Electronic Supplementary Materials

For 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

Section S1 Vehicle and thermal management system modEL

Section S1.1 Description of HETVs

Table S1 Key parameters of the HETVs

Parameters	Value	Unit
Mass	25000	[kg]
Effective frontal area	4.9	[m^2]
Aerodynamic coefficient	0.6	[-]
Sprocket radius	0.31	[m]
Coefficient of rolling friction	0.065	[-]
Driveline efficiency	0.88	[-]
Final drive ratio	4	[-]
Energy storage unit	28kWh, 800V	[kWh, V]
Motor maximum power	390	[kW]
ICE-Gen maximum power	790	[kW]
Top speed (Road)	85	[km/h]
Top speed (Off-road)	65	[km/h]

Table S1 presents the key parameters of the HETVs. The platform employs an 800 V DC bus architecture with a 28 kWh energy storage unit. Each track is driven by an independent motor-gearbox-sprocket assembly. The powertrain core integrates an engine-generator set rated at 790 kW peak power and traction motors delivering 390 kW maximum power each. To accommodate dynamic operational loads while maintaining energy storage state-of-charge (SOC), a power-following control strategy governs energy flows: The engine operates at constant speed, while the generator's output torque modulates in real time according to total DC bus power demand to achieve dynamic energy equilibrium. During transient high-load operations (e.g., climbing or steering), the generator instantly increases torque output, while the energy storage unit provides compensatory power via the DC/DC converter. Under reduced loads, excess energy charges the storage unit to stabilize SOC. This strategy ensures instantaneous power response during extreme maneuvers while optimizing energy utilization efficiency.

Section S1.2 Modeling the powertrain system of HETVs

During operation, the vehicle primarily encounters the following resistances: track rolling resistance F_{roll}, air resistance F_{air}, acceleration resistance F_{acc}, and gradient resistance F_{grad}. Therefore, the vehicle propulsion force F_{ranc} is expressed as:

$$\begin{aligned} F_{ranc} &= F_{roll} + F_{\mathrm{air}} + F_{\mathrm{acc}} + F_{\mathrm{grad}} \\ &= \mu_{\mathrm{roll}} M g \cos\alpha + \frac{1}{2}\rho C_d A v^2 + \delta M \frac{dv}{dt} + M g \sin\alpha \end{aligned} \tag{S1}$$

where M denotes the total vehicle mass; g represents gravitational acceleration; μ_{roll} is the track-ground rolling friction coefficient; α indicates the road gradient angle; ρ is air density; C_d denotes the aerodynamic drag coefficient; A is the vehicle frontal area; v represents the travel speed; and δ is the equivalent coefficient of rotational mass.

Based on the driving sprocket radius $r_{sprocket}$, propulsion force F_{ranc}, and travel speed v, the traction motor speed ω_m and torque T_m are expressed as:

$$\omega_m = \frac{v}{r_{sprocket}} \cdot i_{final} \tag{S2}$$

$$T_m = \frac{F_{trac} \cdot r_{sprocket}}{i_{final} \cdot \eta_{driveline}} \tag{S3}$$

where i_{final} denotes the final reduction ratio of the side transmission assembly, and $\eta_{driveline}$ represents the mechanical efficiency of the driveline.

Figs. S1(a) and **S1(b)** depict the efficiency maps of the ICE-Generator set and the drive motors. As the primary mechanical power source, the drive motors deliver propulsion torque or recovers braking energy. This study employs a Permanent Magnet Synchronous Motor (PMSM) rated at 390 kW peak power. Its efficiency map $\eta_m(T_m, \omega_m)$ is derived from experimental data fitting. The motor's electrical power demand P_{elec} is defined as:

$$P_{elec} = \frac{T_m \omega_m}{\eta_m(T_m, \omega_m)} \tag{S4}$$

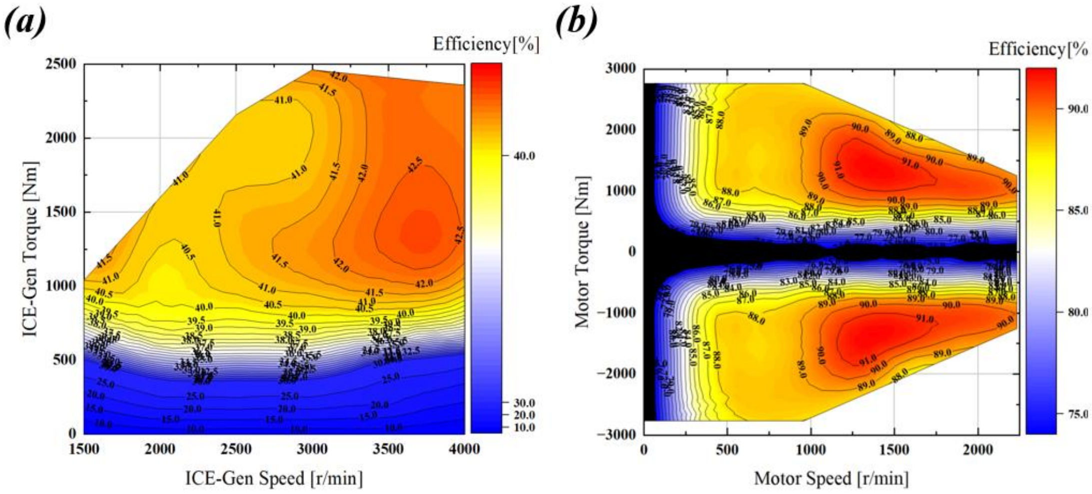

Fig. S1 **(a)** The efficiency maps of the ICE-Generator set. **(b)** The efficiency maps of the drive motors.

The turbocharged diesel engine in this study is mechanically coupled to a generator for electricity generation. The engine delivers mechanical power, P_{ICE}, which is transmitted via a transmission mechanism to drive the generator. The magnitude of this power is determined by both torque T_{ICE} and rotational speed ω_{ICE} :

$$P_{ICE} = \frac{T_{ICE} \cdot \omega_{ICE}}{9550} \tag{S5}$$

The generator converts mechanical energy into electrical energy, which is subsequently rectified and fed into the DC bus. The generator mechanical power input P_{gen} is given by:

$$P_{gen} = \eta_{gen} \cdot P_{ICE} + P_{\text{loss},gen} \tag{S6}$$

where: η_{gen} represents the generator's electromechanical conversion efficiency. $\eta_{gen} \cdot P_{ICE}$ is the electrical power output, and $P_{\text{loss},gen}$ denotes the power loss of the generator due to copper and iron losses.

The energy storage unit serves as a power buffer, with its output/input power determined by the DC bus power balance:

$$P_{batt} = \frac{1}{\eta_{DC/DC}} (P_{bus} - \eta_{DC/AC} \cdot P_{gen}) \tag{S7}$$

$$P_{bus} = P_{motor,L} + P_{motor,R} + P_{ACS} \tag{S8}$$

where: P_{bus} represents the total power demand on the DC bus, $\eta_{DC/AC}$ denotes the DC-AC conversion efficiency, and $P_{motor,L}$, $P_{motor,R}$, P_{ACS} correspond to the power consumption of the left motor, right motor, and auxiliary systems, respectively.

The dynamic model of the energy storage unit is:

$$SOC(t) = SOC_0 - \frac{1}{C_b} \int_0^t \frac{P_{batt}(\tau)}{V_{oc}(SOC)} d\tau \tag{S9}$$

where: C_b is the energy storage unit capacitance, and V_{oc} is the open-circuit voltage.

The vehicle-level power balance must satisfy the following constraint:

$$\eta_{DC/AC} \cdot P_{gen} + \eta_{DC/DC} \cdot P_{batt} = P_{motor,L} + P_{motor,R} + P_{ACS} \tag{S10}$$

That means the net power injected into the DC bus from the generator-produced AC power after rectification via the DC-AC converter, represented as $\eta_{DC/AC} \cdot P_{gen}$, combined with the net power injected from the energy storage unit output power P_{batt} after DC-DC conversion, expressed as $\eta_{DC/DC} \cdot P_{batt}$, equals the total power drawn from the DC bus by the left and right drive motors as well as the auxiliary systems.

