Bi-Level Lifelong Learning-based Transferable Energy Management of Hybrid Electric Vehicles

Hao Zhang, Nuo Lei, Wang Peng, Bingbing Li, Boli Chen, Zhi Wang

***Abstract*—Automotive energy management systems (EMS) are advancing towards comprehensive lifecycle intelligence, spanning from product development level to customer usage level. This paper proposes a bi-level transfer approach with model-agnostic meta-learning (MAML) to realize cross-platform transferable and online-adaptive EMS. During the development level, MAML is adopted to calibrate the heuristic control maps of an instantaneous optimization-based EMS, with a per-unit state and action space design facilitating knowledge transfer. Leveraging usage data, online adaptation is carried out with reliance on digital twin-based cloud computing, updating onboard controller parameters via over-the-air technology. The effectiveness of the proposed lifecycle intelligent EMS is validated through real vehicle experiments. Firstly, the entire MAML-assisted V-cycle development process is demonstrated to validate the optimality and knowledge transfer of the EMS, resulting in zero-shot transfer for EMS calibration on new vehicle products. Additionally, a correction of 8.0%~9.5% fuel economy is improved against the convention reinforcement learning-based EMS during usage via online-adaptation, effectively bridging the optimality gap between the control policy learned in development and the global optimal control in actual driving scenarios.**

***Index Terms*—Energy management, lifecycle intelligence, bi-level transfer, reinforcement learning, real vehicle experiments.**

I. INTRODUCTION

THE escalating demand for eco-friendly logistics has spurred the widespread adoption of range extended electric trucks (REETs) [1], which effectively enhance the energy efficiency and effectively alleviate range anxiety [2]. As a result, optimizing energy management system (EMS) in these vehicles has emerged as a vital research area [3]. Nowadays, the artificial intelligence (AI)-assisted EMS development of connected REETs holds tremendous promise and value, aiming to reduce calibration efforts of engineers [4, 5], shorten vehicle development cycles and enable consecutive optimization under real application scenarios [6].

Currently, EMS is generally categorized into rule-based, optimization-based and learning-based approaches [7]. The former two types find extensive applications in REETs and are relatively easy to implement [8], but the optimality cannot be guaranteed when encountering incompatible parameter settings [9]. In recent years, a compelling approach to tackle control optimality and development efficiency involves the integration of deep reinforcement learning (DRL) and transfer reinforcement learning (T-RL) [10]. Several researchers have employed double deep Q-network (DDQN) [11] to establish EMS with discretized action space, specifically engine output power. Additionally, the actor-critic (AC) framework allows for continuous action-based EMS [12]. For instance, the adoption of deep deterministic policy gradient (DDPG) has proven successful in creating competent EMS with continuous output of engine power [13]. Furthermore, improved DRL algorithms capable of handling high-dimensional state variables [14], such as twin-delayed DDPG (TD3), can be combined with multivariate trip information for traffic-aware EMS, enhancing the optimality of EMSs in complex driving cycles [15]. Nonetheless, the learning-based EMS faces two primary challenges spanning from research and development (R&D) stage to usage phase, respectively [16].

A significant challenge during the R&D phase is to ensure the EMS offline transferability across different vehicle platforms without necessitating re-training [17], denoted as level-1 (L1) transfer in this work. This involves the utilization of high-fidelity simulation environment and the effective transferable EMS model [18]. The digital twin (DT) model has been recognized as the foundation for RL reliable transfer from simulations to real-world applications [19]. Various research endeavors have employed DTs in the design and maintenance of automotive powertrains [20]. Lei et al. proposed a combined powertrain sizing and EMS calibration scheme for hybrid powertrain using a high-fidelity digital twin model, improving the fuel economy by 7.4% [21]. Spiryagin et al. introduced a digital-twin solution in the V-cycle development approach for electric vehicles as a virtual test facility [22]. Zhang et al. utilized a DT model in the development of EMS for a hybrid powertrain. The findings indicate that the DT-based EMS optimization led to improvement of fuel economy with the help of model accuracy [23]. With the help of DTs, Lian et al. applied TL

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 BYD Auto Industry Company Ltd. under Grants 20232000360. (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).

Wang Peng is with the BYD Auto Industry Company Ltd., Shenzhen 518118, China (e-mail: pengwang1916@sina.com).

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).

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

techniques to transfer neural networks (NNs) structure and parameters of a well-trained EMS agents in source domain to other three vehicles with different topologies in target domain [24]. Xu et al. proved that adding noise in the parameter space can increase robustness and achieve faster convergence rate of transfer process [25]. However, the state-of-the-art transfer DRL still needs to go through a time-consuming fine-tuning process, while the notable benefit is that it can avoid the more outrageous exploration process [26].

The other challenge is how to enhance the real-world adaptability of EMS for online transfer from R&D stage to usage phase, indicated as level-2 (L2) transfer [27]. This is because that the EMS trained based on OEM's existing databases of driving data may not necessarily be suitable for actual users [28]. There could be discrepancies between estimated velocity from digital map traffic forecasting and the actual driving velocity due to factors like driver habits [29]. Moreover, the global optimality of EMS is closely related to the commonly used freight routes by truck users, significantly impacting the control strategy on battery state of charge (SOC) depletion [30]. Therefore, by leveraging the wealth of user-generated data, the DT can be used to continually adapt the control system to real-world application scenarios [31]. Yue et al. employed TL to design a DT model for fuel cell degradation prediction. This approach resulted in enhanced accuracy when compared to the prediction model without TL [32]. He et al. investigated the method to transfer prior valid knowledge trained with a hybrid bus in source domain to Prius based on deep transfer learning, accelerating the new EMS convergence [33]. The resulting over-the-air (OTA) updates ensure that the onboard controllers of connected logistics vehicles remain perpetually equipped with the latest and most effective EMS [34]. Nonetheless, these DRL and TL strategies are still a long way from being practically adopted by original equipment manufacturers (OEMs), since directly employing NNs to control engine output power without any interpretability and transferability is not acceptable.

Hence, it is crucial to investigate the AI-based transferrable EMS spanning from R&D to customer usage. To the authors' best knowledge, there is a scarcity of research on the lifecycle intelligent EMS, and no industrial applications of sim-to-real transfer have been found in any publications about RL-based EMS. In response to these gaps, this study proposes a bi-level RL transfer framework to realize transferable and online self-evolving EMS for logistics REETs, as depicted in Fig. 1. The entire RL-assisted V-cycle development process in this work showcases the optimality, transferability and robustness of the lifecycle intelligent EMS. The proposed lifecycle intelligent EMS encompasses several key elements: (1) this work pioneers real vehicle experiments of transferrable RL-based EMS. To satisfy the explainable necessity in OEMs, the equivalent consumption minimization strategy (ECMS) [35] serves as the fundamental EMS framework to enable interpretability, transferability and robustness; and (2) the conventional heuristic control maps for equivalent factor (EF) estimation are replaced by a neural network calibrated by multistate DRL. By utilizing per-unit state and action spaces, the knowledge from the trained DRL agent can be shared directly among different vehicles types from passenger cars to logistics trucks. (3) the DRL agent is designed to be traffic aware where the real-time traffic information from Gaode map has been fed into the state space. By leveraging usage data feedback and DT-based cloud computing, the strategy adapts to discrepancies between map-forecasted speed and actual vehicle speed, effectively addressing deviations in the optimal policy for specific logistics scenarios.

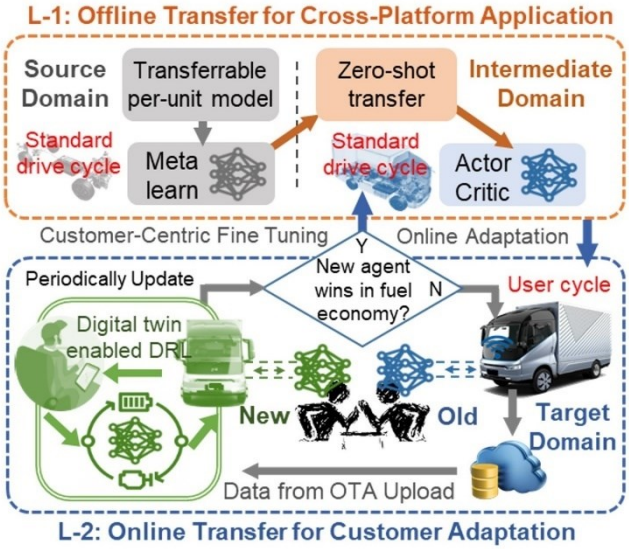


Fig. 1. Bi-level lifelong learning-based transferable EMS.

The rest of this article is structured as follows. Section II describes the powertrain models. Section III presents the design of the bi-level transfer framework for lifecycle intelligent EMS. In Section IV, the entire V-cycle development process of the REET is explained. Section V analyses the transferability, optimality and robustness of the proposed EMS. Section VI summarizes the key findings.

II. Powertrain Model in Source and Target Domain

A. Powertrain Configurations and Energy Flow Modes

In the source domain, the vehicles are passenger cars that utilize series-parallel hybrid systems. Conversely, the target domain features series hybrid trucks. An illustration of the powertrain architectures and engine efficiency characteristics for both vehicle types can be found in Fig. 2. Additionally, the energy flow dynamics for these powertrain systems are presented in Eq. (1).

$$\begin{cases} if\ D_m = EV, \begin{cases} P_{DM} = P_{dem}/\eta_T \\ P_{bat} = P_{DM}/\eta_{DM} + P_{aux} \\ P_{GEN} = P_{ICE} = 0 \end{cases} \\ if\ D_m = PH, \begin{cases} P_{DM} = P_{dem}/\eta_T - P_{ICE} \\ P_{bat} = P_{DM}/\eta_{DM} + P_{aux} \\ P_{GEN} = 0, P_{ICE} = u \end{cases} \\ if\ D_m = SH, \begin{cases} P_{DM} = P_{dem}/\eta_T \\ P_{bat} = P_{DM}/\eta_{DM} - P_{GEN} + P_{aux} \\ P_{GEN} = \eta_{GEN} P_{ICE} = \eta_{GEN} u \end{cases} \end{cases} \quad (1)$$

where P_{dem} refers to the tractive power demand. Besides, P_{DM}, P_{GEN} and P_{ICE} represents the output power of drive motor, generator and combustion engine. Also, D_m means the operation mode, including pure electric (EV), parallel (PH) and series hybrid (SH). In addition, η_T, η_{DM} and η_{GEN} refer to the efficiency of transmission, drive motor and generator. The P_{ICE} is defined as the control variable denoted as u.

