Journal of Zhejiang University-SCIENCE A in press
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.A2500493

A thermal management strategy for hybrid electric drive tracked vehicles considering system safety and energy consumption based on the GMA-TD3-MPC algorithm

Xiao ZHENG, Junliang ZHAO, Yuchao YAN, Zhentao LIU✉

Power Machinery and Vehicular Engineering Institute, Zhejiang University, Hangzhou 310027, China

Abstract: The development of efficient Thermal Management Strategies (TMS) is critical for Hybrid Electric Drive Tracked Vehicles (HETVs) due to the severe thermal safety and energy consumption challenges encountered during complex operations. Conventional strategies struggle to balance high-precision temperature control with multi-objective collaborative optimization, while also requiring long development cycles and exhibiting weak generalization capabilities. To address these issues, we propose a hierarchical thermal management framework integrating a Gated Recurrent Unit Multi-Head Attention Twin Delayed Deep Deterministic Policy Gradient with Model Predictive Control (GMA-TD3-MPC). This framework dynamically integrates Reinforcement Learning (RL) and Model Predictive Control (MPC), utilizing a Gated Recurrent Unit with Multi-Head Attention (GRU-MHA) module to optimize energy consumption and temperature control precision under cyclic conditions; meanwhile, it implements a dynamic threshold triggering mechanism to seamlessly transfer control to the MPC controller when approaching thermal safety limits. Our simulation results demonstrate that compared to baseline strategies, the proposed method accelerates convergence by approximately 28% and 40% over DDPG and TD3, respectively. In a standard temperature environment (25 ℃) under off-road conditions, compared to standalone MPC, the proposed strategy reduces temperature fluctuation ranges in high-temperature and low-temperature circuits by 44.19% and 6.45%, respectively, while also achieving a 5.53% reduction in total energy consumption and a 10.63% decrease in peak power demand. Furthermore, under high-temperature conditions (45 ℃), the strategy reduces the total energy consumption by 13.41%.

Key words: Thermal management strategy; Hybrid electric drive tracked vehicles; Gated recurrent unit with multi-head attention; Hierarchical control; Energy consumption optimization

1 Introduction

Driven by the global energy transition and the advancement of carbon peak and neutrality goals, hybrid electric powertrain technology has emerged as a major focus in tracked vehicle applications involving construction machinery and agricultural equipment (Scolaro et al., 2021; Wang et al., 2021; Zhao et al., 2022; He et al., 2024a; Tian et al., 2026). Such technology is attractive because of its high energy efficiency, low pollution levels, and reduced noise emissions. Unlike conventional powertrains that are reliant on complex mechanical transmissions, hybrid electric tracked vehicles (HETVs) employ an independent dual-motor drive architecture for dynamic torque distribution. This configuration cuts the powertrain response time from 300-500 ms to within 20 ms. This millisecond-level response capability enables instantaneous torque adjustment to suppress track slippage on soft soils, and ensures continuous power delivery during abrupt resistance changes, thereby enhancing off-road and obstacle-crossing capabilities (Hori, 2004; Dersu and Kılıç, 2024). Furthermore, due to the electrification characteristics of their power output, these vehicles operate in pure electric mode at noise levels below 65 dB with zero exhaust emissions, markedly improving environmental friendliness (Karaoglan, 2021; Çeliksóz and Kiliç, 2024).

✉ Zhentao LIU, liuzt@zju.edu.cn

Received Oct. 8, 2025; Revision accepted Mar. 17, 2026;
Crosschecked

However, the highly integrated powertrain of HETVs introduces a complex thermal environment with multiple heat sources, posing challenges to its Thermal Management System (TMS). This system incorporates diverse heat sources—including the engine, generator, drive motors, and power electronics—while contending with issues arising from intricate heat distribution, conflicting thermal control objectives, and dynamic operating conditions (Ghate et al., 2024b). The engine coolant temperature must be maintained within a high-temperature range of 80–125 °C to ensure combustion efficiency, while the motor temperature requires control between 55–75 °C to prevent the demagnetization of permanent magnets. Furthermore, multi-stage intercoolers must sharply reduce the intake air temperature from approximately 250 °C to below 80 °C for effective cooling (Darici et al., 2012; Dong et al., 2020; Wu et al., 2021; Mancarella and Marello, 2023; Zhai et al., 2025).

The traditional single-loop TMS struggles to achieve decoupled control across multiple temperature zones due to its reliance on fixed-threshold control strategies and the strongly coupled relationship between flow and temperature. To meet the cooling demands of high-temperature components, it often overcools low-temperature parts, creating conflicting heat dissipation requirements and wasting significant energy (Chen et al., 2024b; Nam and Ahn, 2025). During extreme operations such as off-road climbing or rapid acceleration, sudden power surges cause the generated heat to jump, while heat transfer processes exhibit a significant time lag (Yuan et al., 2020). Existing mainstream thermal management strategies—including Lookup Table-based, PID (Proportional-Integral-Derivative), and simple rule-based methods—fail to account for this time-delayed characteristic. Consequently, under abrupt load changes, such methods tend to overshoot or delay the response, causing component overheating risks (Han et al., 2017; Liu et al., 2021; Guo et al., 2024). Research indicates that TMS consumes 12–18% of the total vehicle energy while facing thermal runaway risks under extreme conditions (Golubev et al., 2024; Tang et al., 2024; Wang et al., 2025b). For example, HETV off-road missions typically involve highly transient and extreme power loads, making powertrain control more complex than in standard driving cycles (Zhu et al., 2022). Moreover, thermal management in such vehicles is constrained by restricted ventilation pathways, due to armored protection features such as grilles and louvers. These limitations – in tandem with high power levels under extreme temperature conditions – can result in the cooling system consuming up to 11–16% of the total fuel energy (Sundar et al., 2023). Therefore, developing advanced thermal management strategies to synergistically optimize temperature control precision, system energy consumption, and operational safety has become paramount for enhancing the comprehensive performance of HETVs.

In terms of intelligent control algorithms, research is progressively shifting from conventional PID and rule-based methods towards data-driven and optimization-based approaches (Chen et al., 2024a; He et al., 2024b; Arrinda et al., 2025; Guo et al., 2025; Li et al., 2025). For instance, Ghate et al. proposed an Integrated Energy and Thermal Planner (IETP) for series hybrid off-road autonomous tracked vehicles, demonstrating that the synergistic operation of the ICE-generator and thermal actuators could improve fuel efficiency by at least 10% in power-demanding scenarios as compared to traditional separated management strategies (Ghate et al., 2024a). Also, Naini et al. introduced a model reference adaptive controller for thermal management of autonomous vehicles, showing that Nonlinear MPC (NMPC) could reduce fan energy consumption by 73% compared to PI control (Naini et al., 2023). Furthermore, for multi-objective optimization in hybrid systems, Zhang et al. developed a hierarchical framework combining the Twin Delayed Deep Deterministic Policy Gradient (TD3) with optimization algorithms, achieving a 17.99% reduction in operating costs while maintaining thermal stability (Zhang et al., 2025). Moreover, Jiang et al. created a Deep Reinforcement Learning (DRL)-based co-optimization strategy for dynamic coordination between energy and thermal management. Testing across a temperature range of -5 °C to 30 °C showed 5% and 2% hydrogen consumption reductions versus rule-based and Genetic Algorithm (GA) methods, respectively, alongside a 40% reduction in temperature fluctuations (Jiang et al., 2025).