Section S1.3 Modeling of the TMS for HETVs

The high-temperature circuit primarily includes a high-temperature radiator, engine, first-stage intercooler, interstage intercooler, expansion tank, and circulating water pump. This circuit primarily serves heat sources such as the engine cylinder liner and intercoolers. Coolant absorbs heat while flowing through parallel heat sources (including the engine), then is driven by the pump through the circuit. The low-temperature circuit mainly consists of a low-temperature radiator, expansion tank, circulating water pump, second-stage intercooler, left/right drive motors, oil cooler, and multiple parallel heat sources. Electronically controlled proportional valves are installed at the inlets of the parallel branches. Coolant flows from the outlet of the low-temperature radiator, is driven by the pump, first passes through the second-stage intercooler, then enters the parallel heat source cluster (including motors, generator, controllers, transmission oil heat exchangers, etc.). During hybrid operation, all electronic proportional valves on parallel branches remain open. Finally, coolant converges, flows through the oil cooler, and returns to the low-temperature radiator. Both radiators (low-temperature radiator positioned front, high-temperature radiator rear) share two fans driven by fan motors. Forced airflow first passes through the low-temperature radiator for cooling, then the partially warmed airflow continues through the high-temperature radiator for secondary cooling.

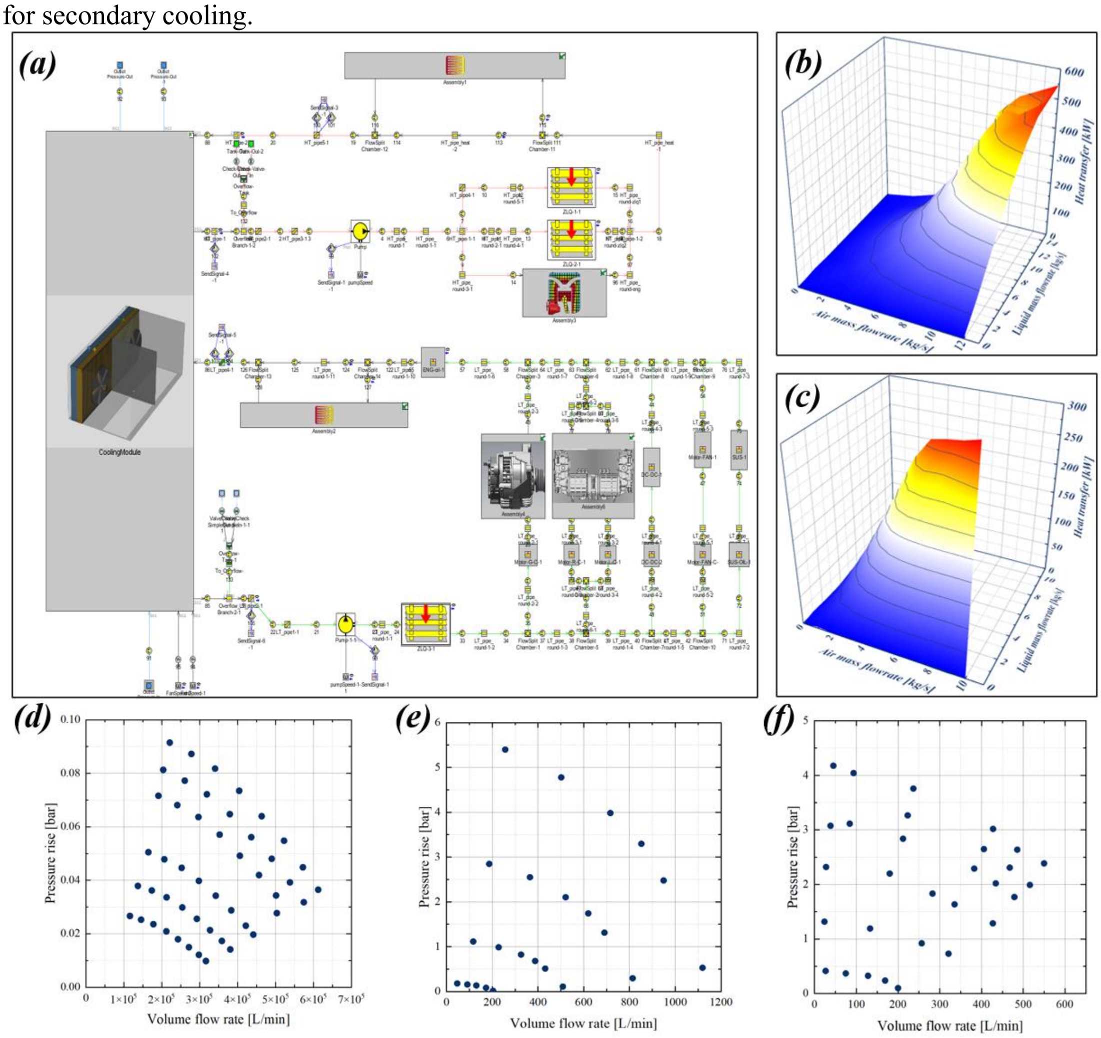

Fig. S2 **(a)** TMS simulation model. **(b)** High-temperature circuit radiator coolant-to-air flow rate–heat dissipation MAP. **(c)** Low-temperature circuit radiator coolant-to-air flow rate–heat dissipation MAP. **(d)** Fan flow rate–pressure rise MAP. **(e)** High-temperature circuit pump head–flow rate MAP. **(f)** Low-temperature circuit pump head–flow rate MAP.

Fig. S2(a) depicts the established TMS simulation model. The GT-SUITE TMS model adopts a modular modeling philosophy, with all key components constructed based on their physical mechanisms. Coolant properties are simulated using GT-SUITE's egl-5050 solution model, while air properties employ a standard air model.
The high-temperature and low-temperature circuit radiators are meticulously modeled using COOL3D. The model thoroughly incorporates three-dimensional core geometries, louver fin characteristics, and coolant channel layouts. Key input parameters include radiator core dimensions, fin specifications, flat tube parameters, and empirical radiator test data. Coupled heat transfer calculations between the air-side and coolant-side are performed using the Effectiveness-Number of Transfer Units method, grounded in fundamental heat transfer and fluid dynamics equations. The resulting flow rate–heat dissipation MAPs for coolant-side and air-side in both temperature circuits are illustrated in **Fig. S2(b)** and **S2(c)**.
Fan performance modeling utilizes the MatrixFan module within COOL3D, incorporating experimentally measured performance data obtained from test benches. Geometric parameters and performance curves are input to the model, which dynamically computes actual airflow and efficiency during simulations through interpolation algorithms based on real-time motor speed and pressure rise. The flow rate–pressure rise MAP for the fan is shown in **Fig. S2. (d)**. The calculation methodology for fan power $\dot{W_{fan}}$ is as follows:

$$\Delta h_{airs} = c_{air} T_{inlet} (PR^{\frac{\gamma-1}{\gamma}} - 1) \tag{S11}$$

$$\dot{W}_{fan} = \dot{m}_{air} \frac{\Delta h_{airs}}{\eta_{fan}} \tag{S12}$$

where Δh_{airs} denotes the isentropic enthalpy change of air, c_{air} represents the specific heat of air, T_{inlet} is the inlet total temperature, PR stands for pressure ratio, γ is the specific heat ratio of air $\dot{m}_{air}$, indicates the air mass flow rate, and η_{fan} signifies the fan's isentropic efficiency.
The high and low-temperature circuit water pumps are modeled using Pump Elements, incorporating experimentally measured performance data from test benches. By inputting pump speed, flow rate, pressure rise, and efficiency data at each operating point, interpolation calculations are performed. The head–flow rate MAPs for the pumps in both circuits are shown in **Fig. S2. (e)** and **(f)**. During simulation, the inputs for each time step are pump speed and pressure rise. Based on the interpolated pump MAP, the flow rate and efficiency are determined. The flow rate is then directly applied to the pump's flow boundary. The efficiency influences the pump's mechanical power consumption, calculated as follows:

$$\Delta h_{liquids} = \frac{P_{out} - P_{in}}{\rho_{in}} \tag{S13}$$

$$\dot{W}_{pump} = \frac{\dot{m}_{liquid} \Delta h_{liquids}}{\eta_{pump}} \tag{S14}$$

where P_{out} and P_{in} denote the static pressure at the pump outlet and inlet, respectively, ρ_{in} represents the coolant density at the inlet, $\dot{m}_{liquid}$ indicates the mass flow rate of coolant through the pump, and η_{pump} signifies the pump's isentropic efficiency.

