

Article

Not peer-reviewed version

---

# Thrust and Pressure Control in Solid Propulsion System via Reinforcement Learning

---

[Zuohao Hua](#), [Zhuang Fu](#)<sup>\*</sup>, Lu Niu

Posted Date: 31 October 2024

doi: 10.20944/preprints202410.2461.v1

Keywords: Solid propulsion system; Attitude and divert nozzles; Thrust and pressure control; Reinforcement learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

## Article

# Thrust and Pressure Control in Solid Propulsion System via Reinforcement Learning

Zuohao Hua <sup>1,2</sup>, Zhuang Fu <sup>1,\*</sup> and Lu Niu <sup>2,\*</sup>

1 School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

2. Shanghai Space Propulsion Technology Research Institute, Shanghai 201109, China

\* Correspondence: zhfu@sjtu.edu.cn; niulu\_sh@139.com

**Abstract:** A reinforcement learning control method for a solid attitude and divert propulsion system is proposed. The system in this research includes 4 divert thrust nozzles, 6 attitude thrust nozzles, and a common combustion chamber. To achieve the required thrust, the pressure in the combustion chamber is first adjusted by controlling the total opening of the nozzles to generate the gas source. Next, by controlling the opening of nozzles at different positions, the required thrust is produced in five-axis directions. Finally, the motor speed is regulated to drive the valve core to the specified position, completing the closed-loop control of the nozzle opening. The control algorithm used is the Proximal Policy Optimization (PPO) reinforcement learning algorithm. Through system identification and numerical modeling, the training environment for the intelligent agent is created. To accommodate different training objectives, multiple reward functions are implemented. Ultimately, through training, a multi-layer intelligent agent architecture for pressure, thrust, and nozzle opening is established, achieving effective system pressure and thrust control.

**Keywords:** Solid propulsion system; Attitude and divert nozzles; Thrust and pressure control; Reinforcement learning

## 1. Introduction

The Attitude and Divert Propulsion System (ADPS) is a propulsion system in aerospace engineering designed to adjust spacecraft attitude, ensuring three-axis stability and enabling orbital maneuvers [1]. ADPS functions by ejecting gas through multiple nozzles: attitude thrust nozzles at the spacecraft's rear control pitch, yaw, and roll, while divert thrust nozzles at the center of mass provide lateral thrust [2]. ADPS types include the liquid attitude and divert propulsion system (LADPS) and the solid attitude and divert propulsion system (SADPS), each with specific advantages and disadvantages tailored to particular spacecraft design objectives and performance requirements [3]. The LADPS uses liquid oxidizer and fuel as propellants, allowing for repeated thrust start and stop and providing a higher specific impulse. However, it requires complex fuel storage and delivery systems that need regular maintenance [4]. In contrast, the SADPS uses solid oxidizers and fuels, which eliminates the need for complex pipeline delivery systems, offering ease of maintenance, safety, and storage. Variable thrust in SADPS is achieved by altering the throat area of the nozzle in real-time [5]. Precise control of nozzle openings and the propellant combustion rate enables pressure and thrust regulation in the combustion chamber [6]. However, the shared combustion chamber among multiple nozzles and the pressure-dependent combustion rate result in a gradual increase in free volume, necessitating intricate control algorithms for safe and reliable operation.

The dynamics of the Solid Attitude and Divert Propulsion System (SADPS) are characterized by nonlinear and time-varying behavior due to nozzle control, pressure regulation, and thrust management. These characteristics make traditional control methods like Proportional Integral Derivative (PID) inadequate for meeting performance standards [7]. Sergio's review examines closed-loop pressure control methods in rocket propulsion, typically using linearized models around specific operating points to derive steady-state controllers, predominantly of the PID type. Sergio emphasizes that future control strategies must robustly achieve the desired thrust levels. Potential