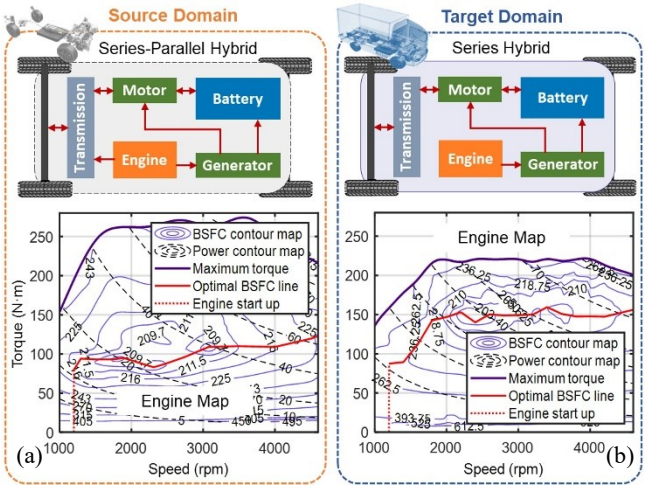


Fig. 2. Powertrain and engine map profiles of the vehicles in source and target domains.

The electric machines (EM) model is illustrated in Eq. (2) showing the relationship between electrical power and mechanical torque and speed, while Eq. (3) describes the engine fueling rate $\dot{m}_f$ with start/stop penalties based on steady-state maps detailed in Fig. 2.

$$P_{EM} = T_{EM}\omega_{EM}\eta_{EM}(T_{EM}, \omega_{EM})^{\kappa} \quad (2)$$

$$\dot{m}_f = P_{ICE}BSFC(T_{ICE}, \omega_{ICE}) + m_{st} \cdot \Lambda \quad (3)$$

where κ denotes the EM's operational mode, either as a motor or a generator. A value of -1 is allocated to κ when the EM is in motor mode. When functioning as a generator, κ adopts a value of 1. Additionally, the static fuel rate is obtained with the product of engine power P_{ICE} and engine brake specific fuel consumption (BSFC), dependent on engine torque T_{ICE} and speed ω_{ICE}. The term m_{st} signifies fuel consumption during engine startup, and Λ indicates the engine's initiation event. Further, the battery modeling is as follows:

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

$$\xi = \xi_0 - \frac{1}{3600C_{bat}}\int_0^t I_{bat}dt \quad (5)$$

where ξ_0 and ξ represent the battery's initial and present SOC, while P_{bat} stands for battery power. Concurrently, R_{bat} refers to resistance, C_{bat} to capacity, and U_{oc} and I_{bat} symbolize the open-circuit voltage and current, respectively.

B. Vehicle Specifications and Digital Twin Validation

Vehicle specification details for both the source and target domains can be observed in Table 1. This work refined the model by collating extensive real-vehicle testing data and combining the physical model with its data-driven counterpart. Digital twins of the vehicles were crafted for both domains, depicted in Fig. 5 (a). The EMS employs a finite state mechanism for mode determination, coupled with a lookup table to ascertain the engine and battery's output power. The engine's digital model power output closely mirrors the real-world strategy, and the battery's SOC aligns well with bench test data, as demonstrated in Fig. 5 (b) and Fig. 5 (c). For the context of this study, the energy management tactic employed in the logistics REET of BYD is termed the benchmark (BMK) strategy, serving as the reference for subsequent energy management strategy assessments.

TABLE I
SPECIFICATIONS OF VEHICLES IN DIFFERENT DOMAINS

Component	Parameters	Source domain	Target domain
Vehicle	Total mass	1920 kg	4495 kg
	Frontal area	2.63 m^2	6.81 m^2
	Drag coefficient	0.325	0.613
Engine	Displacement type	1.5 T	2.0
	Maximum power	120 kW	95 kW
Generator	Maximum power	50 kW	95 kW
Motor	Maximum power	70 kW	150 kW
Battery	Capacity	85 A·h	46 A·h
	Nominal voltage	345 V	385 V

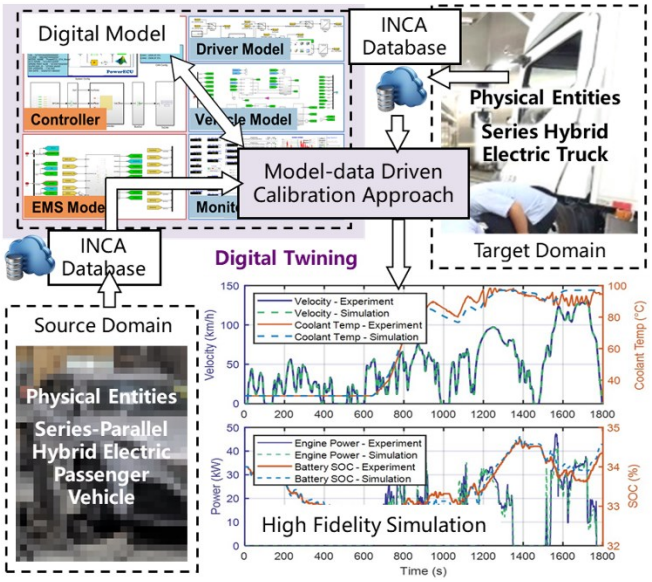


Fig. 3. Validation of powertrain model and benchmark control system based on the digital twin-based modeling approach.

III. BI-LEVEL TRANSFER FOR REINFORCEMENT LEARNING-BASED ENERGY MANAGEMENT STRATEGY

A. DRL-based Energy Management Problem Formulation

The algorithmic framework in this paper constructs the Markov Decision Process (MDP) problem by considering the entire hybrid power system, which uses the Pontryagin's Maximum Principle (PMP) [36] solution for energy allocation, as the environment. This paper defines the energy management of passenger cars under standard conditions as

the source domain $\mathcal{D}_S$ and the energy management of commercial vehicles ultimately in actual user use as the target domain $\mathcal{D}_T$. The MDP in source and target environments are modeled as $\mathcal{M}_i(s_i, a_i, p_i, r_i) \in \{\mathcal{M}_S, \mathcal{M}_T\}$. Here, s_i, a_i, p_i, and r_i represent their unique state spaces, action spaces, transition probability and reward function, respectively. The transition probability $p_i(s_i'|s_i, a_i)$ describes the distribution of the next state s_i' as a function of current s_i and a_i.

The transfer RL is designed to formulate an optimal strategy π_T for co-state estimation in target domain $\mathcal{D}_T$ using external knowledge learned in $\mathcal{D}_S$, both experience and policy, and the internal information obtained in $\mathcal{D}_T$. Based on the EMS task constructed shown form Eq. () to Eq. (), the RL problem is formulated as an infinite-time optimal control process:

$$\pi_d^* = \underset{\pi_d}{\operatorname{argmax}}\, \mathbb{E}_{(s\sim p_d, a\sim\pi_d, d=S,T)}\left[\sum_{i=1}^{\infty} \gamma^{i-1} r(s,a)\right] \tag{6}$$

where $r(s,a)$ denotes the reward associated with the state and action at time step t. Besides, p means the transition probability of MDP. Also, γ is defined as the discount factor belongs to [0,1].

The transfer process from $\mathcal{D}_S$ to $\mathcal{D}_T$ is unobservable. Therefore, an intermediate domain $\mathcal{D}_I$ is set up to showcase the intermediate transfer process. The transfer process of this paper includes two levels: L1 involves the transfer of EMS form the powertrain in $\mathcal{D}_S$ to that in $\mathcal{D}_I$. L2 deals with the transfer of vehicles from standard driving cycles to user-centric actual usage scenarios, namely from $\mathcal{D}_I$ to $\mathcal{D}_T$.

L1 aims to realize a zero-shot transfer among different powertrains, relying on a transferrable modular agent design. We introduce a per-unit agent model into the PMP-based optimal control framework, which ensures that the MDPs of $\mathcal{M}_S$ and $\mathcal{M}_I$ share the same transition probability:

$$p_S = p_I, p_S \in \mathcal{M}_S, p_I \in \mathcal{M}_I \tag{7}$$

L2 is a sample-efficient transfer involving the transition of vehicles from regulatory standard driving conditions to user-centric actual scenarios. Compared to learning from scratch, transfer learning optimizes the agent's initialization process based on knowledge transferred from the source domain, thereby helping the agent converge in the target domain with fewer interactions.

B. Modular RL agent design for cross-platform transfer

The optimization problem is defined with the goal of minimizing global fuel consumption. Conventional PMP employs a shooting method to estimate the initial value of the equivalent factor to ensure that the SOC remains within a reasonable range. However, this method is only valid when the global operating conditions are known. The Hamiltonian operator is defined to transform the global optimization problem into a collection of instantaneous optimization problems, as shown in Eq. ().

$$\mathcal{H}(P_{bat}, P_{dem}, p) = \dot{m}_f(P_{bat}, P_{dem}) - p\dot{\xi}(P_{bat}) \tag{8}$$

where p refers to the co-state of the PMP problem and $\dot{\xi}$ represents the state equation. When transformed into instantaneous optimization problem, the global optimality cannot be guaranteed. Therefore, this paper utilizes DRL for the optimal estimation of the co-state. Within the environment, PMP-based instantaneous optimization determines the unique engine output power based on the Hamiltonian operator $\mathcal{H}$ as presented in Eq. (), thereby affecting fuel economy, as illustrated in Fig. X.

$$P_{bat}^* = \underset{P_{bat}}{\operatorname{argmin}}\left(\dot{m}_f(P_{bat}, P_{dem}) + p^{pu}\frac{P_{bat}}{H_L\eta_{ICE}^*}\right) \tag{9}$$

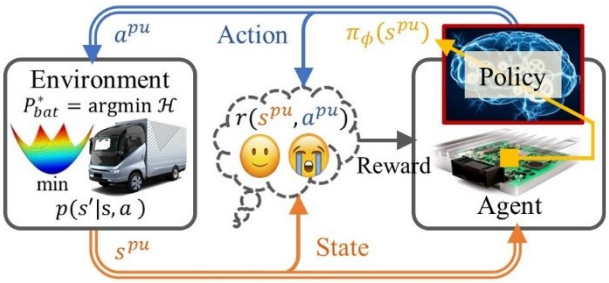


Fig. 4. EMS problem formulation with modular RL agent.