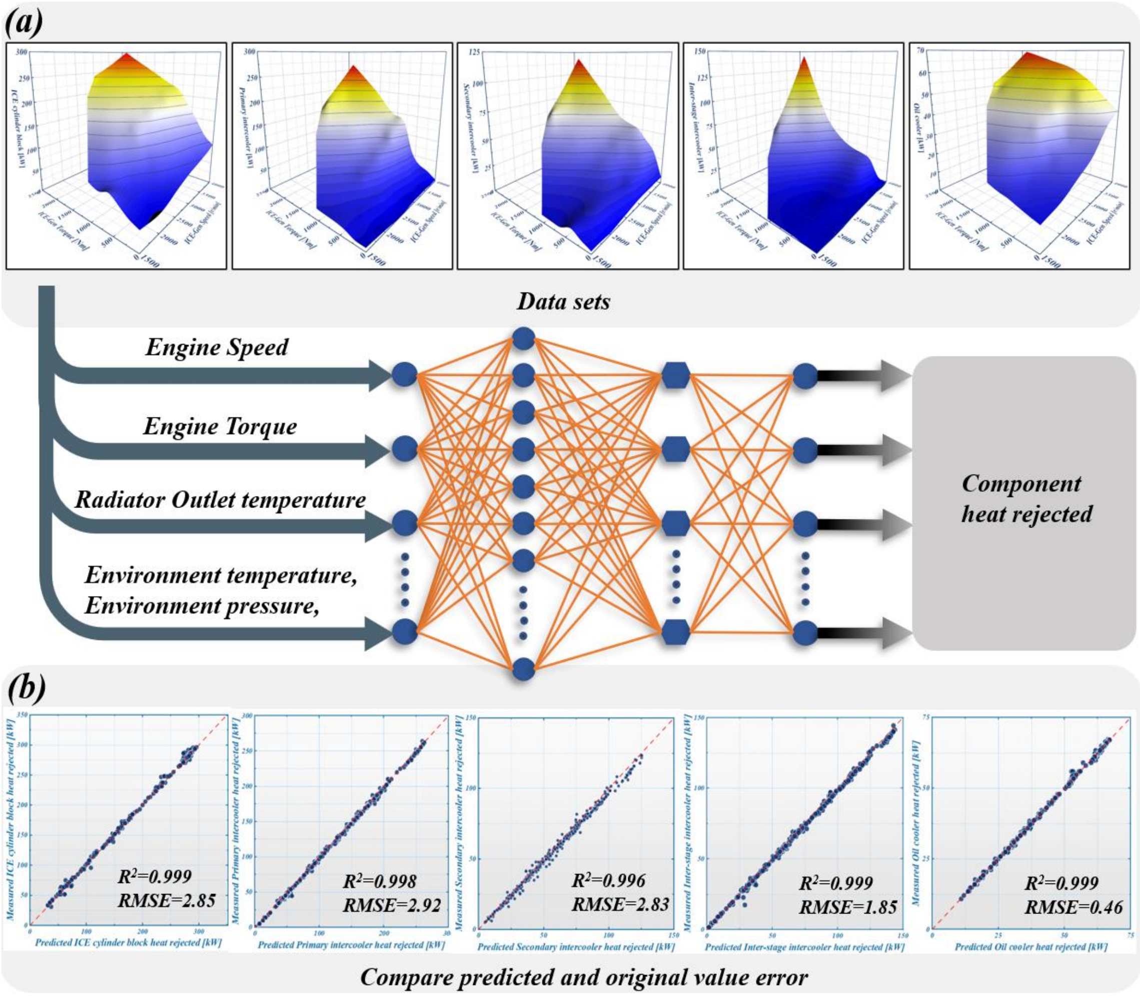

Fig. S3 **(a)** The Heat rejection data to coolant and predictive model schematic for the engine-related heat sources. **(b)** Scatter plot of predicted values and measured values

Fig. S3(a) illustrates the Heat rejection data to coolant and predictive model schematic for the engine-related heat sources (including the engine block, first-stage intercooler, second-stage intercooler, interstage intercooler, and oil cooler). The core input for the thermal management system model is the heat rejected by these components to the coolant. Given the significant nonlinearity and multivariable coupling characteristics inherent in engine-related heat sources, this study employs a Generalized Regression Neural Network (GRNN) to establish the prediction model. Network parameters are detailed in the **Table S2**. Compared to conventional empirical formulas, the GRNN delivers high-precision regression through non-parametric kernel density estimation [1]. The conditional mean at the output layer is expressed as:

$$\hat{Q}_{heat} = \frac{\sum_{i=1}^{N} Q_i \, exp(-\frac{|E - E_i|^2}{2\sigma^2})}{\sum_{i=1}^{N} exp\,(-\frac{|E - E_i|^2}{2\sigma^2})} \tag{S15}$$

where: $E = [T_{eng}, \omega_{eng}, Temp_{out}, Temp_{env}, Pres_{env}]$ denotes the input eigenvector (with T_{eng} engine torque, ω_{eng} engine speed, $Temp_{out}$ heat dissipation outlet temperature, $Temp_{env}$ ambient temperature, $Pres_{env}$ ambient pressure). E_i represents the i-th training sample input point, Q_i is the corresponding measured heat rejection data to coolant, σ is the smoothing parameter, and N is the number of training samples.

Table S2 General Regression Neural Network model parameters

P (input vectors)	T (target class vectors)	σ(Spread)
[1146x5]	[1146x5]	0.13

The heat rejection dataset to coolant is derived from bench test measurements on the powertrain platform, with operating conditions extended via a GT-Suite high-fidelity combustion-heat transfer coupled model. The neural network comprises an input layer, pattern layer, summation layer, and output layer for predicting heat rejection of each engine-related component. Its topology adheres to the Parzen window kernel density estimation framework, maintaining smooth nonlinear mapping and generalization capabilities with limited samples. Effectiveness of this architecture in thermal management prediction tasks has been empirically validated in comparable studies by Ghate et al. and Koli et al. [2, 3]. As shown in **Fig. S3. (b)**, the predicted values exhibit a coefficient of determination R^2 exceeding 0.996 against test values, with the maximum root mean square error (RMSE) constrained within 2.92. This meets the accuracy requirements for thermal state estimation in the TMS. For components demonstrating quasi-static linear or mildly nonlinear thermal characteristics, such as left/right drive motors, fan motors, and controllers, where monotonic mappings exist between coolant heat rejection and operating parameters, high-precision look-up tables are established through limited calibration points. Linear interpolation within operational envelopes maintains errors within acceptable thresholds. Pressure drops across all heat source components are determined via look-up tables.