solutions include enhanced nonlinear approaches and optimal closed-loop control [8]. Ming Zhu designed a pressure control system for pintle solid rocket motors using parameter identification through adaptive control methods, demonstrating significant advantages in terms of stability, robustness, and precision [9]. Bergmans and Di Salvo introduced an adaptive matching technique that adjusts pre-designed controllers based on evolving system dynamics [10], while Joner and Quinquis proposed adaptive controllers to mitigate uncertainty effects [11]. Waxenegger-Wilfing proposed a data-driven method for early warning of thermoacoustic instabilities in rocket engine thrust chambers, successfully using machine learning to predict these instabilities in advance [12]. Previous research extensively investigated spacecraft propulsion, aiming to minimize fuel consumption for effective thrust command allocation and stable attitude control [13–15]. Lee and Lim applied adaptive feedback linearization techniques to achieve pressure regulation and conducted extensive work on five-axis thrust distribution [16–18].

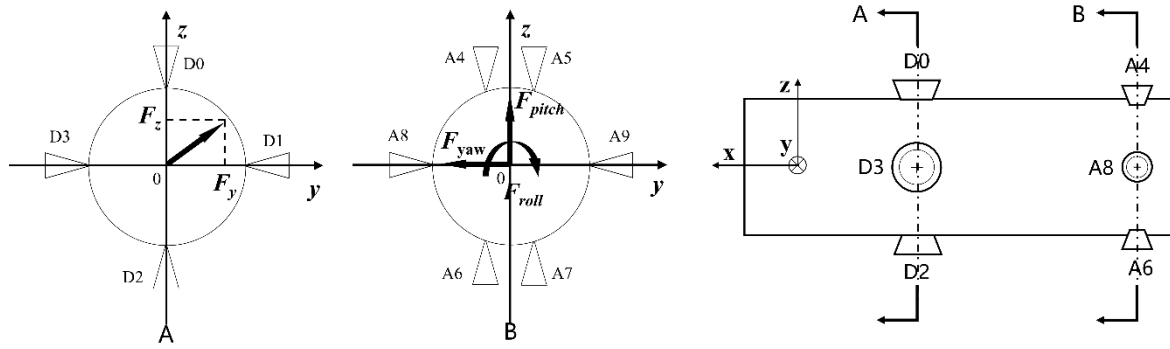
The integration of artificial intelligence into control systems has led to significant advancements, with Reinforcement Learning (RL) emerging as a prominent approach for addressing sequential decision problems [19]. RL offers advantages such as learning from experience, flexibility, autonomous decision-making, robustness, and resource efficiency [20]. The development of an RL-based mission planning strategy for a Partial Space Elevator (PSE) addresses the challenge of maintaining the main satellite's orbital stability during climber operations, demonstrating that using a deep Q-network to optimize the waiting intervals between transfer missions can minimize orbital radius changes effectively [21]. Skarka's [22] review explores the application of Machine Learning (ML) and RL techniques in enhancing the navigation and obstacle avoidance capabilities of Unmanned Aerial Vehicles (UAVs), revealing the effectiveness of various RL algorithms in enabling autonomous navigation through real-time environmental interactions. Additionally, Pham's [23] research introduces an RL framework for robotic manipulators to adapt to joint malfunctions during tasks, achieving a high success rate of 93.6% in diverse failure conditions. Furthermore, a proposed trajectory tracking control algorithm for Unmanned Surface Vehicles (USVs) employs an improved Deep Deterministic Policy Gradient (DDPG) method [24], demonstrating its effectiveness in dynamic ocean environments through experimental validation."

In aerospace vehicle control, RL has demonstrated potential in addressing complex and uncertain environments, enhancing autonomous decision-making capabilities, and handling nonlinear control problems. However, there is a lack of literature on the application of RL in aerospace propulsion systems. Leveraging RL, we developed an intelligent control method and architecture for the Solid Attitude and Divert Propulsion System (SADPS). Considering the system's strong non-linearity, multiple inputs and outputs, disturbances, and coupling characteristics, we established a multi-agent architecture for pressure, thrust, and nozzle opening control. We devised an adaptive reward function and trained an expert neural network, demonstrating superior control effects compared to PID. Simulation tests validated the effectiveness of the intelligent control method.

