**ORIGINAL RESEARCH** **Open access**

# Deep Reinforcement Learning-Enhanced Rule-Based Energy Management for Hybrid Electric Vehicles with Organic Rankine Cycle

Mariana Lopes[1*], Felipe Rocha[1], Beatriz Souza[2]

## Abstract

Freight trucks are highly energy-intensive. By preserving an extensive driving range while simultaneously reducing fuel usage, hybrid powertrains serve as a highly effective energy conservation solution for these vehicles. Capturing discarded engine heat via the organic Rankine cycle elevates overall engine performance and supplies efficient thermal regulation. Nevertheless, integrating this system substantially complicates the powertrain's energy management architecture. To construct a highly efficient and resilient energy management system, this research introduces a rule-based framework embedded with deep reinforcement learning. This technique refines the crucial parameters of the conventional rule-based setup by incorporating deep reinforcement learning capabilities. Consequently, the proposed model merges the superior optimization traits of deep reinforcement learning with the dependable stability of established rules. To confirm the viability of this approach, the authors developed a dynamic system model and executed simulation analyses. Following this, a semi-physical test bench for the hybrid powertrain was assembled to conduct rapid control prototype experiments. Simulation outcomes indicate that, compared to a standard rule-based setup, the newly embedded system reduces energy usage by 4.31% during the C-WTVC driving cycle. Furthermore, the system successfully maintains safe operations and conserves energy even across untested, unfamiliar driving cycles. The rapid control prototype tests reveal a strong alignment between experimental and simulated data, highlighting its viability for actual vehicle engineering and encouraging the broader industrial adoption of deep reinforcement learning.

**Keywords** Deep reinforcement learning, Energy management system, Hybrid electric vehicle, Organic Rankine cycle, Rapid control prototype

*Correspondence:
Mariana Lopes
mariana.lopes@gmail.com
[1] Department of Artificial Intelligence and Renewable Energy, School of Engineering, Federal University of Minas Gerais, Belo Horizonte, Brazil
[2] Department of Smart Energy Engineering, Faculty of Engineering, University of Campinas, Campinas, Brazil

## Introduction

Data from the International Energy Agency indicates that medium- and heavy-duty trucks are responsible for approximately 10% of worldwide CO2 emissions [1]. Across Europe, these heavy-duty transports contribute to over 25% of overall CO2 output [2]. Within China, such vehicles consume upwards of 60% of the energy used for road transport [3]. In the United States, trucks move 72.5% of all cargo, establishing them as the primary mode of freight transport. Consequently, American trucks account for 15.3% of national fuel usage and 24% of greenhouse gas emissions. Focusing on energy efficiency and lowering emissions in medium and heavy-duty trucks is crucial for realizing global

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

carbon neutrality and peak carbon targets. For long-haul trucking, hybrid powertrains represent a vital technological strategy to maintain operational range while successfully cutting fuel use and emissions [4, 5].

Research highlights that in highly variable urban and suburban driving environments, hybrid electric vehicles (HEVs) deliver superior energy conservation owing to the electric motor's high efficiency [6]. Conversely, during highway operations—where trucks spend the majority of their time—the driving dynamics are stable, allowing the internal combustion engine to operate efficiently over long distances. Consequently, standard hybrid systems struggle to provide significant benefits at highway speeds [7]. To effectively elevate engine efficiency, waste heat can be captured and converted into electricity. Transforming this thermal waste into electrical power subsequently extends the vehicle's electric driving range [8].

Furthermore, in city and suburban settings, hybrid systems help optimize the engine's operational parameters, thereby facilitating better integration of waste heat recovery systems. Thus, merging hybrid setups with waste heat recovery creates a synergistic energy strategy effective across all driving scenarios. Due to its flexibility, affordability, and compatibility with low-grade heat, the organic Rankine cycle (ORC) is widely considered a highly efficient method for harvesting engine exhaust heat [9]. Utilizing the ORC to reclaim thermal energy simultaneously boosts engine performance [10] and acts as a substitute for traditional engine thermal management configurations [11]. The ORC extends electric range by transforming wasted thermal energy into electricity, while the hybrid system enables the ORC's electrification. Within HEVs, optimal power distribution is the primary driver of overall efficiency. Investigations demonstrate that varying energy management systems (EMS) can cause up to a 20% variation in fuel consumption for identical vehicles operating on identical routes [12]. Incorporating the ORC introduces even more intricacy to the powertrain, making the role of the EMS critically vital.

Extensive investigations have explored various EMS architectures for HEVs [13-15]. Biswas and Emadi [16] provided a comprehensive review of how HEV energy management strategies evolved from 1993 to 2018. Early EMS designs relied predominantly on fixed rules [17]. Later, scholars refined these rules based on empirical insights derived from dynamic programming (DP) outcomes [18]. Nevertheless, DP's massive computational demands make it impractical to account for every potential driving condition. Following this, engineers implemented instantaneous optimal control strategies [19, 20], though these approaches were constrained by their focus on local, rather than global, optimization. Researchers then shifted toward securing global optimal control for predetermined routes using techniques like model predictive control (MPC) [21, 22]. The drawback of MPC is its heavy reliance on forecasting upcoming traffic and road parameters. More recently, advancements in computing and artificial intelligence (AI) have facilitated the integration of deep reinforcement learning (DRL) algorithms into HEV energy management [23-25]. This innovative approach holds the promise of reaching near-global optimization without requiring prior knowledge or predictions of future driving conditions.

Energy management systems for HEVs integrated with ORC technologies have experienced a comparable developmental trajectory. Gao *et al*. [26] constructed a hybrid truck simulation incorporating an ORC framework designed to capture thermal energy from engine exhaust. Using a rule-based EMS, their quantitative assessment demonstrated that the ORC setup could lower energy consumption by 5.42% during highway driving. Likewise, Andwari *et al*. [27] assessed the efficiency gains of an HEV-ORC combination across three distinct driving cycles, also employing a rule-based EMS. Kruijt *et al*. [28] simulated a heavy-duty parallel hybrid truck featuring an integrated ORC system. By merging the equivalent consumption minimization strategy (ECMS) with the Newton-Raphson shooting technique, they achieved instantaneous optimal control, yielding a 2.5% boost in fuel efficiency during the UDDS cycle. Mansour *et a*l. [29] analyzed the conservation capabilities of a mild hybrid equipped with a coolant-based ORC waste heat recovery setup, leveraging a globally optimal DP management strategy. Their findings showed a 2.4% fuel economy enhancement over the Worldwide Harmonized Light Vehicles Test Cycles. While this research offered a theoretical optimum for the HEV-ORC configuration, it lacked practical real-world applicability. To date, the only openly published literature detailing a DRL-centric EMS for HEV-ORC systems is authored by Wang *et al.* [30]. They introduced a DRL-driven EMS that proved to be 2% more fuel-efficient than traditional rule-based alternatives, effectively optimizing both motor and engine performance.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

The specific DRL architecture implemented in their work closely mirrors the algorithms conventionally used in standard HEVs lacking ORC components.

Traditional DRL-EMS frameworks typically select primary variables—like battery SOC, acceleration, and vehicle speed—as state inputs, while using fuel usage as the reward signal to train the agent to minimize consumption. While this methodology meets training goals under specific routes, contemporary learning models encounter several hurdles, including extended training durations [30], complicated reward function formulations [31], weak systemic robustness [30], and tricky hyper-parameter calibration [32]. These issues are particularly pronounced in intricate architectures with multiple power sources, such as HEV-ORC platforms. Lian *et al*. [31] noted that balancing multiple targets within the reward function heavily dictates the ultimate behavior of the control strategy. Research by Xu *et al*. [32] indicated that while DRL-based systems successfully converge and conserve energy under trained scenarios, their performance drops sharply when confronted with novel driving environments. Furthermore, Zhang *et al*. [33] demonstrated that as the state space expands, the necessary DRL training iterations grow exponentially. Consequently, despite the vast academic focus dedicated to HEV and HEV-ORC energy management, the overwhelming majority of commercial vehicles on the road continue to rely on traditional rule-based controllers [34].

Practical engineering studies seeking to optimize HEV controllers using DRL are predominantly confined to the hardware-in-the-loop (HIL) evaluation stage. For instance, Zou *et al*. [35] introduced an online-updatable EMS using deep Q-learning, which delivered fuel conservation comparable to DP and was successfully validated via HIL testing. Similarly, Deng *et al*. [36] conceptualized a DRL-driven energy management protocol for hybrid fuel cell buses, utilizing HIL test benches for verification. Their strategy demonstrated a 7.8 % enhancement in energy economy over standard rule-based methods. However, to the authors' current understanding, no existing publications demonstrate the deployment of DRL-based EMS for the direct control of a physical hybrid test bench or an actual vehicle. This absence of real-world implementation is largely due to stringent economic and safety constraints.

To summarize, contemporary deep reinforcement learning (DRL) approaches for optimizing the energy management systems (EMS) of intricate hybrid powertrains continue to be hindered by demanding training procedures, deficient robustness (adaptability) when subjected to unfamiliar driving patterns, and a general lack of experimental validation. To formulate an HEV-ORC EMS viable for actual industrial implementation, this research presents an original deep reinforcement learning-embedded rule-based (DRLERB) EMS. This approach optimizes the strategy's key parameter settings by blending a DRL framework directly into the rule-based EMS core. Operating on the continuous DRL model known as Deep Deterministic Policy Gradient (DDPG), the DRLERB method enables continuous refinement of key operational variables. Consequently, this method maintains structural reliability and adaptability under changing real-world traction cycles while simultaneously optimizing control parameters to boost fuel efficiency. To confirm the performance of the DRLERB EMS, this work executes a comprehensive evaluation of its fuel-saving potential across diverse driving environments. Additionally, a scaled-down, hardware-in-the-loop experimental platform representing the hybrid powertrain was constructed, enabling rapid control prototype (RCP) testing to substantiate the practical utility of this approach.

The primary novelty of this investigation lies in the introduction of the deep reinforcement learning embedded rule-based (DRLERB) EMS, which targets the prolonged learning phases, weak generalization capabilities under unencountered road routes, and insufficient empirical testing found in existing DRL-optimized control methods for complex multi-source powertrains. This strategy ensures stable control and adaptability across varied driving schedules while simultaneously refining control actions to minimize fuel consumption. Furthermore, a scaled-down hardware-in-the-loop configuration of the hybrid architecture was established to perform RCP experimental runs, thereby demonstrating the setup's real-world efficacy. To the best of the authors' knowledge, this marks the first instance in which a DRL-reliant energy management system has been successfully deployed to regulate a physical hybrid test apparatus or operating vehicle. Ultimately, this work advances the engineering readiness of DRL for HEV EMS optimization, providing structural insights that carry over to parallel technical domains.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

# Structure, Parameters and Modelling of HEV-ORC

## Structure and parameters of HEV-ORC

**Figure 1** illustrates the layout of the integrated HEV-ORC framework. The vehicle is configured as a parallel hybrid commercial truck utilizing a P2 layout. The physical prototype consists of a 42-tonne heavy-duty tractor unit. Powering this vehicle is a 6-cylinder diesel engine with a rated output of 360 kW coupled alongside a 77 kW electric motor. To capture the thermal energy rejected by the engine through the waste heat recovery loop and ensure optimal engine temperature control, the dual-pressure organic Rankine cycle (DPORC) configuration pictured in **Figure 1** is deployed as the vehicle's thermal management setup. This specialized DPORC architecture optimizes exploitation of waste heat from the engine exhaust, charge air cooler, and engine coolant loop under all driving profiles via the interactive control of bypass valves, working fluid pump settings, and expander rotational velocities. The specific geometric sizing and full-range dynamic simulation of this DPORC arrangement are described in comprehensive detail within a previous study [11].

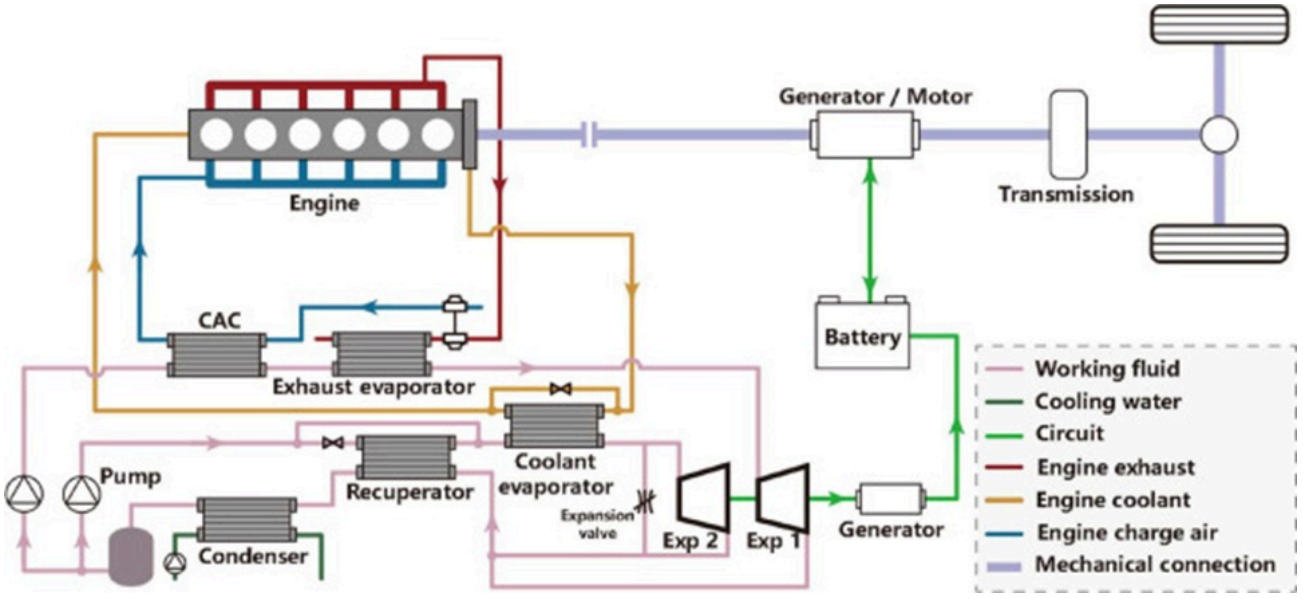



**Figure 1.** Structure of P2 configuration parallel hybrid truck and dual pressure organic Rankine cycle.

## HEV model

The mathematical representation of the HEV constructed for this research focuses explicitly on vehicle longitudinal dynamics. This comprehensive model integrates individual components for the diesel engine, electric motor, energy storage battery, mechanical driveline, driver control behavior, and total chassis aerodynamics. Each subsystem is parameterized according to the physical characteristics of the target truck. Because the engine and the electric motor serve as the prime movers of the hybrid platform, capturing their behavior precisely is vital. The core task of the EMS is to allocate power splits based on the high-efficiency operational islands of both units, securing a coordinated balance between the dual powerplants. Here, the engine and motor models are built using empirical efficiency maps obtained through direct laboratory characterization. Since both components represent the factory-installed power units of the actual base vehicle, they offer an exceptionally high level of integration fidelity. The vehicle's battery pack is simulated using an equivalent circuit model framework, whereas the driver's response is replicated with an equivalent PID feedback control scheme. The mechanical powertrain and structural model of the truck assume a single-degree-of-freedom longitudinal linear translation. The driveline incorporates the main reduction assembly and the gear shifting unit, featuring a final reduction ratio of 2.81 mated to a 16-speed automated manual transmission (AMT). Power flows from both the engine and electric motor directly to the drive wheels through this driveline layout. The total tractive power requested by the

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

driver model is derived by calculating the instantaneous aerodynamic drag, tire rolling resistance, gradient gravitational load, and inertial acceleration force.