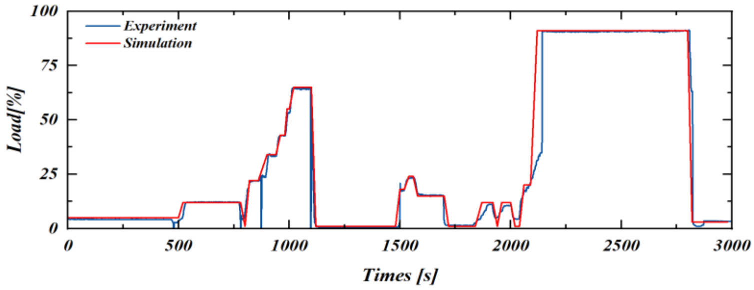

Fig. S4 Experimental vs. simulation loading profile comparisons

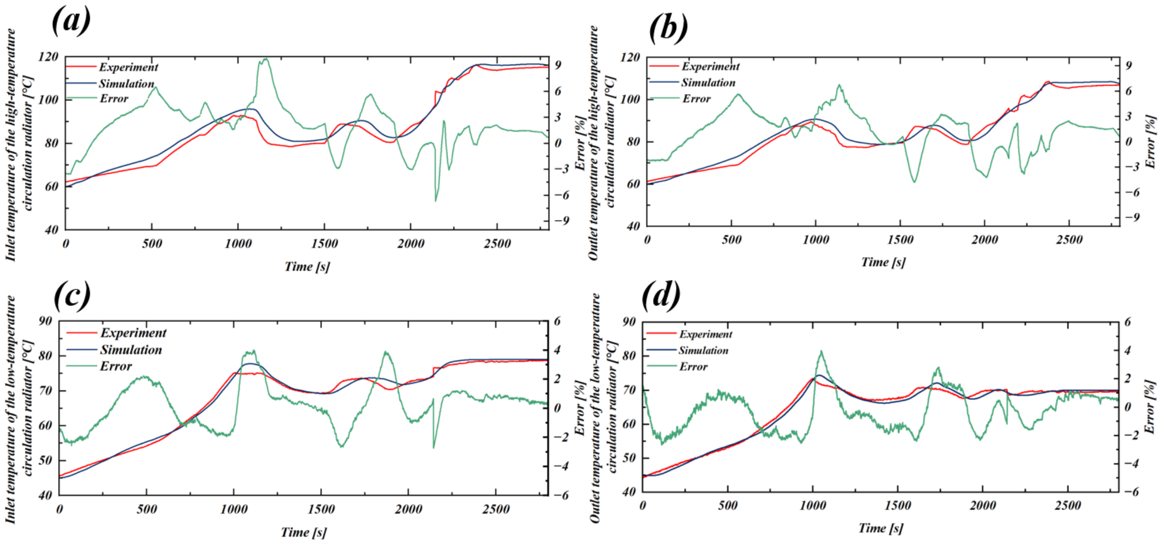

Fig. S5 **(a)** Experimental vs. simulated inlet temperature comparison for the high-temperature circuit radiator. **(b)** Experimental vs. simulated outlet temperature comparison for the high-temperature circuit radiator. **(c)** Experimental vs. simulated inlet temperature comparison for the low-temperature circuit radiator. **(d)** Experimental vs. simulated outlet temperature comparison for the low-temperature circuit radiator.

Fig. S4 and **Fig. S5** present comparative results of loading profiles and temperature distributions between experimental measurements and simulation outputs to validate model accuracy. Under full-load conditions defined by the maximum ICE-Gen output, identical loading profiles were applied to both experimental tests and simulations. Data acquisition occurred at 25°C ambient temperature with 1s sampling intervals. As illustrated in **Fig. S5(a)** and **S5(b)**, simulations demonstrated average temperature deviations of 4.42% at the inlet and 3.75% at the outlet of the high-temperature circuit radiator compared to experimental values. Similarly, **Fig. S5(c)** and **S5(d)** show average deviations of 1.32% and 1.23% for the low-temperature circuit radiator. Comparative analysis revealed smaller discrepancies between simulated and actual measurements during steady-state loading conditions relative to transient loading operations. Notably, 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 subsequent control strategy research utilizing simulation models.

Section S2 Safety- and energy-conscious thermal management strategy

Section S2.1 GMA-TD3 algorithm

RL abstractly frames the closed-loop control problem as a Markov Decision Process (MDP), defined by the tuple:

$$M = \langle S, A, p, r, \gamma \rangle, \tag{S16}$$

where S is the state space, A is the action space, p denotes the state transition probability, r is the immediate reward, and γ represents the discount factor [4]. The agent interacts with the environment through a parameterized policy $\pi_\theta(a|s)$, aiming to maximize expected cumulative discounted returns. However, conventional RL methods exhibit limitations when handling high-dimensional continuous states with temporal dependencies [5]. While the DDPG algorithm addresses this via an Actor-Critic architecture with function approximators, it suffers from Q-value overestimation, ineffective capture of temporal state correlations, and hyperparameter sensitivity [6]. To overcome these drawbacks, the TD3 algorithm employs two structurally identical yet independently parameterized Critic networks $Q_{\theta_1}, Q_{\theta_2}$, computing the target Q-value using their minimum:

$$y = r + \gamma \cdot \min_{i=1,2} Q'_{\theta_i}(s', \tilde{a}') \tag{S17}$$

where Q'_{θ_i} denotes target network parameters, and $\tilde{a}'$ represents the noisy target action. This action is generated by the target Actor with additive truncated Gaussian noise $\varepsilon \sim N(0, \sigma)$ to smooth policy updates. The min operator updates Critic networks, while delayed policy updates mitigate sensitivity to policy drift. Together, these suppress Q-value overestimation and enhance convergence stability [7].

To address temporal data modeling challenges and enhance feature representation capabilities, GMA-TD3 introduces a Gated Recurrent Unit (GRU) and Multi-Head Attention (MHA) mechanism atop TD3's twin-Critic architecture, forming a hierarchical feature extraction structure [8]. A GRU layer is embedded after the input layer to process historical state sequences $s_{t-k:t}$.Leveraging update and reset gate mechanisms, the GRU calculates the hidden state $h_t = GRU(s_{t-k:t})$, effectively capturing temporal dependencies in historical heat dissipation rates, temperatures, and thermal fluctuations. This GRU-derived hidden state sequence is then fed into the Multi-Head Attention (MHA) module, which performs self-attention computation:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{S18}$$

where queries (Q), keys (K), and values (V) are obtained via linear projections and divided into multi-head structures to capture critical temporal information across multiple scales. The multi-head outputs are concatenated and fused through a linear layer, generating a weighted feature representation h_{MHA}. Finally, h_{MHA} is concatenated with the current state s_t and integrated into a perceptive feature vector via fully connected layers:

$$s_{int} = \phi([h_{MHA}; s_t]) \tag{S19}$$

The Actor network generates continuous actions a_t while Critics evaluate $Q(s_{int}, a_t)$. By preserving local sequential patterns via GRU and strengthening long-term dependencies while highlighting critical decision points via MHA, this architecture explicitly fuses temporal-correlated features. This enhances policy generalizability, robustness, and energy efficiency in dynamic TMS.

In this study, the state vector includes outlet temperatures of high/low-temperature circuits radiators, ambient temperature/pressure, total heat rejection to coolant in both circuits, real-time fan and pump power. The trained Actor policy network generates actions. The output layer uses Tanh activation followed by a linear scaling layer to normalize actions to [0,1], subsequently rescaled into speed commands for both fans and high/low-temperature circuit pumps.