1) Per-unit state space: The environment states of the MDP primarily encompass the power allocation relationships determined by PMP, which can be denoted as the engine power P_{ICE}. Additionally, they include states relevant to the operating conditions. In this paper, the state space S is defined to incorporate P_{ICE}, the average power demand P_{avg} of the entire vehicle over a past time domain N_{avg}, the difference between the battery's SOC and its target value $\Delta\xi$, and the projected vehicle speed V_{FX} over a future time horizon N_{FX} extracted from maps. Further, a per-unit state space is proposed, where states related to the powertrain are divided by their respective reference or rated values. This method explicitly decouples the vehicle's physical attributes from the DRL agent. As a result, the powertrain parameters can be externally integrated, circumventing implicit embedding within the DRL agent's neural network, thus paving the way for seamless transfer across diverse vehicle platforms.

$$s^{pu} = \left[P_{ICE}^{pu}, P_{avg}^{pu}, \Delta\xi^{pu}, V_{FX}^{pu}\right] \tag{10}$$

$$\begin{cases} P_{ICE}^{pu} = \dfrac{P_{ICE}}{P_{ICE}^*|\eta_{ICE} = \eta_{ICE}^{max}} \\ P_{avg}^{pu} = \dfrac{P_{avg}}{\omega_n^{wl} r\left(m_v a_n + \mu m_v g + \frac{1}{2}\rho C_D A v^2\right)} \\ \Delta\xi^{pu} = \dfrac{\Delta\xi}{\Delta\xi^n} \\ V_{FX}^{pu} = \dfrac{V_{FX}}{V_{FX}^n} \end{cases} \tag{11}$$

where P_{ICE}^* denotes to mechanical power of the highest BTE point denoted by η_{ICE}^{max}. Also, ω_{DM} means the drive motor speed. T_{dem}, ρ, g, μ, C_d and σ represent the torque demand, air density, local gravitational acceleration, rolling resistance, air resistance coefficient and the road slope, respectively. Further, m_v, v and A donate to the vehicle mass, velocity and frontal area, respectively. Moreover, $\Delta\xi^n$ and V_{FX}^n represent the value of maximum SOC deviation and maximum forecasted velocity. Also, ω_n^{wl} refers to the nominal wheel angular speed

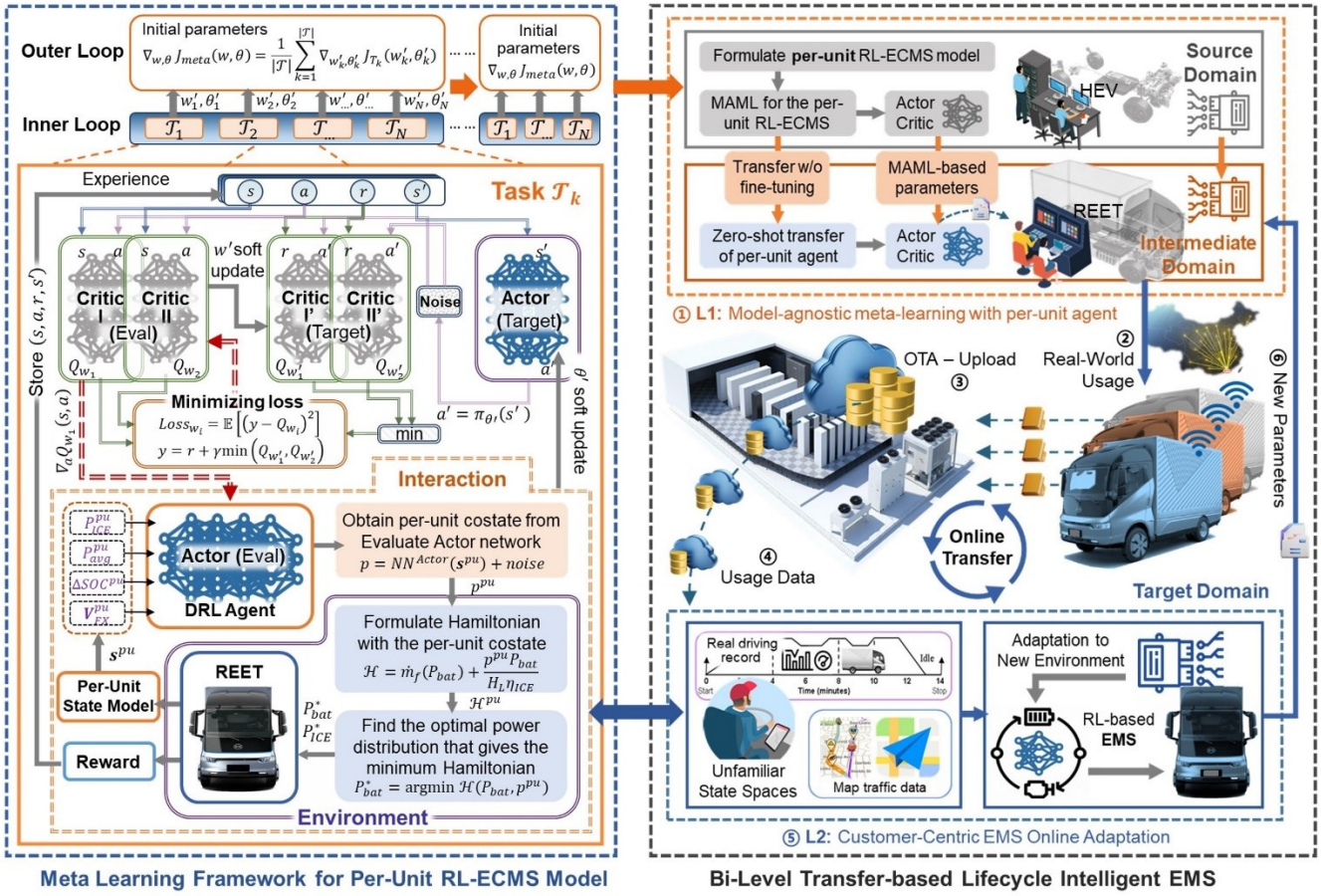


Fig. 5. Algorithm design and application framework of the bi-level lifelong learning-based transferable EMS using MAML.

determined by vehicle velocity limit. Besides, a_n is the nominal value of acceleration which is determined by its 0 to 100 km/h acceleration time.

2) Per-Unit Action Space: In Eq. (X), this work adopts the engine optimal brake thermal efficiency (BTE) η^*_{ICE} and battery capacity C_{bat} to explicitly decouple the powertrain characteristics from the actions of DRL, facilitating a modular agent design using a per-unit co-state shown in Eq. (), where H_L refers to the fuel lower heating value.

$$p^{pu} = -\frac{100\%H_L\eta^*_{ICE}}{3600C_{bat}U_{oc}}p \in [1-\varepsilon, 1+\varepsilon] \tag{12}$$

The action space a is defined by the per-unit costate p^{pu} as presented in Eq. (X) to facilitate modular RL design and obviates the need for fine-tuning process during L1 transfer.

$$a = p^{pu} \in [1-\varepsilon, 1+\varepsilon] \tag{13}$$

3) Reward Formulation: The reward function r_t is designed to include the engine fueling rate (mL/s), denoted as $\dot{m}_f(a)$, and the battery SOC penalty $\varsigma(s)$, which are both set to negative as shown in Eq. (11). The weight λ balances the engine consumption and battery charge depletion.

$$r(s,a) = -\left[\lambda\dot{m}_f(a) + (1-\lambda)\varsigma(s)\right] \tag{14}$$

$$\varsigma(s) = \begin{cases} 0, & \Delta\xi \geq 0 \\ \Delta\xi^2, & \Delta\xi < 0 \end{cases} \tag{15}$$

C. Model-agnostic meta-learning for adaptability

The integration of twin-delayed DDPG (TD3) with MAML allows for a stable and efficient policy adaptation that can quickly converge to optimal policies for new tasks using a meta-learned initialization that encodes prior experiences. For the inner loop, the TD3 algorithm is utilized to adapt to each task $\mathcal{T}_k$ from a batch of tasks $\mathcal{T}$. The adaptation process involves updating the policy parameters θ and critic parameters w_1, w_2 using task-specific data.