However, critical limitations persist in existing thermal management strategy research, constraining applications in complex multi-source HETVs systems. In particular, the efficacy of MPC hinges heavily on precise mathematical models of controlled objects. At the same time, HETV thermal management is a complex nonlinear system involving strong Multiphysics coupling (combustion-electrical-thermal-fluidic), making accurate modelling difficult. Accumulated model errors consequently degrade control precision in complex operating scenarios (Xi et al., 2013; Wang et al., 2025a). RL methods possess strong adaptive optimization potential. However, conducting full online training and decision-making throughout the operational control process typically demands excessive time and computational resources. Moreover, when system states deviate from training distributions, the agents may output hazardous actions, causing transient temperature overshoots in components and potentially triggering thermal runaway (Cobbe et al., 2018; Nguyen et al., 2020; Atkinson et al., 2021).

To address the aforementioned challenges, we propose a hierarchical thermal management framework that is explicitly distinguished from existing purely learning-based or rule-based strategies. Focusing on the TMS of HETVs, we develop an intelligent hybrid thermal management strategy based on a Gated Recurrent Unit-enhanced Multi-Head Attention Twin Delayed Deep Deterministic Policy Gradient algorithm integrated with Model Predictive Control (GMA-TD3-MPC). This approach aims to synergistically optimize temperature control precision, system energy consumption, and operational safety. The core contributions of this work are threefold:

(1) By utilizing a GMA-TD3 agent for global energy-optimal planning and MPC for safety-constrained local tracking, this framework bridges the gap between data-driven optimization and model-based robustness. An intelligent switching mechanism dynamically integrates the strengths of both methods, ensuring high-precision and low-energy operation within safe boundaries.

(2) To improve exploration in high-dimensional spaces, a Gated Recurrent Unit-enhanced Multi-Head Attention (GMA) mechanism is integrated into the TD3 algorithm to filter out suboptimal actions. Furthermore, Gated Recurrent Units (GRUs) and Multi-Head Attention (MHA) are incorporated to capture strong temporal dependencies and emphasize critical decision-making features, addressing the limitations of conventional RL methods.

(3) The framework is tailored to handle the strong coupling and dynamic constraints of HETVs. Simulation results demonstrate that the proposed strategy outperforms conventional baselines in temperature accuracy, energy conservation, and actuator smoothness.

The findings of this study provide a framework for balancing high precision, energy efficiency, and safety in thermal management for HETVs. Subsequent sections detail the powertrain architecture and thermal management system modelling of the studied HETVs (Section II), elaborate on the architecture, algorithmic modules, and methodology of the proposed GMA-TD3-MPC hybrid control strategy (Section III), and validate the control performance and energy efficiency of the strategy through comparative simulation experiments (Section IV).

2 Vehicle and thermal management system model

2.1 Description of HETVs

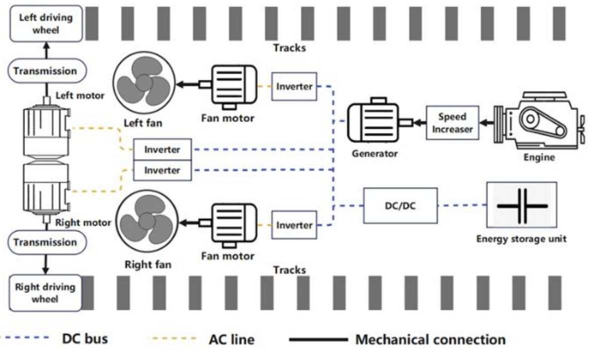

Fig. 1 The powertrain composition of HETVs

Fig. 1 illustrates the powertrain composition of the HETVs investigated in this study. The powertrain core is comprised of an Internal Combustion Engine (ICE) that drives a generator via a speed-increasing gearbox. The electrical energy generated is transmitted to a DC bus. A dedicated energy storage unit connects to the DC bus through a DC/DC converter, enabling bidirectional energy flow for charging or discharging based on operational demands. Propulsion is achieved through two independently driven tracks; each track is powered by a traction motor coupled to a final reduction gearbox that drives the sprocket; each traction motor is supplied by an individual inverter, which converts the DC bus voltage into a three-phase AC to actuate the motor. This left-right independent inverter-motor-gearbox-sprocket architecture enables flexible steering and traction torque control. Cooling is provided by two fans that facilitate heat exchange between the thermal management system and the ambient air. Each fan motor is powered by a dedicated inverter, drawing energy from the DC bus to drive the cooling fans and ensure thermal dissipation for critical components. Key parameters of the HETVs are detailed in **Table S1** in the Electronic Supplementary Materials (ESM).

2.2 HETV powertrain system modeling

We next concentrate on modeling the thermal management strategy for HETVs. For the vehicle dynamic model, the effects of lateral and vertical dynamics on motion are neglected, focusing exclusively on longitudinal dynamics (Yu et al., 2023; Ma et al., 2024). It should be noted that tracked vehicles experience unique steering loads arising from skid-steering mechanisms and track-ground friction (Sundar et al., 2026). Although the longitudinal dynamic model used here simplifies the explicit calculation of steering resistance F_{steer}, these loads are accounted for through an equivalent modeling approach—specifically, by adopting a relatively high rolling friction coefficient μ_{roll} to simulate the comprehensive resistance encountered during off-road driving. This ensures that the simulated powertrain power demand and the associated thermal load effectively capture the typical high-load characteristics, so as to provide an adequate safety margin for validating the thermal management system strategy. The detailed mathematical formulations of the vehicle dynamic modeling are provided in **Section S1.2** of the ESM.

2.3 Modeling of the HETV thermal management system

This section elaborates on the physics-based simulation model established on the GT-SUITE platform, constructed according to the thermal management system (TMS) architecture of the HETVs shown in **Fig. 2**. This vehicle employs a dual-loop cooling system, comprised of independently operating high-temperature and low-temperature circuits. The energy storage unit in this HETV configuration consists of a supercapacitor pack used for transient power buffering. The supercapacitor module employs an independent air-cooling scheme, and is thermally decoupled from the liquid cooling networks. Therefore, the thermal management of the energy storage unit is considered outside the scope of the coupled liquid TMS model established in this study. More details on the physical modeling of the TMS are provided in **Section S1.3** of the ESM. Notably, the maximum errors during transient loading remained within 10%. Therefore, the developed model is deemed suitable for characterizing TMS thermal behavior, and meets the accuracy requirements for our subsequent control strategy simulation investigations.