During training, the agent learns state-to-action mapping motivated by optimization of a designed reward function. In the initial exploration phase, high uncertainty due to limited prior knowledge necessitates active environmental exploration to discover high-reward actions. Action selection directly links to received rewards, requiring efficient trial-and-error learning. The core objective is to minimize total power consumption under temperature constraints. The immediate reward function is designed as:

$$\begin{gathered} r_t = \omega_1 |T_{high,target} - T_{high,in}| + \omega_2 |T_{low,target} - T_{low,out}| + \omega_3 (P_{fan} + P_{pump}) \\ + I_1 \omega_4 (T_{high,max} - T_{high,in}) + I_2 \omega_5 (T_{low,max} - T_{low,out}) + I_3 \omega_6 \left(\frac{du_{fanspeed}}{dt}\right) \\ I_4 \omega_7 \left(\frac{du_{pumpspeed}}{dt}\right) + I_5 C \end{gathered} \tag{S20}$$

In which ω_i is the weight for each reward component; $T_{high,target}$ and $T_{low,target}$ are the temperature target values for the high-temperature circuit and low-temperature circuit, respectively; $T_{high,max}$ and $T_{low,max}$ are the maximum temperature limits for the high-temperature circuit and low-temperature circuit, respectively; $T_{high,in}$ and $T_{low,out}$ 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; $\frac{du_{fanspeed}}{dt}$ and $\frac{du_{pumpspeed}}{dt}$ 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 0 otherwise. This equation serves as the core mechanism for handling the coupling of conflicting objectives. 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 weights for each part and the judgment function are detailed in the **Table S3**. The selection of weights and thresholds follows a magnitude balancing principle to address the scale disparity between different physical units [9, 10]. Since the numerical value of power consumption is significantly larger than temperature deviation, the weight ω_3 is set to a lower order of magnitude compared to temperature weights ω_1 and ω_2 to normalize their contributions to the total reward. Furthermore, the penalty weights ω_4 and ω_5 are assigned the highest relative values to establish a safety-priority hierarchy, ensuring that the agent prioritizes avoiding thermal violations over energy optimization during the exploration phase. While these specific numerical weights are tailored to the current platform, the strategy exhibits strong portability as long as the relative priority hierarchy is maintained. Consequently, applying this strategy to a different vehicle configuration would simply necessitate re-scaling the weights according to the same magnitude balancing principle to match the new physical parameter ranges, rather than requiring exhaustive searches for precise absolute values. Sensitivity tests conducted during the pilot phase confirmed that small variations in these balanced weights do not significantly alter the convergence behavior.

Table S3 The weights for each part and the judgment function

Weights	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
Values	-0.5	-0.5	-0.04	-1	-1	-0.008	-0.006
Judgment	I_1	I_2	I_3	I_4	I_5		
Function	$T_{high,in}$ ⩾120℃	$T_{low,out}$ ⩾75℃	$\frac{du_{fanspeed}}{dt}$⩾ 500	$\frac{du_{fanspeed}}{dt}$⩾ 400	$T_{high,in}$⩾ 130℃; $T_{low,out}$ ⩾85℃		

The interaction experience data tuples between the agent and environment are stored in the experience replay memory. During model updates, mini-batches are randomly sampled from it for iterative optimization of the parameters of the Critic network and the actor network. The hyperparameters of the network are detailed in the **Table S4**.

Table S4 The hyperparameters of the network

Parameters	**Values**	**Parameters**	**Values**
Learning rate for the actor network	2e-4	Learning rate for the critic network	5e-4
Learning rate decay factor	5e-3	Discounting factor	0.99
Exploration Noise Standard Deviation	0.2	Exploration Noise Mean Attraction Constant	0.1
Experience Buffer Length	2e6	Minibatch size	512
Policy Update Frequency	3	Sequence Length	256

Section S2.2 MPC Algorithm

MPC predicts dynamic behavior over a future prediction horizon using the system's state-space model. For the thermal management system, augmented state vectors are established respectively for the fan and the high/low-temperature circuit pumps, integrating temperature errors and historical control inputs:

$$\xi_1(k) = [\Delta T_{high}(k), \Delta T_{low}(k), u_{fan}(k-1)]^T \tag{S21}$$

$$\xi_2(k) = [\Delta T_{high,in-out}(k), u_{high,pump}(k-1)]^T \tag{S22}$$

$$\xi_3(k) = [\Delta T_{low,in-out}(k), u_{low,pump}(k-1)]^T \tag{S23}$$

where $\Delta T_{high}(k)$ (k)and $\Delta T_{low}(k)$ denote the weighted errors between the outlet temperatures of the high-temperature circuit radiator and the low-temperature circuit radiator relative to their target temperatures, respectively:

$$\Delta T_{high}(k) = \left(T_{high,target} - T_{high,out}\right) \cdot K_1, \tag{S24}$$

$$\Delta T_{low}(k) = \left(T_{low,target} - T_{low,out}\right) \cdot K_2 \tag{S25}$$

$\Delta T_{high,in-out}(k)$ and $\Delta T_{low,in-out}(k)$ denote the weighted errors between the actual and target inlet-to-outlet temperature differences of the high-temperature circuit radiator and low-temperature circuit radiator, respectively:

$$\Delta T_{high,in-out}(k) = (\left(T_{high,out} - T_{high,in}\right) - T_{high,in-out,target}) \cdot K_3 \tag{S26}$$

$$\Delta T_{low,in-out}(k) = (\left(T_{low,out} - T_{low,in}\right) - T_{low,in-out,target}) \cdot K_4 \tag{S27}$$

The terms K_1, K_2, K_3 and K_4 represent the scaling coefficients designed to adjust their relative weights within the state vector, to ensure balanced tracking performance between the two cooling circuits.

The discretized state space equation are:

$$\xi(k+1) = A_d\xi(k) + B_d\Delta u(k) \tag{S28}$$

$$y(k) = C_d\xi(k) \tag{S29}$$

Among them, A_d, B_d, C_d denote the discrete system matrices obtained by discretizing the continuous model using the forward Euler method. The continuous-time state-space model is discretized with a sampling time of $T_s = 1$s. Although thermal components exhibit different time constants, the prediction model focuses on the macroscopic temperature dynamics of the coolant and component masses, which are characterized by large thermal inertia (i.e., relatively long time constants) [11]. The selected sampling step T_s satisfies the numerical stability condition. Moreover, compared to implicit discretization methods (such as the Tustin or backward Euler method), the explicit Euler formulation significantly reduces the computational complexity of updating the Jacobian matrix in each optimization iteration, thereby facilitating the real-time execution of the algorithm [12].

The system output is predicted over N_p future steps, and an optimal control sequence is generated by solving a Quadratic Programming (QP) problem. Since the control horizon N_c is typically chosen to be smaller than the prediction horizon N_p ($N_c < N_p$) to reduce computational burden, it is explicitly assumed that the control actions remain constant beyond the control horizon. That is, the control increments $\Delta u(k+i)$ are set to zero for all i in the range $[N_c, N_p - 1]$. The objective function minimizes the state tracking error within the prediction horizon:

$$J = \sum_{i=1}^{N_p} \left\| y(k+i) - y_{ref} \right\|_Q^2 + \sum_{i=0}^{N_c-1} \|\Delta u(k+i)\|_R^2 + \rho \cdot \varepsilon^2 \tag{S30}$$

Here, Q is the state error weighting matrix, R the control increment weighting matrix, ρ is the weighting coefficient for the slack variable, and ε denotes the non-negative slack variable. The inclusion of ε converts hard state constraints into soft constraints, thereby ensuring the feasibility of the QP solver even under extreme disturbance conditions.

The quadprog QP solver is invoked at each control cycle:

$$min_{\Delta U} \frac{1}{2} \Delta U^T H \Delta U + f^T \Delta U \tag{S31}$$

where $H = \Theta^T Q \Theta + R, f = 2(\Phi\xi)^T Q \Theta$, with Θ and Φ being the prediction model matrices. After solving, the first element of the control sequence $\Delta u_i^*(k)$ is applied to update the actuators:

$$u_{fan}(k) = u_{fan}(k-1) + \Delta u_{fan}^*(k) \tag{S32}$$

$$u_{high,pump}(k) = u_{fan}(k-1) + \Delta u_{high,pump}^*(k) \tag{S33}$$

$$u_{low,pump}(k) = u_{fan}(k-1) + \Delta u_{low,pump}^*(k) \tag{S34}$$

Here, $u_{fan}(k), u_{high,pump}(k), u_{low,pump}(k)$ represent the output values of the fan, high-temperature pump, and low-temperature pump, respectively. The parameters for MPC are detailed in the **Table S5**.

