

SAC-Optimized Active Disturbance Rejection Controller for Thermal Regulation in Aerospace Applications

Dong Li¹, Yanping Gu^{2*}, Li Sun¹

1 School of Energy and Environment, Southeast University, Nanjing 210096, China

2 Shanghai Institute of Satellite Engineering, Shanghai 201109, China

(Corresponding Author: gyp0523@163.com)

ABSTRACT

Efficient temperature regulation is essential for maintaining the sensitivity and noise performance of spaceborne infrared detectors under varying observation modes. This paper proposes an Active Disturbance Rejection Controller (ADRC) with dynamic parameter tuning based on the Soft Actor-Critic (SAC) reinforcement learning algorithm, aiming to optimize the temperature control performance and disturbance rejection capability of spacecraft systems. A thermal transfer model was established, and the Soft Actor-Critic (SAC) algorithm was introduced to dynamically tune parameters (system gain, controller bandwidth and observer bandwidth) by maximizing a weighted sum of the reward and policy entropy. This approach enables real-time estimation and compensation of total disturbances. Simulation results demonstrate that, in temperature tracking tasks, the proposed method reduces the integral of absolute error (IAE) by 14.61% and 14.87%, and shortens the settling time by 36.36% and 58.47%, compared to fixed-parameter ADRC and PID controllers, respectively. Under external periodic disturbances, the proposed controller improves control accuracy by 15.6% and 13.6%. Monte Carlo robustness tests further show that, under $\pm 5\%$ parameter perturbations, the method exhibits small fluctuations in IAE, settling time, and overshoot. The proposed SAC-ADRC strategy provides a promising solution for rapid and high-precision thermal regulation of infrared detectors in aerospace applications.

Keywords: Aerospace thermal control, High-Precision Temperature Control, Active Disturbance Rejection Controller (ADRC), Soft Actor-Critic (SAC), Infrared detector

1. INTRODUCTION

Spaceborne infrared detectors are crucial in fields such as Earth observation, astronomical surveying, and space surveillance[1]. The sensitivity and noise

performance of these detectors are largely dependent on their ability to switch operating temperatures rapidly and accurately under different observation modes[2].

Thermoelectric Coolers (TECs) have been widely employed in temperature control systems for spaceborne infrared detectors due to their high efficiency, compact size, and tunability, with most systems relying on conventional PID controllers for temperature regulation[3]. While PID control is stable and simple, it struggles with the large thermal inertia and low conductivity of the Peltier block. Unlike traditional controllers, the Active Disturbance Rejection Controller (ADRC) does not rely on an accurate mathematical model of the plant. Instead, it treats unmeasured disturbances and unmodeled dynamics as a "total disturbance," which is estimated and compensated for in real time[4]. This study applies a first-order Active Disturbance Rejection Controller (ADRC) to the thermoelectric cooler (TEC)-based temperature control system of a spaceborne infrared detector, with the controller parameters optimized using the Soft Actor-Critic (SAC) algorithm. This approach enables dynamic optimal regulation under varying thermal disturbances, enhances response speed, and improves temperature control accuracy for aerospace applications.

2. PROBLEM DESCRIPTION

2.1 Overview of the Spaceborne Infrared Detector Temperature Control System

As shown in *Fig. 1*, the thermal control structure of the spacecraft infrared detector consists of a detector chip, a thermoelectric cooler (TEC), and a radiator. The detection performance is highly sensitive to temperature, requiring precise regulation to maintain the target value. The semiconductor TEC, composed of multiple PN junctions, has its upper surface in close contact with the detector chip. Serving as both a heating and cooling element[5], the TEC achieves bidirectional thermal control by utilizing the Peltier effect, through

Foundation item: National Key R&D Program of China (No.2022YFB3902902)

This is a paper for the 5th Applied Energy Symposium and Forum: Renewable Energy Matrix (REM2025), Oct. 29-31, 2025, Yancheng, China.

reversing the current direction flowing in the semiconductor material[6].