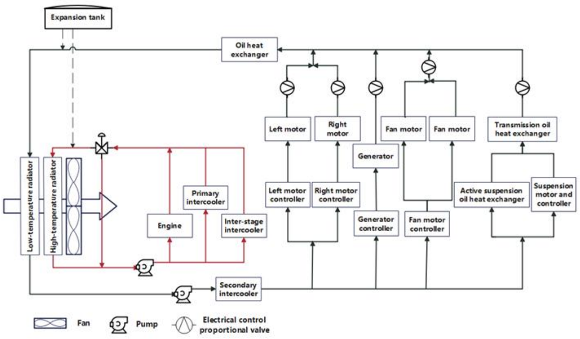

Fig. 2 The thermal management system architecture of the HETVs

3 Safety and energy-conscious thermal management strategy

3.1 Control strategy framework

This section details the proposed integrated GMA-TD3-MPC control strategy for HETVs. GMA-TD3-MPC synergizes the GMA-TD3 algorithm with an MPC algorithm. Its design objectives are precision temperature regulation, energy consumption optimization, and thermal safety assurance for critical components. The architecture's core innovation is dynamically harmonizing the strengths of the GMA-TD3 and MPC through an Action Selection Module, which generates final actuator commands.

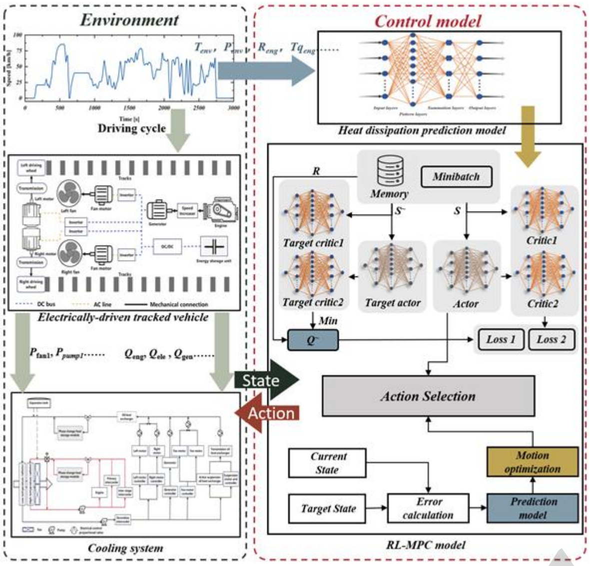

Fig. 3 Framework of the control strategy based on GMA-TD3-MPC

The control strategy comprises three interconnected modules: an "environment" module, a "heat rejection data prediction" module, and the GMA-TD3-MPC module. Environment provides the system state vector required for decision-making, including engine speed, engine torque, left/right drive motor speeds, left/right drive motor torques, ambient temperature, and ambient pressure. The heat rejection data prediction module processes environmental inputs to generate heat rejection data across TMS components using interpolation and neural network predictions. These predictions augment the system state vector transmitted to the GMA-TD3-MPC module. The GMA-TD3-MPC module dynamically computes optimal actuator commands by coordinating GMA-TD3's adaptive learning capabilities with MPC's predictive robustness. The overall framework is illustrated in **Fig. 3**.

3.2 GMA-TD3 algorithm

Reinforcement learning (RL) abstracts the closed-loop control problem as a Markov Decision Process (MDP), whose detailed mathematical formulations are provided in **Section S2.1** of the ESM. During training, the agent learns a state-to-action mapping motivated by optimization of a designed reward function. In the initial exploration phase, the high uncertainty due to limited prior knowledge necessitates active environmental exploration to discover high-reward actions. Action selection provides direct links to received rewards, requiring efficient trial-and-error learning. The core objective is to minimize the total power consumption under temperature constraints. The immediate reward function is designed as:

$$\begin{aligned} r_t = {} & \omega_1 |T_{\text{high,target}} - T_{\text{high,in}}| \\ & + \omega_2 |T_{\text{low,target}} - T_{\text{low,out}}| \\ & + \omega_3 \left(P_{\text{fan}} + P_{\text{pump}}\right) \\ & + I_1 \omega_4 \left(T_{\text{high,max}} - T_{\text{high,in}}\right) \\ & + I_2 \omega_5 (T_{\text{low,max}} - T_{\text{low,out}}) \\ & + I_3 \omega_6 \left(\frac{du_{\text{fanspeed}}}{dt}\right) \\ & + I_4 \omega_7 \left(\frac{du_{\text{pumpspeed}}}{dt}\right) + I_5 C \end{aligned} \tag{1}$$

in which ω_i is the weight for each reward component; $T_{\text{high,target}}$ and $T_{\text{low,target}}$ are the temperature target values for the high-temperature circuit and low-temperature circuit, respectively; $T_{\text{high,max}}$ and $T_{\text{low,max}}$ are the maximum temperature limits for the high-temperature circuit and low-temperature circuit, respectively; $T_{\text{high,in}}$ and $T_{\text{low,out}}$ are the inlet temperature of the high-temperature circuit radiator and the outlet temperature of the low-temperature circuit radiator, respectively; P_{fan} and P_{pump} are the real-time power of the fan and pump, respectively; $\frac{du_{\text{fanspeed}}}{dt}$ and $\frac{du_{\text{pumpspeed}}}{dt}$ are the rate of change of fan speed and pump speed, respectively; C is the boundary penalty; and I_i is the judgment function, where $I_i = 1$ if the condition is satisfied, and is 0 otherwise. This equation serves as the core mechanism for handling coupling of conflicting objectives. Notably, the optimization of temperature tracking precision and the minimization of energy consumption are inherently contradictory. By integrating them into a unified scalar reward function with calibrated weights, the GMA-TD3 agent learns to implicitly navigate the trade-off surface between thermal performance and power usage; this soft coupling allows the system to sacrifice a marginal amount of temperature precision for significant energy savings within the safe operating range. The specific weights and judgment functions used here are provided in **Table S3**, while the hyperparameters of the neural network are detailed in **Table S4** in the ESM.

3.3 MPC algorithm

MPC is a multivariable optimization control strategy based on dynamic models. Its core principle is to achieve optimal tracking of system states through rolling optimization and feedback correction (Norouzi et al., 2023). In the thermal management system of an HETV, MPC dynamically regulates pump and fan speeds to ensure that component temperatures stabilize within target ranges. Distinguished from conventional feedback control, MPC leverages the system's mathematical model to predict future states and optimize control inputs over a finite time horizon, so as to meet the control objectives. The configurations for the MPC are detailed in **Section S2.2** of the ESM.