1) MAML Inner Loop Adaptation with TD3: The Actor network is designed to yield a per-unit estimation of the co-state as $a = \pi_\theta(s) + N(0,\delta^2)$ with Gaussian noise. Prior to initiating the learning process, the replay buffer $\mathcal{B}$ is initialized with the memory of M to store the transition tuples e harvested from the interaction process.

a) Critic update: the two Critic networks, parameterized by w_1 and w_2 for network I and II respectively, assess the Q-values from given state-action pairs and are updated using the following loss function to minimize the mean squared Bellman error (MEBE):

$$J(w_i) = \frac{1}{M}\sum_{j=1}^{M}\left[y - Q_{w_i}(s,a)\right]^2, \qquad i = 1,2 \tag{16}$$

$$y = r + \gamma\, min\left[Q_{w_1'}(s',\tilde{a}'), Q_{w_2'}(s',\tilde{a}')\right] \tag{17}$$

where $Q_{w_1'}(s',\tilde{a}')$ and $Q_{w_2'}(s',\tilde{a}')$ are derived from the Target Critic networks. The clipped action, symbolized as $\tilde{a}$, originates from the Target Actor network and is enhanced with

bounded random noise for improved value estimating as defined in Eq. (24), where $\pm c$ demarcates its range. It should be noted that the Target networks has a slower update rate compared to the Evaluate networks.

$$\tilde{a}' = \pi_{\theta'}(s') + clip[N(0, \delta^2), -c, c] \tag{18}$$

b) Actor update: two distinct Evaluate Critic networks are adopted for Actor update. The Evaluate Actor network π_θ, , responsible for the policy update, utilizes the gradient from the Critic network w_1 to adjust its parameters. Here, the gradient indicates how the policy should be changed to maximize the expected reward. Moreover, it should be noted that the update of the Target Actor network is carried out every d steps, which means that the weights update of the Target network is slower than that of the Evaluate network.

$$\begin{aligned} \nabla_\theta J(\theta) &= \mathbb{E}\left[\nabla_a Q_{w_1}(s,a)|_{a=\pi_\theta(s)} \nabla_\theta \pi_\theta(s)\right] \\ &\approx \frac{1}{M}\sum_{j=1}^{M} \nabla_a Q_{w_1}(s,a)|_{a=\pi_\theta(s)} \nabla_\theta \pi_\theta(s) \end{aligned} \tag{19}$$

2) MAML Outer Loop Meta-Optimization: After individual task adaptations in the inner loop, the outer loop optimizes the initial policy and critic parameters θ, w_1 and w_2 across all tasks. The meta-optimization aims to find parameters from which the inner loop adaptations lead to effective task-specific policies with minimal learning steps.

a) Meta-Critic update: The meta-optimization for the critic parameters uses the adapted parameters w'_{1k} and w'_{2k} of the Target Critic networks from each task in $\mathcal{T}_k$ and updates the initial parameters to minimize the MEBE errors across tasks:

$$\nabla_w J_{meta}(w) = \frac{1}{|\mathcal{T}|}\sum_{k=1}^{|\mathcal{T}|} \nabla_{w'_k} J_{T_k}(w'_k) \tag{20}$$

where J_{T_k} denotes the loss function of task $\mathcal{T}_k$, and $|\mathcal{T}|$ represents the cardinality of the set $\mathcal{T}$.

b) Meta-Actor update: Similarly, the meta-optimization for the Actor network updates the initial policy parameters θ using the gradients from the task-specific adapted policies θ'_k:

$$\nabla_\theta J_{meta}(\theta) = \frac{1}{|\mathcal{T}|}\sum_{k=1}^{|\mathcal{T}|} \nabla_{\theta'_k} J_{T_k}(\theta'_k) \tag{21}$$

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 ratio of σ_w and σ_θ, respectively [].

3) Hybrid Transfer Scheme: this work employs a hybrid transfer methodology that integrates both experience transfer (ET) and policy transfer (PT), which obtained from MAML. Within the scope of ET, experiences $e_S = (s, a, r, s')_S$ recorded from the source domain $\mathcal{D}_S$ are stored in a replay buffer $\mathcal{B}$ and then are inherited to the target domain, allowing the model to assimilate past experiences before encountering new situations within the target environment $\mathcal{M}_S$. In terms of PT based on MAML, the trained policy in domain $\mathcal{D}_S$ is firstly transferred to the intermediate domain $\mathcal{D}_I$, and then introduced into the target domain $\mathcal{D}_T$ for fast adaptation, namely the L2 transfer process, to align with the new environment. For the hybrid transfer mechanism, the model undergoes parallel training, leveraging both the ET and PT techniques. This dual mechanism, capitalizing on experiences and policy from $\mathcal{D}_S$, anticipates a swifter convergence in $\mathcal{D}_T$.

TABLE II
TD3-BASED MAML FOR POLICY ADAPTATION

Model-Agnostic Meta-Learning Algorithm with TD3	
1	**For** outer loop episode = 1,2, … , $\|\mathcal{T}\|$, **do**
2	Initialize task-specific buffers $D_i (i = 1,2,\dots,\|\mathcal{T}\|)$ with capacity M, number of tasks $\|\mathcal{T}\|$, number of adaptation steps K, Actor and Critic parameters: θ and $w_i (i = 1,2)$
3	**For** inner loop task $\mathcal{T}_k = \mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{\|\mathcal{T}\|}$, **do**
4	**For** task episode = 1,2, … , E, **do**
5	Sample transitions using π_θ and add to buffer D_i
6	Initialize task-specific parameters by $w'_i \leftarrow w_i (i = 1,2)$ and $\theta' \leftarrow \theta$
7	Sample N_s mini-batch of $e_n = (s_n, a_n, r_n, s_{n+1})$, $(n = 1,2 \dots, N_s)$, from buffer D_i
8	Obtain $\tilde{a}_{t+1} = \pi_{\theta'}(s_{t+1}) + clip[N(0, \delta'^2), -c, c]$, $y = r + \gamma \min\left[Q_{w'_1}(s', \tilde{a}'), Q_{w'_2}(s', \tilde{a}')\right]$
9	Update Critics $w_i (i = 1,2)$ by minimizing the loss: $J(w_i) = \frac{1}{N_s}\sum_{n=1}^{N_s}\left[y - Q_{w_i}(s,a)\right]^2, (i = 1,2)$
10	**If** Actor network update is called, **then**
11	Update θ by the deterministic policy gradient: $\nabla_\theta J(\theta) = \frac{1}{N_s}\sum_{n=1}^{N_s} \nabla_a Q_{w_1}(s,a)|_{a=\pi_\theta(s)} \nabla_\theta \pi_\theta(s)$
12	Update target networks: $w'_i \leftarrow \sigma_w w_i + (1-\sigma_w) w'_i, (i = 1,2)$ $\theta' \leftarrow \sigma_\theta \theta + (1-\sigma_\theta)\theta'$
13	**End If**
14	**End For**
15	**End For**
16	Meta-gradients for initial value update of w_1, w_2 and θ
17	**For** task $\mathcal{T}_k = \mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{\|\mathcal{T}\|}$, **do**
18	Update Critic network using $\nabla_w J_{meta}(w)$ and σ_w
19	Compute adapted policy gradient $\nabla_\theta J_{meta}(\theta)$ and update the Actor network with σ_θ
20	**End For**
21	Output the adapted parameters θ and $w_i (i = 1,2)$
22	**End for**

IV. RESULTS AND DISCUSSIONS

A. Testing and Validation Setup

The V-cycle in Fig. 6 illustrates a systematic approach to the development and validation of the proposed transferable EMS, integrating both offline and online processes. After the system function definition, the framework starts the MAML for EMS training with reliance on a per-unit model suitable for offline cross-platform transfer (L1), symbolized as the left part. As the V-cycle ascends on the right, it represents the sim-to-real application, beginning with online transfer through a digital twin-based validation, where the EMS undergoes control system validation using software-in-the-loop (MIL) and hardware-in-the-loop (HIL) simulation, ensuring that the system interacts correctly with actual hardware without the risk and expense of a full physical prototype. Then the real vehicle-in-the-loop (VIL) chassis tests are carried out, which

subject the EMS to actual powertrain to confirm its performance, reliability, and efficacy. Finally, the vehicles are dispatched to real-world driving conditions to collect customer-oriented usage data for online transfer (L2). This comprehensive V-cycle approach ensures that the T-RL-based energy management is thoroughly tested and validated at every stage from conceptualization to real-world deployment.

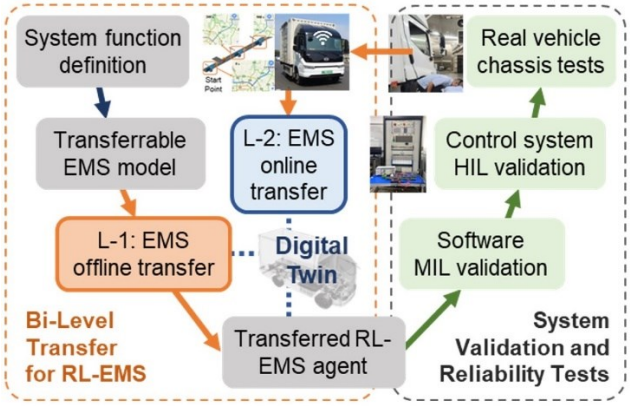


Fig. 6. The validation approach of the bi-level lifelong learning-based transferable EMS.

For L1 offline learning, the training process is randomly initialized with composite homologation driving cycles using by OEMs including the worldwide harmonized light vehicles test cycle (WLTC), China light-duty vehicle test cycle (CLTC), China heavy-duty vehicle test cycle (CHTC), and the China version world transient vehicle cycle (C-WTVC). The forecasted velocity of the state input is replaced by a stepwise curve generated according to the average value of the real velocity every 500 seconds. The L2 online transfer shifted to accommodate real-world conditions, where the EMS was exposed to actual driving cycles, diverging from the standardized patterns used in offline learning. The input of traffic information is directly fed with real data from Gaode Map, reflecting the individual pattern in daily customer usage. The REET operative data in customer usage is compared to those of the OEM's homologation driving cycles as shown in Fig. 9 and Fig. 10, which indicates the necessity of customer-centric policy adaptation in the L2 online transfer.

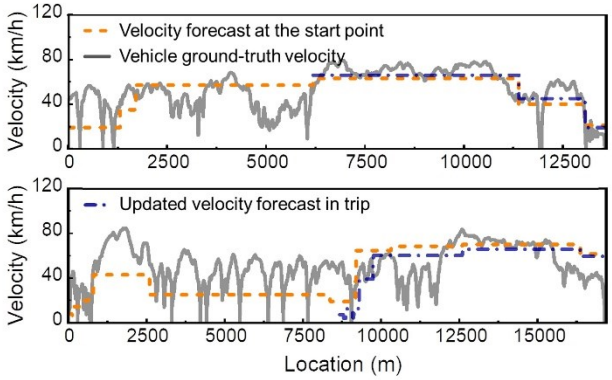


Fig. 7. Examples of ground truth velocity and Gaode map-based forecasted velocity in the user driving cycle database.