## 2. Model and Algorithm

### 2.1. Solid Attitude and Divert Propulsion System (SADPS)

As shown in Figure 1, the four divert thrust nozzles are arranged at the spacecraft's center of mass, generating thrust in the y-axis and z-axis directions. Additionally, six attitude thrust nozzles are positioned at a certain distance from the center of mass, providing pitch, yaw, and roll control along the three axes.

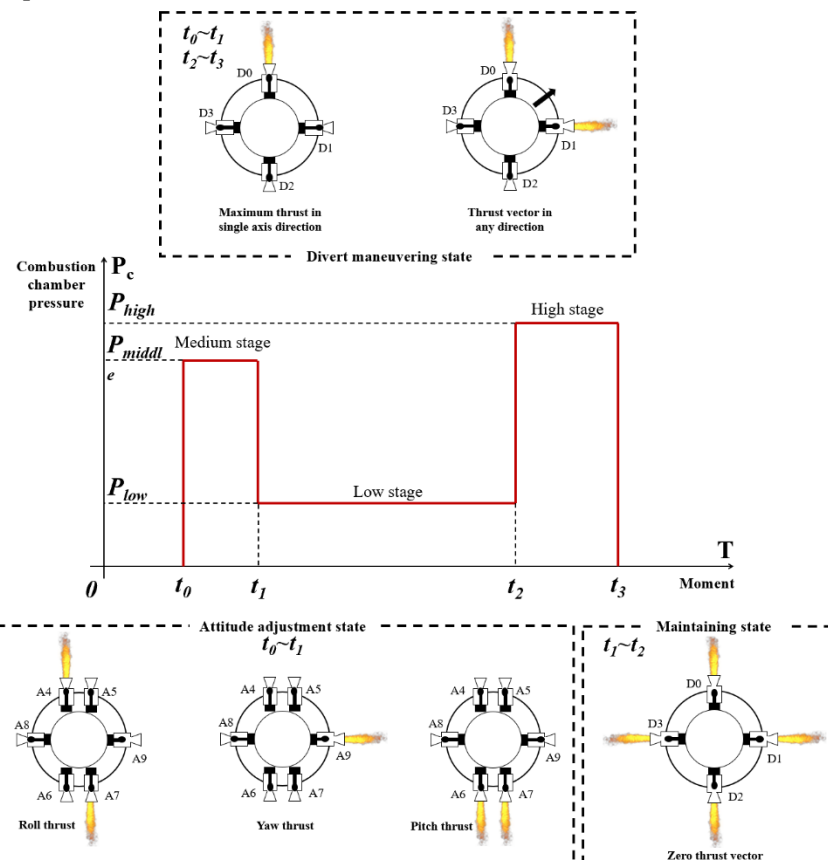


**Figure 1.** Nozzle layout of the solid attitude and divert propulsion system.

The SADPS operates in three modes: power-saving mode, performance mode, and overload mode.

- **Performance mode** (Figure 2) requires a medium level of thrust to adjust the spacecraft's orbit and pitch angle.
- **Power-saving mode** minimizes propellant loss by reducing pressure as much as possible while maintaining flight attitude and compensating for gravity.
- **Overload mode** necessitates rapid changes in the spacecraft's trajectory and flight attitude.

Typically, chamber pressure levels are set to constant values to ensure stability and predictability in performance.