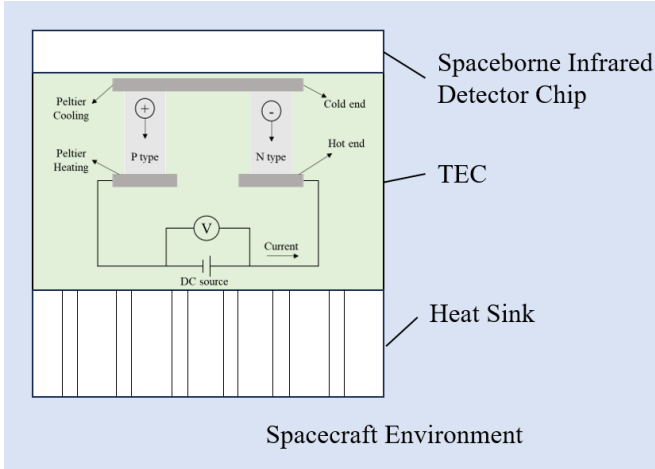


Fig. 1 Heat Transfer Structure of TEC in Spaceborne Infrared Detector

2.2 Mathematical Model of the Temperature Control System

The model of the temperature control system for the spaceborne infrared detector is derived based on the principle of energy conservation:

$$Q = C \frac{dT_c}{dt} + Q_L \quad (1)$$

Here, Q represents heat exchange caused by current through the semiconductor, C denotes the thermal capacity of the system, T_c is the temperature of the detector chip, and Q_L represents the heat dissipated to the surroundings through radiation and loss during the heat exchange process between the TEC and the chip.

Assuming that the sum of the contact thermal resistance, convective thermal resistance, and equivalent radiative thermal resistance is Θ , T denotes the temperature difference between the detector chip and the space environment, the heat loss exchanged with the environment can be expressed as:

$$Q_L = \frac{T}{\Theta} \quad (2)$$

By expressing T and Q in incremental form, the following can be obtained:

$$C \frac{d\Delta T}{dt} + \frac{\Delta T}{\Theta} = K_0 \Delta I \quad (3)$$

Here, K_0 denotes the thermal conductivity, and ΔI represents the incremental current flowing through the semiconductor material under steady-state conditions.

Equation (3) describes the mathematical model of the spaceborne infrared detector temperature control system. By applying the Laplace transform, the transfer function of the temperature control system is obtained:

$$G(s) = \frac{\Delta T(s)}{\Delta I(s)} = \frac{K_0}{C_s + \frac{1}{\Theta}} = \frac{K}{Ts + 1} \quad (4)$$

Here, $K = K_0 \Theta$, $T = C \Theta$. Considering the temperature lag caused by heat transfer, a time delay L is introduced, and the final transfer function is obtained:

$$G(s) = \frac{K}{Ts + 1} e^{-Ls} \quad (5)$$

3. FIRST-ORDER ADRC CONTROL SCHEME

The schematic diagram of the Linear Active Disturbance Rejection Controller is shown in Fig. 2[7], where v denotes the reference input, y represents the system output, and $\omega(t)$ indicates the external disturbance acting on the system.

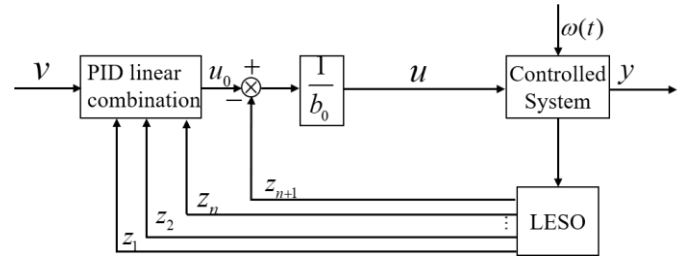


Fig. 2 Schematic Diagram of Linear Active Disturbance Rejection Control

The controlled plant in this study is a first-order system, and a first-order Active Disturbance Rejection Controller (ADRC) is employed.

By defining the total disturbance f and augmenting it as a new state, the extended system can be expressed as follows:

$$\begin{cases} \dot{x}_1 = b_0 u(t - L) + f \\ \dot{x}_2 = \dot{f} \end{cases} \quad (6)$$

The extended state-space equations of the first-order system are:

$$\begin{cases} \dot{x} = Ax + Bu + Ef \\ y = Cx \end{cases} \quad (7)$$

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} b_0 \\ 0 \end{bmatrix}, \quad E = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Where u is the system input, x_1 and x_2 are the state variables.

