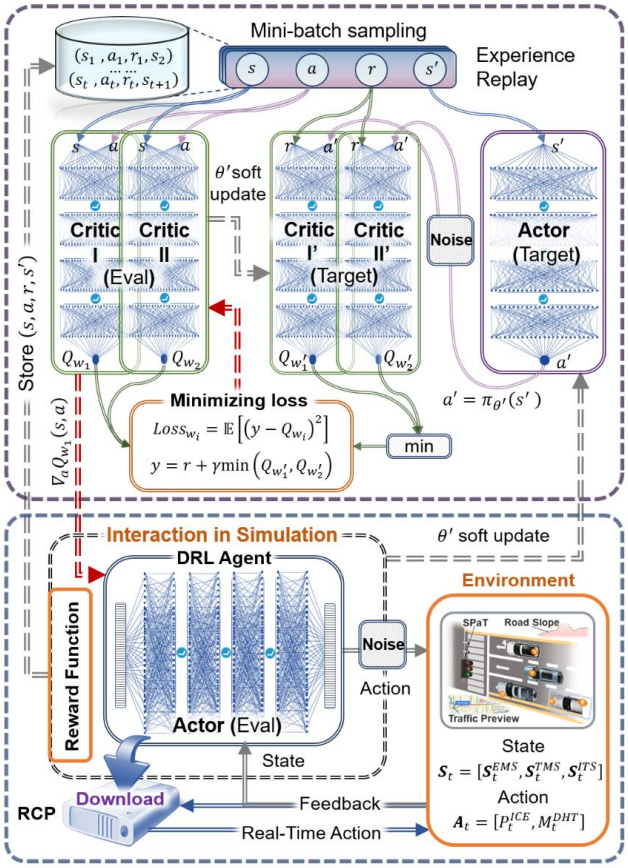


Fig. 6. Learning framework of the RL-based ITEM system.

Actor network, modified by a clipped random noise for improved precision in value estimating, defined as shown in Eq. (24), where $\pm c$ refers to the upper and lower range.

$$\tilde{a}_{t+1} = \pi_{\theta'}(s_{t+1}) + clip[N(0, \delta^2), -c, c] \tag{24}$$

Also, $Q_{w_i}(s_t, a_t)$ refers to the action-value function representing the expected accumulated reward following policy π at time step t which is defined by Eq. (25).

$$Q_{w_i}(s_t, a_t) = \mathbb{E}_\pi\left[\sum_{k=t}^{M} \gamma^{k-t} r_k(s_k, a_k)\right] \tag{25}$$

Each transition tuple $e_t = (s_t, a_t, r_t, s_{t+1})$ feeds s_t into the Evaluate Actor network to produce action $a_t = \pi_\theta(s_t)$, which is employed by the pair of Evaluate Critic networks independently to compute the Q-values, represented as $Q_{w_1}(s_t, \pi_\theta(s_t))$ and $Q_{w_2}(s_t, \pi_\theta(s_t))$, with w_1 and w_2 denoting the parameters of Evaluate Critic network I and II, respectively. Moreover, $Q_{w_1}(s_t, \pi_\theta(s_t))$ is used for updating the weights of the Evaluate Actor network using Eq. (26). It's worth noting that the Target networks are updated every d steps, implying that the pace of weight updates in the Target networks lags behind that of the Evaluate networks.

$$\begin{aligned} \nabla_\theta J(\theta) &= \mathbb{E}\left[\nabla_a Q_{w_1}(s_t, a)|_{a=\pi_\theta(s_t)} \nabla_\theta \pi_\theta(s_t)\right] \\ &\approx \frac{1}{M}\sum_{m=1}^{M} \nabla_a Q_{w_1}(s_j, a)|_{a=\pi_\theta(s_j)} \nabla_\theta \pi_\theta(s_j) \end{aligned} \tag{26}$$

Concurrently, the parameters of the Evaluate networks are optimized synchronously with the training progress, while the Target networks' weights take after the corresponding Evaluate networks with soft updates with a ratio of σ as detailed in Eq. (27). The pseudo code for the RL-based offline training for the ITEM agent is provided in Table II. Upon the completion of the targeted episodes E, the optimal policy is chosen and loaded into the online controller.

$$\begin{cases} w_i' \leftarrow \sigma_w w_i + (1 - \sigma_w) w_i', & i = 1, 2 \\ \theta' \leftarrow \sigma_\theta \theta + (1 - \sigma_\theta)\theta' \end{cases} \tag{27}$$