## DPORC system model

The five distinct fluid-to-fluid heat exchangers within the DPORC framework are formulated utilizing the finite volume method [37], which provides excellent mathematical stability. The analytical modeling of each heat-transfer unit enforces compliance with fluid mass conservation, fluid energy conservation, and solid wall thermal energy balance equations.

Fluid mass conservation relation:

$$\dot{m}_{in} - \dot{m}_{out} = V\left(\frac{d\bar{\rho}}{d\bar{P}}\bigg|_{\bar{h}}\frac{dP}{dt} + \frac{d\bar{\rho}}{d\bar{h}}\bigg|_{p}\frac{d\bar{h}}{dt}\right) \quad (1)$$

Fluid energy conservation relation:

$$\dot{m}_{in}h_{in} - \dot{m}_{out}h_{out} + \alpha V(T_w - \bar{T}) + V\frac{dP}{dt} = V\left[\bar{h}\left(\frac{d\bar{\rho}}{dP}\bigg|_{\bar{h}}\frac{dP}{dt} + \frac{d\bar{\rho}}{d\bar{h}}\bigg|_{p}\frac{d\bar{h}}{dt}\right) + \bar{\rho}\frac{d\bar{h}}{dt}\right] \quad (2)$$

Solid wall thermal energy balance:

$$\alpha_1 A_{w,1}(\bar{T}_1 - T_w) + \alpha_2 A_{w,2}(\bar{T}_2 - T_w) = \rho_w V_w c_{pw}\frac{dT_w}{dt} \quad (3)$$

Where $m_{in}$, $m_{out}$, $h_{in}$, and $h_{out}$ denote the respective fluid mass flow rates and enthalpies at the entry and exit cross-sections. The variables $\rho$, P, V, and \alpha represent the fluid density, working pressure, internal channel volume, and heat transfer coefficient, respectively. Lastly, $T_1$, $T_2$ *and* $T_w$ designate the mean values for the cold-fluid temperature, hot-fluid temperature and solid wall temperature.

Given the fast transient response dynamics exhibited by both the working fluid pump and the power expander, quasi-steady or localized transient models are deployed. The comprehensive transient mathematical architecture of the complete DPORC layout was fully delineated in an earlier publication [11]. Furthermore, the control logic governing the DPORC across all driving regimes was established in that same prior work [11] and remains fully integrated here. To confirm the prediction accuracy of this mathematical approach, a transient simulation of the ORC loop was built and calibrated against an actual laboratory ORC experimental installation previously engineered by the research group [9]. **Figure 2** displays the physical laboratory test bench for the ORC installation. Calibration of critical operational variables—specifically evaporation pressure and evaporation temperature—was conducted across a transient testing interval lasting approximately 750 s. The resulting correlation curves are presented in **Figure 3.** These findings reveal that the mean relative variances for the operating pressure and temperature remain below 4% and 0.89%, respectively, validating the high fidelity of the mathematical ORC framework.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

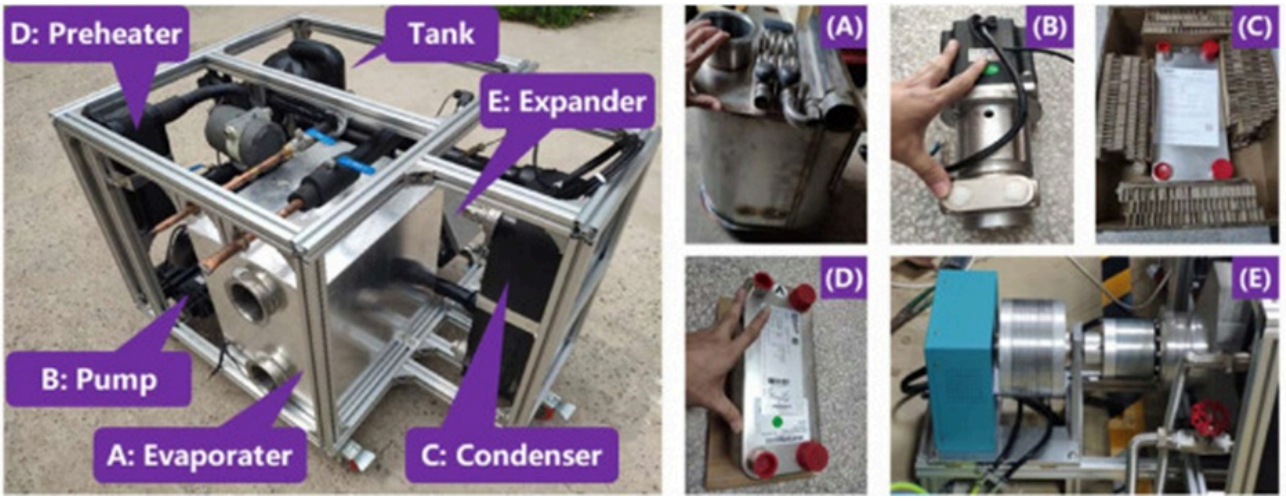



**Figure 2.** The ORC system experimental bench.

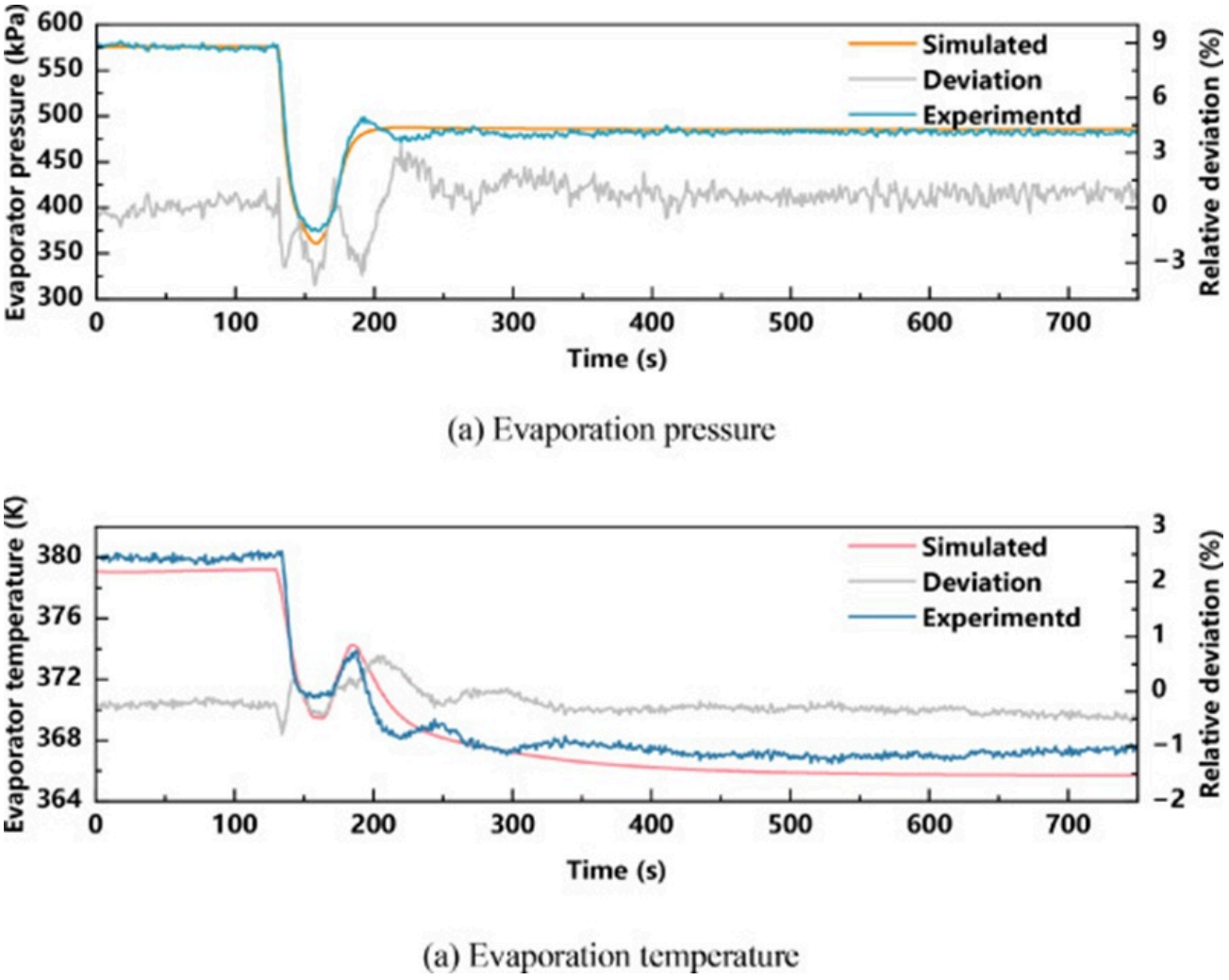



**Figure 3.** Calibration results of the ORC system.

Executing DRL algorithms demands massive processing capacity, which conflicts with the typically slow execution speeds of fully transient thermodynamic equations. Consequently, this research utilizes a surrogate neural network model built

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

using training data extracted from the full-range DPORC transient simulation. This neural network replacement was developed utilizing 61,000 s of operational data generated by the physical model. **Figure 4** highlights the input heat source properties ($T_{g_in}$, $T_{c_in}$, $T_{w_in}$, $m_{g_in}$, $m_{c_in}$, $m_{w_in}$) along with the corresponding net power generation ($W_{net}$) within the training set. The variable $W_{net}$ is determined via the following relationship:

$$W_{net} = W_{exp} - W_p \quad (4)$$

Where $W_{exp}$ represents the mechanical work produced by the expander, and $W_p$ is the electrical work consumed by the fluid pump. An extensive characterization of this specific formulation is documented in a previous report [38]. The underlying dataset spans the entire spectrum of possible internal combustion engine states. The fundamental architecture of this neural network substitute is shown in **Figure 5**, mapping the inlet heat source variables directly to the resulting $W_{net}$ output of the DPORC system. To evaluate the prediction capabilities of this neural network approach, its outputs were cross-examined against the full physical model over a 5000 s driving profile in **Figure 6**. The mean relative discrepancy between the neural network surrogate and the detailed transient model settled at 9.8%, which stays well within acceptable engineering tolerances for system-level simulations.

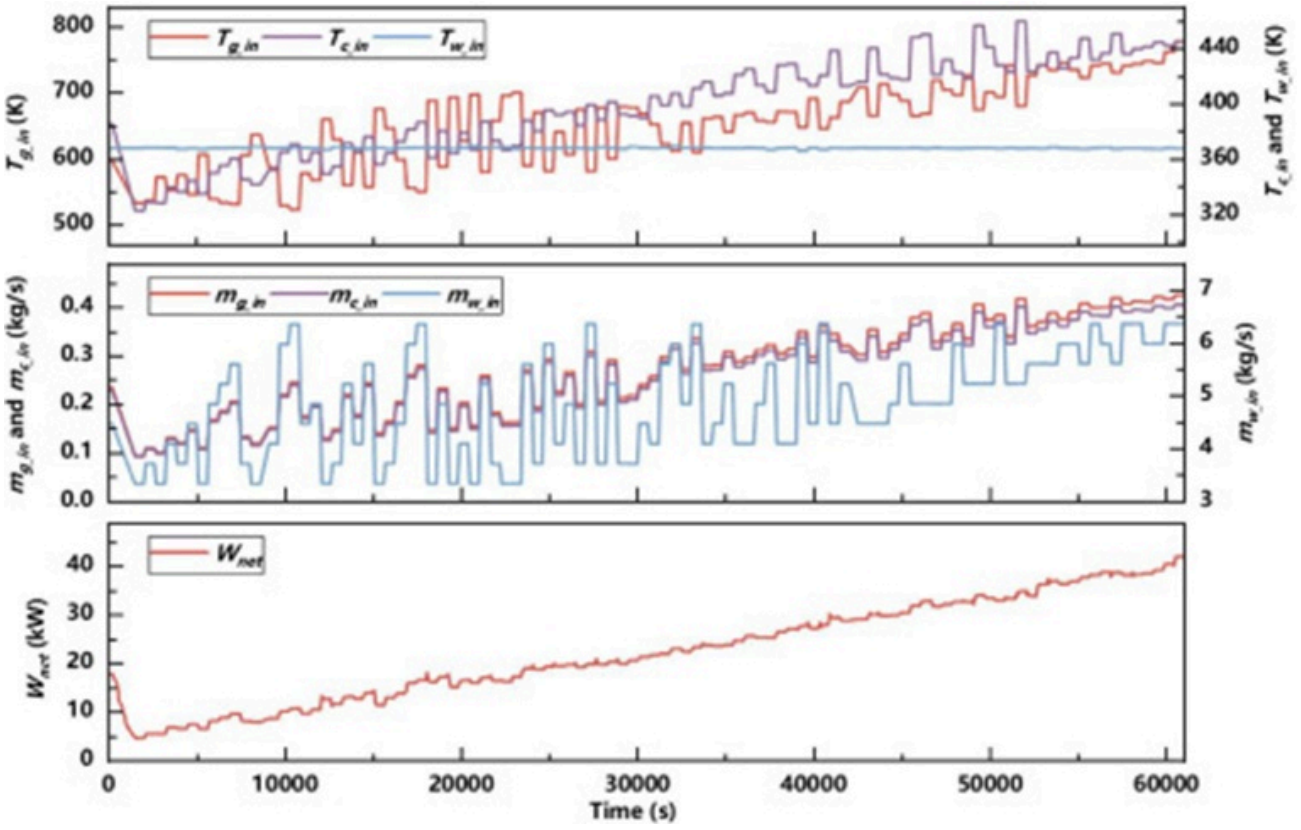



**Figure 4.** Simulation of the dynamic model at all working conditions for 61,000 s.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