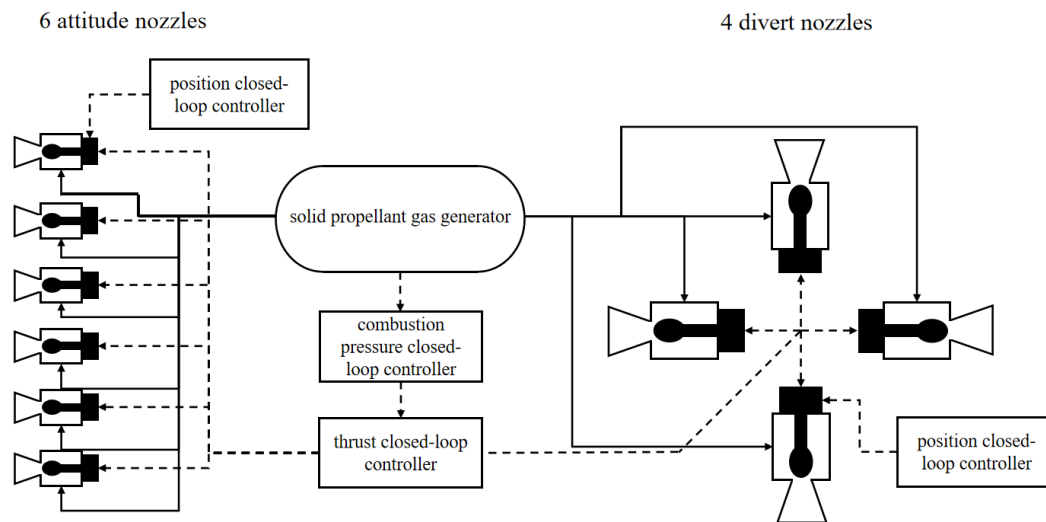


**Figure 2.** Three operating states of the solid attitude and divert propulsion system.

From time  $t_0$  to  $t_1$ , the SADPS operates in performance mode, during which the spacecraft adjusts its attitude and changes its orbit to aim at the target. From time  $t_1$  to  $t_2$ , the SADPS switches to power-saving mode, maintaining orbital altitude and overcoming Earth's gravity. From time  $t_2$  to  $t_3$ , the

engine enters overload mode, performing the final rapid maneuvers to correct any deviation from the target.

The control system model is depicted in Figure 3. The gas generator primarily produces the working medium required for thrust. The hot gas can be expelled simultaneously through four divert thrust nozzles and six attitude thrust nozzles. The pressure controller regulates the gas generator's pressure and provides the total opening command for the nozzles. The thrust controller allocates the opening of different nozzles and, within the constraints of the total nozzle opening, generates the required five-axis thrust. The opening controller drives the motor to rotate according to each valve's opening instructions, completing the closed loop of nozzle operation. Through the coordinated hierarchical work of the three controllers, the switching of the engine's operating modes and the output of thrust are achieved.