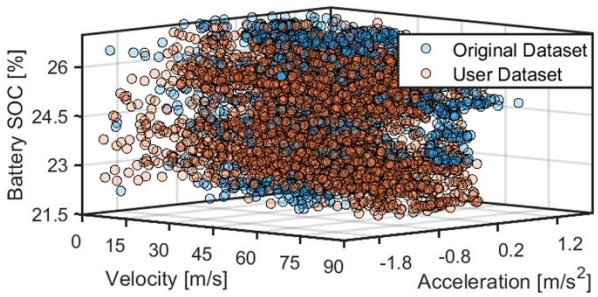


Fig. 8. Data plots of the original and user datasets in terms of the state space deviation.

B. Parameter Design and Transfer Ability

The specific hyperparameters are listed in Table III, and a total of 300 episodes have been designated for the L1 training. 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.

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

Fig. 9 contrasts the efficacy of bi-level transfer learning approach with a modular agent against conventional RL. The purple curve signifies the training process for passenger cars under OEM's homologation driving cycles. The T-RL benefiting from the per-unit design leverages existing network structures and parameters, initiating with a superior cumulative reward compared to the conventional RL which learns from scratch in the target domain, thereby enhancing transferability to new powertrains and user driving scenarios.

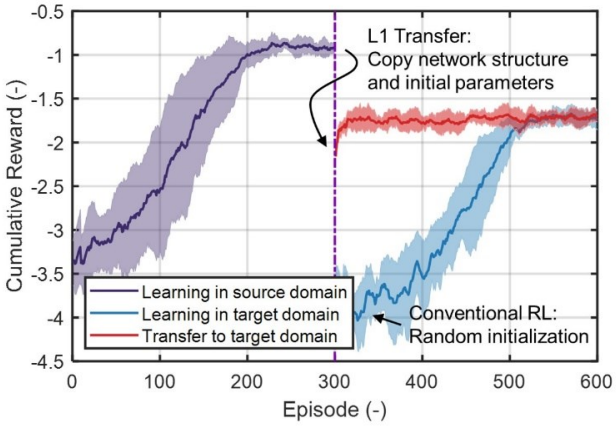


Fig. 9. Learning curve of the training process in source domain (purple line); T-RL with bi-level transfer (red line) and retraining (blue line) in target domain of user A.

C. Simulation Validation of the Bi-Level MAML

1) Comparison Study of Level-1 Cross-Platform Transferability: To analyze the L1 transfer-based EMS under homologation cycles shown in Fig. 10, the L1 transfer performance from the passenger HEV (source domain) to the commercial REET (intermediate domain) is validated. The engine output power, battery SOC and the equivalence factor are compared for PMP, T-RL with L1 transfer, and the strategy obtained in source domain with standard RL without transfer. The L1 transfer-based T-RL demonstrates improved fuel economy (FE), as shown in Table IV, with managing engine power and battery charge with similar trend of offline PMP, suggesting successful cross-platform EMS transfer.

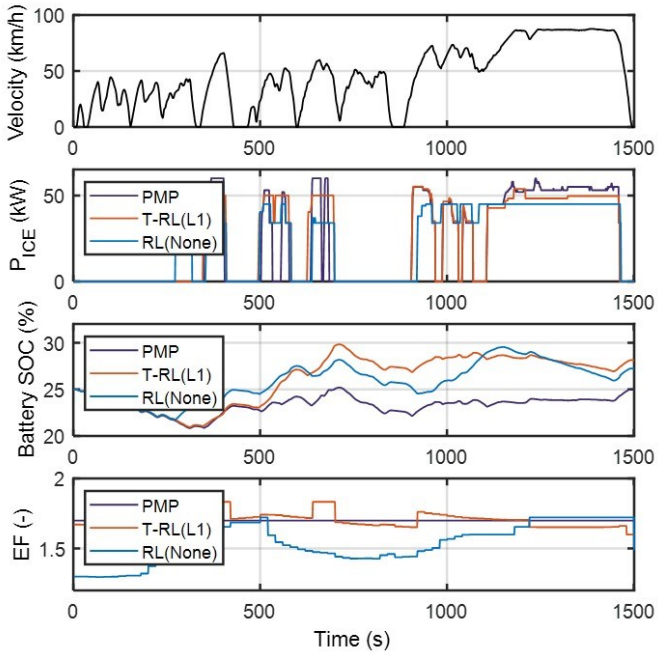


Fig. 10. Illustration of the effectiveness of L1 transfer of T-RL in terms of the engine output power, battery SOC and equivalent factor under standard driving cycle.

TABLE IV
SIMULATION OF L1 CROSS-PLATFORM TRANSFERABILITY

Performance	PMP	T-RL (L1)	RL
Initial SOC (%)	25.0	25.0	25.0
Final SOC (%)	25.0	28.1	27.1
Corrected FE (L/100km)	10.9	11.3	13.1
Fuel savings (%)	16.2	13.7	-

2) HIL Test of Level-2 Scenario Adaptation: The hardware-in-the-loop (HIL) testing system, shown in Fig. 11, is utilized to validate the real-time control performance of the EMSs. This platform 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 RCP are both connected to HIL terminals and the cloud platform, enabling real-time operation of the powertrain loaded in the real-time computer (RTPC).

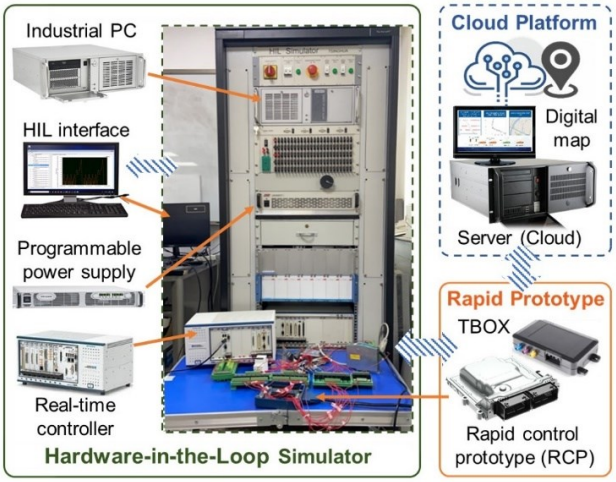


Fig. 11. Hardware-in-the-loop simulator with cloud platform.

Fig. 12 demonstrates the HIL-based assessment of different EMSs for REET, comparing engine output power, battery SOC and battery power across real-world driving cycle of user A. The results are summarized in Table V. The velocity profile serves as the driving cycle context. The graphs depict how each strategy performs in real-time, with the T-RL methods expected to adjust energy distribution more similar to that of the offline optimality, improving efficiency over the standard BMK approach by 10.4% while the conventional RL fails to improve the BMK fuel economy.

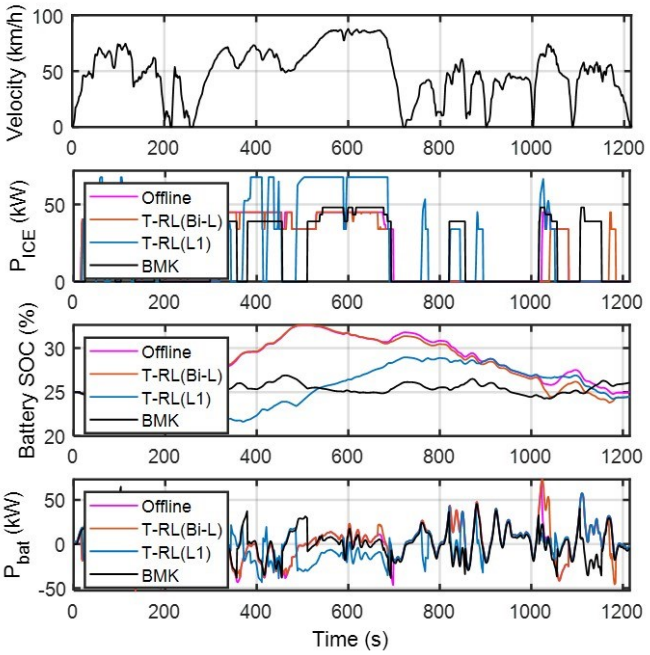


Fig. 12. HIL-based comparisons of the offline optimal strategy, T-RL with bi-level transfer, T-RL with L1 transfer and benchmark strategy in terms of the engine output power, battery SOC and battery power under user driving cycle A.

Fig. 13 compares different EMSs during driving cycle of user B using HIL simulation, as shown in Table V. The powertrain dynamics including engine power, battery SOC

and battery power over time. The performance of bi-level T-RL strategy is highlighted to show more adaptive and efficient energy management compared to the T-RL-based EMS without L2 transfer, which offers 8.6% enhancement of fuel economy in real-world scenario, especially noticeable in the engine power and SOC graphs where T-RL is more consistent with the offline PMP.

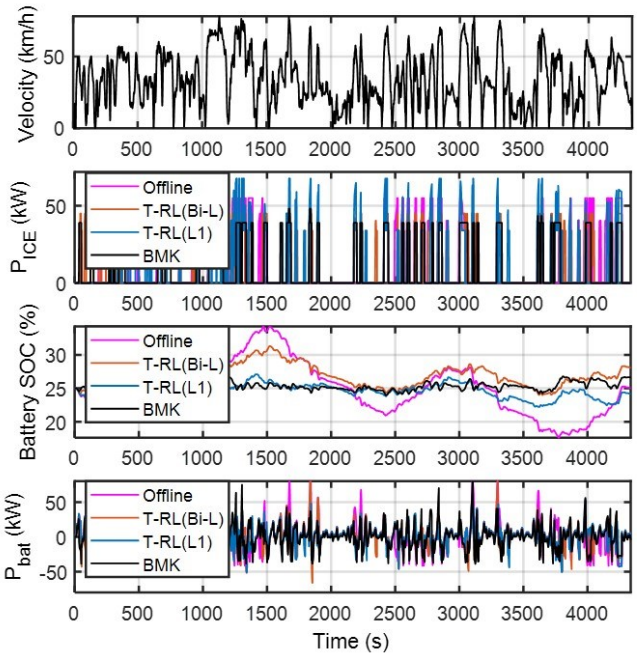


Fig. 13. HIL-based comparisons of the offline optimal strategy, T-RL with bi-level transfer, T-RL with L1 transfer and benchmark strategy in terms of the engine output power, battery SOC and battery power under user driving cycle B.