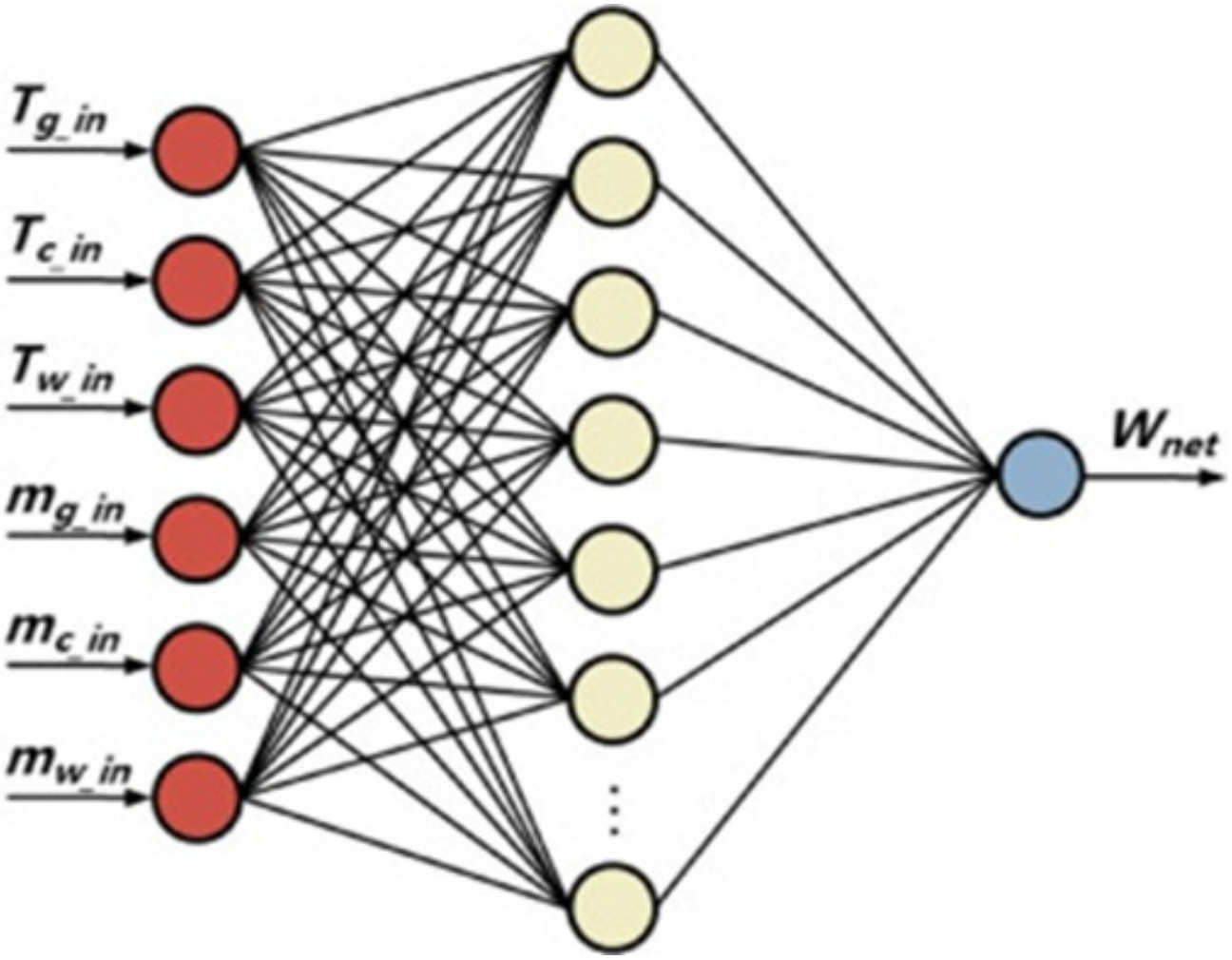
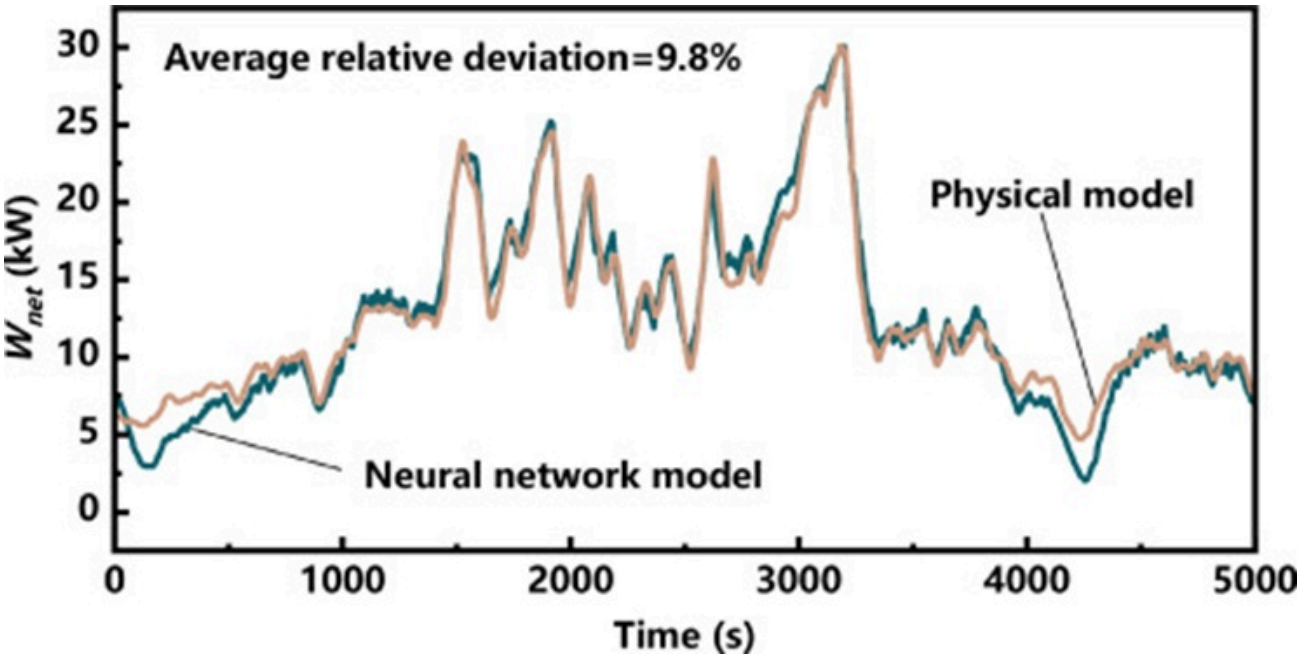



**Figure 5.** Schematic of neural network mathematical model.




© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

**Figure 6.** Comparison of neural network model and dynamic model simulation results.

# Energy Management Strategy

This section provides an analytical breakdown of both the standard rule-based EMS and the newly formulated Deep Reinforcement Learning Embedded Rule-Based EMS.

## Rule-based EMS

Traditional rule-based energy management strategies (EMS) represent the standard control approach deployed in production vehicles today. In this investigation, the rule-based EMS tailored for the HEV-DPORC architecture illustrated in **Figure 7** was formulated by consulting extensive literature [39-43]. Grounded in the optimal efficiency islands of the specific internal combustion engine and electric motor, this EMS follows the mainstream synthesis methodologies detailed in the referenced works.

The energy management system encompasses 6 distinctive operational states:

- Brake energy recovery / mechanical braking mode
- Pure electric motor drive mode
- Engine-only drive mode
- Combined hybrid drive mode
- In-motion driving charging mode
- Stationary parking mode

It coordinates balanced power distribution by dynamically toggling between these regimes based on real-time driving conditions and vehicle metrics. **Table 1** outlines the active operational boundaries and functional descriptions for each mode.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

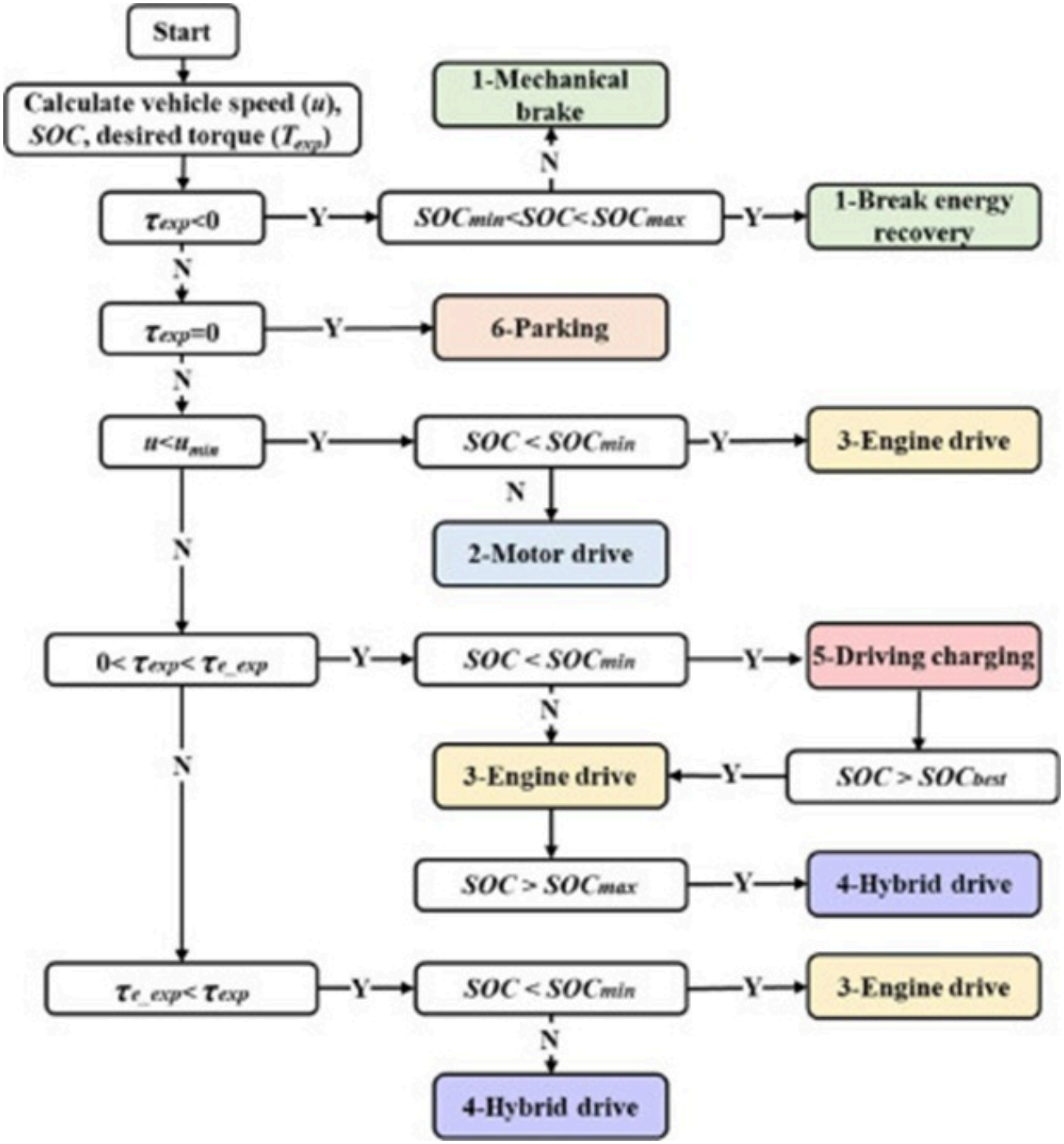



**Figure 7.** The logic of rule-based EMS.

**Table 1.** Description of the operating status in each operating mode.

| Mode | Applicable range | Description |
|---|---|---|
| 1- Break energy recovery | $\tau_{exp} < 0$<br>SOC< $SOC_{max}$ | $\tau_{eng}= 0$<br>$\tau_{motor}= \tau_{exp}$ |
| 2-Motor drive | $\tau_{exp} > 0$<br>$u < u_{min}$ | $\tau_{eng}= 0$<br>$\tau_{motor}= \tau_{exp}$ |

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

| | | |
|---|---|---|
| | SOC> $SOC_{min}$ | |
| 3-Engine drive | $\tau_{exp} \approx \tau_{e_exp}$ or SOC < $SOC_{min}$ | $\tau_{eng} = \tau_{exp}$<br>$\tau_{motor} = 0$ |
| 4-Hybrid drive | $\tau_{exp} > \tau_{e_exp}$<br>SOC > $SOC_{min}$ | $\tau_{eng} = \tau_{e_exp}$ or $\tau_{eng} = \tau_{exp} - \tau_{motor}$<br>$\tau_{motor} = \tau_{exp} - \tau_{e_exp}$ or $\tau_{motor} = \tau_{m_max}$ |
| 5-Driving charging | $0 < \tau_{exp} < \tau_{e_exp}$<br>$\tau_{exp} < \tau_{e_exp}$<br>SOC < $SOC_{min}$ | $\tau_{eng} = \tau_{e_exp}$<br>$\tau_{motor} = \tau_{exp} - \tau_{e_exp}$ |
| 6-Parking | $\tau_{exp} = 0$ | $\tau_{eng} = 0$<br>$\tau_{motor} = 0$ |

## Deep reinforcement learning embedded rule-based EMS

Deep reinforcement learning (DRL) operates as a goal-oriented framework, marrying the feature perception capabilities of deep learning with the strategic decision-making traits of reinforcement learning. An autonomous agent is trained to manipulate critical system variables using environmental input states and a targeted reward function to satisfy specific operational objectives.

**Figure 8** delineates the organizational framework of the deep reinforcement learning-embedded rule-based EMS introduced in this research. The primary target of the DRLERB EMS is minimizing the total cumulative equivalent energy consumption ($f_{equ}$). The training environment comprises the physical HEV-DPORC system. Here, the agent acquires states and reward feedback from the environment and transmits its calculated action—defined as the engine torque adjustment value—back to the rule-based core.

Subsequently, the rule-based EMS outputs ultimate torque split commands to the environment, synthesized from environmental states, the high-efficiency operating zones of the engine and motor, and the corrective action supplied by the DRL agent. Through iterative training loops, the agent identifies the parameter adjustment policy that maximizes the received reward.

In this study, the DRLERB EMS is designed for a parallel hybrid vehicle with a P2 configuration. For series, power-split (parallel-serial), or alternative parallel arrangements of HEVs, the baseline energy management strategy remains inherently rule-based. Consequently, the development methodology of this DRLERB EMS can be readily transferred to other hybrid topologies to optimize their respective rule-based modules, showcasing excellent versatility across diverse hybrid vehicle configurations.

The cornerstone of this DRLERB EMS is leveraging DRL to tune the fundamental parameters of a rule-based algorithm dynamically. Because rule-based systems have demonstrated reliable structural stability and operational resilience across real-world driving environments, the DRLERB EMS architecture retains these long-term safety and stability benefits. This enduring robustness is verified in subsequent sections under various traction profiles. The mathematical formulation of the DRLERB framework is detailed below.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

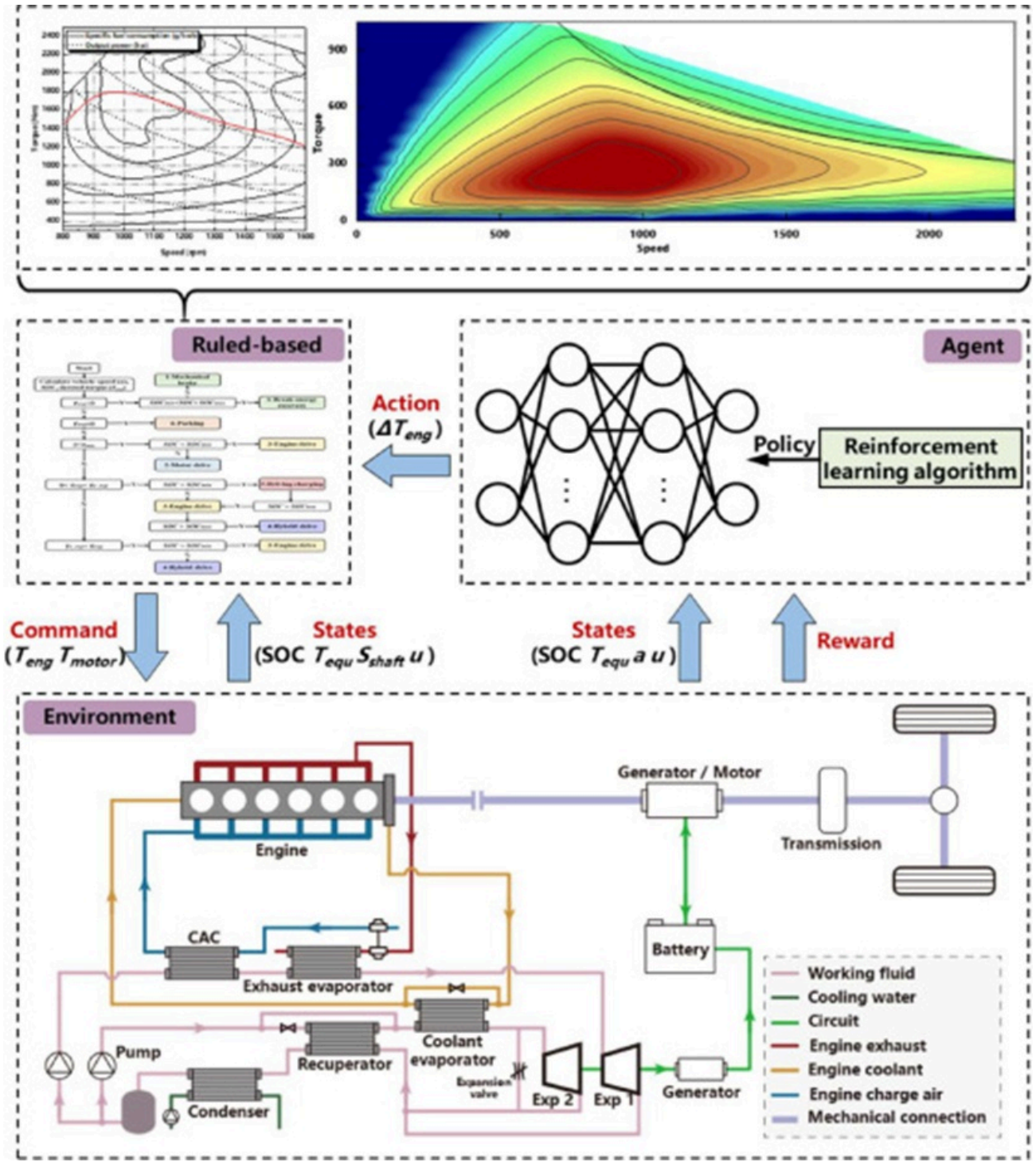



**Figure 8.** The logic of deep reinforcement learning embedded rule-based EMS.

## States, reward and action

The action variable within the DRLERB framework is defined as the engine torque modification value ($\Delta\tau_{eng}$). This action input is passed directly to the rule-based EMS, which then applies the modified torque command to the vehicle environment. The mathematical equation for the modified engine torque command is formulated as follows:

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

$$\tau_{eng} = \tau_{eng_0} + \Delta\tau_{eng} \quad (5)$$

$$\tau_{exp} = \tau_{eng} + \tau_{motor} \quad (6)$$

Where $\tau_{eng}$ represents the modified engine torque output, and $\tau_{eng_0}$ is the baseline torque computed by the rule-based EMS before adjustment. The corresponding electric motor torque $\tau_{motor}$ is derived by subtracting $\tau_{eng}$ from the gross vehicle demand torque $\tau_{exp}$, thereby finalizing the vehicle torque distribution sequence. For this investigation, the operational boundaries for $\Delta\tau_{eng}$ are restricted to the interval of [-300, 300].

The state vectors within the DRL algorithm must comprehensively depict the vehicle's instantaneous operational status. In this research, battery state of charge (SOC), vehicle torque demand $\tau_{exp}$, vehicle velocity (u), and longitudinal acceleration (a) are chosen as the active state metrics. The functional state variables utilized by the global EMS correspond directly with the inputs tracked by the learning agent. While the SOC captures the real-time energy reserve of the battery pack, the variables $\tau_{exp}$, u, and a quantify the immediate tractive demand of the truck, which dictates the high-efficiency operating zones for the engine and electric motor.

The reward function acts as the decisive element governing the efficacy of the DRL training process. Because the objective of any reinforcement learning mechanism is maximizing the total cumulative reward across an execution sequence, a highly effective reward function must reflect the primary operational targets of the EMS. In this work, the EMS seeks to minimize the cumulative energy consumption of both the motor and engine while concurrently stabilizing the battery SOC within a prescribed window. Consequently, the reward formulation incorporates both criteria.

Given that the underlying rule-based structure inherently bounds overall SOC fluctuations, the reward function prioritizes energy efficiency. The analytical calculation sequence for energy consumption is established as follows:

$$f_{equ,per} = \frac{P_{eng}}{\eta_{eng}} + \frac{P_{motor} - P_{DPORC} - P_{braking}}{\eta_{eng}\eta_{motor}} \quad (7)$$

$$f_{equ} = \int_0^{T_i} f_{equ,per}\,dt \quad (8)$$

Where $f_{equ,per}$ and $f_{equ}$ represent the instantaneous equivalent energy consumption rate and the cumulative equivalent energy consumption over an epoch $T_i$, respectively. The variables $P_{eng}$, $P_{motor}$, $P_{DPORC}$ and $P_{braking}$ denote the engine output work, electric motor output work, net mechanical work generated by the DPORC system and captured regenerative braking energy, respectively. Furthermore, $\eta_{eng}$ and $\eta_{motor}$ translate to the instantaneous operational efficiencies of the engine and motor, extracted dynamically from their respective empirical efficiency maps.

The global reward function comprises two distinct elements: a positive incentive for minimizing fuel consumption and a negative penalty for trailing outside the standard SOC boundaries. The mathematical evaluation of the reward function proceeds as follows:

$$R = \begin{cases} b - f_{equ,per} - c|SOC - SOC_{best}| & SOC > SOC_{\max} or SOC < SOC_{\min} \\ b - f_{equ,per} & SOC_{\min} < SOC < SOC_{\max} \end{cases} \quad (9)$$

Where b and c represent positive scaling constants; b ensures that the overall reward value R remains positive, thereby transforming the $f_{equ,per}$ minimization target into an R maximization problem. To constrain the battery within safe bounds,

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

an aggressive penalty is applied whenever the SOC drifts from its standard operating corridor.

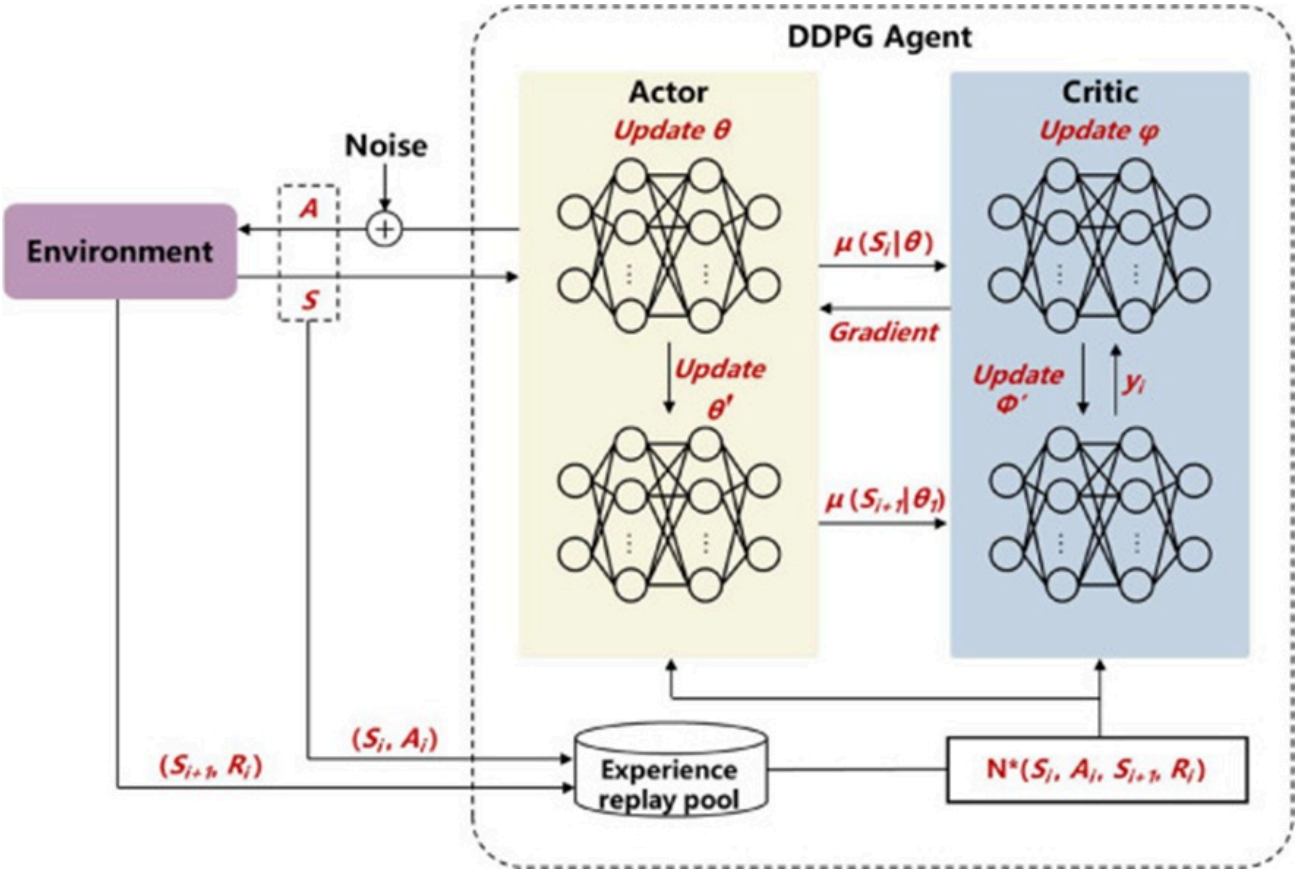
The relative weighting of the SOC penalty within the total reward function R can be tuned by altering c, making the calibration of b and c paramount for stable training convergence. Lower values of $f_{equ,per}$ directly correspond to higher values of R. Within this formulation, the maximum boundary $SOC_{max}$ is set to 0.8, and the minimum limit $SOC_{min}$ is defined as 0.3.

## Agent

The specific DRL architecture implemented in this research is the Deep Deterministic Policy Gradient (DDPG) framework introduced by Lillicrap *et al*. [44]. DDPG operates as a model-free, actor-critic algorithm engineered explicitly for continuous action environments. It serves as an advanced adaptation of Deep Q-Learning (DQN), expanding its capabilities into the continuous control domain.

During training execution, DDPG leverages an experience replay buffer and decoupled target networks; while the experience replay mirrors the DQN format, the target network update mechanism follows a different protocol. The experience replay memory logs historical transitions and samples them stochastically to break data correlation and maximize sample efficiency. To bolster learning stability, DDPG implements fixed target structures that undergo periodic, soft synchronization to minimize tracking target divergence.

**Figure 9** maps out the structural mechanics of the DDPG algorithm. The core DDPG engine maintains actor networks denoted by $\mu(S_i|\theta)$ and critic networks denoted by $Q(S_i, A|\varphi)$. To secure agent training stability, DDPG instantiates two corresponding target neural networks initialized with identical starting parameters for both the actor and critic, labeled as ($\mu'(S_{i+1}|\theta')$ and $Q'(S_{i+1}, A|\varphi')$), respectively. In this notation, A, S, and R signify the action, state, and reward elements, whereas $\theta$, $\theta'$, $\varphi$ and $\varphi'$ represent the underlying weight parameters of each respective network.




© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

**Figure 9.** The logic of the DDPG algorithm.

Table 2 shows the key parameters of the DDPG agent.

**Table 2**. Key parameters of the DDPG agent.

| Hyperparameter | Configured value |
|---|---|
| Reward discount coefficient | 0.99 |
| Critic network learning rate | $10^{-3}$ |
| Actor network learning rate | $10^{-4}$ |
| Number of hidden units in actor and critic networks | 100 |
| Target network soft-update factor | $10^{-3}$ |
| Replay memory capacity | $10^{6}$ |
| Mini-batch sample size | 256 |
| Exploration noise variance | 0.12 |
| Noise variance reduction rate | $10^{-5}$ |

# Rapid Control Prototype Bench Construction and Testing

Rapid control prototyping (RCP) represents a vital milestone in the development lifecycle of advanced control frameworks. By pairing a virtual computing controller with physical hardware components, RCP establishes a comprehensive semi-physical hardware-in-the-loop simulation environment. This configuration enables rigorous verification of the execution reliability and accuracy of control algorithms under real-world physical constraints.

To confirm the operational performance of the proposed DRLERB EMS, a scaled-down semi-physical hardware platform was constructed around the parameters of the target truck. Figure 10 illustrates the global architecture of this semi-physical simulation platform. The setup is segregated into four main modules:

- The hybrid powertrain test bench
- The hardware actuation system
- A real-time simulation processor
- A supervisory control and data acquisition (SCADA) interface

The physical hybrid powertrain bench integrates 3 separate electric machines and 1 mechanical clutch coupling. Specifically, Motor 1 acts as the engine emulator; Motor 2 is configured as a dual-extension-shaft machine; and Motor 3 operates as a transient dynamometer tasked with applying programmable road load resistances. The mechanical clutch assembly is positioned intermediate to Motor 2 and Motor 3.

The supporting hardware control system bundles variable-frequency drives for all **3** electric machines, specialized communication interfaces to connect with the main simulator (via an EtherCAT card), and necessary auxiliary accessories. The central real-time simulator compiles and executes the DRLERB EMS alongside the mathematical representations of the remaining vehicle components. The SCADA node handles compiling and downloading the control models into the real-time processor while supervising the active telemetry of the entire installation.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

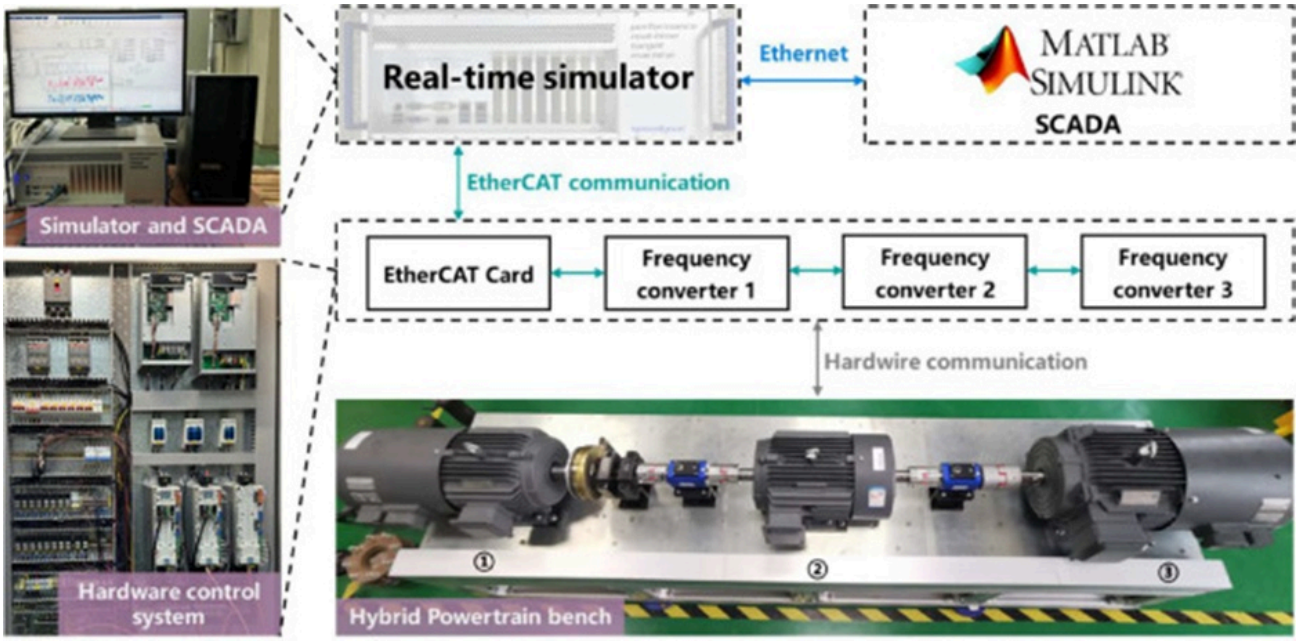



**Figure 10.** Semi-physical simulation system architecture (Real-time simulator, Hybrid Powertrain bench, Hardware control system, SCADA).