TABLE II
PROCEDURES OF TRAINING THE DRL-BASED ITEM

Offline Training Algorithm	
1	Initialize replay memory D with capacity M
2	Initialize two critic networks Q_{w_i} and actor network π_θ with random parameters $w_i (i = 1,2)$ and θ
3	Initialize target networks by $w_i' \leftarrow w_i (i = 1,2)$ and $\theta' \leftarrow \theta$
4	**For** episode = 1,2, …, E, **do**
5	Initialize $\boldsymbol{S}_1 = [\boldsymbol{S}_1^{TMS}, \boldsymbol{S}_1^{EMS}, \boldsymbol{S}_1^{STDP}]$
6	**For** step = 1,2, …, T, **do**
7	Select an action with $a_t = \pi_\theta(s_t) + N(0, \delta^2)$
8	Execute a_t to get next state s_{t+1} reward r_t, and store the transition tuple $e_t = (s_t, a_t, r_t, s_{t+1})$
9	Sample mini-batch of N_s transitions $e_n = (s_n, a_n, r_n, s_{n+1})$, $(n = 1,2 \ldots, N_s)$, from D
10	Calculate $\tilde{a}_{t+1} = \pi_{\theta'}(s_{t+1}) + clip[N(0, \delta'^2), -c, c]$, $y_t = r_t + \gamma \min\left[Q_{w_1'}(s_{t+1}, \tilde{a}_{t+1}), Q_{w_2'}(s_{t+1}, \tilde{a}_{t+1})\right]$
11	Update critics $w_i (i = 1,2)$ by minimizing the loss: $J(w_i) = \frac{1}{M}\sum_{t=1}^{M}\left[y_t - Q_{w_i}(s_t, a_t)\right]^2, (i = 1,2)$
12	**If** (t mod d) **then**
13	Update θ by the deterministic policy gradient: $\nabla_\theta J(\theta) = \frac{1}{M}\sum_{m=1}^{M} \nabla_a Q_{w_1}(s_j, a)|_{a=\pi_\theta(s_j)} \nabla_\theta \pi_\theta(s_j)$
14	Update target networks: $w_i' \leftarrow \sigma w_i + (1 - \sigma) w_i', (i = 1,2)$ $\theta' \leftarrow \sigma\theta + (1 - \sigma)\theta'$
15	**End If**
16	**End For**
17	**End for**

IV. TESTING AND VALIDATION SETUP

The ITEM system is trained under the WLTC driving cycle during the charge sustaining (CS) stage, where both the initial and target SOC values are set to 34% according to parameter setting of the real vehicle controller in Dongfeng Motor. Since the homologation driving cycle does not have real-time traffic information, the author's previous study proposed a synthetic construction of traffic data [33] and it is adopted in this paper. In addition, this work mainly focuses on the PHEV performance during charge sustaining stage, thus each episode of the training process is initialized with a fixed SOC starting value of 34%, while the ambient temperature is sampled according to the probability shown in Fig. 7, a fitted data of Beijing average temperature from November to April.

A detailed capture of a 27 km driving during peak hours was carried out in Beijing, China, employing the vehicle diagnostics system OBD-II and dashboard-mounted cameras. Concurrently, real-time traffic information was recorded from

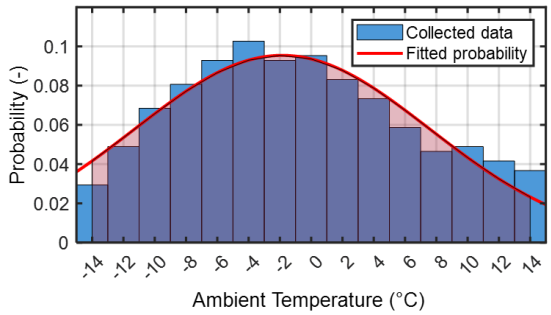


Fig. 7. Winter ambient temperature occurrence probability.

the Gaode Map. To develop the real-world driving model, PreScan is adopted to create the vehicle-in-the-loop simulation environment, as depicted in Fig. 8, integrating terrain details extracted from Google Earth. To enhance the authenticity of the traffic model, data from the subject vehicle and traffic light condition are incorporated. This setup allows to gather precise multivariate traffic data, as discussed in Section III, for the training and validation of ITEM controller.

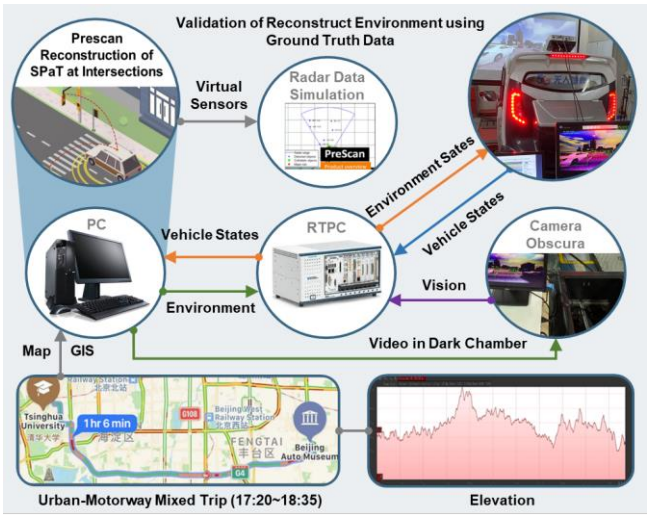


Fig. 8. Platform for real-world driving scenario reconstruction.

The reconstructed traffic data can be seen in Fig. 9. Besides, the comparative study is carried out against the DP, RL-based EMS and the rule-based benchmark control strategies. And the state space of benchmark RL-EMS $\boldsymbol{S}_t^{RL-EMS}$ also adopts multivariate except for TMS related states, as shown below

$$\boldsymbol{S}_t^{RL-EMS} = [\boldsymbol{S}_t^{EMS}, \boldsymbol{S}_t^{STDP}] \tag{28}$$

Its action space is consistent with that of the ITEM agent. Furthermore, the SOC compensation procedure is performed to obtain the corrected fuel consumption before conducting the comparison of control results with the identical terminal SOC, which aligns with the SAE J1711 standard [38].