3.4 Action selection module

The Action Selection Module is the execution unit of the GMA-TD3-MPC strategy, which is responsible for dynamically scheduling control actions generated by the GMA-TD3 and MPC. When the system state deviates from the suitable temperature range, it switches to the MPC controller to ensure safe operation of the thermal management system. When the temperature stabilizes within the suitable range, the GMA-TD3 strategy is activated to balance temperature control and energy consumption. The temperature deviation is defined as the decision variable for action selection:

$$\Delta T_{\mathrm{high}} = |T_{\mathrm{high,target}} - T_{\mathrm{high,out}}| \tag{2}$$

$$\Delta T_{\mathrm{low}} = |T_{\mathrm{low,target}} - T_{\mathrm{low,out}}| \tag{3}$$

To eliminate abrupt rotational speed changes during mode switching, a proportional mixing function based on temperature deviation is designed to achieve smooth action transitions:

$$\begin{gathered} u_{\mathrm{pumphigh,output}} = \alpha_{\mathrm{high}} \cdot u_{\mathrm{TD3}} \\ +(1-\alpha_{\mathrm{high}}) \cdot u_{\mathrm{MPC}} \end{gathered} \tag{4}$$

$$\begin{gathered} u_{\mathrm{pumplow,output}} = \alpha_{\mathrm{low}} \cdot u_{\mathrm{TD3}} + \\ (1-\alpha_{\mathrm{low}}) \cdot u_{\mathrm{MPC}}, \end{gathered} \tag{5}$$

$$\begin{gathered} u_{\mathrm{fan,output}} = \alpha_{\mathrm{high}} \cdot \alpha_{\mathrm{low}} u_{\mathrm{TD3}} \\ +(1-\alpha_{\mathrm{high}} \cdot \alpha_{\mathrm{low}}) \cdot u_{\mathrm{MPC}} \end{gathered} \tag{6}$$

where $\alpha_{\mathrm{high}} = f\left(\Delta T_{\mathrm{high}}\right), \alpha_{\mathrm{low}} = f(\Delta T_{\mathrm{low}})$ satisfy the following piecewise conditions:

$$a = \begin{cases} 0 & \Delta T > 5 \\ \frac{1}{2}\left[1+\cos\left(\pi \cdot \frac{\Delta T - 5}{2}\right)\right] & 3 < \Delta T \leq 5 \\ 1 & \Delta T \leq 3 \end{cases} \tag{7}$$

While the GMA-TD3 module handles the coupled optimization of energy and precision, the Action Selection Module implements an objective decoupling and prioritization mechanism to guarantee safety. When the temperature deviation ΔT exceeds the threshold, the strategy recognizes that the safety objective outweighs the energy considerations. Consequently, the control logic decouples the safety constraint from the multi-objective optimization, switching dominance to the MPC controller. This ensures that maintaining the component temperatures within strict thermal limits takes absolute precedence over energy minimization during critical thermal events. Through this real-time scheduling and smooth switching, the proposed framework coherently optimizes both operational safety and energy efficiency in the thermal management system.

4 Results and discussion

In this section we analyze TMS strategies considering safety and energy consumption, so as to validate the effectiveness of the proposed approach. To comprehensively evaluate the control performance, testing was conducted using a typical driving cycle for tracked vehicles. Detailed information regarding the input boundary conditions, the specific driving cycle proportions, and the dynamic response characteristics of the powertrain (including the rotational speed and torque of the engine and electric motors) are provided in **Fig. S6**, **Table S6**, and **Fig. S7** in the ESM, respectively.

4.1 Algorithm training and convergence analysis

Fig. 4 presents a comparison of the learning stability of the DDPG, TD3, and GMA-TD3 algorithms in developing TMS strategies. To ensure the statistical reliability of the results and eliminate the bias caused by random initialization, robustness tests were conducted. Each algorithm was trained using five independent random seeds; the reward curves presented in Fig. 12 denote the average performance across these independent runs. The results indicate that while DDPG and standard TD3 showed larger performance variance across different seeds, GMA-TD3 maintained consistent convergence trajectories with narrow standard deviations, confirming the repeatability and robustness of the strategy.

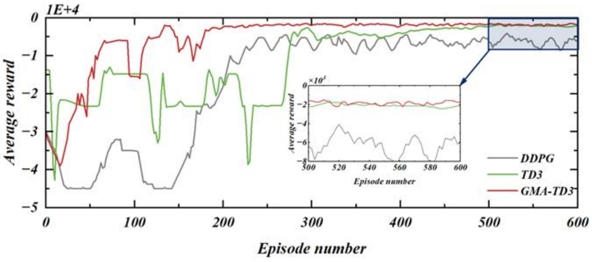

Fig. 4 Reward convergence results of different reinforcement learning algorithms

To isolate the benefits of the proposed architecture, we utilize the standard TD3 algorithm as an ablation baseline. It is important to note that the GRU and MHA are evaluated together as a combined GMA module. This is a deliberate design choice based on their synergistic mechanism. Thermal dynamics exhibit significant temporal delays: while the GRU efficiently extracts temporal dependencies from historical state sequences, the MHA mechanism is structurally coupled to directly process the GRU's hidden states, dynamically assigning attention weights to critical historical moments. Separating them would disrupt this coupled temporal-feature extraction process. Therefore, our ablation study explicitly focuses on the global contribution of the integrated GMA module compared to the baseline TD3 algorithm. During the initial training phase, all RL algorithms frequently caused component temperatures to exceed safety thresholds due to insufficient learning of dynamic thermal loads, leading to early termination of the episode. As training progressed, the algorithms gradually optimized their strategies. However, significant performance disparities emerged: the simple TD3 agent suffered from slower convergence (stabilizing after 300 steps) compared to GMA-TD3 (stable around 180 steps), while DDPG exhibited significant oscillations even beyond step 250. This disparity highlights the limitation of standard MLP-based RL in handling Partially Observable Markov Decision Processes (POMDP) caused by significant thermal delays. The inclusion of the GRU module effectively bridges this gap by retaining historical thermal state information to compensate for the system's inertia, while the Multi-Head Attention (MHA) mechanism enhances feature extraction efficiency by selectively prioritizing critical thermal risk signals. Consequently, GMA-TD3 outperformed DDPG and the baseline TD3 in convergence speed by about 28% and 40%, respectively. This advantage primarily stems from its embedded GRU module and augmented attention mechanism, which identify and prioritize critical thermal risk signals. By leveraging time-series thermal state data, the algorithm captures temporal dependencies and learns patterns of thermal delay effects. These capabilities accelerate the learning process while improving the robustness of the final control strategy.

4.2 Performance analysis under normal temperature conditions

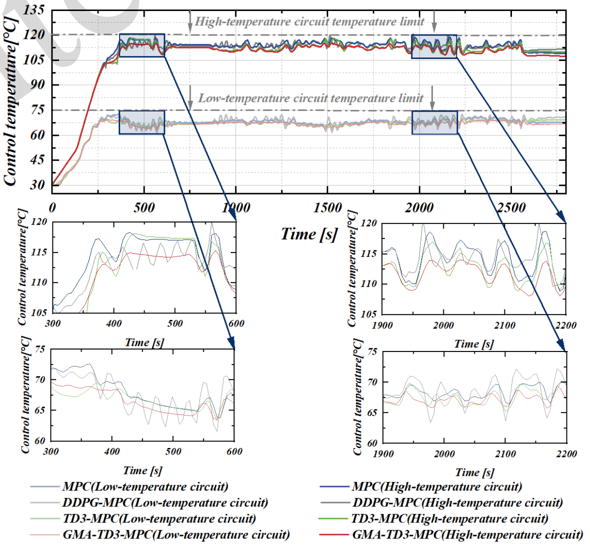

Fig. 5 The comparison of controlled temperature results in high-/low-temperature circuits under different control algorithms

To evaluate the proposed strategies, a comparative study was conducted at an ambient temperature of 25 ℃ and a pressure of 1 bar. The detailed dynamic heat rejection profiles from key components to the coolant under this driving cycle are provided in **Fig. S8** of the ESM. These profiles demonstrate severe thermal load surges and fluctuations that are highly coupled with off-road driving conditions, posing significant challenges to the TMS. Note that detailed pump and fan speed variations across different control algorithms are provided in **Fig. S9** in the ESM.