The real-time simulator is linked in a series daisy-chain configuration to the EtherCAT communication card and the individual frequency converters within the hardware control enclosure. Concurrently, high-speed Ethernet lines facilitate data exchange between the real-time simulator and the SCADA workstation. The physical motor control signals are distributed from the drive enclosure directly to the hybrid test bed via industrial hardwired connections, completing the signaling pathways across all subsystems.

**Figure 11** maps out the step-by-step communication initialization protocol spanning the SCADA workstation, real-time processor, and physical hybrid powertrain bench. This configuration sequence involves the following actions:

1. Define a new EtherCAT project topology within the TwinCAT3 software suite on the SCADA machine.

2. Pair the master interface of the EtherCAT interface board to a designated communication port on the SCADA PC that supports native EtherCAT protocols.

3. Scan the active hardware line to automatically discover the physical I/O blocks, configure their parameters, and export the topology as a standardized XML configuration file.

4. Map out and parameterize the network interfaces connecting the EtherCAT card to the variable-frequency motor controllers.

5. Integrate the structured XML device configuration file directly into the Simulink model workspace.

6. Set up data acquisition pipelines, establish control loop routines, and run baseline validation passes.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

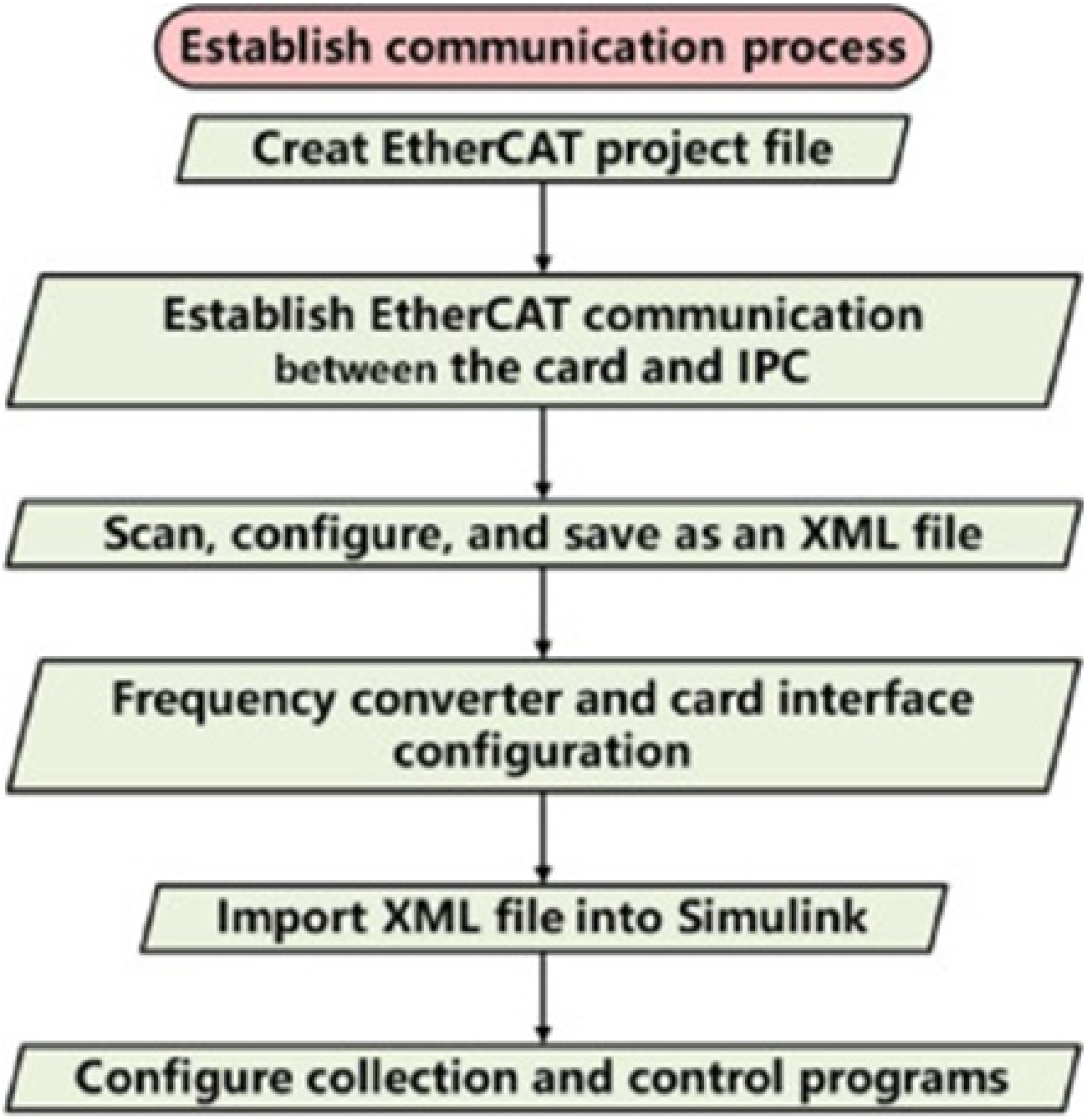



**Figure 11.** The process of establishing communication.

Following successful hardware installation and signal validation, primary system diagnostic testing was initialized**.** **Figure 12** tracks the bidirectional signal paths active between the physical hybrid test bed and the real-time simulator core. Operating as the core decision-making unit, the EMS hosted within the simulator functions as the system “brain,” orchestrating the behavior of the real-world powertrain bench and all virtual simulator blocks. Both the engine and motor emulators are set to torque-control mode, where the EMS routes instantaneous torque demands directly to the bench drives.

In return, the bench motors feed real-time velocity and torque measurements back into the simulator to calculate system states. Simultaneously, the absorption dynamometer reproduces real-time load torques, encompassing both inertial and

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

non-inertial resistance forces.

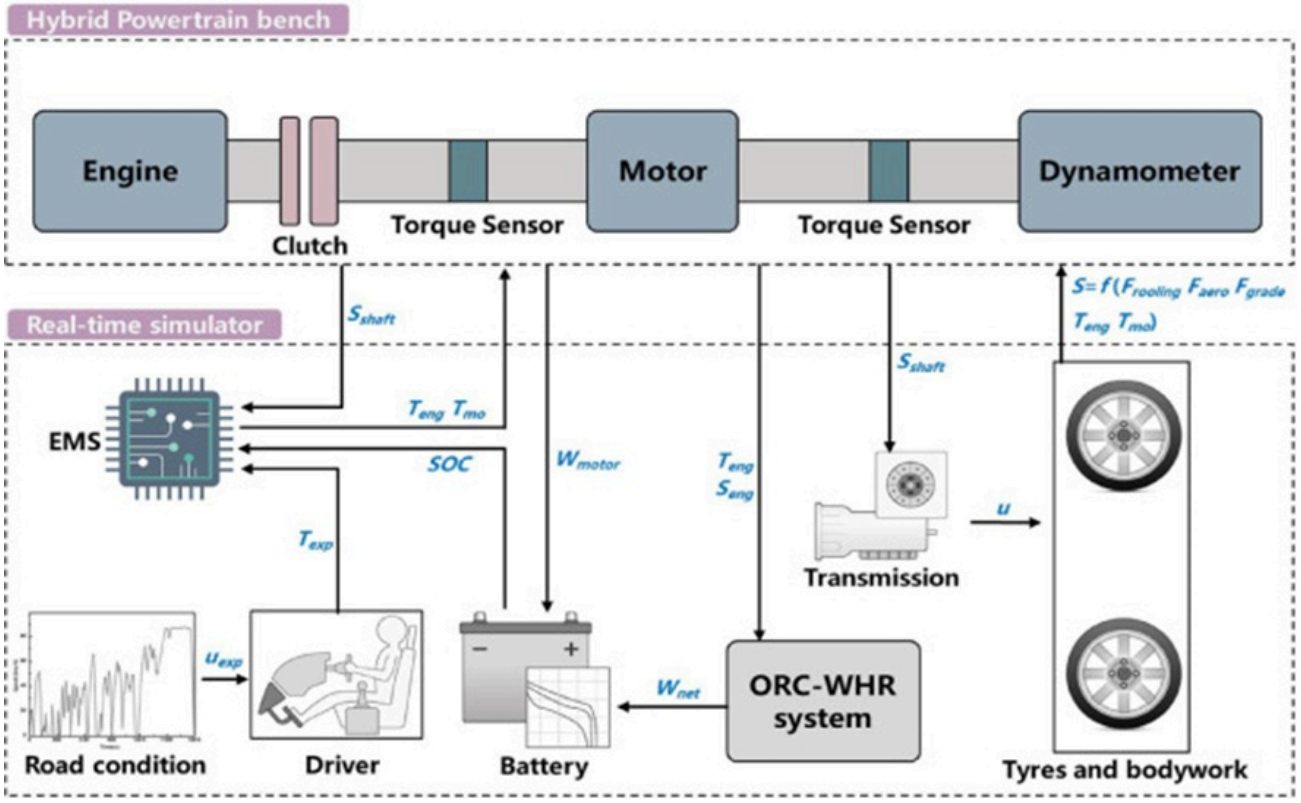



**Figure 12.** Schematic of signal transmission between Hybrid Powertrain bench and Real-time simulator during the HEV-DPORC system rapid control prototype experiment.

Next, functional evaluations were performed across 4 principal hybrid operating modes: motor-only propulsion, hybrid drive, engine-only propulsion, and in-motion battery charging. **Figure 13** compiles the empirical observations recorded across these 4 states. Each discrete mode was sustained for a 30 s evaluation window under dynamic variations in rotational speed and torque, showing smooth and stable system responses.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

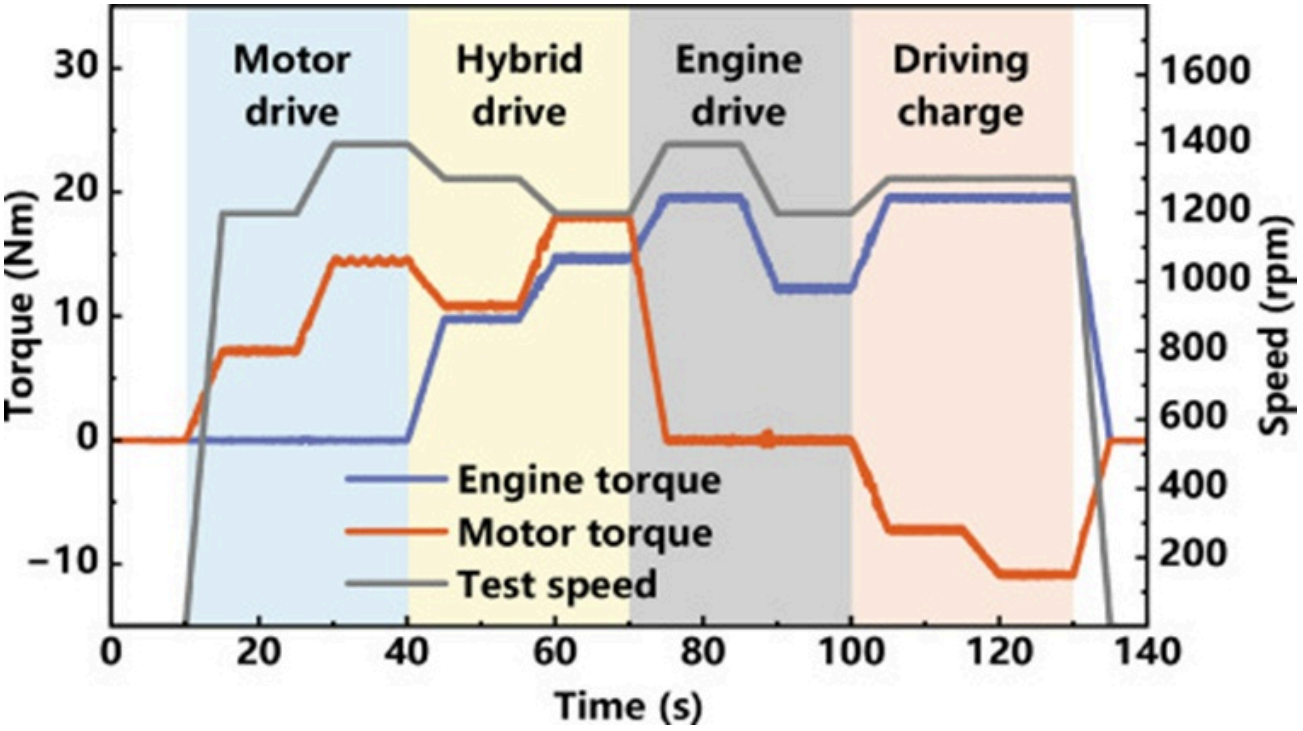



**Figure 13.** Testing of HEV system in various operating modes.

# Results Analysis

This section provides a comparative evaluation between the baseline rule-based EMS and the proposed DRLERB EMS under both trained driving schedules and entirely unfamiliar road profiles. Additionally, the practical viability of the DRL methodology is confirmed through physical RCP experimental runs.

## DRL training process and simulation result analysis

Training of the HEV-DPORC energy management strategy was conducted over the standard C-WTVC driving schedule. **Figure 14** records the target speed profile characterizing the C-WTVC test route, which incorporates a comprehensive mix of urban, suburban, and highway driving segments.