D. Real Vehicle Tests of Transferred Energy Management

Fig. 14 illustrates a vehicle-in-the-loop (VIL) setup for validating the EMS of REET, where the vehicle chassis is mounted on a dynamometer with a cooling fan to mimic natural airflow. The vehicle control unit (VCU) interacts with the powertrain and a cloud platform to simulate traffic forecasts of digital map. These inputs, alongside data recorded by the dynamometer's measurement systems, allow for comprehensive testing and calibration of the vehicle's energy management system under varying driving scenarios.

TABLE V
HIL VALIDATION OF L2 SCENARIO ADAPTATION

Drive cycle	Method	Initial SOC (%)	Final SOC (%)	Corrected FE (L/100km)	Saving (%)
User A	PMP	25.0	25.0	9.8	14.7
	Bi-L	25.0	24.5	10.3	10.4
	L1	25.0	24.5	11.3	1.7
	BMK	25.0	26.1	11.5	-
User B	PMP	25.0	25.0	8.1	16.5
	Bi-L	25.0	28.1	8.5	12.3
	L1	25.0	24.1	9.3	4.1
	BMK	25.0	26.7	9.7	-

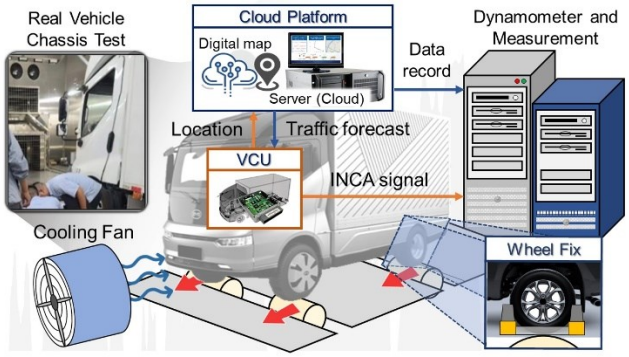


Fig. 14. Testing setup of vehicle-in-the-loop validation.

1) VIL-based Validation of Level-1 Cross-Platform Transferability: Fig. 15 depicts the VIL testing comparison for T-RL system with L1 transfer against a benchmark strategy under standard driving cycle, and the results are summarized in Table VI. The recorded velocity profiles with respect to the two control strategies have been compared which confirms the effectiveness of comparison. The T-RL (L1) demonstrates the corrected fuel economy of 11.24 L/100km and higher terminal SOC, which outperforms the benchmark strategy with 11.91 L/100km by 5.6%.

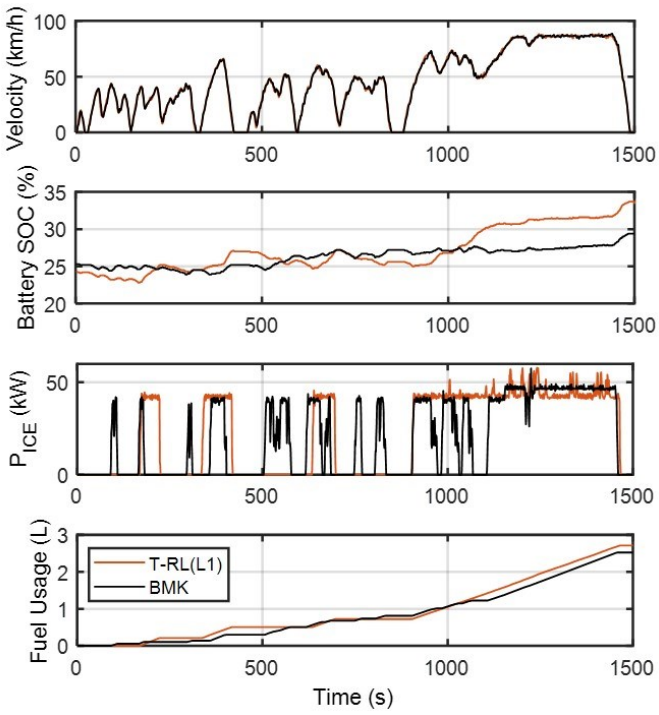


Fig. 15. VIL test-based comparisons of T-RL with L1 transfer and benchmark strategy in terms of battery SOC, engine output power and fuel usage under standard driving cycle.

Fig. 16 presents engine operation conditions from real vehicle tests, contrasting the T-RL strategy after L1 transfer against the benchmark strategy. The BMK points (purple) cluster along lower efficiency BSFC contours, while the T-RL

points (orange) are more centered among the high-efficiency region. This suggests that the T-RL strategy can be effectively transferred to the intermediate domain, the REET powertrain under standard driving cycle.

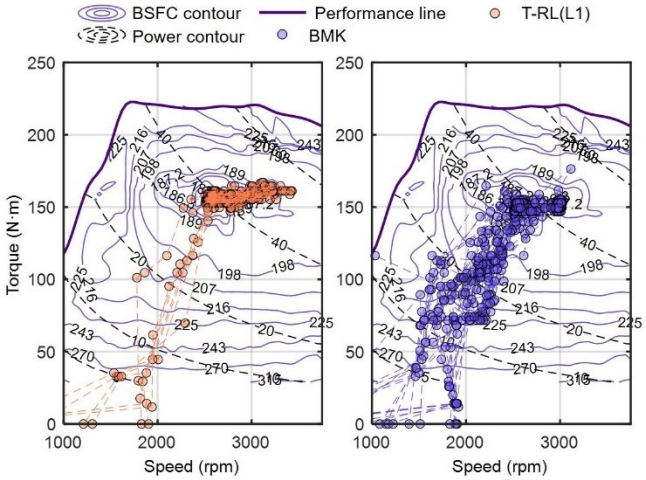


Fig. 16. Comparisons of engine operative situation in T-RL with L1 transfer and benchmark strategy under standard driving cycle with VIL tests.

TABLE VI
EXPERIMENTS OF L1 CROSS-PLATFORM TRANSFERABILITY

Performance	T-RL (L1)	BMK
Initial SOC (%)	24.3	25.3
Final SOC (%)	33.7	29.4
Corrected FE (L/100km)	11.2	11.9
Fuel savings (%)	5.9	-

2) VIL validation of Level-2 Scenario Adaptation: Fig. 17 visualizes the results from VIL tests comparing the two T-RL strategies, one with bi-level transfer and one with L1 transfer only, against a benchmark strategy during driving cycle of user A. The driving speed for testing the three control strategies can be found consistent which ensures the fairness of comparison. As observed in Table VII, the bi-level T-RL strategy obviously amends the damage of optimality of T-RL-based EMS caused by state space deviation between standard driving cycles (intermediate domain) and the customer driving cycles (target domain). The bi-level transfer tends to outperform the benchmark strategy in terms of fuel economy by 9.8%, and outperforms the benchmark system by 12.9%.

The VIL test-based engine operating conditions of the T-RL with bi-level transfer, T-RL with L1 transfer, and the benchmark strategy are compared as shown in Fig. 18. Displayed on engine speed-torque maps, the data points illustrate each strategy's engine load distribution relative to BSFC contours. The T-RL with bi-level transfer (orange) mostly congregate around the low BSFC levels, while T-RL with L1 transfer (blue) spread across various operation regions with low thermal efficiency. The benchmark (purple) shows a similar distribution to that of bi-level T-RL but witnesses a long-time operation in the BSFC states from 215 g/kWh to 270 g/kWh, covering the power range from 15 kW to 20 kW.

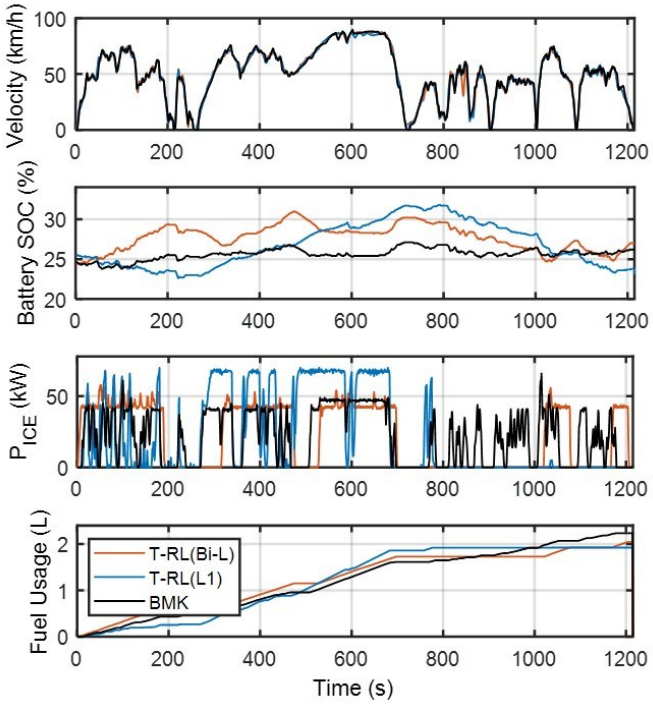


Fig. 17. VIL test-based comparisons of T-RL with bi-level transfer, T-RL with L1 transfer and benchmark strategy under user driving cycle A

The VIL test-based engine operating conditions of the T-RL with bi-level transfer, T-RL with L1 transfer, and the benchmark strategy are compared as shown in Fig. 18. Displayed on engine speed-torque maps, the data points illustrate each strategy's engine load distribution relative to BSFC contours. The T-RL with bi-level transfer (orange) mostly congregate around the low BSFC levels, while T-RL with L1 transfer (blue) spread across various operation regions with low thermal efficiency. The benchmark (purple) shows a similar distribution to that of bi-level T-RL but witnesses a long-time operation in the BSFC states from 215 g/kWh to 270 g/kWh, covering the power range from 15 kW to 20 kW.