Table S5 The parameters for MPC

Parameters	Values
Prediction Horizon (Np)	50
Control Horizon (Nc)	15
$K_1/K_2/K_3/K_4$	-0.0022/ -0.002/ -0.0017/ -0.0015
State Weighting Matrix	10*eye(Nx*Np)
Control Weighting Matrix	3*eye(Nu*Nc)
Slack Factor	100

Section S3 Results and discussion

This section analyzes TMS strategies incorporating safety and energy consumption considerations, validating the superiority of the proposed approach. Parameters for all strategies are uniformly configured and tested under identical conditions to ensure comparability. The selection of baseline methods follows a strict comparative logic. Specifically, DDPG and TD3 are selected to represent the evolutionary predecessors of the proposed algorithm, verifying the architectural benefits of the GMA module. It is important to note that the standalone MPC baseline is configured to be structurally identical to the bottom-layer controller of the proposed GMA-TD3-MPC framework. Although utilizing a Nonlinear MPC (NMPC) with an explicit energy cost function could serve as an alternative benchmark, this study deliberately selects the Linear MPC as the baseline. This design isolates the contribution of the RL agent, allowing for a precise quantification of the global energy optimization benefits provided by the learning-based algorithm compared to the local tracking capabilities of the underlying MPC [13, 14]. Furthermore, the convex formulation of Linear MPC ensures computational stability in the solution process from a mathematical standpoint. Unlike NMPC, which may suffer from local minima or convergence issues, the convex QP problem guarantees a global optimum within a bounded time, strictly meeting the requirements for computational reliability in safety-critical thermal management systems [15].

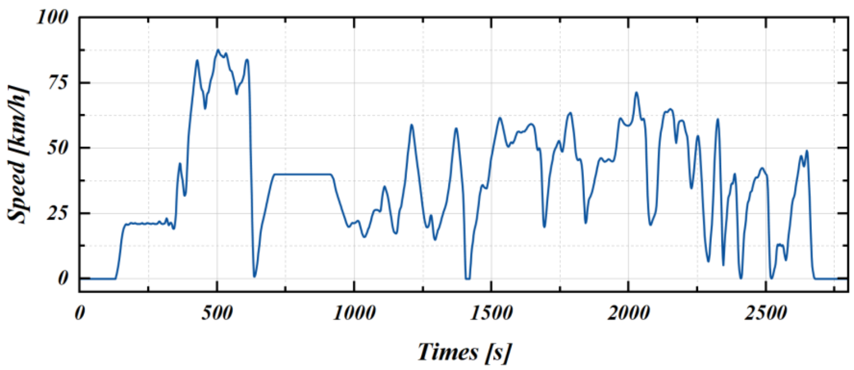

Fig. S6 Typical driving cycle for tracked vehicles

Table S6 Proportion and duration of each operating state in the driving cycle

State	Starting	Accelerating	Decelerating	Cruising	Climbing	Descending	Off-road	Idling
Duration (s)	135	700	185	545	120	65	800	250
Proportion (%)	4.8	25.0	6.6	19.5	4.3	2.3	28.6	8.9

To comprehensively evaluate the performance of different TMS strategies across various operational states of hybrid electric tracked vehicles, testing was conducted using a typical driving cycle for tracked vehicles, as illustrated in **Fig. S6**. The performance of TMS under scenarios including acceleration, deceleration, off-road driving, and grade climbing was examined, with the proportion and duration of each scenario detailed in the **Table S6**. It is important to note that this test cycle represents a 'new' and unseen operating condition that serves as a validation set. It is distinct from the randomized stochastic segments used during the training phase, thereby verifying the generalization capability and robustness of the proposed GMA-TD3-MPC strategy under unknown sequential loads.

The simulation model was established using a co-simulation framework integrating MATLAB/Simulink and GT-SUITE. The training process was executed on a workstation equipped with an Intel Core i7-14700KF CPU (@ 3.80 GHz), an NVIDIA RTX 3080 GPU, and 64 GB of RAM. To validate the environmental adaptability of the proposed strategy, comparative tests were conducted under two ambient temperature conditions—standard (25℃) and high temperature (45℃)—with a simulation sampling time of 1 second.

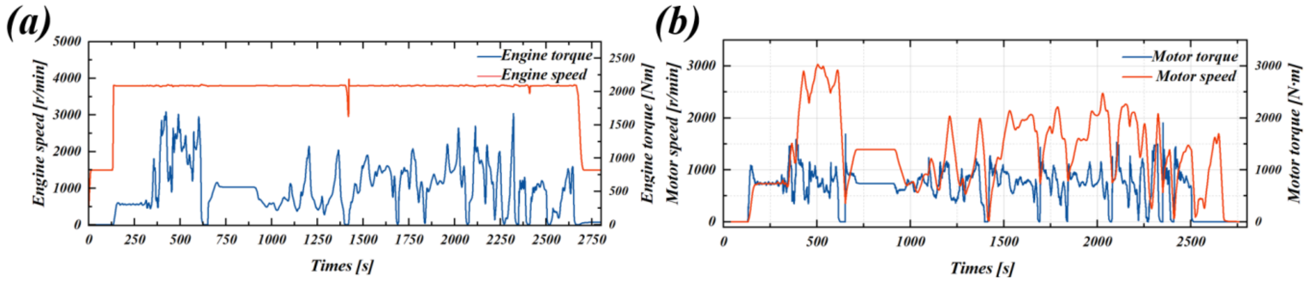

Fig. S7 **(a)** Engine speed and torque variations under driving cycle. **(b)** Speed and torque variations of left/right motors under driving cycle.

Fig. S7(a) and **S7(b)** illustrate the rotational speed and torque of the engine and left/right electric motors under varying driving cycle, The engine operation exhibits a distinct 'constant-speed, variable-torque' characteristic. Upon vehicle startup, the engine speed is rapidly established and maintained at approximately 1500 r/min. Subsequently, in response to load demands, it swiftly increases and stabilizes at 3800 r/min. In contrast to the rotational speed, the engine torque demonstrates oscillatory characteristics, responding to transient power demands on the DC Bus to maintain the system's energy balance. Meanwhile, the response characteristics of the traction motors are highly coupled with the vehicle's driving conditions. The motor speed curves reflect frequent starts, rapid accelerations, and cruising phases. Notably, during high-dynamic intervals (e.g., 300-700 s and 1750-2500 s), the motor torque exhibits dense peak pulses, corresponding to the power required to overcome acceleration resistance and track running resistance over off-road terrain.

Section S3.1 Performance analysis under normal temperature conditions

Fig. S8 presents Heat rejection from key components to the coolant under the driving cycle at standard ambient temperature. These plots reveal strong coupling between cooling demands and powertrain loads. During high-torque operations and rapid load transients, the powertrain generates substantial and rapidly fluctuating heat rejection. Due to the power-following characteristics of the engine-generator set, engine speed predominantly operates within the high-rpm range of 3800 r/min. Engine torque exhibits intense fluctuations, frequently reaching peak values during periods corresponding to vehicle acceleration, high-speed travel, steep grade climbing, and off-road maneuvers. These high-torque phases induce instantaneous thermal load surges in engine-related components.

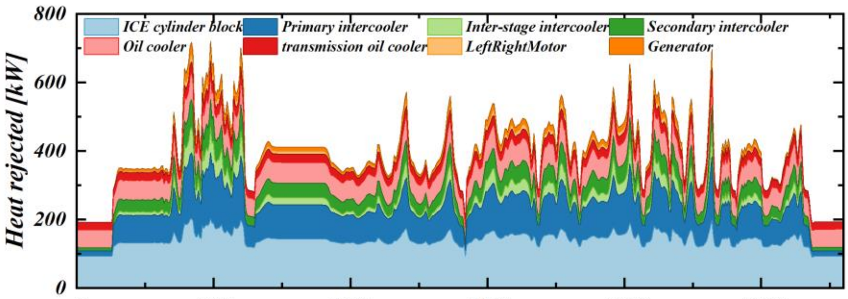

Fig. S8 Heat rejection from key components to the coolant under the driving cycle at standard ambient temperature.