**Figure 15** tracks the progression of the EMS training sequence. The predefined benchmark target reward was set at 610.5. As illustrated in **Figure 15**, the learning curve successfully converges around the 80th episode. By way of comparison, the standalone DRL architecture implemented by Wang *et al*. [30] to optimize an HEV-DPORC controller without external constraints required roughly 1100 episodes to reach convergence. Because the learning agent in this study is guided by the direct intervention of a rule-based EMS structure, convergence happens rapidly, ultimately attaining a peak reward score of 611.07 at the 285th episode.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

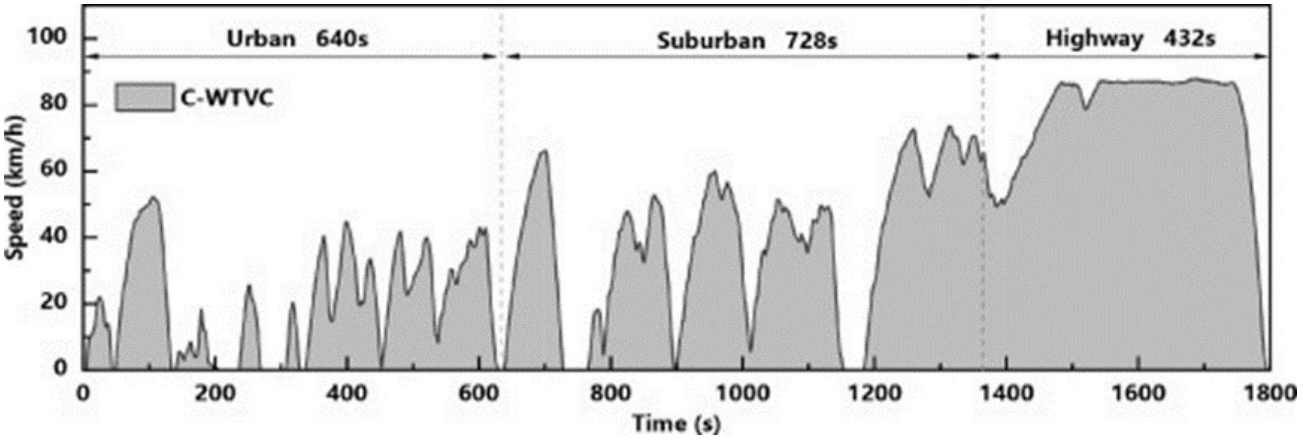



**Figure 14.** C-WTVC driving cycle.

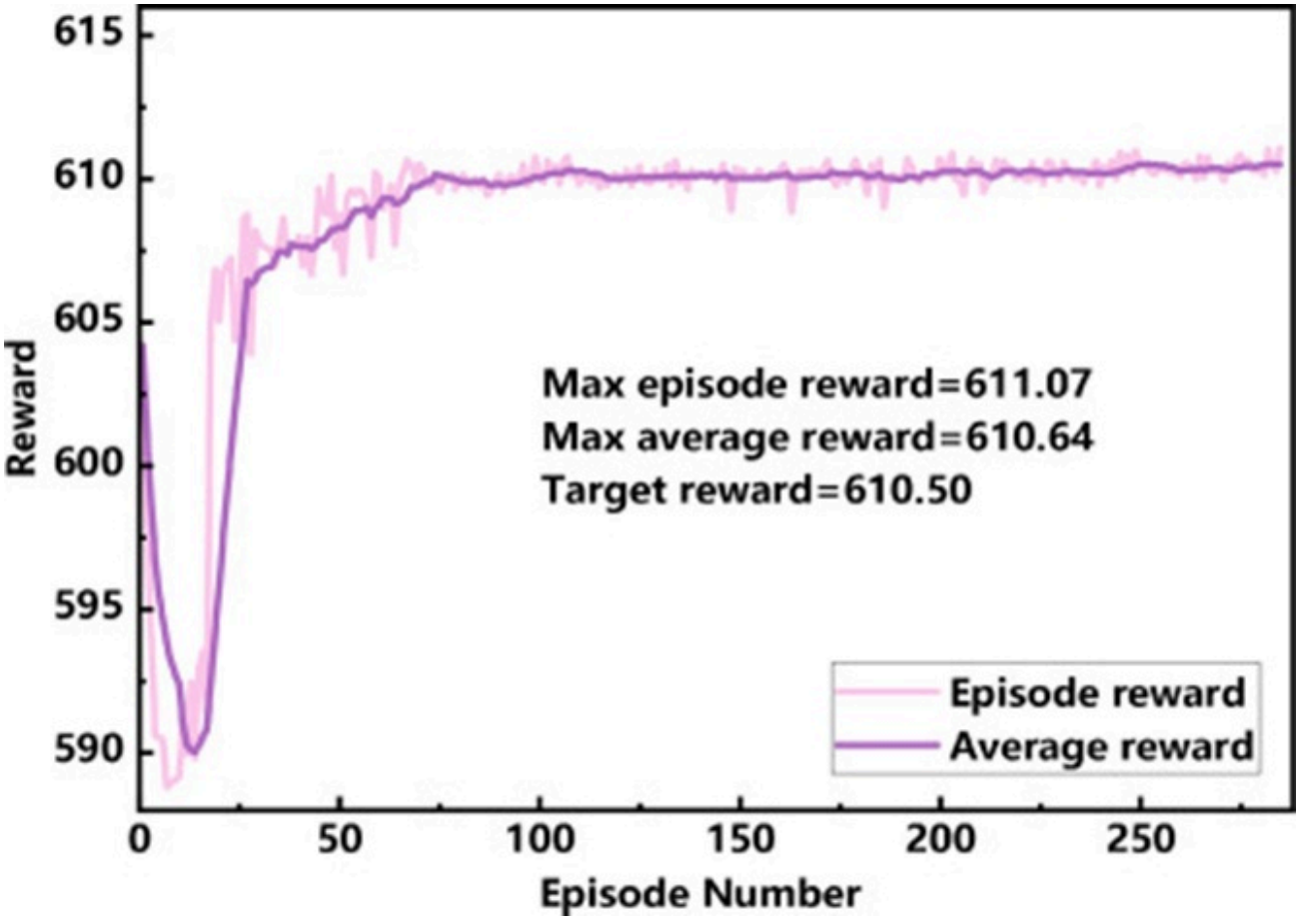



**Figure 15.** The DRLERB EMS training process.

Following this, the performance outputs and primary operational metrics of the standard rule-based controller and the DRLERB EMS are compared across the C-WTVC driving profile. **Figure 16** captures the transient profiles of these critical parameters under both control frameworks.

As plotted in **Figure 16a**, both strategies closely track the target velocity trajectory, proving the physical validity of the integrated HEV-DPORC system model parameters. **Figures 16b and 16c** map out the dynamic torque trends for the engine and electric machine, respectively. These results clarify how the DRL layer corrects the baseline rule-based

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

commands under the DRLERB EMS; notably, the torque modifications are more pronounced during medium-to-high engine load periods, primarily manifesting as a reduction in internal combustion engine strain.

**Figure 16d** details the state-of-charge (SOC) profiles of the battery pack. Under DRLERB EMS control, the operational load factor of the electric motor is elevated during suburban and highway driving windows relative to the standard rule-based baseline. Conversely, within urban zones, the battery SOC shares a highly consistent trend between both management strategies, implying that the DRL module focuses its torque redistribution efforts mostly across urban and suburban driving periods.

**Figure 16e** plots the cumulative equivalent fuel consumption (f_{\text{equ}}) curves, proving that the DRLERB EMS yields a 4.31% reduction in total energy consumption compared to the conventional rule-based controller. This conservation benefit is predominantly realized within suburban and highway sections. Comprehensive efficiency breakdowns for each component are provided in the subsequent analysis.

Lastly, **Figure 16f** displays the net power output (W_{\text{net}}) generated by the DPORC loop; because the DRL algorithm actively shifts the operational points of both the engine and motor, the corresponding thermodynamic input conditions driving the DPORC loop are modified accordingly.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

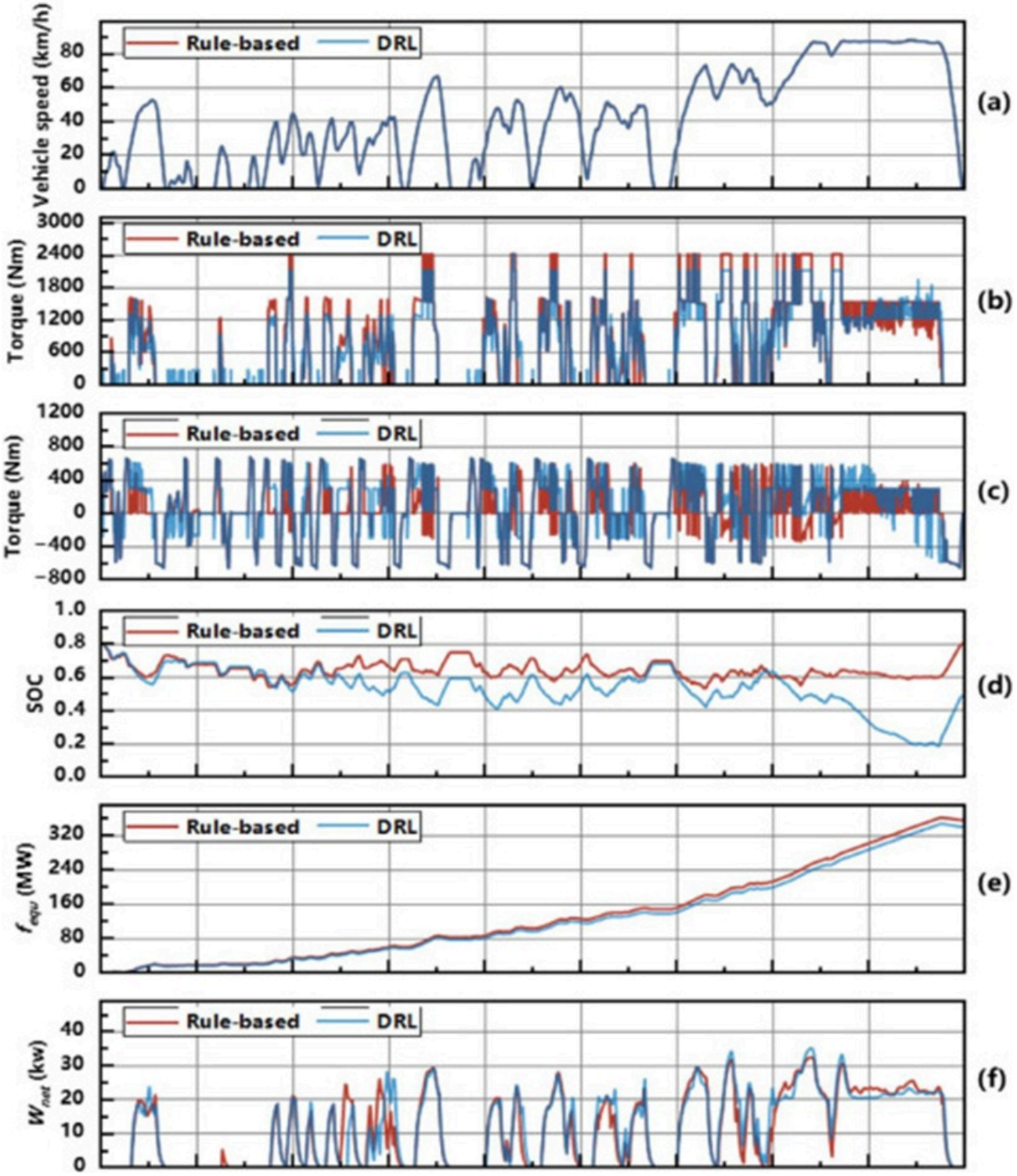



**Figure 16.** The dynamic characteristics of the key parameters under the C-WTVC driving cycle: (a) Vehicle speed; (b) Engine torque; (c) Motor torque; (d) Battery SOC; (e) Equivalent energy consumption; (f) DPORC system net output power.

Following this, an assessment of the energy conservation capabilities of individual components is conducted. **Table 3** details the energy utilization for each component utilizing both energy management strategies. The rule-based EMS

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

serves as the standard controller for HEV-DPORC architectures. **Figure 17** illustrates the enhanced energy-saving capabilities of the DRLERB EMS relative to the traditional rule-based approach.

The computations presented in **Table 3** and **Figure 17** are derived utilizing **Eqs. 7 and 8.** Since the net mechanical work delivered by the powertrain remains identical across both EMS configurations, the underlying model is proven logical, ensuring a fair comparative analysis. In contrast to the standard rule-based EMS, implementing the DRLERB EMS caused a rise in the electric motor's energy usage alongside a drop in the internal combustion engine's energy demand. At the same time, the net work $W_{net}$ generated by the DPORC loop remained largely unchanged. Collectively, this dynamic produced a 4.31% drop in total equivalent energy usage during the C-WTVC driving cycle. Overall powertrain efficiency climbed from 55.24% to 57.72%.

In summary, the DRL framework successfully refines the vehicle's torque allocation over the C-WTVC cycle, sequentially minimizing overall fuel consumption. This decline in equivalent energy demand directly translates to substantial reductions in CO2 output. Research by Coronado *et al*. [45] indicates that combusting 1 $m^3$ of diesel fuel within a commercial truck yields 2.683 tons of CO2. The annual CO2 mitigation for a truck utilizing the proposed DRLERB EMS is evaluated using the corresponding formula:

$$E_{CO2} = 2.683 \frac{f_{rule} - f_{DRL}}{\rho Ca} T_{year} \quad (10)$$

Where $f_{rule}$ and $f_{DRL}$ denote the mean equivalent energy consumption rates for the rule-based and DRLERB EMS setups, respectively. The parameter Ca represents diesel's heating value, established at 42,600 kJ/kg, while \rho signifies the diesel fluid density, recorded as 830 kg/ $m^3$. The variable $T_{year}$ defines the annual operational hours for the truck, assumed here to be 3000 h. Computations reveal that under the C-WTVC protocol, a single truck deploying the DRLERB EMS can eliminate 6.96 tons of CO2 emissions annually relative to the baseline rule-based controller.

**Table 3.** Energy consumption parameters of power source under rule-based and DRLERB EMS.

| Energy management method | System efficiency | Effective output power (kW) | Total equivalent energy consumption (kW) | Average equivalent energy consumption – DPORC (kW) | Average equivalent energy consumption – Motor (kW) | Average equivalent energy consumption – Engine (kW) |
|---|---|---|---|---|---|---|
| Rule-based strategy | 55.24% | 108.93 | 197.20 | -24.65 | 20.81 | 201.05 |
| DRLERB approach | 57.72% | 108.93 | 188.71 | -25.27 | 23.56 | 190.42 |

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

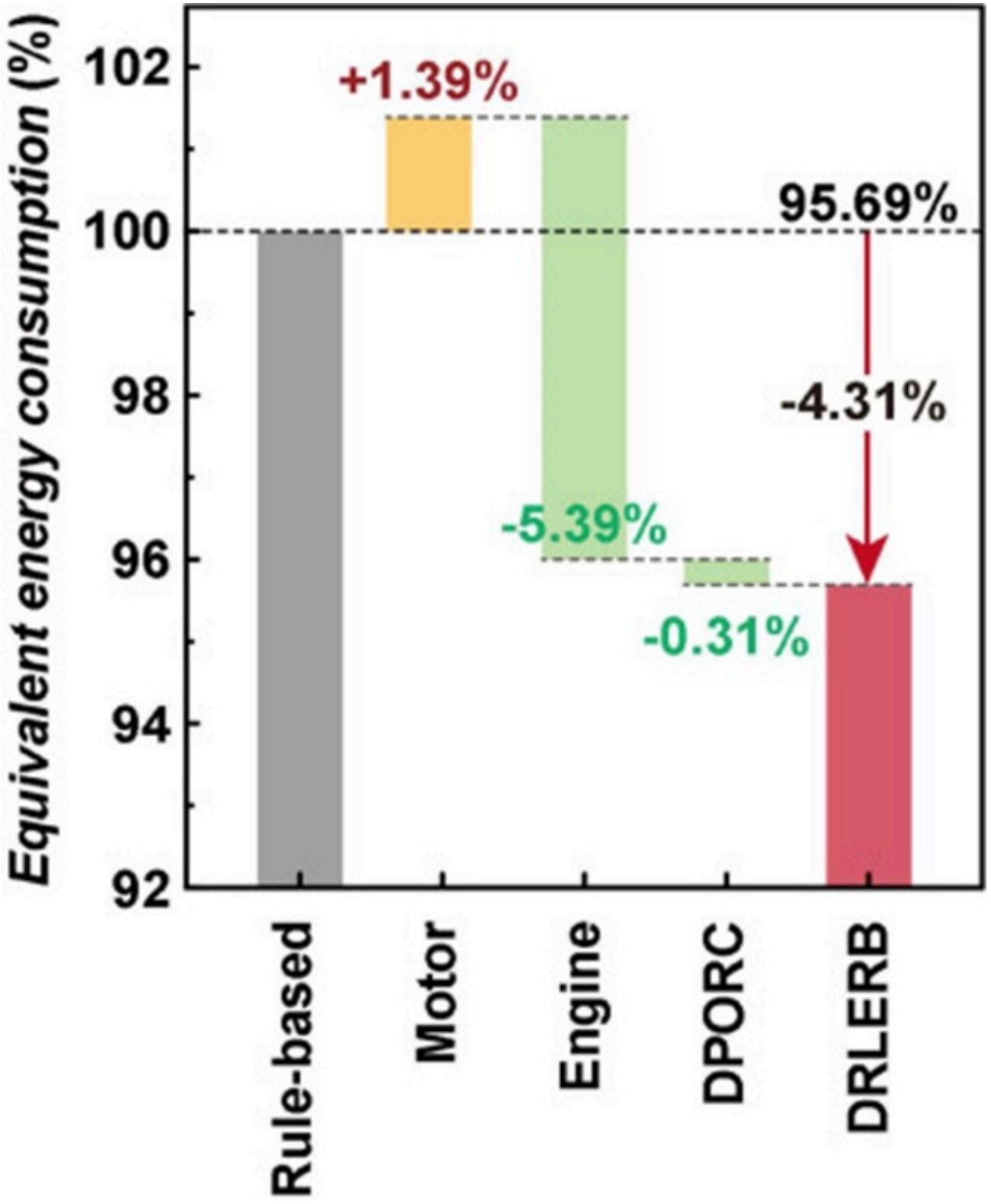



**Figure 17.** Analysis of energy saving potential of DRLERB EMS compared to rule-based EMS.

## Driving condition adaptability analysis

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Because the C-WTVC profile encompasses a blend of urban, suburban, and highway environments, the fully trained DRLERB EMS should theoretically translate effectively to alternative driving schedules. Consequently, the performance of the DRLERB EMS is evaluated across several additional routes.