Fig. 5 presents comparative results of controlled temperatures in the circuits. It can be seen that GMA-TD3-MPC exhibits minimal fluctuations in both the fan and pump speed curves, indicating superior convergence in action space exploration. The embedded GMA module enhances the attention mechanisms and memory capabilities for historical thermal states, enabling smoother actuator adjustments. As depicted in the 300–600 s magnified view, when the system approaches the ideal temperature range, GMA-TD3-MPC executes more precise actions than other algorithms, achieving smoother temperature transitions. At the maximum vehicle speed, it maintains the high-temperature circuit control temperature at 2.1 °C lower than MPC and TD3-MPC with comparable rotational speeds, thus preventing overheating risks. The 1900–2200 s interval (off-road driving with significant load fluctuations) further reveals that GMA-TD3-MPC delivers faster and more stable control outputs to suppress temperature oscillations. It restricts temperature fluctuations to 4.8 °C (in the high-temperature circuit) and 2.9 °C (low-temperature circuit), representing reductions of 44.19% and 6.45% compared to MPC's 8.6 °C and 3.1 °C fluctuations, respectively. Relative to TD3-MPC, the fluctuation reductions reach 38.46% and 23.68%. While TD3-MPC outperforms the baseline MPC in temperature stability, it remains inferior to GMA-TD3-MPC. DDPG-MPC demonstrates the poorest performance: during critical operating conditions, the pumps and fans exhibit oscillatory behavior, with high-temperature circuit temperatures repeatedly reaching the 120 °C safety threshold. This necessitates frequent MPC intervention for correction. Overall, the deterministic policy of DDPG lacks exploratory robustness and fails to handle nonlinear heat exchange dynamics under varying vehicle operating conditions.

Fig. 6(a) and **(b)** illustrate the variations in power consumption and cumulative energy consumption over time for different control algorithms, while **Fig. 6(c)** presents a comparative analysis of their energy consumption outcomes. The GMA-TD3-MPC algorithm achieves the lowest peak power demand during the cycle; compared to MPC, its maximum power demand decreases by 10.63%, effectively mitigating abrupt fluctuations in demand. Notably, at the 504 s mark, peak power reductions of 18.08% compared to MPC were observed, with the peak power advantage persisting thereafter. This demonstrates the algorithm's accurate sequence modeling capability, predicting power system demands and preemptively adjusting cooling power, and thus avoiding overshoot caused by response lag. Additionally, GMA-TD3-MPC maintains the lowest cumulative energy consumption rate throughout the driving cycle. Regarding global energy consumption, GMA-TD3-MPC achieves the lowest total energy consumption of 18.08 kWh for the entire cycle. This represents a 5.53% reduction compared to MPC's 19.14 kWh, and a further 1.08% improvement over the second-best TD3-MPC at 18.29 kWh. While this 5.53% reduction might appear modest, its practical implications for HETVs are substantial. First, this improvement was achieved against a high-performance optimization-based controller rather than a basic rule-based strategy, indicating a significant algorithmic gain. Second, and perhaps more critically, the 10.63% reduction in peak power is vital for practical engineering applications. By smoothing high-frequency power surges, this strategy alleviates thermal and electrical stress on the onboard energy storage system and power electronics. Thus, GMA - TD3 - MPC not only reduces energy consumption, but also enhances system reliability under extreme operating conditions.

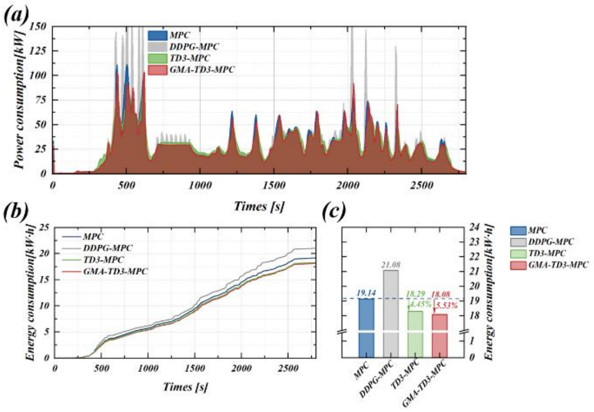

Fig. 6 (a) Comparison of power consumption over time among different control algorithms; **(b)** Comparison of cumulative energy consumption over time among different control algorithms; **(c)** Energy consumption results across different control algorithms

4.3 Performance analysis under high-temperature

conditions

To further validate the robustness of the proposed strategy under extreme conditions, the ambient temperature was increased to 45 ℃ for testing under the same driving cycle. This elevated ambient temperature significantly narrows the temperature gradient for radiator heat exchange, thereby imposing greater challenges on both the operational safety and energy efficiency of the thermal management system. The detailed distribution of heat rejection power from key components to the coolant under this elevated temperature is illustrated in **Fig. S10** in the ESM. Also, the control trajectories for the high/low-temperature circuit pumps and radiator fans under various strategies at an ambient temperature of 45 ℃ are presented in **Fig. S11** in the ESM.

Fig. 7 illustrates the regulation performance of various control strategies on the coolant temperatures in both the high-temperature and low-temperature circuits, within a 45 ℃ ambient environment. Compared to the standard temperature, the high ambient temperature leads to a sharp narrowing of the available heat transfer temperature difference between the radiator and the ambient air. This physical constraint is particularly severe for the low-temperature circuit, and thus the performance disparities among the different control algorithms are amplified in this case. With the safety threshold for the low-temperature circuit set at 75 ℃, the effective temperature difference is reduced to only 30 ℃ at an ambient temperature of 45 ℃ , which imposes stringent requirements on the response speed and precision of the controllers. While the MPC maintains the temperatures within safe limits, it exhibits a significant "overreaction." During the acceleration phase (400-500 s) and the off-road segment (after 2000 s), the MPC forces the temperature down to approximately 67 ℃, suggesting excessive control effort. In contrast, DDPG-MPC struggles in the high-load region (500-600 s), with local peaks approaching 74 ℃. This leaves a safety margin of less than 1 ℃, which implies substantial thermal runaway risk.