Consequently, heat rejection from the three-stage intercoolers (Primary/Inter-stage/Secondary intercoolers) escalates synchronously, a direct result of the high-temperature exhaust gases from turbocharger compressors under heavy loads. Intercooler thermal loads demonstrate positive correlation with engine power output, while lubricant cooling requirements rise with increasing engine torque (peak dissipation≈80kW). Notably, engine torque experiences severe oscillations between 500-550 s and 2250-2375 s, coinciding with frequent gear shifting in off-road conditions, leading to dramatic shifts in cooling demands. During idling phases (0-125 s and 2750-3000 s), engine block heat rejection dominates significantly, while intercoolers maintain minimal heat rejection. Motor speed and torque fluctuate intensely throughout 0-3000 s, with pronounced peaks reflecting rapid shifts in power requirements during various driving scenarios and frequent acceleration/deceleration. These peaks generate substantial motor heat during acceleration, regenerative braking, or short-duration high-power operation. Multiple heat rejection peaks around 2250-2375 s directly correspond to high-torque outputs, validating the thermal impact of extreme operational demands.

Fig. S9(a) and **S9(b)** compare pump speed variations and fan speed variations in high/low-temperature circuits across different control algorithms, As can be seen that GMA-TD3-MPC exhibits minimal fluctuation in both fan and pump speed curves, indicating superior convergence in action space exploration. The embedded GMA module enhances attention mechanisms and memory capabilities for historical thermal states, enabling smoother actuator adjustments.

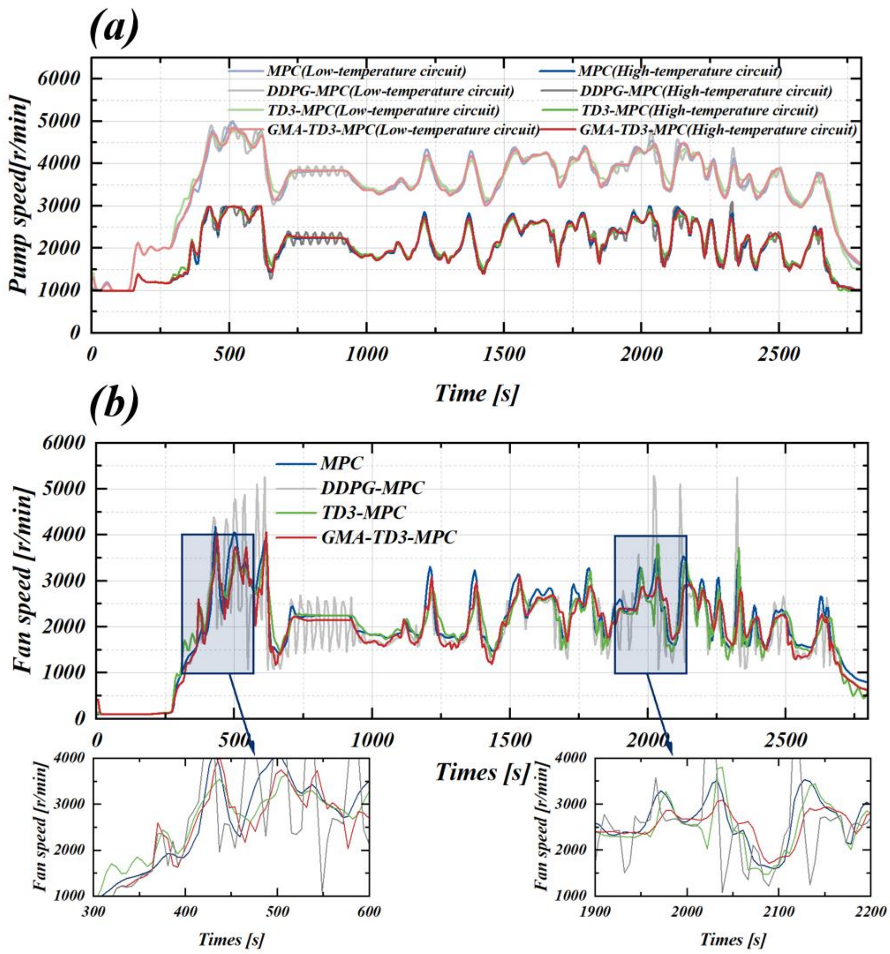

Fig. S9 Comparison of actuator speed profiles under different control algorithms at standard ambient temperature (a) High- and low-temperature circuit pumps; (b) Radiator fan.

As shown in the 300–600s 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 maximum vehicle speed, it maintains the high-temperature circuit control temperature 2.1°C lower than MPC and TD3-MPC with comparable rotational speeds, effectively preventing overtemperature risks. The 1900–2200s 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 (high-temperature circuit) and 2.9°C (low-temperature circuit), representing reductions of 44.19% and 6.45% respectively compared to MPC's 8.6°C and 3.1°C fluctuations. Relative to TD3-MPC, fluctuation reductions reach 38.46% and 23.68%. While TD3-MPC outperforms baseline MPC in temperature stability, it remains inferior to GMA-TD3-MPC. DDPG-MPC demonstrates the poorest performance: during critical operating conditions, 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. The deterministic policy of DDPG lacks exploratory robustness and fails to handle nonlinear heat exchange dynamics under varying vehicle operating conditions

Section S3.2 Performance analysis under high temperature conditions

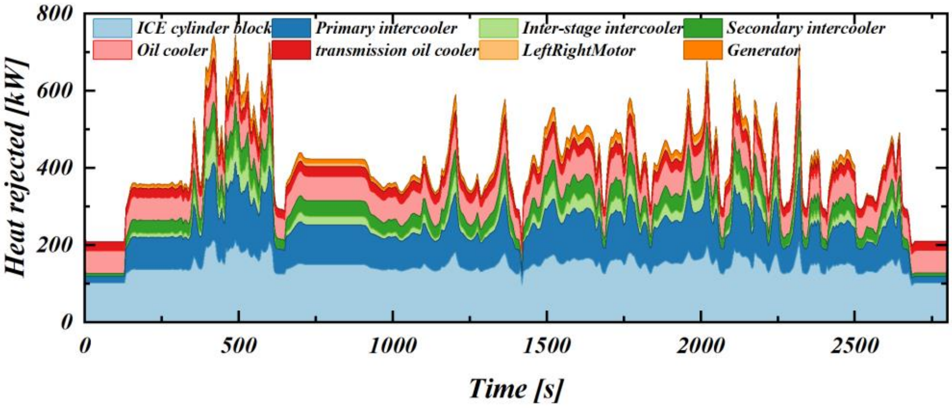

Fig. S10 Heat rejection from key components to the coolant under the driving cycle at high ambient temperature.

Fig. S10 illustrates the distribution of heat rejection power from key components to the coolant for the HETVs under a typical driving cycle at an elevated ambient temperature of 45℃. Compared to the standard condition (25℃), the high-temperature environment not only deteriorates the heat exchange boundary conditions of the radiator but also significantly alters the heat generation characteristics of the powertrain. The most prominent thermodynamic shift observed at 45□ is the increased heat load within the three-stage intercooling system. As the ambient intake temperature rises from 25□ to 45□, the resulting decrease in air density compels the turbocharger to operate at a higher-pressure ratio or elevated initial temperatures to maintain the required intake mass flow and engine power output. This leads to a marked increase in the compressor outlet temperature. Consequently, the cumulative amplitude of the intercooling zones in the figure exhibits significant expansion during acceleration and high-load phases. The intercoolers must remove a greater amount of sensible heat to reduce the intake air temperature to the necessary combustion boundary conditions, thereby driving up the intercooling thermal load. From a global perspective, the peak total heat rejection power repeatedly exceeds 700 kW, surpassing that of the standard condition and presenting more stringent challenges to the thermal management system.