V. Results and Discussions

A. Parameter design and Learning Ability

The specific hyperparameters utilized during the training process of the ITEM agent are listed in Table III, and a total of 500 training episodes have been designated. Additionally, the

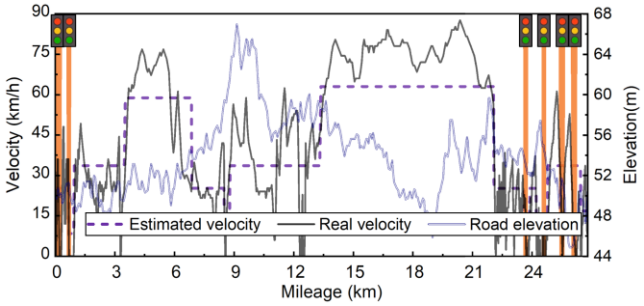


Fig. 9. Trip profile of the reconstructed real driving cycles.

parameters T_b and T_f corresponding to the velocity vector and the slope vector of the state inputs, have been set to 10 seconds and 50 seconds, respectively. The neural network architecture consists of four fully connected layers, with each layer comprising 200, 150, 100, and 50 neurons respectively. These values have been determined based on comprehensive experiments and the latest literatures in this field. Fig. 10 displays the convergence curves of the offline training of the benchmark EMS and the proposed ITEM system. The RL-based EMS shows a faster convergence rate as it does not consider thermal variables in its state space. Despite the ITEM is slower in convergence speed, it achieves a higher return.

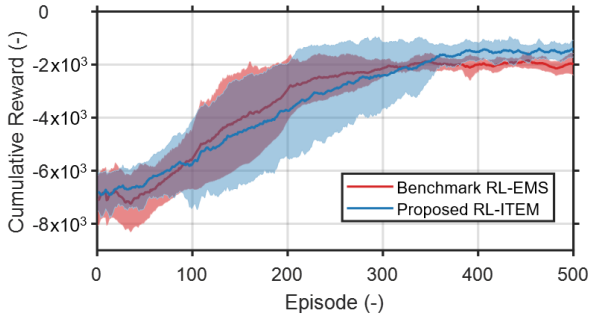


Fig. 10. Learning curve of episode return in ITEM and EMS.

TABLE III
HYPERPARAMETERS OF TD3 ALGORITHM

Hyperparameter	Value	Hyperparameter	Value
Discount factor γ	0.995	Actor learning rate	1e-4
Buffer capacity M	1e6	Critic learning rate	1e-4
Batch size N_s	256	Update rate σ_w, σ_θ	0.005
Delayed update d	4	Noise clip range c	0.5

B. Optimality Validation in Homologation Driving Cycle

To understand the training process and performance after convergence, Fig. 11 shows the results from the 450th to the 500th episode. The fuel consumption after correction has been illustrated, since each episode has a different terminal SOC. Further, the ITEM demonstrates better fuel economy than conventional RL-EMS as the temperature decreases. This is because that the increased necessity of substituting electric heating with coolant heating at lower temperatures. The training results of both algorithms were separately fitted, and fuel consumption is recorded with initial coolant temperature ranging from -10°C to 10°C, as shown in Table IV.

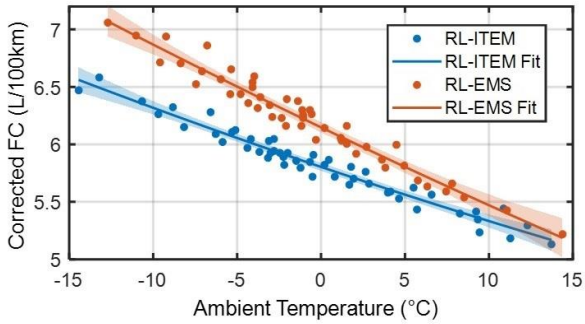


Fig. 11. Training results in terms of fuel economy for ITME and EMS during the 450th to the 500th episodes.

TABLE IV
COMPARISON OF FITTED AVERAGE FUEL CONSUMPTION (FC)

Ambient Temp (°C)	EMS FC (L/100km)	ITEM FC (L/100km)	Difference (%)
10	5.47	5.32	2.82
5	5.78	5.55	4.14
0	6.14	5.79	6.04
-5	6.50	6.05	7.43
-10	6.88	6.33	8.69

To analyze the performance under homologation cycles, the ITEM is compared against the RL-EMS and BMK strategies. The results of SOC, cumulative fuel consumption, and coolant temperature are illustrated in the plots of Fig. 12. In this test, the ambient temperature is set as 10°C and the initial coolant temperature is set as 23°C. Fig. 12 (a) presents the comparison of SOC dynamics, revealing that the ITEM strategy achieves a shallower depth of discharge compared to RL-EMS and BMK. This is attributed to the fact that ITEM learns to expedite the engine warm-up process. In terms of the discharge process and the shape of the SOC trajectory, ITEM exhibits results closest to those of DP, outperforming the other two control systems and leading to improved fuel economy. In comparison to the RL-EMS and BMK control strategies, ITEM achieves fuel savings of 3.4% and 4% respectively, while maintaining the highest average water temperature. The comparison of different control systems can be found in Table V.