Based on this, the Extended State Observer (ESO) is formulated as:

$$\dot{z} = \mathbf{A}z + \mathbf{B}u + \mathbf{L}(y - z_1) \quad (8)$$

Where y denotes the output, z represents the state vector of the observer, $\mathbf{L} = [l_1 \quad l_2]^T = [\beta_1 \quad \beta_2]^T$ is the error feedback gain matrix.

Based on the above, the formula for LADRC algorithm is:

$$\begin{cases} \dot{z}_1 = z_2 + l_1(y - z_1) \\ \dot{z}_2 = l_2(y - z_1) \\ u_0 = k_p(v - z_1) \\ u = \frac{u_0 - z_2}{b_0} \end{cases} \quad (9)$$

For the characteristic equation of ESO, by placing the closed-loop pole at $-\omega_o$, the observer parameter $\beta_1 = 2\omega_o$ and $\beta_2 = \omega_o^2$ can be obtained. For the characteristic equation of the ADRC controller, placing the closed-loop pole at $-\omega_c$ yields the controller parameter $k_p = \omega_c$. Therefore, the parameters that need to be tuned include b_0 , the observer bandwidth ω_o , and the controller bandwidth ω_c .

4. PARAMETER TUNING BASED ON SAC

In this study, the system is modeled as a Markov Decision Process (MDP), and the Soft Actor-Critic (SAC) algorithm is employed for decision-making optimization.

4.1 Construction of Markov Decision Process (MDP)

In the MDP framework, the agent chooses actions in different states, leading to state transitions and rewards. The goal is to pick actions that maximize the total expected reward over time. By modeling the problem as an MDP, an optimal policy sequence can be determined, providing a theoretically optimal solution for complex decision-making tasks.

- State Space

$$\mathbf{S} = [e_t, st_t]^T \quad (t = 1, 2, 3 \dots n) \quad (10)$$

Here, t denotes a time step, n denotes the total number of time steps required for a complete interaction sequence between the agent and the environment, and e_t represents the difference between the input signal at the next time step and the simulated system output at the current time step. st_t denotes the mean value of the input signal at the next time step.

- Action Space

The action consists of the three parameters to be tuned, with the action space defined as:

$$\mathbf{A} = [b_0, \omega_o, \omega_c]^T \quad (11)$$

- Reward Function

The design of the reward function considers four performance metrics: Integral of Absolute Error (IAE), Maximum Sensitivity (M_s), Gain Margin (G_m), Phase Margin (P_m). The coefficients k_1, k_2, k_3, k_4 can be adjusted according to task requirements to emphasize different control objectives. In this study, the coefficients were initially set based on scale balancing and then fine-tuned through simulation comparisons to achieve a balanced trade-off in overall control performance. The reward function at the i -th time step can be expressed as:

$$r_i = -k_1 IAE_i - k_2 \left(|M_{s_i} - 1.2| \right)^2 + k_3 G_{m_i} + k_4 P_{m_i} \quad (12)$$

4.2 Design of SAC-Based Policy Optimization Algorithm

Within the MDP framework, this study proposes a SAC-based method for online tuning of ADRC parameters by maximizing the weighted sum of cumulative reward and policy entropy.

The policy entropy H is calculated as:

$$H(\pi(\cdot | s_t)) = \mathbb{E}_{a_t \sim \pi} [-\log \pi(a_t | s_t)] \quad (13)$$

SAC adopts an Actor-Critic architecture, consisting of one Actor network, two Critic networks, and two target Critic networks. The network parameters are φ , θ_i and $\bar{\theta}_i$ respectively. The Actor network generates an action a_t based on the current state s_t , aiming to maximize the expected return. The update rule for the Actor network is given by:

$$\pi^* = \arg \max_{\pi} \sum_{t=0}^T \mathbb{E}_{(s_t, a_t) \sim \rho_{\pi}} [r(s_t, a_t) + \alpha H(\pi(\cdot | s_t))] \quad (14)$$