**Figure 18** illustrates a comparative analysis of the equivalent energy consumption between the DRLERB and standard rule-based strategies across four unencountered driving profiles. The equivalent energy usage dropped by 3.63%, 4.72%, 2.55%, and 3.19% throughout the CHTC, WLTC, HWY, and ARTEMIS test cycles, respectively. These results confirm the strong adaptability of the DRLERB EMS when subjected to entirely novel operational environments. **Figure 19** tracks the fluctuations in the battery's state of charge under these four untrained route profiles. The data demonstrates that the battery SOC remains consistently within safe operational limits across every driving cycle, verifying the robust nature of the DRLERB EMS when navigating untested scenarios.

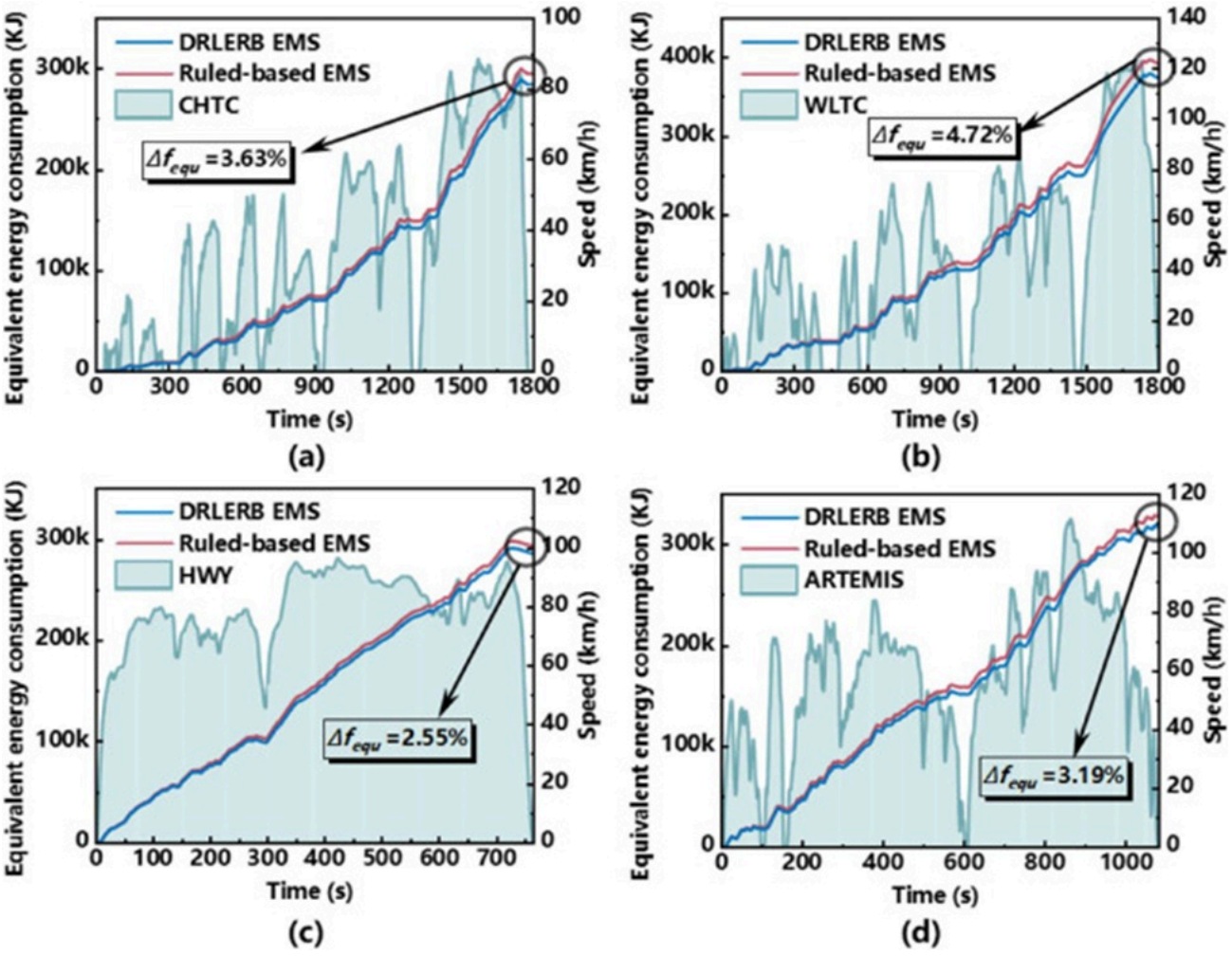



**Figure 18.** Comparison of equivalent energy consumption of DRLERB and Rule-based EMS under other driving cycles: (a) CHTC; (b) WLTC; (c) HWY; (d) ARTEMIS.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

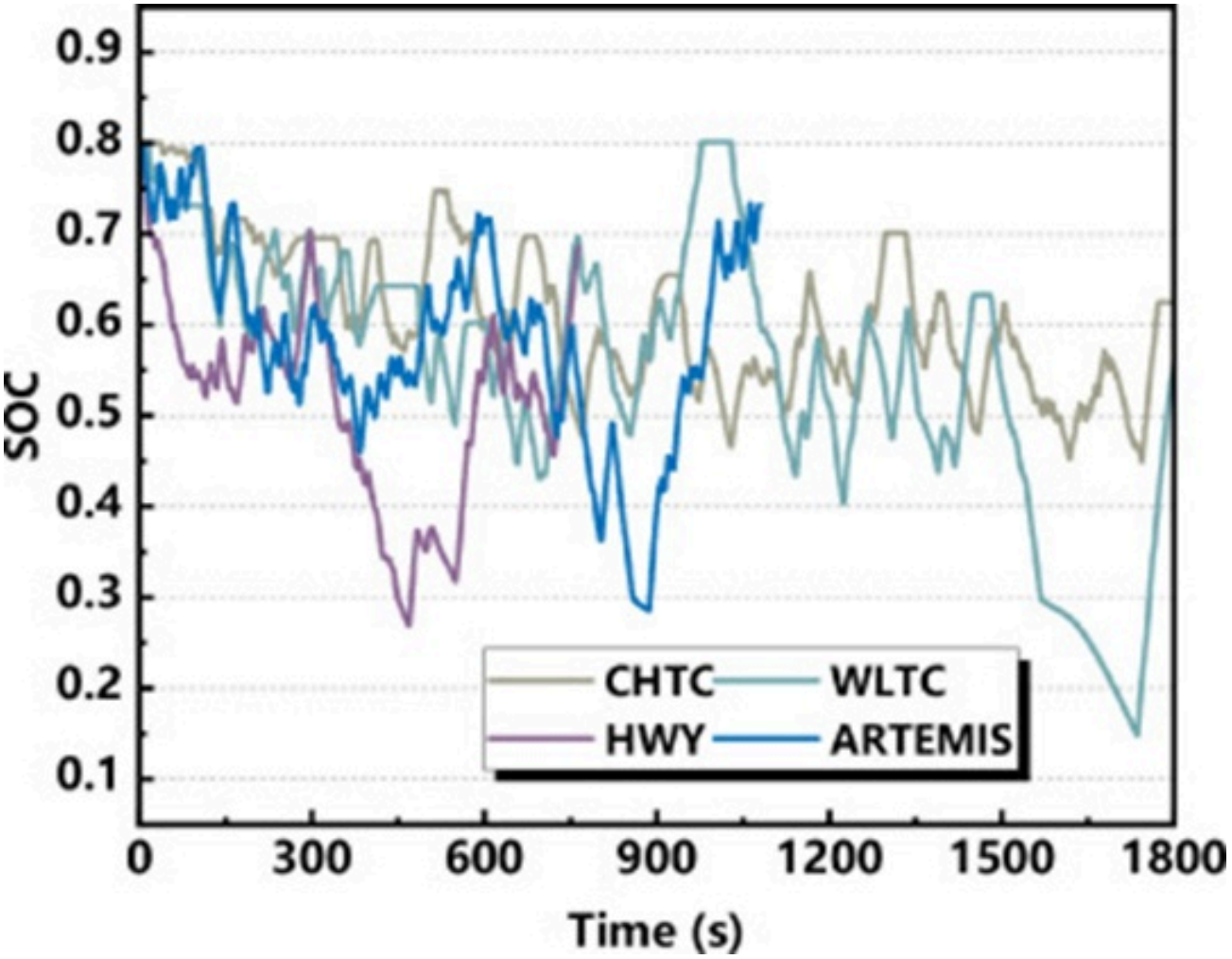
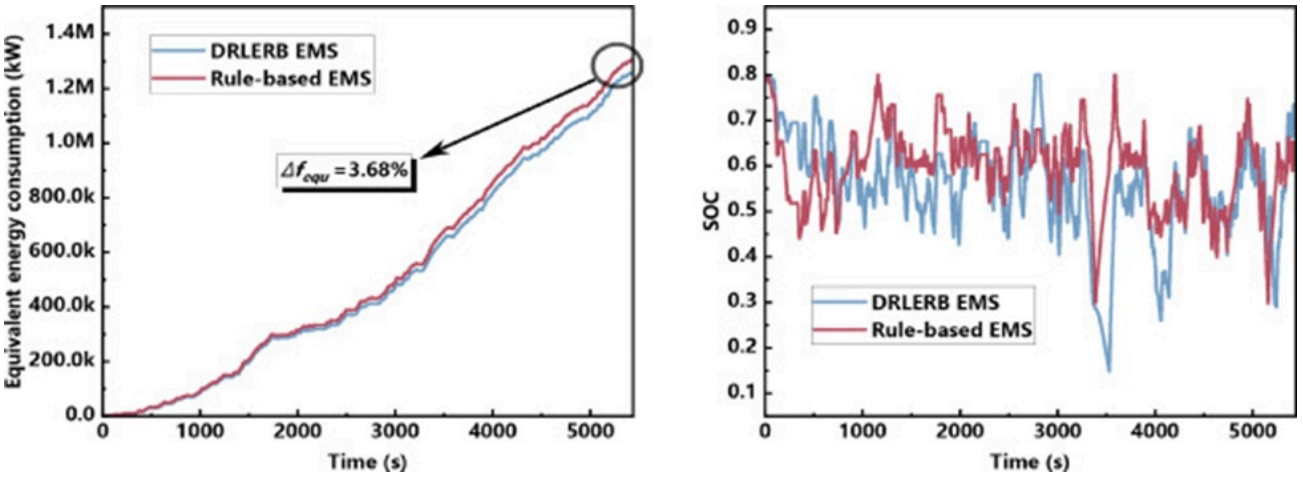



**Figure 19.** Battery SOC changes under four driving cycles.

To further confirm the sustained stability and long-haul adaptability of the DRLERB controller, the four independent traction cycles outlined in **Figure 18** were spliced together to execute an extended comparative simulation against the rule-based EMS. **Figure 20** plots the battery SOC and cumulative energy demand spanning a continuous 5440 s interval. These findings demonstrate that the DRLERB EMS maintains reliable stability and adaptive control over prolonged operating durations.




© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

**Figure 20.** Comparison of equivalent energy consumption and SOC of DRLERB and Rule-based EMS over 5440 s.

## RCP experimental results analysis

The real-world viability of the DRLERB EMS was subsequently evaluated utilizing the RCP hardware platform. Because native reinforcement learning toolboxes in Simulink currently lack direct code generation support for real-time compilation, the code generation process in this investigation was achieved by feeding the trained scripting files and output data directly into a MATLAB Function block. Consequently, the RCP empirical trials exhibit a slight margin of error when contrasted against the pure software simulation. The hardware-in-the-loop RCP test was executed over the C-WTVC profile. **Figure 21** displays the comparative alignment of core metrics between the simulation model and the physical experiment.

The testing procedure adheres to the methodology specified before. The empirical speed and torque telemetry depicted in **Figure 21** stems directly from the physical hybrid test bed, with the torque recordings scaled proportionally to match the full-size vehicle. As observed in **Figure 21a**, the divergence between the physically measured vehicle spindle speed and the simulated velocity is negligible. **Figures 21b and 21c** map the simulated versus experimental torque responses for the internal combustion engine and electric motor, respectively. A marginal discrepancy exists within the torque splitting behavior, primarily because the DRL algorithm's code generation relies on the MATLAB Function workaround during live testing. Nevertheless, the individual torque outputs for both prime movers align strongly with the theoretical simulation profiles.