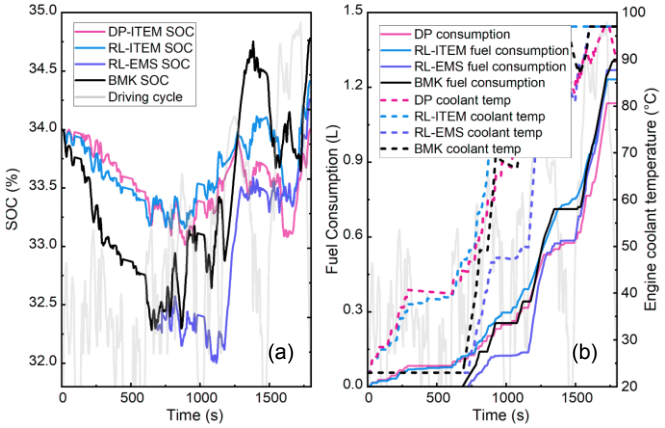


Fig. 12. Algorithm comparisons in 10°C with SOC profiles (a) and cumulative fuel usage and coolant temperature (b).

TABLE V
OPTIMALITY VALIDATION UNDER WLTC IN 10°C

Algorithm	DP	ITEM	EMS	BMK
Final SOC (%)	34.0	34.4	34.3	34.8
Avg coolant temp (°C)	62.9	65.8	51.7	56.3
FC (L/100km)	4.93	5.32	5.49	5.62
Corrected FC(L/100km)	4.93	5.26	5.44	5.47
Fuel savings (%)	10.9	4.0	0.6	-

Fig. 13 illustrates the influence of state variables on the output of the Actor network. It explains the degree to which each state variable affects the output. The feature importance of a state variable is determined by its frequency of use at the decision tree split points. The data generated during the simulation training process are collected and fed into the decision tree, and the results are obtained using the Gradient Boosting Regression Tree (GBRT) software package [33]. The results indicate that SOC has the highest feature importance, followed by coolant temperature and velocity. Besides, traffic and terrain data also play a significant role in policy formation. The results demonstrate that incorporating TMS-related states in this work leads to distinct decision-making processes for the agent. This expansion helps the agent acquire necessary observations to improve control performance.

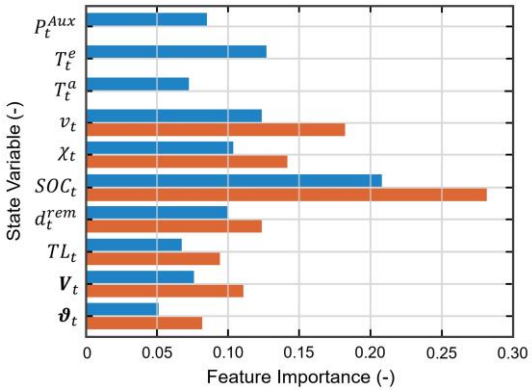


Fig. 13. The feature importance of the state variables of the RL-based IETM and EMS.

C. Online HIL Experiment in Real-World Driving Condition

A hardware-in-the-loop (HIL) testing system is utilized to validate the real-time control performance of the IETM system under the CS stage. In the HIL test, the ambient temperature and the initial coolant temperature are both set as -10°C, and a higher initial SOC state of 45% is selected. The testing platform, depicted in Fig. 14, comprises an upper computer that manages and configures input and output (I/O) interfaces, communication interfaces, and test cases using NI VeriStand software. The I/O signals from the rapid control prototype (RCP) are connected to the HIL terminals, enabling real-time operation of the automotive powertrain within the real-time computer (RTPC).

The detailed control results are compared, as depicted in Figure 15. The engine output power curve serves as a critical

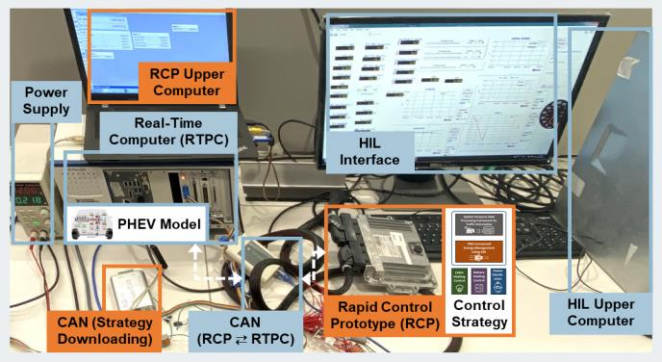


Fig. 14. Facilities for the hardware-in-the-loop test.

indicator for evaluating the performance of the control systems. It is evident that both the ITEM and BMK algorithms initiate engine preheating prior to 1000 s. However, the ITEM algorithm promptly commences engine operation once the engine reaches a temperature of 40°C, thereby raising the coolant temperature. Although BMK achieves a higher coolant temperature than RL-EMS before 2000 s through rapid preheating, it struggles to sustain the coolant temperature above 75°C for an extended duration. This limitation hampers its capability to substitute cabin electric heating.