And the Actor network gradient is:

$$J(\varphi) = \mathbb{E}_{s_t \sim D} \left[\alpha \log \pi_t(a_t | s_t) - \min_{i=1,2} Q_{\theta_i}(s_t, a_t) \right] \quad (15)$$

Here, ρ_{π} represents the current policy distribution, $Q_{\theta_i}(s_t, a_t)$ denotes the Q -value obtained by inputting the current state-action pair into the Critic network, and α is the temperature coefficient used to balance the trade-off between immediate rewards and

entropy. The loss function for the temperature coefficient is calculated as[8]:

$$L(\alpha) = \mathbb{E}_{a_t \sim \pi_t} [-\alpha \log \pi_t(a_t | s_t) - \alpha \bar{H}] \quad (16)$$

Here, \bar{H} denotes the target entropy, which is defined as the negative of the action space dimension.

The Critic networks evaluate the potential of the selected actions, guiding the policy updates toward increasing the expected return. To mitigate overestimation bias, SAC employs two Critic networks, Q_{θ_1} and Q_{θ_2} , and updates them by minimizing the Bellman error as follows[9]:

$$L(\theta_i) = \mathbb{E}_{(s_t, a_t) \sim D} \left[\left(Q_{\theta_i}(s_t, a_t) - (r(s_t, a_t) + \gamma \min_{\theta_j} Q_{\theta_j}(s_{t+1}, a_{t+1}) - \alpha \log \pi(s_{t+1}, a_{t+1})) \right)^2 \right] \quad (17)$$

s_{t+1} and a_{t+1} denote the predicted next state and action, $Q_{\theta_j}(s_{t+1}, a_{t+1})$ is the target Q -value obtained by inputting the next state-action pair into the target Critic networks, γ is the discount factor, which balances the trade-off between immediate and future rewards.

Table 1 presents the pseudocode for the parameter optimization procedure of the ADRC based on SAC.

Table 1 Pseudocode of Parameter Optimization Procedure

Algorithm 1 The solution process of the SAC-based method

Initialize: The parameters of the Actor network φ , Critic networks θ_1, θ_2 , the target Critic networks $\bar{\theta}_1, \bar{\theta}_2$, and temperature parameter α

The learning rate λ , discounting factor γ and soft update coefficient τ

Total training episodes M , maximum steps per episode n

Begin the training process:

for episode = 1 to M **do**

 Initialize the environment and obtain initial state s_0

for t = 1 to n **do**

 Sample actions a_t using Gaussian policy

 Simulate system response under ADRC with parameters a_t

 Observe next state s_{t+1} and compute reward

r_t , entropy bonus H_t

 Store transition $(s_t, a_t, r_t, s_{t+1}, H_t)$ in buffer

if buffer has enough samples **then**

 Sample minibatch from buffer

for each sample (s, a, r, s', H) in batch **do**

$\theta_i \leftarrow \theta_i - \lambda \nabla_{\theta_i} L(\theta_i), i=1,2$

$\varphi \leftarrow \varphi + \lambda \nabla_{\varphi} J(\varphi)$

$\alpha \leftarrow \alpha - \lambda \nabla_{\alpha} L(\alpha)$

$\bar{\theta}_i \leftarrow \tau \theta_i + (1-\tau) \bar{\theta}_i, i=1,2$

end for

end if

end for

end for

5. SIMULATION RESULTS

The parameters of reinforcement learning are listed in Table 2

Table 2 Reinforcement Learning Parameters Table

| Parameter Name | Parameter Settings |
|------------------------------|--------------------|
| Actor Network Learning Rate | 0.00001 |
| Critic Network Learning Rate | 0.0001 |
| Discount Factor γ | 0.95 |
| Training episodes | 250 |
| b_0 Constraint Range | [0.58,0.64] |
| ω_o Constraint Range | [1.2,1.6] |
| ω_c Constraint Range | [0.15,0.55] |

The comparative simulation control system was implemented in SIMULINK, as illustrated in Fig. 3