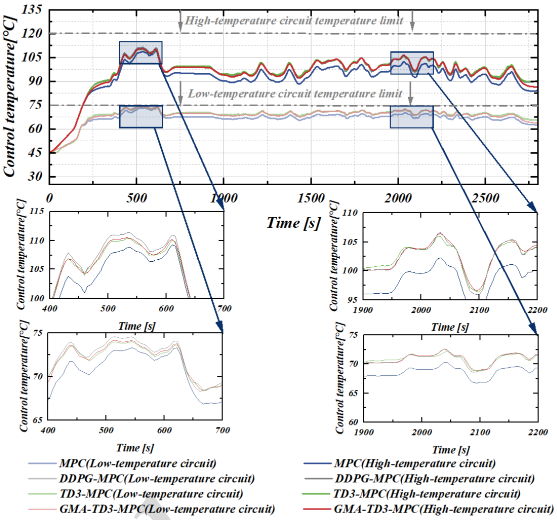

Fig. 7 A comparison of controlled temperature results in the high/low-temperature circuits under different control algorithms

Although TD3-MPC achieves a temperature regulation trend similar to the proposed method, it does so at the cost of higher energy consumption. Overall, GMA-TD3-MPC demonstrates the best comprehensive performance. During the identical off-road segment (1900-2200 s), GMA-TD3-MPC restricts the temperature fluctuation amplitudes in the high-temperature and low-temperature circuits to within 8.8 ℃ and 3.5 ℃, respectively. Compared to the fluctuations for MPC (9.9 ℃ and 3.8 ℃), this represents significant reductions of 11.11% and 7.89%, respectively.

Fig. 8(a)-(c) illustrate the instantaneous power demand and cumulative energy consumption of various control strategies at an ambient temperature of 45 °C. In comparison to standard conditions (25 °C), the high-temperature environment triggers a massive surge in the energy demand of the thermal management system. Specifically, the total energy consumption of the baseline MPC increases from approximately 19.18 kWh at 25 °C, to 63.65 kWh at 45 °C—an increase of more than threefold. To compensate for the diminished heat transfer temperature gradient, the actuators are compelled to operate within high-speed ranges for extended durations to maintain thermal equilibrium. The energy consumption differences among the algorithms are also quite pronounced; the MPC strategy frequently allows power peaks to hit the upper limit, keeping the actuator in an inefficient high-power operating state for prolonged periods. Compared to MPC, DDPG-MPC (57.87 kWh) and TD3-MPC (56.19 kWh) achieved energy savings of 9.08% and 11.72%, respectively. However, power spikes remain evident in their profiles. When lacking accurate predictions of temporal thermal loads, these algorithms still resort to intermittent high-power output to correct temperature deviations.

The proposed GMA-TD3-MPC strategy exhibits the most significant potential for energy conservation under extreme conditions. Its total energy consumption over the full cycle is only 55.11 kWh, representing a 13.41% reduction compared to the baseline MPC. While ensuring that temperature limits are not exceeded, it consistently operates the actuator within a relatively efficient range. This approach addresses thermal safety concerns under high-temperature conditions, and also achieves conserves electrical energy through intelligent management. This outcome has promising implications for enhancing the endurance of HETVs in high-temperature environments.

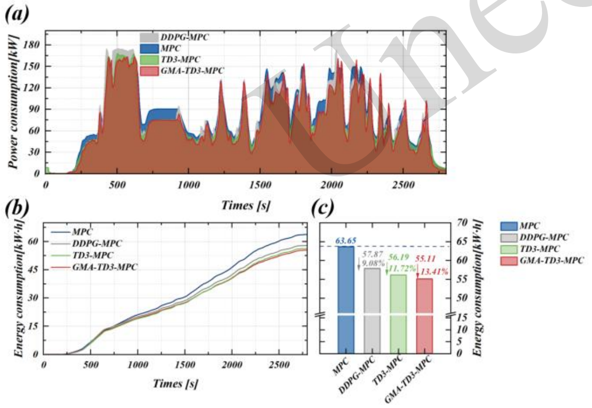

Fig. 8 (a) Comparison of power consumption over time among different control algorithms; **(b)** Comparison of cumulative energy consumption over time among different control algorithms; **(c)** Energy consumption results across different control algorithms

5 Conclusions

This study presented a GMA-TD3-MPC control strategy for hybrid electric tracked vehicles (HETVs), integrating thermal management safety and energy consumption considerations. The strategy dynamically schedules controllers through an action selection module that combines GMA-TD3 with MPC, optimizing energy consumption while ensuring the thermal safety of critical components. Comparative simulations against benchmark strategies (traditional MPC, TD3-MPC, and DDPG-MPC) demonstrate the method's advantages in temperature control accuracy, system energy consumption, and operational stability.

The GMA-TD3 algorithm exhibits exceptional training convergence and stability. Benefiting from the temporal modeling capabilities of the GRU and multi-head attention (MHA) mechanism, GMA-TD3 efficiently captures the dynamic dependency characteristics of the thermal management system. During training, the GMA-TD3 model achieved stable convergence in approximately 180 steps, which represents 28% and 40% faster convergence than DDPG and TD3, respectively.

Moreover, the GMA-TD3-MPC strategy significantly enhances temperature control accuracy, suppressing overheating risks. By leveraging reinforcement learning (RL) for temporal analysis of historical states, the actuator action smoothness is markedly improved. Under maximum vehicle speed conditions at standard ambient temperature, the resulting high-temperature circuit temperature is reduced by 2.1 ℃ compared to the TD3-MPC and MPC strategies, effectively preventing overheating. In off-road conditions, the amplitude of temperature fluctuations decreases by 44.19% (high-temperature circuit) and 6.45% (low-temperature circuit) compared to MPC, and by 38.46% and 23.68% compared to TD3-MPC. Furthermore, under high-temperature conditions (45 ℃), the strategy demonstrates superior robustness; in comparison to MPC, it reduces temperature fluctuation amplitudes in the high-temperature and low-temperature circuits by 11.11% and 7.89%, respectively, thereby ensuring that component temperatures remain within safe thresholds.

The GMA-TD3-MPC strategy accurately predicts thermal load variations and proactively adjusts the cooling power, thereby reducing peak power demand and cumulative energy consumption. Under typical tracked vehicle driving cycles at a standard temperature (25 ℃), the total energy consumption decreases to 18.08 kWh, representing reductions of 5.53% and 1.08% compared to MPC and TD3-MPC, respectively. At the same time, the peak power demand decreases by 10.63% compared to MPC, demonstrating the helpfulness of the temporal predictions in stabilizing energy consumption. Under extreme high-temperature conditions (45 ℃), the energy conservation advantage widens; the total energy consumption is constrained to 55.11 kWh, achieving a 13.41% reduction compared to the baseline MPC strategy.