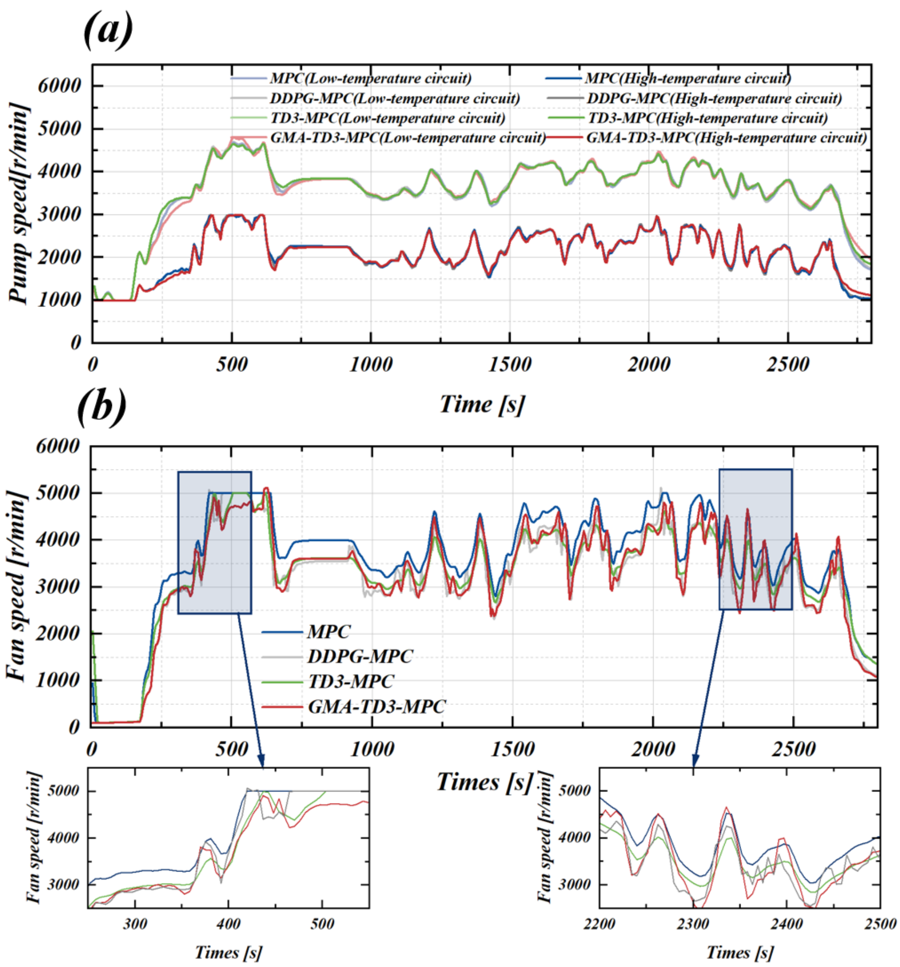

Fig. S11 Comparison of actuator speed profiles under different control algorithms at high ambient temperature **(a)** High- and low-temperature circuit pumps; **(b)** Radiator fan.

Fig. S11(a) and **S11(b)** illustrates the comparison of the control trajectories for the pumps and fans under various strategies at an ambient temperature of 45℃. To counteract the elevated thermal load compared to the 25°C scenario, all control strategies significantly raise the baseline actuator speeds. However, the conventional MPC strategy exhibits pronounced conservatism under these conditions. As illustrated in **Fig. S11(b)**, during the high-load acceleration phase (400–600 s), the MPC-controlled fan speed rapidly saturates, maintaining the upper limit of 5,000 r/min for an extended duration.

Regarding the RL-based methods, while DDPG-MPC converges, the enlarged insets reveal persistent high-frequency fluctuations. The TD3 strategy mitigates these oscillations via its dual-critic mechanism, yet it displays a perceptible lag in response to transient thermal shocks. In contrast, the proposed GMA-TD3-MPC demonstrates superior dynamic regulation. Specifically, during the 400–600 s interval, the proposed method achieves thermal equilibrium with a peak fan speed of approximately 4,200–4,800 r/min, effectively avoiding the full-load saturation seen in MPC. Furthermore, in the variable load phase between 2,200 s and 2,500 s, the GMA-TD3-MPC trajectory remains smooth, reducing the peak speed by roughly 300–500 r/min compared to the MPC baseline. By proactively anticipating thermal load trends and leveraging the system's thermal inertia, the proposed algorithm eliminates both the stiff saturation characteristic of MPC and the jitter inherent in DDPG, thereby optimizing the trade-off between thermal safety and energy consumption.

References

[1] X. Zheng, R.M. Yang, Q.F. Wang, Y.C. Yan, Y. Zhang, J.H. Fu, Z.T. Liu, Comparison of GRNN and RF algorithms for predicting heat transfer coefficient in heat exchange channels with bulges, Applied Thermal Engineering, 217 (2022).

[2] A. Ghate, A. Sundar, Q.L. Zhu, R. Prucka, M. Figuero-Santos, M. Barron, Development of an

integrated energy and thermal planner for a series hybrid off-road autonomous tracked vehicle☆, Energy Conversion and Management, 322 (2024).
[3] R. Koli, D. Egan, Q. Zhu, R. Prucka, Quantification of Linear Approximation Error for Model Predictive Control of Spark-Ignited Turbocharged Engines, SAE Technical Paper Series, (2019).
[4] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks, 61 (2015) 85-117.
[5] S. Fujimoto, D. Meger, D. Precup, Off-Policy Deep Reinforcement Learning without Exploration, in: 36th International Conference on Machine Learning (ICML), Long Beach, CA, 2019.
[6] R.C. Huang, H.W. He, X.Y. Zhao, Y.L. Wang, M.L. Li, Battery health-aware and naturalistic data-driven energy management for hybrid electric bus based on TD3 deep reinforcement learning algorithm, Applied Energy, 321 (2022).
[7] J. Zhang, B. Liu, Y. Zhou, X. Huang, S. Chen, D. Xuan, Hierarchical energy management strategy for fuel cell hybrid electric vehicles based on improved TD3 considering interior temperature comfort, Applied Thermal Engineering, 276 (2025) 126965.
[8] W. Rong, J. Zheng, Y. Chen, Y. Liu, Z. Zhang, Autonomous collision avoidance decision-making method with human-like attention distribution for MASSs based on GMA-TD3 algorithm, Ocean Engineering, 330 (2025) 121118.
[9] H.P. Van Hasselt, A. Guez, M. Hessel, V. Mnih, D. Silver, Learning values across many orders of magnitude, Advances in neural information processing systems, 29 (2016).
[10] J. García, F. Fernández, A comprehensive survey on safe reinforcement learning, J. Mach. Learn. Res., 16 (2015) 1437-1480.
[11] A. Karnik, Vehicle Powertrain Thermal Management System Using Model Predictive Control, SAE International Journal of Materials and Manufacturing, 9 (2016) 525-533.
[12] X. Liu, A. Yebi, P. Anschel, J. Shutty, B. Xu, M. Hoffman, S. Onori, Model Predictive Control of an Organic Rankine Cycle System, Energy Procedia, 129 (2017) 184-191.
[13] R. Lian, J. Peng, Y. Wu, H. Tan, H. Zhang, Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle, Energy, 197 (2020) 117297.
[14] T. Liu, Y. Zou, D.X. Liu, F.C. Sun, Reinforcement Learning-Based Energy Management Strategy for a Hybrid Electric Tracked Vehicle, Energies, 8(7) (2015) 7243-7260.
[15] D.Q. Mayne, Model predictive control: Recent developments and future promise, Automatica, 50(12) (2014) 2967-2986.