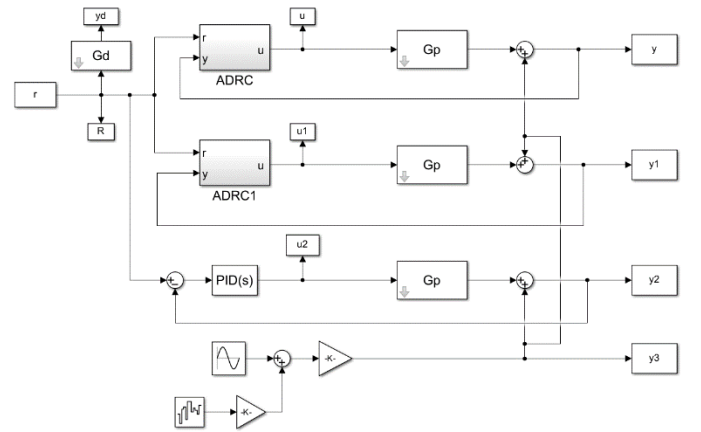


Fig. 3 Diagram of the comparative simulation control system

The training curve of the SAC-ADRC algorithm is shown in Fig. 4, where the "average" curve represents the smoothed reward value over every five training steps. As illustrated by the curve, the training demonstrates good convergence performance.

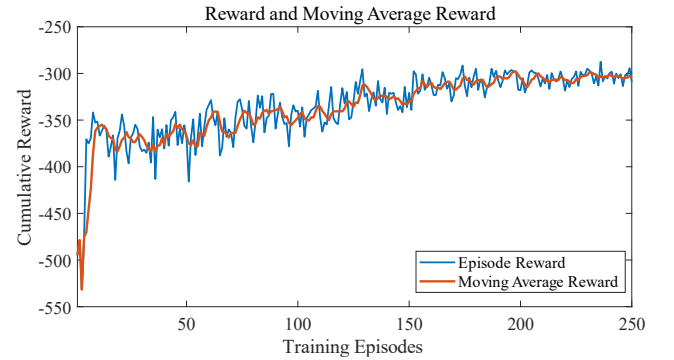


Fig. 4 Training curve

5.1 Setpoint Tracking Response Comparison

The trained SAC network is applied to offline simulation and compared with fixed-parameter ADRC and PID controllers. The setpoint tracking response comparison is shown in Fig. 5

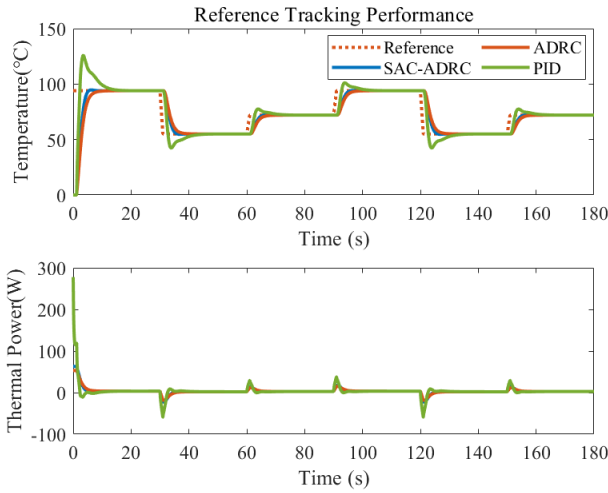


Fig. 5 Comparison of setpoint tracking responses

The performance metrics are summarized in Table 3.

Table 3 Comparison of metrics in Setpoint Tracking

| Controller | IAE | First Stage Regulation Time | Overshoot σ (%) |
|----------------------|--------|-----------------------------|------------------------|
| SAC-ADRC | 592.1 | 4.9 | 0.72 |
| Fixed-parameter ADRC | 693.43 | 7.7 | -0.39 |
| PID | 695.53 | 11.8 | 33.4 |

It can be observed that the ADRC controller based on the SAC algorithm meets the speed and accuracy requirements of infrared detector thermal control effectively.