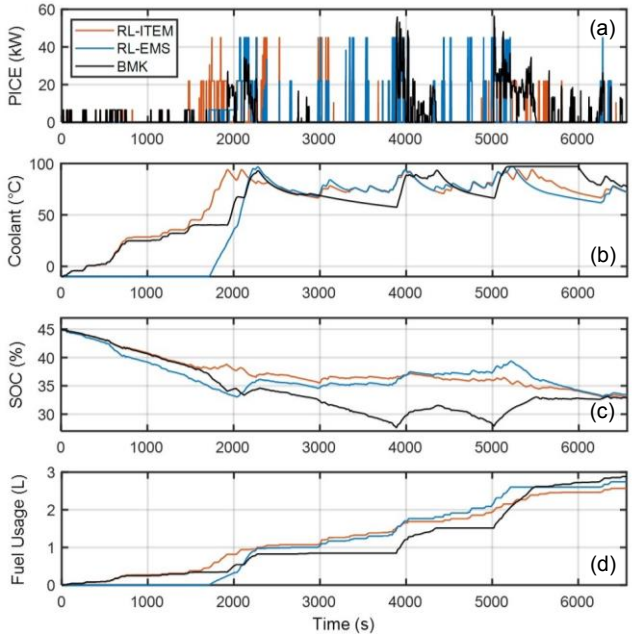


Fig. 15. Real-time performance: (a) engine output power; (b) coolant temperature; (c) battery SOC; (d) fuel usage.

As a result, the fuel economy of RL-EMS exceeds that of BMK in the later stages, while ITEM demonstrates the best performance. To provide a more comprehensive depiction of the powertrain control results, a zoomed-in view of the output power of engine, coolant temperature, power of the drive motor, and battery SOC is presented during the driving cycle from 1500 s to 3500 s, as illustrated in Fig. 16. The ITEM efficiently accomplishes engine preheating, initiating engine operation and promptly elevating the coolant temperature to a level required for heating, while maintaining it consistently. Conversely, BMK refrains from engine operation mode after completing the preheating process due to the logic of rule-based energy management strategy, since the battery SOC has not reached its threshold for engine operation. The RL-EMS, lacking thermal management-related information in terms of heating load and coolant temperature in its state space, commences engine preheating around 1600 s. However, even after completing preheating, RL-EMS fails to raise the coolant temperature adequately. Hence, despite RL-EMS showing commendable energy management performance, its overall energy-saving effectiveness is compromised.

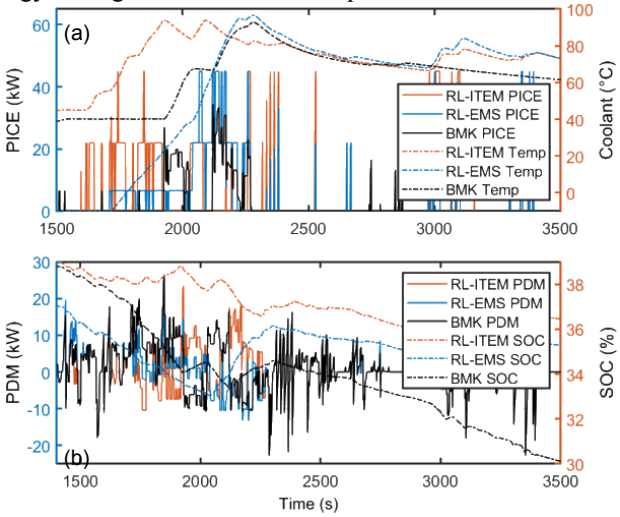


Fig. 16. Magnified illustration of (a) engine power and coolant temperature; (b) drive motor power and battery SOC.

The HIL test results of RL-ITEM, RL-EMS and BMK can be found in Fig. 17 to illustrate the corrected fuel economy, energy consumption caused by electric heating, as well as the duration of engine warm state allowed for cabin heating. The fuel economy of ITEM outperforms the state-of-the-art RL-based EMS and rule-based benchmark system by 7.1% and 12.4%, which is evidenced by the less energy usage on electric heating and longer duration in the required coolant temperature for cabin heating, and the detailed results are summarized in Table VI. Therefore, the proposed method can be modified and applied to the calibration of intelligent ITEM, to tackle the complex real-world driving conditions and explore the concealed fuel-saving potentials in climate-adaptive energy management systems.

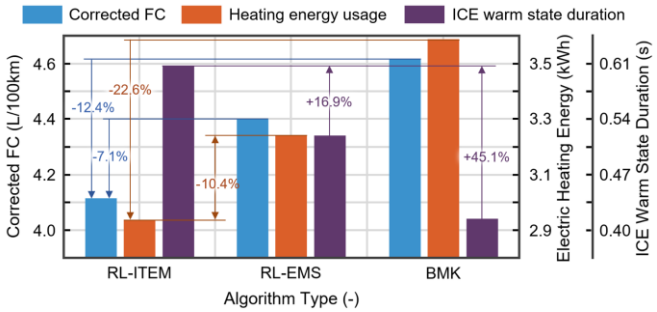


Fig. 17. HIL test results in terms of fuel economy, energy consumption of electric heating, as well as the duration of engine warm state allowed for cabin heating.