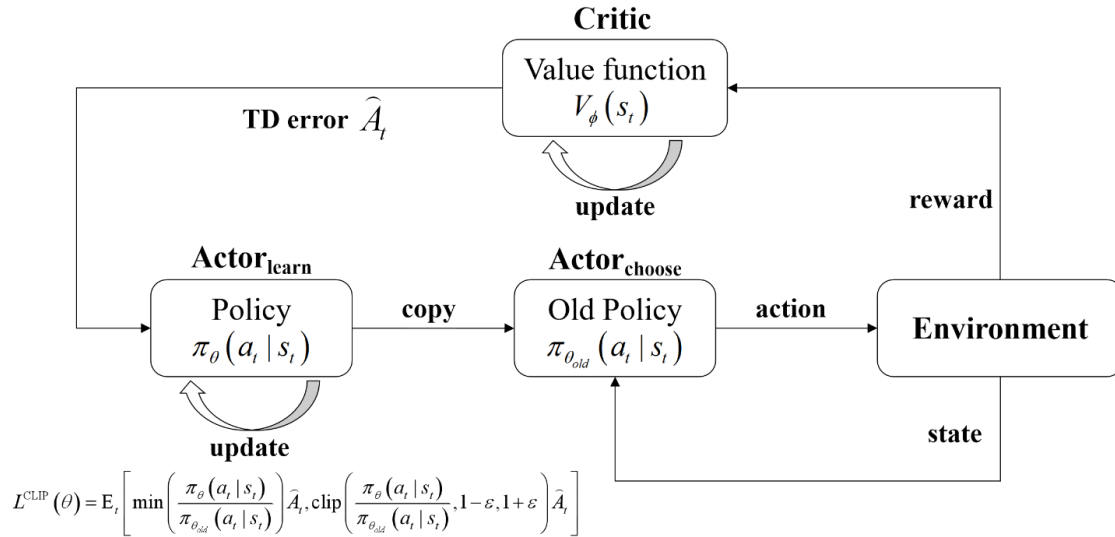


**Figure 3.** Control system model of the SADPS (solid line: gas path; dashed line: signal path).

Each nozzle's aperture size is adjusted by a throat plug-type gas valve. This valve comprises a plug-shaped element located in the center of the valve throat. When this element interacts with the inner wall of the nozzle, it forms an annular gas channel known as the equivalent throat area. By altering the position of the throat plug, the equivalent throat area is adjusted, thereby controlling the nozzle's output thrust [25]. The throat plug regulating valve consists of several components, including a brushless DC motor, a decelerator, a valve body, a nozzle, thermal protection components, and transmission components. The throat plug is driven by the rotary motor through a rotary-to-linear transmission mechanism, causing it to reciprocate horizontally.

## 2.2. Reinforcement Learning Trainer

The PPO algorithm consists of three neural networks: Actor<sub>learn</sub>, Actor<sub>choose</sub>, and Critic. It aims to improve strategy actions by computing action mean and variance, then selecting the final action via random sampling. Advantage clipping limits strategy updates to ensure robustness [26]. The basic PPO structure is shown in Figure 4.



**Figure 4.** The basic structure of PPO algorithm.

The  $\pi_{\theta_{old}}(a_t | s_t)$  chooses action based on the  $s_t$  and receives reward. The  $V_{\phi}(s_t)$  evaluates state values  $V_t$  and output TD error. During updates,  $\hat{A}_t$  is backpropagated and clipped within the range  $(1-\epsilon, 1+\epsilon)$  to refine both the Critic and Actor<sub>learn</sub> networks. Periodically, the Actor<sub>learn</sub> parameters are copied to Actor<sub>choose</sub> for better action selection.

To accelerate training, we applied several techniques to improve the efficiency of PPO:

- **Network Simplification:** Replaced neural network hidden layers of widths 256 and 128 with widths 64 and 64, reducing parameters by 87% and inference time by over 20% without compromising control effectiveness.
- **Generalized Advantage Estimation [27]:** Utilized generalized advantage estimation during network weight updates. Calculating the mean and variance of the advantage function across the entire batch made the training process smoother and more stable.
- **Learning Rate Adjustment:** Set initial learning rates for the Actor and Critic to  $1e-4$  and  $3e-4$ , respectively. Reduced them exponentially by an order of magnitude over 1000 training epochs for more stable convergence.
- **Random Initialization and Gradient Clipping:** Initialized network parameters with samples from a standard normal distribution and applied gradient clipping once before updating the network.
- **Activation Function Change [28]:** Replaced ReLU with Tanh in the hidden layers to avoid the vanishing gradient problem associated with negative values.

### 2.3. Training Environment

#### 2.3.1. Chamber Pressure Dynamics

To regulate pressure, a combustion chamber pressure dynamics model is established for training the pressure control agent. The pressure-throat area relationship in a solid rocket motor is described by a zero-dimensional internal ballistic equation [29]:

$$\frac{V_c}{\sqrt{k} \left( \frac{2}{k+1} \right)^{\frac{k+1}{k-1}} c^{*2}} \frac{dp_c}{dt} = \rho_p A_b a p_c^n - \frac{p_c A_{tsum}}{c^*} \quad (1)$$