The dynamic parameter adjustment process of the SAC-ADRC is illustrated in Fig. 6

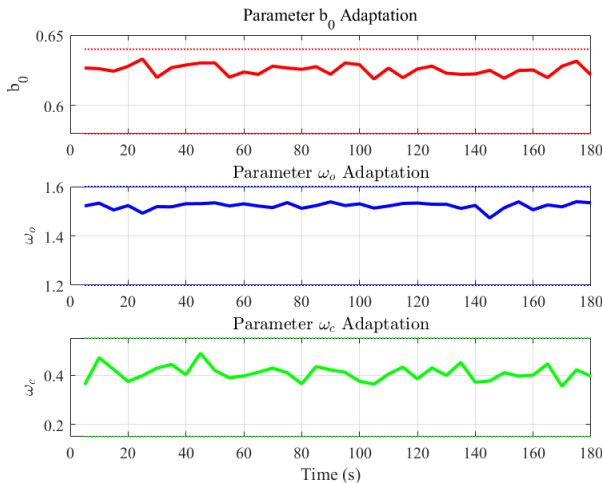


Fig. 6 Dynamic parameter adjustment process of SAC-ADRC

5.2 Comparison under External Disturbance

Due to the thermal resistance of the Peltier element varying with temperature, fluctuations in heat transfer efficiency occur. To evaluate disturbance rejection

performance, a direct temperature disturbance is applied to the system:

$$\Delta T_{dist} = Q \cdot \Delta R_{th}(t) \quad (18)$$

Here, Q represents the current heating power, and $\Delta R_{th}(t)$ denotes the variation in thermal resistance. A sinusoidal signal simulates periodic disturbances, while a random noise term represents stochastic variations like heat loss and resistance. The comparison under external disturbances is shown in Fig. 7

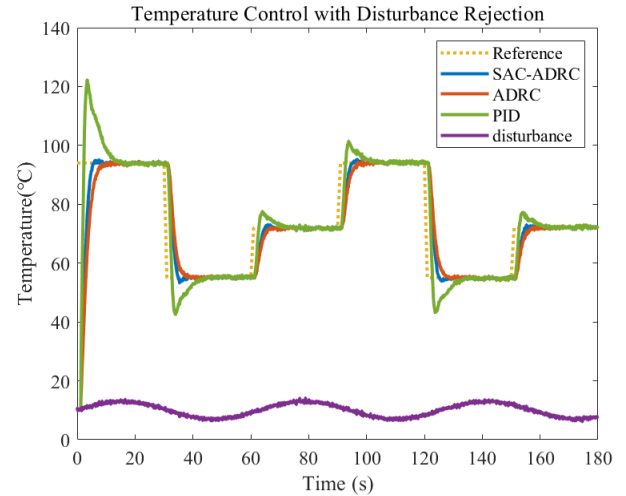


Fig. 7 Temperature control comparison under external disturbance

The IAE for control using SAC-ADRC, fixed-parameter ADRC, and PID controllers are 597.7, 708.4, and 691.9, respectively. It can be seen that under external disturbances, the SAC-ADRC controller is still able to track the setpoint with high speed and accuracy, demonstrating strong disturbance rejection capability.

5.3 Robustness Analysis Based on the Monte Carlo Method

System robustness is tested using Monte Carlo sampling. The system gain K , the inertia time constant T_s , and the time delay L of the controlled plant undergo $\pm 5\%$ perturbations around their nominal values. A total of 2000 randomized trials are conducted. In addition, a one-time $\pm 25\%$ abrupt variation was introduced into these parameters every 100 simulations to emulate component aging or parameter drift. Meanwhile, a solar-irradiance-induced transient disturbance was superimposed on the reference signal to replicate sudden changes in external thermal loads. The distributions of Overshoot (σ) - Settling Time (t_s) and Integral of Absolute Error (IAE) - Settling Time (t_s) are shown in Fig. 8

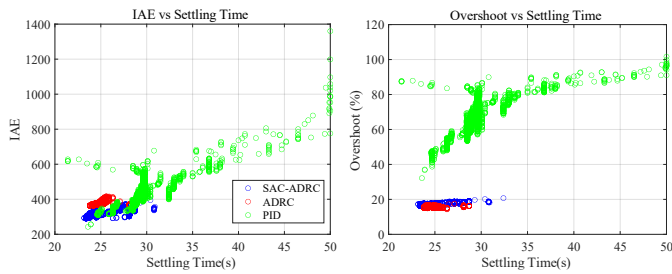


Fig. 8 Distributions under Plant Parameter Perturbations

It can be observed that when the plant parameters are perturbed, the SAC-ADRC controller achieves superior performance in terms of both overshoot and IAE, with smaller variability. This clearly indicates that reinforcement learning can further enhance the ADRC controller's ability to observe unmodeled dynamics and increases its tolerance to model inaccuracies.