TABLE VI
HIL VALIDATION FOR GENERALIZATION TEST IN -10°C

Algorithm	ITEM	EMS	BMK
Final SOC (%)	33.1	33.5	32.9
Avg coolant temp (°C)	63.1	51.0	61.4
Electric heating energy (kWh)	2.94	3.24	3.60
Corrected FC (L/100km)	4.11	4.40	4.62
Fuel savings (%)	12.4	4.8	-

VI. CONCLUSIONS

This paper proposes the trip-oriented thermal and energy management for a multi-mode connected PHEV. The performance of the proposed ITEM has been validated by comprehensive experiments under homologation and real-world drive cycles with different SOC initial conditions, and the findings of the study are summarized as follows:

1) A multi-mode PHEV model with high-fidelity thermal and energy consumption characteristics is calibrated using experimental data. Based on this, a model-free multistate RL algorithm with a hybrid action space is designed for the trip-oriented ITEM system, featuring with engine smart warm-up and engine-assisted heating.
2) The proposed ITEM controller incorporates techniques including a STDP framework, bounded double Q-values, and delayed policy updates. Offline tests confirm the ITEM achieving fuel economy enhancement from 2.2% to 9.6% when ambient temperature drops from 15℃ to -15℃, compared with the state-of-the-art RL-based EMS, and realizes on average 93.7% of the performance of DP.
3) The adaptability of the ITEM system is facilitated by integrating coolant temperature, ambient temperature as well as multisource traffic and terrain data processed by STDP framework into the state space. The HIL experiments conducted in -10℃ demonstrate the real-time implementation in real-world driving scenarios, which reduces fuel cost by 7.1% and 12.4% compared to RL- EMS and rule-based control strategies, respectively.

REFERENCES

[1] L. Xinglong, F. Zhao, and Z. Liu, "Energy-saving cost-effectiveness analysis of improving engine thermal efficiency and extending all-electric range methods for plug-in hybrid electric vehicles," *Energy Conversion and Management,* p. 115898, 01/01 2022.

[2] W. Zhou, N. Zhang, and H. Zhai, "Enhanced Battery Power Constraint Handling in MPC-Based HEV Energy Management: A Two-Phase Dual-Model Approach," *IEEE Transactions on Transportation Electrification,* vol. 7, no. 3, pp. 1236-1248, 2021.

[3] Z. Zhao, P. Tang, and H. Li, "Generation, Screening, and Optimization of Powertrain Configurations for Power-Split Hybrid Electric Vehicle: A Comprehensive Overview," *IEEE Transactions on Transportation Electrification,* vol. 8, no. 1, pp. 325-344, 2022.

[4] X. Tang, J. Zhang, X. Cui, X. Lin, L. Grzesiak, and X. Hu, "Multi-Objective Design Optimization of a Novel Dual-Mode Power-Split Hybrid Powertrain," *IEEE Transactions on Vehicular Technology,* vol. PP, pp. 1-1, 11/25 2021.

[5] H. Zhang, S. Liu, N. Lei, Q. Fan, and Z. Wang, "Leveraging the benefits of ethanol-fueled advanced combustion and supervisory control optimization in hybrid biofuel-electric vehicles," *Applied Energy,* vol. 326, p. 120033, 2022/11/15/ 2022.

[6] J. Wu, Y. Zou, X. Zhang, G. Du, G. Du, and R. Zou, "A Hierarchical Energy Management for Hybrid Electric Tracked Vehicle Considering Velocity Planning With Pseudospectral Method," *IEEE Transactions on Transportation Electrification,* vol. 6, no. 2, pp. 703-716, 2020.

[7] L. Kai, Q. Liu, Y. Hu, J. Gao, and H. Chen, "A Coupling Thermal Management Strategy Based on Fuzzy Control for a Range Extended Electric Vehicle Power System," *IEEE Transactions on Vehicular Technology,* vol. 71, no. 3, pp. 1-1, 12/24 2021.

[8] M. R. Amini, H. Wang, X. Gong, D. Liao-McPherson, I. Kolmanovsky, and J. Sun, "Cabin and Battery Thermal Management of Connected and Automated HEVs for Improved Energy Efficiency Using Hierarchical Model Predictive Control," *IEEE Transactions on Control Systems Technology,* vol. 28, no. 5, pp. 1-16, 07/03 2019.

[9] S. E. Vore, M. Kosowski, M. L. Reid, Z. Wilkins, J. Minicucci, and T. H. Bradley, "Measurement of Medium-Duty Plug-In Hybrid Electric Vehicle Fuel Economy Sensitivity to Ambient Temperature," *IEEE Transactions on Transportation Electrification,* vol. 4, no. 1, pp. 184-189, 2018.

[10] C. Yang, M. Zha, W. Wang, L. Yang, S. You, and C. Xiang, "Motor-Temperature-Aware Predictive Energy Management Strategy for Plug-In Hybrid Electric Vehicles Using Rolling Game Optimization," *IEEE Transactions on Transportation Electrification,* vol. 7, no. 4, pp. 2209-2223, 2021.