The proposed GMA-TD3-MPC strategy not only addresses the challenge of multi-zone decoupled regulation, but also demonstrates superiority in energy consumption optimization. However, some limitations persist, such as the model's dependency on the idealized simulation environment of the *GT-SUITE* platform, which does not consider real-world factors like sensor noise and abrupt environmental changes. Future research efforts will focus on validating the real-time capability of the proposed strategy. Notably, this strategy was designed with computational efficiency in mind: online execution involves only the fast inference of the RL agent and the solution of a convex quadratic programming problem for the linear MPC, both of which can be readily embedded into the 1-second-long control cycle of the system. Conceptually, the proposed framework would be deployed as a supervisory thermal controller within the Vehicle Control Unit (VCU) or a dedicated Thermal Management Control Unit (TMCU). In the architecture of a physical vehicle, this supervisory unit would acquire real-time system states via the Controller Area Network (CAN) bus, compute the global optimal actuator speed commands, and transmit these targets to the local actuator controllers for high-frequency low-level execution. Building on this, the strategy's adaptability to extreme operating conditions should be investigated further. Additionally, considering the unique skid-steering characteristics of tracked vehicles, follow-up studies will integrate a dual-track dynamic model to explicitly analyze the asymmetric thermal loads on the left and right traction motors during steering maneuvers; this will help enhance the spatial resolution of the thermal management control. Most importantly, to bridge the gap between simulations and deployment in real vehicles, real-time hardware-in-the-loop testing should be introduced to verify the robustness of the algorithm under physical communication delays and hardware constraints.

Acknowledgments

This research was conducted at the Power Machinery and Vehicular Engineering Institute, Zhejiang University (ZJU). The authors sincerely acknowledge the financial support and assistance provided by the Basic Product Innovation Program of the State Administration for Science, Technology, and Industry for National Defence, under the Automotive Power Special Project-Research on XXX Efficient Thermal Energy Integrated Management Methods.

Author contributions

Xiao ZHENG designed the research. Xiao ZHENG, Junliang ZHAO and Yuchao YAN processed the corresponding data. Xiao ZHENG wrote the first draft of the manuscript. Zhentao LIU helped to organize the manuscript. Xiao ZHENG and Zhentao LIU revised and edited the final version.

Conflict of interest

Xiao ZHENG, Junliang ZHAO, Yuchao YAN and Zhentao LIU declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Arrinda M, Batalla J, Vicente I, et al., 2025. Model predictive control as the cloud control strategy for a battery thermal management system. *Energy Reports*, 13:4635-4646. https://doi.org/10.1016/j.egyr.2025.04.032

Atkinson C, Mccane B, Szymanski L, et al., 2021. Pseudo-rehearsal: Achieving deep reinforcement learning without catastrophic forgetting. *Neurocomputing*, 428:291-307. https://doi.org/10.1016/j.neucom.2020.11.050

Çeliksóz D, Kiliç V, 2024. Series-hybrid powertrains: Advancing mobility control in electric tracked vehicle technology. *World Electric Vehicle Journal*, 15(2) https://doi.org/10.3390/wevj15020047

Chen J-H, He P, Cai S-J, et al., 2024a. Modeling and temperature control of a water-cooled pemfc system using intelligent algorithms. *Applied Energy*, 372:123790. https://doi.org/10.1016/j.apenergy.2024.123790

Chen J-H, He P, He Z-H, et al., 2024b. Multi-objective decoupling control of thermal management system for pem fuel cell. *Energy and AI*, 18:100447. https://doi.org/10.1016/j.egyai.2024.100447

Darici S, Canli E, Dogan S, et al., 2012. Determination of heat transfer rate and pressure drop performance of an intercooler for heavy duty engines. *International Journal of Arts & Sciences*, 5:43-57.

Dersu Ç, Kılıç V, 2024. Series-hybrid powertrains: Advancing mobility control in electric tracked vehicle technology.

World Electric Vehicle Journal, 15(2):47. https://doi.org/10.3390/wevj15020047

Dong F, Wang Z, Ni J, 2020. The effect of different coolants on emissions and fuel consumption of a diesel engine. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 42(7) https://doi.org/10.1007/s40430-020-02448-6

Ghate A, Sundar A, Zhu Q, et al., 2024a. Development of an integrated energy and thermal planner for a series hybrid off-road autonomous tracked vehicle. *Energy Conversion and Management*, 322:119163. https://doi.org/10.1016/j.enconman.2024.119163

Ghate A, Sundar A, Zhu QL, et al., 2024b. Development of an integrated energy and thermal planner for a series hybrid off-road autonomous tracked vehicle ☆ . *Energy Conversion and Management*, 322 https://doi.org/10.1016/j.enconman.2024.119163

Guo R, Sun Z, Luo M, 2024. Energy-efficient battery thermal management strategy for range extended electric vehicles based on model predictive control and dynamic programming. *Energy*, 307:132769. https://doi.org/10.1016/j.energy.2024.132769

Guo X, Peng J, He H, et al., 2025. Integrated thermal-energy management for electric vehicles in high-temperature conditions using hierarchical reinforcement learning. *Expert Systems with Applications*, 276:127221. https://doi.org/10.1016/j.eswa.2025.127221

Han J, Yu S, Yi S, 2017. Advanced thermal management of automotive fuel cells using a model reference adaptive control algorithm. *International Journal of Hydrogen Energy*, 42(7):4328-4341. https://doi.org/10.1016/j.ijhydene.2016.10.134

He H, Su Q, Huang R, et al., 2024a. Enabling intelligent transferable energy management of series hybrid electric tracked vehicle across motion dimensions via soft actor-critic algorithm. *Energy*, 294:130933. https://doi.org/10.1016/j.energy.2024.130933

He P, Chen J-H, Zhang C-Z, et al., 2024b. Optimization of modeling and temperature control of air-cooled pemfc based on tlbo-de. *Energy and AI*, 18:100430. https://doi.org/10.1016/j.egyai.2024.100430

Hori Y, 2004. Future vehicle driven by electricity and control-research on four-wheel-motored "uot electric march ii". *IEEE Transactions on Industrial Electronics*, 51(5):954-962. https://doi.org/10.1109/TIE.2004.834944

Jiang KR, Tian ZB, Wen T, et al., 2025. Collaborative optimization strategy of hydrogen fuel cell train energy and thermal management system based on deep reinforcement learning. *Applied Energy*, 393 https://doi.org/10.1016/j.apenergy.2025.126057

Karaoglan MU, 2021. Drive cycle performance analysis for various powertrain alternatives of zero-emission battery electric military vehicles. *International Journal of Global Warming*, 25(3-4):482-497. https://doi.org/10.1504/ijgw.2021.119013

Liu H, Wen M, Yang H, et al., 2021. A review of thermal management system and control strategy for automotive engines. *Journal of Energy Engineering*, 147(2):03121001. https://doi.org/10.1061/(ASCE)EY.1943-7897.0000743

Ma Y, Ma Q, Liu Y, et al., 2024. Two-level optimization strategy for vehicle speed and battery thermal management in connected and automated evs. *Applied Energy*, 361:122928. https://doi.org/10.1016/j.apenergy.2024.122928

Mancarella A, Marello O, 2023. Effect of coolant temperature on performance and emissions of a compression ignition engine running on conventional diesel and hydrotreated vegetable oil (hvo). *Energies*, 16(1):144.