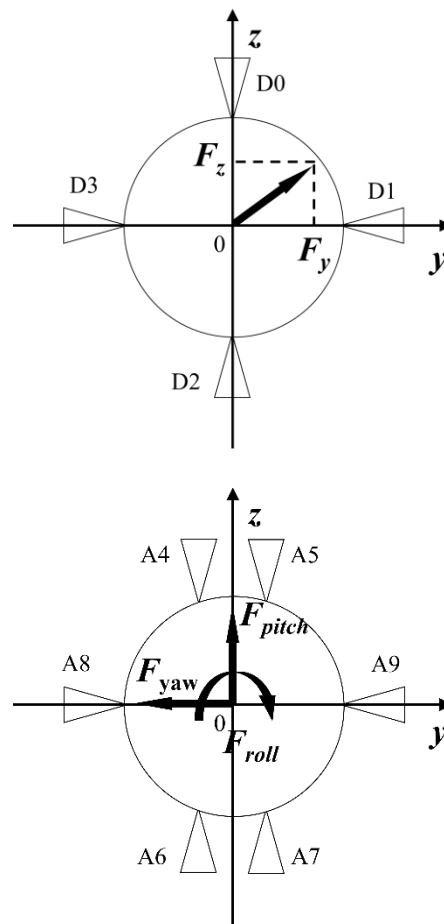
Here,  $V_c$ ,  $c^*$ , and  $k$  denote the combustion chamber free volume, characteristic velocity, and heat capacity ratio of the gas, respectively.  $A_b$  and  $\rho_p$  represent the burning area and density of the propellant, and the burning rate is given by  $a p_c^n$ , where  $P_c$  is the chamber pressure. The term  $A_{tsum}$  denotes the total nozzle opening. Typically, once the propellant grain shape of the engine is confirmed, the thickness of the burned propellant can describe the variation of  $V_c$  and  $A_b$ . By modifying  $A_b$ , it is possible to simulate the dynamic process inside the combustion chamber using

numerical solution methods, thereby creating a training environment for the pressure agent. In this study, we employed the fourth-order Runge-Kutta method to solve the equations.

### 2.3.2. Five-Axis Thrust Calculation

In Figure 5(a), divert nozzles produce lateral acceleration by thrusting at the spacecraft's center of mass, typically in the range of hundreds to thousands of newtons. This allows for rapid orbital adjustments without attitude changes. Nozzles 0 and 2 act in the z-axis, while nozzles 1 and 3 act in the y-axis.

Fig.5(b) illustrates attitude control nozzles at the spacecraft's tail, exerting tens to hundreds of newtons. Nozzles 4 and 7 induce clockwise rolling, while nozzles 5 and 6 induce counterclockwise rolling. Nozzles 4 and 5 pitch up, nozzles 6 and 7 pitch down, nozzle 8 causes left yaw, and nozzle 9 causes right yaw.



a) divert control thrust synthesis b) attitude control thrust synthesis

**Figure 5.** Synthesis method of divert control thrust and attitude control thrust.

The equations for the synthesis of thrust in five-axis direction are as follows:

y-axis thrust:

$$F_y = F_3 - F_1 \quad (2)$$

z-axis thrust:

$$F_z = F_2 - F_0 \quad (3)$$

yaw-axis thrust:

$$F_{yaw} = F_9 - F_8 \quad (4)$$

pitch-axis Thrust:

$$F_{pitch} = (F_6 + F_7) - (F_4 + F_5) \quad (5)$$

roll-axis Thrust:

$$F_{roll} = (F_6 + F_5) - (F_4 + F_7) \quad (6)$$

The thrust generated by each nozzle can be calculated according to the thrust calculation Equation. The thrust can be expressed as the product of the thrust coefficient  $C_f$ , pressure  $P_c$ , and the throat area  $A_t$ .