[11] B. Chen, X. Li, S. Evangelou, and R. Lot, "Joint Propulsion and Cooling Energy Management of Hybrid Electric Vehicles by Optimal Control," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 5, pp. 1-1, 02/28 2020.

[12] H. T. Pham, P. P. J. v. d. Bosch, J. T. B. A. Kessels, and R. G. M. Huisman, "Integrated energy and thermal management for hybrid electric heavy duty trucks," in *2012 IEEE Vehicle Power and Propulsion Conference*, 2012, pp. 932-937.

[13] M. Shams-Zahraei, A. Z. Kouzani, S. Kutter, and B. Bäker, "Integrated thermal and energy management of plug-in hybrid electric vehicles," *Journal of Power Sources,* vol. 216, pp. 237-248, 2012/10/15/ 2012.

[14] L. Lu, H. Chen, Y.-F. Hu, X. Gong, and J. Zhao, "Modeling and optimization control for an electric cooling system to minimize fuel consumption," *IEEE Access,* vol. 7, pp. 1-1, 05/16 2019.

[15] Y. Zhang *et al.*, "Machine Learning-Based Vehicle Model Construction and Validation—Toward Optimal Control Strategy Development for Plug-In Hybrid Electric Vehicles," *IEEE Transactions on Transportation Electrification,* vol. 8, no. 2, pp. 1590-1603, 2022.

[16] A. A. Mamun, Z. Liu, D. M. Rizzo, and S. Onori, "An Integrated Design and Control Optimization Framework for Hybrid Military Vehicle Using Lithium-Ion Battery and Supercapacitor as Energy Storage Devices," *IEEE Transactions on Transportation Electrification,* vol. 5, no. 1, pp. 239-251, 2019.

[17] J. Li, Q. Zhou, Y. He, H. Williams, H. Xu, and G. Lu, "Distributed Cooperative Energy Management System of Connected Hybrid Electric Vehicles With Personalized Non-Stationary Inference," *IEEE Transactions on Transportation Electrification,* vol. 8, no. 2, pp. 2996-3007, 2022.

[18] N. Zhao, F. Zhang, Y. Yang, S. Coskun, X. Lin, and X. Hu, "Dynamic Traffic Prediction-Based Energy Management of Connected Plug-in Hybrid Electric Vehicles with Long Short-Term-State of Charge Planning," *IEEE Transactions on Vehicular Technology,* vol. 72, no. 5, pp. 1-15, 01/01 2023.

[19] H. Zhang, Q. Fan, W. Wang, J. Huang, and Z. Wang, "Reinforcement Learning based Energy Management Strategy for Hybrid Electric Vehicles Using Multi-Mode Combustion," *Qiche Gongcheng/ Automotive Engineering,* vol. 43, pp. 683-691, 05/25 2021.

[20] H. Wang, I. Kolmanovsky, M. R. Amini, and J. Sun, "Model Predictive Climate Control of Connected and Automated Vehicles for Improved Energy Efficiency," in *2018 Annual American Control Conference (ACC)*, 2018, pp. 828-833.

[21] H. Qiuhao, M. R. Amini, I. Kolmanovsky, J. Sun, A. Wiese, and J. Seeds, "Multihorizon Model Predictive Control: An Application to Integrated Power and Thermal Management of Connected Hybrid Electric Vehicles," *IEEE Transactions on Control Systems Technology,* vol. 30, no. 3. PP, 07/08 2021.

[22] Y. He *et al.*, "Multiobjective Co-Optimization of Cooperative Adaptive Cruise Control and Energy Management Strategy for PHEVs," *IEEE Transactions on Transportation Electrification,* vol. 6, no. 1, pp. 346-355, 2020.

[23] G. Du, Y. Zou, X. Zhang, L. Guo, and N. Guo, "Heuristic Energy Management Strategy of Hybrid Electric Vehicle Based on Deep Reinforcement Learning With Accelerated Gradient Optimization," *IEEE Transactions on Transportation Electrification,* vol. 7, no. 4, pp. 2194-2208, 2021.

[24] H. Zhang, Q. Fan, S. Liu, S. E. Li, J. Huang, and Z. Wang, "Hierarchical energy management strategy for plug-in hybrid electric powertrain integrated with dual-mode combustion engine," *Applied Energy,* vol. 304, p. 117869, 2021/12/15/ 2021.

[25] X. Tang, J. Chen, K. Yang, M. Toyoda, T. Liu, and X. Hu, "Visual Detection and Deep Reinforcement Learning-Based Car Following and Energy Management for Hybrid Electric Vehicles," *IEEE Transactions on Transportation Electrification,* vol. 8, no. 2, pp. 2501-2515, 2022.

[26] X. Tang, J. Chen, H. Pu, T. Liu, and A. Khajepour, "Double Deep Reinforcement Learning-Based Energy Management for a Parallel Hybrid Electric Vehicle With Engine Start–Stop Strategy," *IEEE Transactions on Transportation Electrification,* vol. 8, no. 1, pp. 1376-1388, 2022.

[27] F. Zhang, X. Hu, R. Langari, and D. Cao, "Energy management strategies of connected HEVs and PHEVs: Recent progress and outlook," *Progress in Energy and Combustion Science,* vol. 73, 07/01 2019.