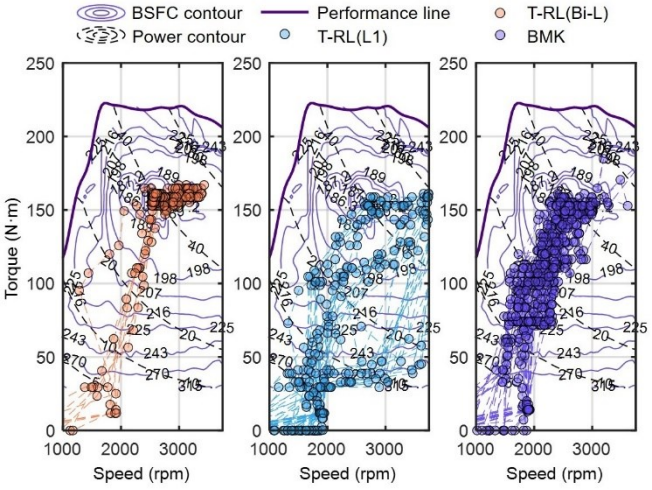


Fig. 18. Comparisons of engine operative situation under user driving cycle A with VIL tests.

The VIL test results of the two transferable EMS, bi-level and L1 transfer-based T-RL, as well as a benchmark strategy during driving cycle of user B are compared in Fig. 19. Notably, the T-RL (bi-level) method shows a lower fuel consumption curve, indicating a more efficient energy management under the tested customer driving conditions than the T-RL (L1) and benchmark methods. The experiment results are summarized in Table VII, where the bi-level transfer results outperform the strategy obtained in the intermediate domain by 8.5%. Fig. 20 contrasts the engine operative points under the three strategies against the backdrop of BSFC contours during the driving cycle of user B in VIL tests. The bi-level T-RL strategy (orange) shows a concentrated pattern suggesting consistent engine operation, the T-RL with L1 transfer only (blue) demonstrates a broader spread indicating wrong decision of engine output power in unfamiliar driving conditions of user B.

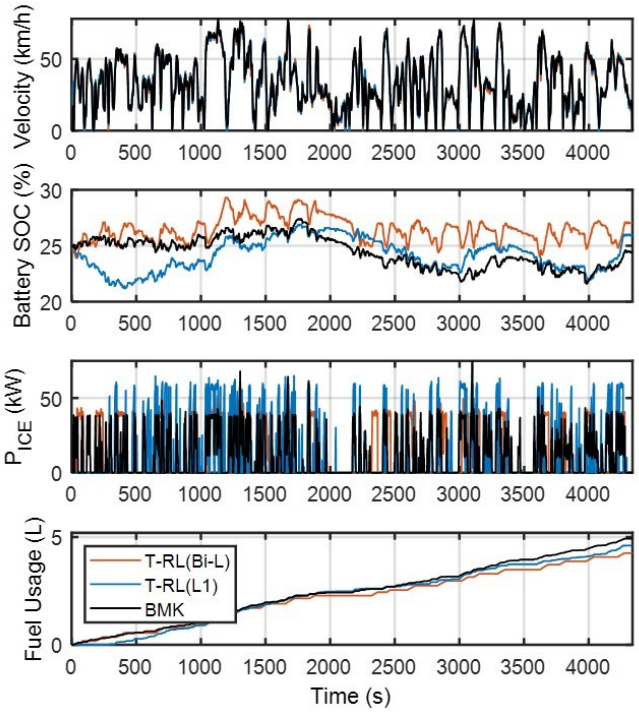


Fig. 19. VIL test-based comparisons of T-RL with bi-level transfer, T-RL with L1 transfer and benchmark strategy under user driving cycle B

TABLE VII
VIL VALIDATION OF L2 SCENARIO ADAPTATION

Test	Performance	Bi-L	L1	BMK
User A	Initial SOC (%)	24.4	25.5	24.6
	Final SOC (%)	27.0	23.8	26.2
	Corrected FE (L/100km)	10.1	11.2	11.6
	Fuel savings (%)	12.9	3.4	-
User B	Initial SOC (%)	24.8	25.1	25.0
	Final SOC (%)	27.0	26.0	24.4
	Corrected FE (L/100km)	8.6	9.4	9.9
	Fuel savings (%)	13.1	5.1	-

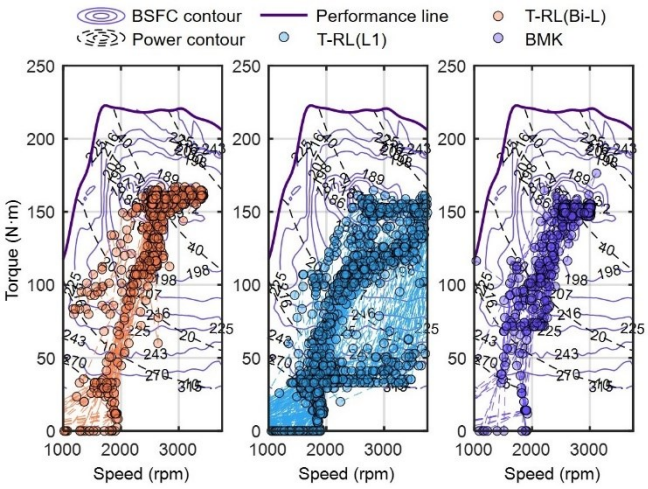


Fig. 20. Comparisons of engine operative situation under user driving cycle B with VIL tests.

The VIL test results of the T-RL with bi-level transfer or L1 transfer are compared against the conventional RL and benchmark strategy in Fig. 16, which illustrate the convergence speed, corrected fuel economy under both standard and user driving cycles. The learning process when shifting from the source domain to the target domain is shortened by 91.7% with the bi-level transfer scheme. Besides, the fuel economy of T-RL is 5.9% less than that of the OEM's benchmark strategy under standard cycle, and outperforms the state-of-the-art RL-based control and OEM's benchmark under users' actual driving scenarios by 8.7% and 12.6%, respectively. Therefore, the proposed method is capable of cross-platform application and fast online fine-tuning to adapt to unfamiliar state space in customers' driving conditions.

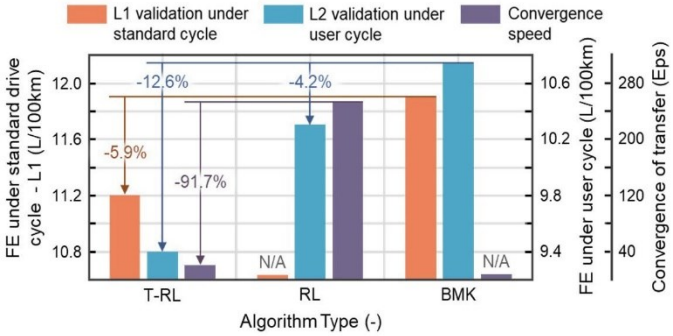


Fig. 16. Real vehicle test results in terms of fuel economy, and convergence speed for new driving scenario adaptation.

VI. CONCLUSIONS

This paper proposes a bi-level T-RL based on model-agnostic meta-learning to realize transferable and self-evolving EMS for logistics REETs. The performance of the proposed T-RL has been validated by comprehensive HIL and VIL experiments under both homologation and real-world drive cycles, and the findings are summarized as follows:

1) This work pioneers in real-vehicle experiments of implementing learning-based EMS. The ECMS serves as

the fundamental framework to enable interpretability and transferability, where the conventional heuristic control maps for equivalent factor estimation are replaced by a neural network calibrated by T-RL. This approach marks a firm advancement in sim-to-real transfer in RL-based EMS.

2) The L1 transfer is validated by the per-unit agent trained in the source domain achieving 96% of the optimality realized by offline PMP in the intermediate domain, demonstrating a zero-shot knowledge transfer from passenger cars to logistics trucks. The T-RL with L1 transfer is featured with a 13.7% gap in FE compared to conventional RL, and initiates with a superior cumulative reward for starting the fine-tuning of L2 transfer.
3) The L2 transfer realizes online adaptation to real-world driving conditions enabled by usage data feedback and DT-based cloud computing, and a correction of 8.0%~9.5% fuel economy is improved against the convention reinforcement learning-based EMS during usage via online-adaptation, effectively bridging the optimality gap between the control policy learned during development and the global optimal operation in actual driving scenarios.

REFERENCES

[1] Y. Lin, X. Na, D. Wang, X. Dai, and F. Y. Wang, "Mobility 5.0: Smart Logistics and Transportation Services in Cyber-Physical-Social Systems," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 6, pp. 3527-3532, 2023, doi: 10.1109/TIV.2023.3286995.

[2] X. Qi, G. Wu, P. Hao, K. Boriboonsomsin, and M. Barth, "Integrated-Connected Eco-Driving System for PHEVs With Co-Optimization of Vehicle Dynamics and Powertrain Operations," IEEE Transactions on Intelligent Vehicles, vol. PP, pp. 1-1, 05/26 2017, doi: 10.1109/TIV.2017.2708599.

[3] Z. Hou et al., "A Learning-and-tube-based Robust Model Predictive Control Strategy for Plug-in Hybrid Electric Vehicle," IEEE Transactions on Intelligent Vehicles, pp. 1-15, 2023, doi: 10.1109/TIV.2023.3331268.

[4] Q. Zhou et al., "Transferable representation modelling for real-time energy management of the plug-in hybrid vehicle based on k-fold fuzzy learning and Gaussian process regression," Applied Energy, vol. 305, p. 117853, 2022/01/01/ 2022, doi: https://doi.org/10.1016/j.apenergy.2021.117853.

[5] J. Chen, Y. Zhang, S. Teng, Y. Chen, H. Zhang, and F. Y. Wang, "ACP-Based Energy-Efficient Schemes for Sustainable Intelligent Transportation Systems," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 5, pp. 3224-3227, 2023, doi: 10.1109/TIV.2023.3269527.

[6] L. Guo et al., "Bi-level Optimization of Speed Trajectory and Power Management for Autonomous HEVs in off-Road Scenarios," IEEE Transactions on Intelligent Vehicles, pp. 1-14, 2023, doi: 10.1109/TIV.2023.3313123.