6. CONCLUSIONS

This paper addresses the stringent requirements for high-efficiency and high-precision temperature regulation in spaceborne infrared detector systems by proposing and designing an Active Disturbance Rejection Control (ADRC) scheme with dynamic parameter tuning based on the Soft Actor-Critic (SAC) reinforcement learning algorithm. Simulation studies were conducted to evaluate the controller's performance in target temperature tracking, rejection of orbital thermal disturbances, and Monte Carlo robustness analysis, with comparisons made against fixed-parameter ADRC and conventional PID control. The results demonstrate that the SAC-ADRC controller outperforms both fixed-parameter ADRC and PID control in terms of response speed, regulation accuracy, overshoot suppression, and tracking performance. Moreover, it exhibits superior disturbance rejection and stability under external perturbations, as well as enhanced robustness in Monte Carlo random disturbance tests. In conclusion, the proposed SAC-based ADRC method significantly improves the temperature control performance of spaceborne infrared detectors, surpassing fixed-parameter ADRC and traditional PID control in terms of temperature stability, steady-state accuracy, and resistance to environmental disturbances. This approach provides a novel technical pathway for achieving high-precision thermal management in spacecraft infrared detection systems. Future work will be carried out on a TEC-based experimental platform to perform hardware-in-the-loop (HIL) and thermal-vacuum tests, aiming to verify the algorithm's real-time performance and engineering applicability under space thermal conditions.

REFERENCE

- [1] H. H. Hogue, M. G. Mlynczak, M. N. Abedin, S. A. Masterjohn, and J. E. Huffman, "Far-infrared detector development for space-based Earth observation," *Proc. SPIE 7082, Infrared Spaceborne Remote Sensing and Instrumentation XVI*, San Diego, CA, USA, Sep. 2008, pp. 70820E, doi: 10.1117/12.797078.
- [2] J. Tang, W. Lei, X. Ren, B. Yang, and X. Weng, "Calculation and analysis of working performance about spaceborne HgCdTe infrared detector affected by temperature," *Proc. SPIE 10846, Optical Sensing and Imaging Technologies and Applications*, Beijing, China, Dec. 2018, pp. 1084610, doi: 10.1117/12.2504237.
- [3] Z. Wang, Y. Hu, C. Ni, L. Huang, A. Zhang, and X. Sun, "The design of high precision temperature control system for InGaAs short-wave infrared detector," *Proc. SPIE 10697, Infrared Technology and Applications*, 2018, pp. E03.
- [4] Han J. From PID to Active Disturbance Rejection Control. *IEEE T Ind Electron*, 2009, 56(3): 900~906.
- [5] J. M. Zamboni, "Integrated thermoelectric cooler/package for infrared detector array temperature stabilization," in *Proc. SPIE 5209, Materials for Infrared Detectors III*, San Diego, CA, USA, Dec. 2003, doi: 10.1117/12.509779.
- [6] T. H. Kwan, X. Wu, and Q. Yao, "Bidirectional operation of the thermoelectric device for active temperature control of fuel cells," *Applied Energy*, vol. 222, pp. 410–422, May 2018.
- [7] Y. Liu, H. Xiong, Y. Zou and G. Hu, "Simulation Research on Active Power Filter Application Based on Auto Disturbance Rejection Control," *2024 8th International Conference on Robotics, Control and Automation (ICRCA)*, Shanghai, China, 2024, pp.263-267.
- [8] T. Haarnoja et al., "Soft Actor-Critic Algorithms and Applications," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019.
- [9] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 1861–1870.