Naini SS, Miller RS, Rizoo D, et al., 2023. A model reference adaptive controller for an electric motor thermal management system in autonomous vehicles. *Sae International Journal of Electrified Vehicles*, 12(1):3-16. https://doi.org/10.4271/14-12-01-0001

Nguyen TT, Nguyen ND, Nahavandi S, 2020. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *Ieee Transactions on Cybernetics*, 50(9):3826-3839. https://doi.org/10.1109/tcyb.2020.2977374

Norouzi A, Heidarifar H, Borhan H, et al., 2023. Integrating machine learning and model predictive control for automotive applications: A review and future directions. *Engineering Applications of Artificial Intelligence*, 120 https://doi.org/10.1016/j.engappai.2023.105878

Scolaro E, Beligoj M, Estevez MP, et al., 2021. Electrification of agricultural machinery: A review. *IEEE Access*, 9:164520-164541. https://doi.org/10.1109/ACCESS.2021.3135037

Sundar A, Vannarath A, Prucka R, et al., 2023. The influence of cooling air-path restrictions on fuel consumption of a series hybrid electric off-road tracked vehicle. *SAE Technical Paper Series*,

Sundar A, Ghate A, Zhu Q, et al., 2026. An integrated power, energy, and thermal management strategy using cascaded control for off-road autonomous hybrid vehicles. *Applied Energy*, 406:127269. https://doi.org/10.1016/j.apenergy.2025.127269

Tang A, Wu Z, Xu T, et al., 2024. Week-level early warning strategy for thermal runaway risk based on real-scenario operating data of electric vehicles. *eTransportation*, 19:100308. https://doi.org/10.1016/j.etran.2023.100308

Tian J, Jing H, Xiong Y, et al., 2026. Experimental study of the effect of electrode geometry on the ignition and flame development of nh3/air mixtures in nanosecond plasma discharges. *Combustion and Flame*, 286:114809. https://doi.org/10.1016/j.combustflame.2026.114809

Wang D, Zheng W, Li S, et al., 2025a. Impact analysis of uncertainty in thermal resistor-capacitor models on model predictive control performance. *Energy and Buildings*, 328:115112. https://doi.org/10.1016/j.enbuild.2024.115112

Wang F, Harindintwali JD, Yuan ZZ, et al., 2021. Technologies and perspectives for achieving carbon neutrality. *Innovation*, 2(4) https://doi.org/10.1016/j.xinn.2021.100180

Wang T, Wu J, Wu J, et al., 2025b. Energy consumption analysis and performance evaluation of electric vehicle integrated thermal management system experiments. *Applied Thermal Engineering*, 269:126002. https://doi.org/10.1016/j.applthermaleng.2025.126002

Wu B, Jia Z, Li ZG, et al., 2021. Different exhaust temperature management technologies for heavy-duty diesel engines with regard to thermal efficiency. *Applied Thermal Engineering*, 186:116495. https://doi.org/10.1016/j.applthermaleng.2020.116495

Xi Y-G, Li D-W, Lin S, 2013. Model predictive control — status and challenges. *Acta Automatica Sinica*, 39(3):222-236. https://doi.org/10.1016/S1874-1029(13)60024-5

Yu Y, Jiang J, Min H, et al., 2023. Integrated energy and thermal management strategy for extended range electric vehicle based on battery temperature and state-of-charge global planning. *Energy Conversion and Management*, 288:117154. https://doi.org/10.1016/j.enconman.2023.117154

Yuan R, Fletcher T, Ahmedov A, et al., 2020. Modelling and co-simulation of hybrid vehicles: A thermal management perspective. *Applied Thermal Engineering*, 180:115883. https://doi.org/10.1016/j.applthermaleng.2020.115883

Zhai L, Yan J, Liu A, et al., 2025. Structure design of stator oil circuit and thermal characteristics analysis for high-speed oil-cooled flat wire motor. *Transactions of Beijing institute of Technology*, 45(6):600-608. https://doi.org/10.15918/j.tbit1001-0645.2023.187

Zhang J, Liu B, Zhou Y, et al., 2025. Hierarchical energy management strategy for fuel cell hybrid electric vehicles based on improved td3 considering interior temperature comfort. *Applied Thermal Engineering*, 276:126965. https://doi.org/10.1016/j.applthermaleng.2025.126965

Zhao X, Ma X, Chen B, et al., 2022. Challenges toward carbon neutrality in china: Strategies and countermeasures. *Resources Conservation and Recycling*, 176 https://doi.org/10.1016/j.resconrec.2021.105959

Zhu Q, Kumar A, Sundar A, et al., 2022. Development of a series hybrid electrified powertrain for a high speed tracked vehicle based on driving cycle simulation. *SAE International Journal of Advances and Current Practices in Mobility*, 4 https://doi.org/10.4271/2022-01-0367

Electronic supplementary materials

Section S1–S3, Table S1–S6, Figs. S1–S11

中文概要

题　目：基于 GMA-TD3-MPC 算法考虑系统安全与能耗的混合动力履带车辆热管理策略

作　者：郑潇，赵俊良，严宇超，刘震涛

机　构：浙江大学，动力机械与车辆工程研究所，中国杭州，310027

目　的：混合动力履带车辆（HETVs）在复杂工况下面临严峻的热安全和能耗挑战，传统热管理策略难以兼顾高精度温度控制与多目标协同优化。本文旨在探讨一种集成门控循环单元、多头注意力、双延迟深度确定性策略梯度与模型预测控制（GMA-TD3-MPC）的分层热管理框架，以协同优化温度控制精度、系统能耗和运行安全。

创新点：1. 将深度强化学习（GMA-TD3）与模型预测控制（MPC）进行动态整合，通过动态阈值触发机制在接近热安全边界时无缝切换至MPC控制器，兼顾了数据驱动的全局寻优与基于模型的局部边界安全性；2. 在TD3算法中引入门控循环单元（GRU）与多头注意力（MHA）机制，有效提取强时间序列依赖特征并强化关键决策信息，克服了传统强化学习过滤次优动作时的局限性。

方　法：1. 通过建立GT-SUITE物理仿真模型，并采用广义回归神经网络（GRNN）建立发动机相关热源组件的散热预测模型，实现对多物理场耦合热环境的精确映射（图1和图2）；2. 通过理论推导与奖励函数设计，构建GMA-TD3-MPC分层架构，运用动作选择模块和混合函数实现兼顾安全及能耗的温度控制（图3、公式(1)-(7)）；3. 通过引入典型履带车辆驾驶循环工况，在标准温度（25℃）和高温（45℃）下进行联合仿真验证，对比分析所提策略在温度控制及能耗上的可行性和有效性（图5、图6、图7、图8）。

结　论：1. 引入GRU与MHA机制的GMA-TD3算法具备优异的训练收敛性，收敛速度较DDPG和TD3分别提升了约28%和40%；2. 运用GMA-TD3-MPC适应性策略后，温度控制精度显著提升，在标准环境越野工况下，高、低温回路的温度波动幅度较单独MPC分别降低了44.19%和6.45%，有效避免了系统过热风险；3. 适应性控制策略实现了能耗的大幅优化，25℃工况下总能耗较MPC降低5.53%，峰值功率需求降低10.63%；而在45℃高温极端工况下，总能耗大幅下降了13.41%，热管理系统整体性能得到显著提高。

关键词：热管理策略；混合动力履带车辆；带有门控循环单元的多头注意力机制；分层控制；能耗优化