[7] E. Taherzadeh, M. Dabbaghjamanesh, M. Gitizadeh, and A. Rahideh, "A New Efficient Fuel Optimization in Blended Charge Depletion/Charge Sustenance Control Strategy for Plug-In Hybrid Electric Vehicles," IEEE Transactions on Intelligent Vehicles, vol. 3, no. 3, pp. 374-383, 2018, doi: 10.1109/TIV.2018.2843173.

[8] C. Chen, J. Wang, Q. Xu, J. Wang, and K. Li, "Mixed platoon control of automated and human-driven vehicles at a signalized intersection: Dynamical analysis and optimal control," Transportation Research Part C: Emerging Technologies, vol. 127, p. 103138, 06/01 2021, doi: 10.1016/j.trc.2021.103138.

[9] G. Li, D. Görges, and M. Wang, "Online Optimization of Gear Shift and Velocity for Eco-Driving Using Adaptive Dynamic Programming," IEEE Transactions on Intelligent Vehicles, vol. 7, no. 1, pp. 123-132, 2022, doi: 10.1109/TIV.2021.3111037.

[10] T. Liu, K. Tan, W. Zhu, and L. Feng, "Computationally Efficient Energy Management for a Parallel Hybrid Electric Vehicle Using Adaptive Dynamic Programming," IEEE Transactions on Intelligent Vehicles, pp. 1-15, 2023, doi: 10.1109/TIV.2023.3285392.

[11] B. Shuai et al., "Heuristic action execution for energy efficient charge-sustaining control of connected hybrid vehicles with model-free double Q-learning," Applied Energy, vol. 267, 04/06 2020, doi: 10.1016/j.apenergy.2020.114900.

[12] N. Cui, W. Cui, and Y. Shi, "Deep Reinforcement Learning Based PHEV Energy Management With Co-Recognition for Traffic Condition and Driving Style," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 4, pp. 3026-3039, 2023, doi: 10.1109/TIV.2023.3235110.

[13] 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, doi: 10.1109/TTE.2023.3263328.

[14] Z. Zhu, N. Pivaro, S. Gupta, A. Gupta, and M. Canova, "Safe Model-Based Off-Policy Reinforcement Learning for Eco-Driving in Connected and Automated Hybrid Electric Vehicles," IEEE Transactions on Intelligent Vehicles, vol. 7, no. 2, pp. 387-398, 2022, doi: 10.1109/TIV.2022.3150668.

[15] 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, doi: https://doi.org/10.1016/j.enconman.2022.115450.

[16] H. Zhang, N. Lei, S. Liu, Q. Fan, and Z. Wang, "Data-driven predictive energy consumption minimization strategy for connected plug-in hybrid electric vehicles," Energy, vol. 283, p. 128514, 2023/11/15/ 2023, doi: https://doi.org/10.1016/j.energy.2023.128514.

[17] Y. Zhang, J. Chen, S. Teng, H. Zhang, and F. Y. Wang, "Sustainable Lifecycle Management for Automotive Development via Multi-Dimensional Circular Design Framework," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 9, pp. 4151-4154, 2023, doi: 10.1109/TIV.2023.3319478.

[18] 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, doi: https://doi.org/10.1016/j.apenergy.2022.120033.

[19] H. Zhang, B. Chen, N. Lei, B. Li, R. Li, and Z. Wang, "Integrated Thermal and Energy Management of Connected Hybrid Electric Vehicles Using Deep Reinforcement Learning," IEEE Transactions on Transportation Electrification, pp. 1-1, 2023, doi: 10.1109/TTE.2023.3309396.

[20] H. Zhang, J. Peng, H. Dong, H. Tan, and F. Ding, "Hierarchical reinforcement learning based energy management strategy of plug-in hybrid electric vehicle for ecological car-following process," Applied Energy, vol. 333, p. 120599, 2023/03/01/ 2023, doi: https://doi.org/10.1016/j.apenergy.2022.120599.

[21] N. Lei, H. Zhang, H. Wang, and Z. Wang, "An Improved Co-optimization of Component Sizing and Energy Management for Hybrid Powertrains Interacting with High-fidelity Model," IEEE Transactions on Vehicular Technology, pp. 1-12, 2023, doi: 10.1109/TVT.2023.3296114.

[22] M. Spiryagin, J. Edelmann, F. Klinger, and C. Cole, "Vehicle system dynamics in digital twin studies in rail and road domains," Vehicle System Dynamics, vol. 61, pp. 1-50, 03/13 2023, doi: 10.1080/00423114.2023.2188228.

[23] C. Zhang et al., "Dedicated Adaptive Particle Swarm Optimization Algorithm for Digital Twin Based Control Optimization of the Plug-In Hybrid Vehicle," IEEE Transactions on Transportation Electrification, vol. 9, no. 2, pp. 3137-3148, 2023, doi: 10.1109/TTE.2022.3219290.

[24] R. Lian, H. Tan, J. Peng, Q. Li, and Y. Wu, "Cross-Type Transfer for Deep Reinforcement Learning Based Hybrid Electric Vehicle Energy Management," IEEE Transactions on Vehicular Technology, vol. 69, no. 8, pp. 8367-8380, 2020, doi: 10.1109/TVT.2020.2999263.

[25] Z. Xu, S. Shi, A. X. Liu, J. Zhao, and L. Chen, "An Adaptive and Fast Convergent Approach to Differentially Private Deep Learning," in IEEE INFOCOM 2020 - IEEE Conference on Computer Communications, 6-9 July 2020 2020, pp. 1867-1876, doi: 10.1109/INFOCOM41043.2020.9155359.

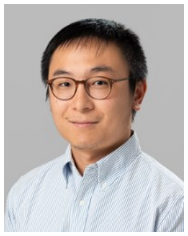
[26] W. Gong, C. Liu, X. Zhao, and S. Xu, "A Model Review for Controller-hardware-in-the-loop Simulation in EV Powertrain Application," IEEE Transactions on Transportation Electrification, vol. PP, pp. 1-1, 01/01 2023, doi: 10.1109/TTE.2023.3290999.
[27] N. Stroe, S. Olaru, G. Colin, K. Ben-Cherif, and Y. Chamaillard, "Predictive Control Framework for HEV: Energy Management and Free-Wheeling Analysis," IEEE Transactions on Intelligent Vehicles, vol. 4, no. 2, pp. 220-231, 2019, doi: 10.1109/TIV.2019.2904416.
[28] W. Hong, I. Chakraborty, H. Wang, and G. Tao, "Co-Optimization Scheme for the Powertrain and Exhaust Emission Control System of Hybrid Electric Vehicles Using Future Speed Prediction," IEEE Transactions on Intelligent Vehicles, vol. 6, no. 3, pp. 533-545, 2021, doi: 10.1109/TIV.2021.3049296.
[29] B. Li, Y. Ouyang, X. Li, D. Cao, T. Zhang, and Y. Wang, "Mixed-Integer and Conditional Trajectory Planning for an Autonomous Mining Truck in Loading/Dumping Scenarios: A Global Optimization Approach," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 2, pp. 1512-1522, 2023, doi: 10.1109/TIV.2022.3214777.
[30] D. Phan, A. Bab-Hadiashar, M. Fayyazi, R. Hoseinnezhad, R. N. Jazar, and H. Khayyam, "Interval Type 2 Fuzzy Logic Control for Energy Management of Hybrid Electric Autonomous Vehicles," IEEE Transactions on Intelligent Vehicles, vol. 6, no. 2, pp. 210-220, 2021, doi: 10.1109/TIV.2020.3011954.
[31] J. Shi, J. Wu, B. Xu, and Z. Song, "Cybersecurity of Hybrid Electric City Bus with V2C Connectivity," IEEE Transactions on Intelligent Vehicles, pp. 1-16, 2023, doi: 10.1109/TIV.2023.3281032.
[32] M. Yue, K. Benaggoune, J. Meng, and D. Diallo, "Implementation of an Early Stage Fuel Cell Degradation Prediction Digital Twin Based on Transfer Learning," IEEE Transactions on Transportation Electrification, vol. PP, pp. 1-1, 01/01 2022, doi: 10.1109/TTE.2022.3229716.
[33] 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, doi: 10.1109/TTE.2021.3054896.
[34] C. Yang, R. Chen, W. Wang, Y. Li, X. Shen, and C. Xiang, "Cyber-Physical Optimization-Based Fuzzy Control Strategy for Plug-in Hybrid Electric Buses Using Iterative Modified Particle Swarm Optimization," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 5, pp. 3285-3298, 2023, doi: 10.1109/TIV.2023.3260007.
[35] H. Kazemi, Y. P. Fallah, A. Nix, and S. Wayne, "Predictive AECMS by Utilization of Intelligent Transportation Systems for Hybrid Electric Vehicle Powertrain Control," IEEE Transactions on Intelligent Vehicles, vol. 2, no. 2, pp. 75-84, 2017, doi: 10.1109/TIV.2017.2716839.
[36] D. Shen, D. Karbowski, and A. Rousseau, "A Minimum Principle-Based Algorithm for Energy-Efficient Eco-Driving of Electric Vehicles in Various Traffic and Road Conditions," IEEE Transactions on Intelligent Vehicles, vol. 5, no. 4, pp. 725-737, 2020, doi: 10.1109/TIV.2020.3011055

Hao Zhang is currently a Ph.D. candidate at School of Vehicle and Mobility, Tsinghua University, China. He was also a Visiting Scholar 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 connected and automated hybrid electric vehicles and distributed energy storage systems.

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 includes component sizing, energy management and eco-driving of connected and automated hybrid electric vehicles.

Wang Peng is the head of research and development for commercial vehicles in BYD. He has served as the project leader in the National Key Research and Development Projects of the 13th and 14th Five Year Plans, and received the China Automotive Industry Awards of Science and Technology, leading and participating in the formulation of more than ten national standards for new energy vehicles. His research focuses on design and industrial application of new energy commercial vehicles.

Bingbing Li is a Ph.D. candidate at School of Mechanical Engineering, Southeast University, China, and a visiting scholar at the Department of Electronic and Electrical Engineering, University College London, U.K. His research interests include connected and automated vehicles, energy-efficient driving, and vehicle dynamics control.



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.

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.