[28] H. Zhang, J. Peng, H. Tan, H. Dong, and F. Ding, "A Deep Reinforcement Learning-Based Energy Management Framework With Lagrangian Relaxation for Plug-In Hybrid Electric Vehicle," *IEEE Transactions on Transportation Electrification,* vol. 7, no. 3, pp. 1146-1160, 2021.

[29] A. Biswas, P. G. Anselma, and A. Emadi, "Real-Time Optimal Energy Management of Multimode Hybrid Electric Powertrain With Online Trainable Asynchronous Advantage Actor–Critic Algorithm," *IEEE Transactions on Transportation Electrification,* vol. 8, no. 2, pp. 2676-2694, 2022.

[30] J. Guo, J. Wang, Q. Xu, B. Wang, and K. Li, "Deep Reinforcement Learning-based Hierarchical Energy Control Strategy of a Platoon of Connected Hybrid Electric Vehicles through Cloud Platform," *IEEE Transactions on Transportation Electrification,* Early Access, 2023.

[31] X. Wang, Y. Yuan, L. Tong, C. Yuan, B. Shen, and T. Long, "Energy Management Strategy for Diesel–Electric Hybrid Ship Considering Sailing Route Division Based on DDPG," *IEEE Transactions on Transportation Electrification,* pp. 1-1, 2023.

[32] H. He, Y. Wang, J. Li, J. Dou, R. Lian, and Y. Li, "An Improved Energy Management Strategy for Hybrid Electric Vehicles Integrating Multistates of Vehicle-Traffic Information," *IEEE Transactions on Transportation Electrification,* vol. 7, no. 3, pp. 1161-1172, 2021.

[33] H. Zhang, S. Liu, N. Lei, Q. Fan, S. E. Li, and Z. Wang, "Learning-based supervisory control of dual mode engine-based hybrid electric vehicle with reliance on multivariate trip information," *Energy Conversion and Management,* vol. 257, p. 115450, 2022/04/01/ 2022.

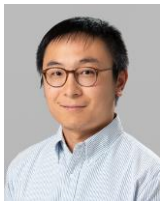
[34] B. Hu and J. Li, "An Adaptive Hierarchical Energy Management Strategy for Hybrid Electric Vehicles Combining Heuristic Domain Knowledge and Data-Driven Deep Reinforcement Learning," *IEEE Transactions on Transportation Electrification,* vol. 8, no. 3, pp. 3275-3288, 2022.

[35] B. Xu *et al.*, "Learning Time Reduction Using Warm-Start Methods for a Reinforcement Learning-Based Supervisory Control in Hybrid Electric Vehicle Applications," *IEEE Transactions on Transportation Electrification,* vol. 7, no. 2, pp. 626-635, 2021.

[36] X. Tian, Y. Cai, X. Sun, Z. Zhu, Y. Wang, and Y. Xu, "Incorporating Driving Style Recognition Into MPC for Energy Management of Plug-In Hybrid Electric Buses," *IEEE Transactions on Transportation Electrification,* vol. 9, no. 1, pp. 169-181, 2023.

[37] C. Hou, M. Ouyang, L. Xu, and H. Wang, "Approximate Pontryagin's minimum principle applied to the energy management of plug-in hybrid electric vehicles," *Applied Energy,* vol. 115, pp. 174–189, 02/01 2014.

Hao Zhang is currently a Ph.D. candidate at School of Vehicle and Mobility, Tsinghua University, China, and a Visiting Researcher at Department of Electronic and Electrical Engineering, University College London, U.K. He was honored with China National Scholarships for three times. His research focuses on the artificial intelligence methods and their applications in the design and control of electrified vehicles.



Boli Chen received the MSc and the Ph.D. in Control Systems from Imperial College London, UK, in 2011 and 2015 respectively. Currently, he is a Lecturer in the Department of Electronic and Electrical Engineering, University College London, U.K. He is also an associate editor of the European Journal of Control and an associate editor of the EUCA Conference Editorial Board. His current research focuses on control, optimization and estimation of complex dynamical systems, mainly from automotive and power electronics areas.

Nuo Lei is currently a Ph.D. candidate in power engineering and engineering thermos-physics with the School of Vehicle and Mobility, Tsinghua University. His current research interests include the component sizing, energy management strategy, and eco-driving of hybrid electric vehicles.

Bingbing Li is a Ph.D. candidate at School of Mechanical Engineering, Southeast University, China, and a visiting researcher at the Department of Electronic and Electrical Engineering, University College London, U.K. His research focuses on energy-efficient driving control of CAVs.

Rulong Li received the M.S. degree from Tsinghua University, Beijing, China, in 2006. He works with the Dongfeng Motor Corporation Ltd., Wuhan, China, directing the company's powertrain development of passenger cars. He has led the development of several hybrid electric vehicle products, and possessed rich experience in system integration.

Zhi Wang received the Ph.D. degree from Tsinghua University, China, in 2005. He is a Professor and deputy dean of the School of Vehicle and Mobility, Tsinghua University, Beijing, China. He authored over 230 papers and 70 patents. He has led over 20 projects for national initiatives and OEMs, and received the China Automotive Industry Awards of Science and Technology. His research focuses on the design and control of low-carbon and carbon-free propulsion systems.