In conclusion, the physical RCP validations utilizing the real-time simulation processor and hybrid powertrain bench confirm that the newly proposed DRLERB EMS is capable of seamless real-time execution and precise regulatory control over all hybrid powertrain subsystems. Given its successful handling of computational loads, data transmission demands, and real-time processing constraints, this framework holds genuine promise for direct integration into commercial HEVs. This research effectively accelerates the transition of DRL technologies into practical HEV energy management, confirming both its utility and real-world viability. From a commercial perspective, the aforementioned tests reveal that the DRLERB EMS requires no supplementary physical hardware; engineers merely need to flash the control code derived from the trained DRL agent directly into the existing vehicle control unit to upgrade the EMS. Thus, the DRLERB EMS achieves substantial fuel conservation at a highly attractive cost-benefit ratio.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

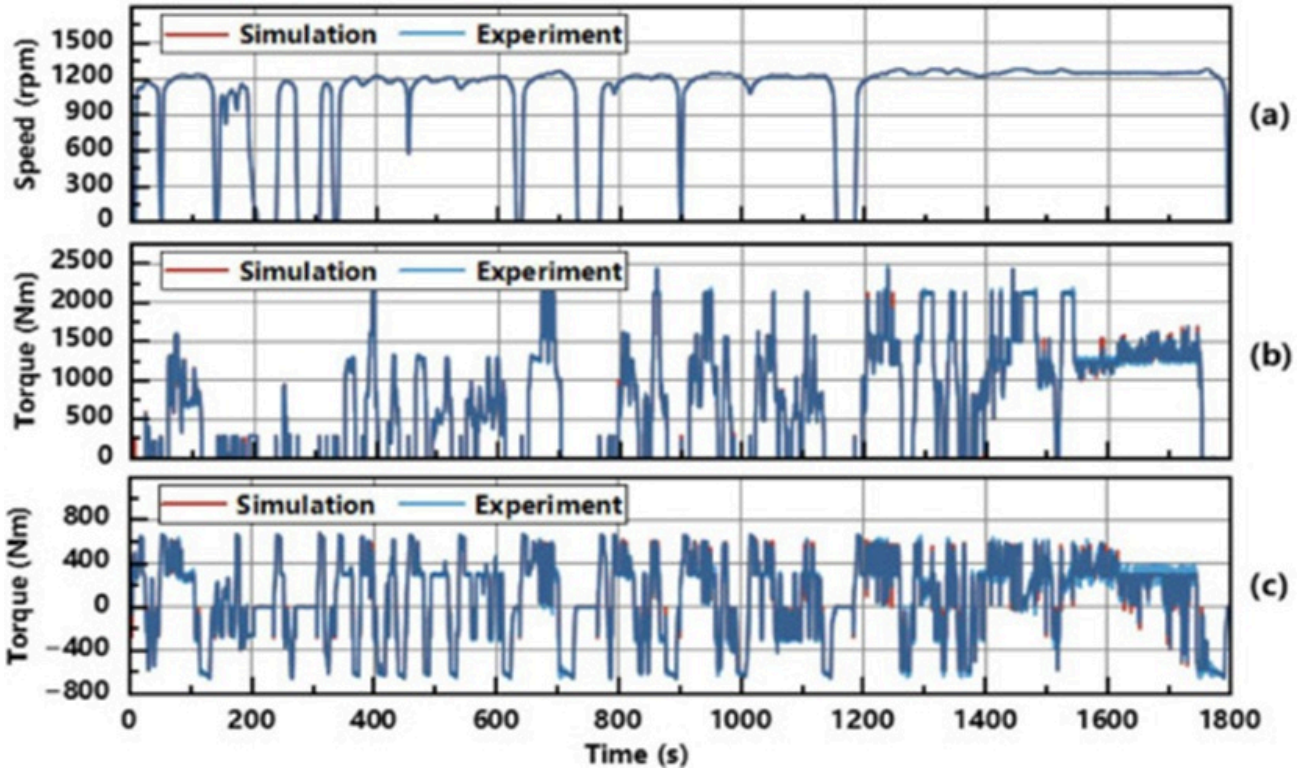



**Figure 21.** Comparison of simulated and experimental dynamic characteristics of the key parameters under C-WTVC driving cycle: (a) Vehicle spindle speed; (b) Engine torque; (c) Motor torque.

# Conclusion

To engineer a highly efficient, commercially viable energy management strategy for hybrid commercial vehicles equipped with dual-pressure organic Rankine cycle waste heat recovery, this investigation introduced a control framework that directly embeds deep reinforcement learning within a rule-based foundation. To validate the structural integrity and operational logic of this approach, a comprehensive transient model of the integrated HEV-DPORC architecture and the corresponding DRLERB EMS was formulated to conduct extensive simulation evaluations. Following the theoretical analysis, a physical hybrid powertrain hardware-in-the-loop test platform was assembled to execute rapid control prototyping experiments. The primary findings are summarized as follows:

(1) Throughout the C-WTVC driving schedule, the DRL agent successfully stabilizes and finalizes the tuning of critical EMS parameters in approximately 80 training episodes. Following this training phase, the fully calibrated DRLERB EMS reduces fuel consumption by 4.31 % relative to the conventional rule-based counterpart over the same traction cycle.

(2) When subjected to unencountered CHTC, WLTC, HWY, and ARTEMIS route profiles, the DRLERB EMS consistently maintains safe vehicle operation. This is primarily because the embedded architecture preserves the inherent structural stability of the baseline rule-based controller. Furthermore, the DRLERB EMS successfully lowered overall fuel usage across all four of these unfamiliar test cycles.

(3) Physical rapid control prototyping validations were successfully executed by pairing a real-time computational simulator with a scaled hybrid powertrain test bed. The physical telemetry closely mirrored the theoretical software

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

simulations. This alignment confirms that the proposed DRLERB EMS can govern physical hybrid powertrain hardware, verifying its readiness for transition into production vehicles.

Ultimately, this research presented the DRLERB EMS architecture and successfully validated it via semi-physical hardware experiments. To bridge the gap toward full-scale industrial deployment in physical vehicles, future efforts will focus on porting the trained DRL agent directly onto commercial automotive control units to regulate the physical hybrid test bed. Concurrently, further investigations will assess whether standard factory sensor suites can adequately supply the telemetry resolution demanded by the DRLERB EMS, with the ultimate goal of finalizing a turn-key DRLERB control package suitable for mass production.

## Nomenclature

$a$ = Acceleration (m/s$^2$)

$A$ = Action

$Ca$ = Battery capacity (Ah)

$F$ = Fuel consumption

$F$ = Force (N)

$I$ = Current (A)

$m$ = Mass flow rate (kg/s)

$P$ = Pressure (kPa)

$R$ = Reward/Resistance (Ω)

$S$ = State

$T$ = Temperature (°C)

$U$ = Voltage (V)

$W$ = Power output (W)

$H$ = Efficiency (%)

$\tau$ = Torque (Nm) Subscripts

*bat* = Battery

*c* = Engine cooling water

*ca* = Charge air

*e* = Engine

*equ* = equivalent

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

*exp* = Expender / Expectation

*g* = Exhaust gas

*in* = Inlet

*net* = Net worth

*out* = Outlet

*p* = Working fluid pump

## Abbreviation

DDPG = Deep deterministic policy gradient

DP = Dynamic programming

DPORC = Dual pressure organic Rankine cycle

DRL = Deep reinforcement learning

DRLERB = Deep reinforcement learning embedded rule-based

EMS = Energy management systems

HEV = Hybrid electric vehicle

HIL = Hardware-in-the-loop

MPC = Model predictive control

Inlet = Organic Rankine cycle

RCP = Rapid control prototype

SCADA = Supervisory control and data acquisition

WHR = Waste heat recovery

# Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFE0100100).

# Conflict of interest

None

# Financial support

None

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

# Ethics statement

None

Received: 19 Jul 2022 Revised: 24 Sep 2022 Accepted: 11 Oct 2022
Published online: 15 November 2023

## Rights and permissions

**Open Access** The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

# References

1. IEA. IEA CO2 emissions from fuel combustion statistics: greenhouse gas emissions from energy (Edition 2021). Paris: OECD Publishing; 2021. Available from: https://www.oecd-ilibrary.org/energy/data/iea-co2-emissions-from-fuel-combustion-statistics_greenhouse-gas-emissions-from-energy_data-00422-en

2. Pili R, Wieland C, Spliethoff H, Haglind F. Numerical analysis of feedforward concepts for advanced control of organic Rankine cycle systems on heavy-duty vehicles. J Clean Prod. 2022;351.

3. Tian H, Liu P, Shu G. Challenges and opportunities of Rankine cycle for waste heat recovery from internal combustion engine. Prog Energy Combust Sci. 2021;84:100906.

4. Xu F, Crawford C, Feng Y, Lin Z, Li S. Environment-economic analysis of diesel, hybrid electric, plug-in hybrid electric trucks in China. Transp Res Part D-Transp Environ. 2023;117.

5. Gao Z, Smith DE, Daw CS, Edwards KD, Kaul BC, Domingo N, et al. The evaluation of developing vehicle technologies on the fuel economy of long-haul trucks. Energy Convers Manag. 2015;106:766-81.

6. Castel-Branco AP, Ribau JP, Silva CM. Taxi fleet renewal in cities with improved hybrid powertrains: life cycle and sensitivity analysis in Lisbon case study. Energies (Basel). 2015;8(9):9509-40.

7. Pielecha I, Cieslik W, Merkisz J. Analysis of the electric drive mode use and energy flow in hybrid drives of SUVs in urban and extra-urban traffic conditions. J Mech Sci Technol. 2019;33(10):5043-50.

8. Jouhara H, Khordehgah N, Almahmoud S, Delpech B, Chauhan A, Tassou SA. Waste heat recovery technologies and applications. Therm Sci Eng Progr. 2018;6:268-89.

9. Zhang X, Wang X, Cai J, He Z, Tian H, Shu G, et al. Experimental study on operating parameters matching characteristic of the organic Rankine cycle for engine waste heat recovery. Energy. 2022;244:122681.

10. Sprouse C III, Depcik C. Review of organic Rankine cycles for internal combustion engine exhaust waste heat recovery. Appl Therm Eng. 2013;51(1-2):711-22.

11. Zhang X, Wang X, Cai J, Wang R, Bian X, Yuan P, et al. Achieving reasonable waste heat utilization in all truck operating conditions via a dual-pressure organic rankine cycle and its operating strategy. J Clean Prod. 2023;419:138302.

12. Pisu P, Rizzoni G. A comparative study of supervisory control strategies for hybrid electric vehicles. IEEE Trans Control Syst Technol. 2007;15(3):506-18.

13. Salmasi FR. Control strategies for hybrid electric vehicles: evolution, classification, comparison, and future trends. IEEE Trans Veh Technol. 2007;56(5):2393-404.

14. Zhou J, Zhao J, Wang L. Cost-based research on energy management strategy of electric vehicles using hybird energy storage system. Tsinghua Sci Technol. 2024;29(3):684-97.

15. Xiong R, Cao J, Yu Q. Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle. Appl Energy. 2018;211:538-48.

16. Biswas A, Emadi A. Energy management systems for electrified powertrains: state-of-the-art review and future trends. IEEE Trans Veh Technol. 2019;68(7):6453-67.

17. Lee HD, Sul SK. Fuzzy-logic-based torque control strategy for parallel-type hybrid electric vehicle. IEEE Trans Ind Electron. 1998;45(4):625-32.

18. Lin CC, Peng H, Grizzle JW, Kang JM. Power management strategy for a parallel hybrid electric truck. IEEE Trans Control Syst Technol. 2003;11(6):839-49.

19. Sciarretta A, Back M, Guzzella L. Optimal control of parallel hybrid electric vehicles. IEEE Trans Control Syst Technol. 2004;12(3):352-63.

20. Lin CC, Peng H, Grizzle JW. A stochastic control strategy for hybrid electric vehicles. In: Proceedings of the 2004 American Control Conference; 2004 Jun 30-Jul 2; Boston, MA, USA. Piscataway (NJ): IEEE; 2004. p. 4710-5.

21. Hellstrom E, Ivarsson M, Aslund J, Nielsen L. Look-ahead control for heavy trucks to minimize trip time and fuel consumption. Control Eng Pract. 2009;17(2):245-54.

22. Borhan H, Vahidi A, Phillips AM, Kuang ML, Kolmanovsky IV, Di Cairano S. MPC-based energy management of a power-split hybrid electric vehicle. IEEE Trans Control Syst Technol. 2012;20(3):593-603.

23. Ahmadian S, Tahmasbi M, Abedi R. Q-learning based control for energy management of series-parallel hybrid vehicles with balanced fuel consumption and battery life. Energy AI. 2023;11:100234.

24. Xu B, Rathod D, Zhang D, Yebi A, Zhang X, Li X, et al. Parametric study on reinforcement learning optimized energy management strategy for a hybrid electric vehicle. Appl Energy. 2020;259:114200.

25. Du G, Zou Y, Zhang X, Liu T, Wu J, He D. Deep reinforcement learning based energy management for a hybrid electric vehicle. Energy. 2020;201:117591.

26. Gao Y, Wang X, Tian H, Cai JW, Shu GQ. Quantitative analysis of fuel-saving potential for waste heat recovery system integrated with hybrid electric vehicle. Int J Energy Res. 2020;44(14):11152-70.

27. Andwari AM, Pesiridis A, Karvountzis-Kontakiotis A, Esfahanian V. Hybrid electric vehicle performance with organic Rankine cycle waste heat recovery system. Appl Sci-Basel. 2017;7(4):375.

28. Kruijt K, Verbruggen FJR, Speetjens MFM, Hofman T. Modeling and control of a waste heat recovery system for integrated powertrain design optimization. IFAC-PapersOnLine. 2019;52(5):598-603.

29. Mansour C, Bou Nader W, Dumand C, Nemer M. Waste heat recovery from engine coolant on mild hybrid vehicle using organic Rankine cycle. Proc Inst Mech Eng Part D J Autom Eng. 2019;233(10):2502-17.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

30. Wang X, Wang R, Shu G, Tian H, Zhang X. Energy management strategy for hybrid electric vehicle integrated with waste heat recovery system based on deep reinforcement learning. Sci China-Technol Sci. 2022;65(3):713-25.

31. Lian R, Peng J, Wu Y, Tan H, Zhang H. Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle. Energy. 2020;197:117297.

32. Xu D, Zheng C, Cui Y, Fu S, Kim N, Cha SW. Recent progress in learning algorithms applied in energy management of hybrid vehicles: a comprehensive review. Int J Precis Eng Manuf-Green Technol. 2023;10(1):245-67.

33. Zhang W, Wang J, Liu Y, Gao G, Liang S, Ma H. Reinforcement learning-based intelligent energy management architecture for hybrid construction machinery. Appl Energy. 2020;275:115401.

34. Torreglosa JP, Garcia-Trivino P, Vera D, Lopez-Garcia DA. Analyzing the improvements of energy management systems for hybrid electric vehicles using a systematic literature review: how far are these controls from rule-based controls used in commercial vehicles? Appl Sci-Basel. 2020;10(23):8584.

35. Zou R, Fan L, Dong Y, Zheng S, Hu C. DQL energy management: an online-updated algorithm and its application in fix-line hybrid electric vehicle. Energy. 2021;225:120174.

36. Deng P, Wu X, Yang J, Yang G, Jiang P, Yang J, et al. Optimal online energy management strategy of a fuel cell hybrid bus via reinforcement learning. Energy Convers Manag. 2024;300:117921.

37. Cai J, Tian H, Wang X, Wang R, Shu G, Wang M. A calibrated organic Rankine cycle dynamic model applying to subcritical system and transcritical system. Energy. 2021;237:121542.

38. Zhang X, Wang X, Cai J, Wang R, Bian X, Wang J, et al. Selection maps of dual-pressure organic Rankine cycle configurations for engine waste heat recovery applications. Appl Therm Eng. 2023;228:120451.

39. Li Z, Huang Y, Wang H. A new system configuration design and power management strategies for a multi-source hybrid truck. Proc Inst Mech Eng Part I J Syst Control Eng. 2018;232(8):1053-62.

40. Ben Halima N, Chaieb M, Ben Hadj N, Abdelmoula R, Neji R. Study of a parallel hybrid electric vehicle performance by means of rules-based control strategies. In: 2020 17th International Multi-Conference on Systems, Signals and Devices (SSD); 2020 Jul 20-23; Sfax, Tunisia. Piscataway (NJ): IEEE; 2020. p. 703-8.

41. Anbaran SA, Idris NRN, Jannati M, Aziz MJ, Alsofyani I. Rule-based supervisory control of split-parallel hybrid electric vehicle. In: 2014 IEEE Conference on Energy Conversion (CENCON); 2014 Oct 13-14; Johor Bahru, Malaysia. Piscataway (NJ): IEEE; 2014. p. 7-12.

42. Cuma MU, Koroglu T. A comprehensive review on estimation strategies used in hybrid and battery electric vehicles. Renew Sustain Energy Rev. 2015;42:517-31.

43. Moulik B, Soeffker D. Optimal rule-based power management for online, real-time applications in HEVs with multiple sources and objectives: a review. Energies. 2015;8(9):9049-63.

44. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. arXiv:1509.02971 [cs.LG]. 2015.

45. Coronado CR, de Carvalho JA, Silveira JL. Biodiesel CO2 emissions: a comparison with the main fuels in the Brazilian market. Fuel Process Technol. 2009;90(2):204-11.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press