$$F = C_f P_c A_t \quad (7)$$

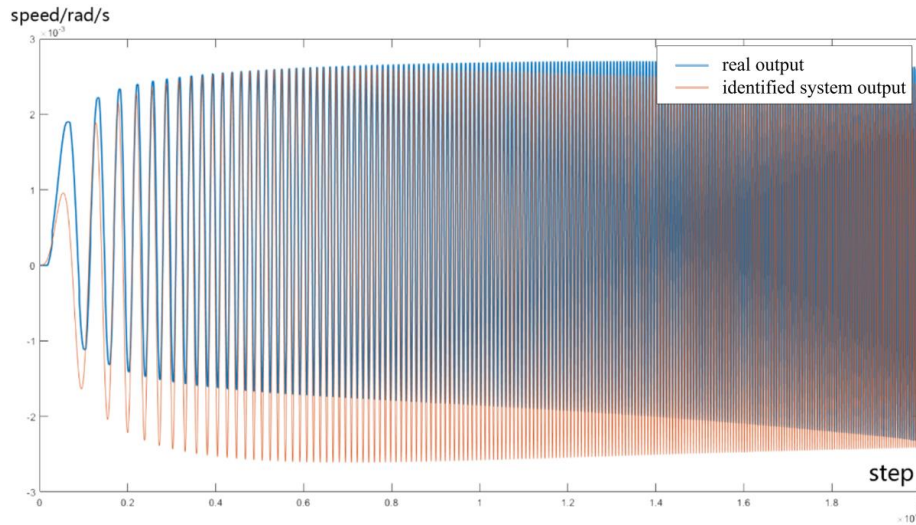
### 2.3.3. Valve Servo Motion

Achieving closed-loop control of nozzle opening by regulating the valve core position is crucial. Modeling the dynamic behavior of the valve accurately is challenging due to the presence of various mechanical components and frictional forces within the gas valve. To address this, we employ system identification techniques to tailor the servo system.

We establish a fourth-order state-space model, achieving a system similarity of 92.1%. The identified observable state space is represented in Equation (8):

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B\omega_{cmd} \\ y(t) &= Cx(t) + D\omega_{cmd} \end{aligned} \quad (8)$$

where  $x(t)$  denotes an n-dimensional state vector,  $A$  is the system matrix,  $B$  represents the control matrix,  $\omega_{cmd}$  denotes the speed command,  $C$  stands for the output matrix,  $D$  indicates the direct transfer matrix, and  $y(t)$  symbolizes the motor speed. Figure 6 illustrates the system output obtained by applying the same swept frequency signal input to both the actual and identification systems, as referenced in prior work [30].



**Figure 6.** Comparison of output between actual system and identification System.

## 2.4. Reward Function

### 2.4.1. Pressure Control Agent's Reward

The pressure control agent utilizes a reward function to manage the total nozzle opening and stabilize chamber pressure near the command, considering three key aspects:

- (1) Deviation Penalty: When chamber pressure significantly deviates from the command, a penalty term is applied to promptly correct the pressure.

$$R_{Pc\_deviation} = -|P_{ccmd}(t_n) - P_c(t_n)| \quad (9)$$

- (2) Synchronization Reward: As chamber pressure approaches the command, a synchronization reward stabilizes the pressure, preventing rapid oscillations.

$$R_{Pc\_synchronization} = -|(P_c(t_n) - P_c(t_{n-1})) - (P_{ccmd}(t_n) - P_{ccmd}(t_{n-1}))| \quad (10)$$

- (3) Steady-State Reward: When chamber pressure consistently tracks the command, a small reward encourages maintaining this state.

$$R_{Pc\_bonus} = 1 \text{ if } |R_{Pc\_deviation}| < \varepsilon_1 \text{ and } |R_{Pc\_synchronization}| < \varepsilon_2 \text{ else } 0 \quad (11)$$

- (4) Dynamic Weights: When the pressure deviation is large, the deviation penalty is intensified; when the pressure deviation is small, the synchronization reward is intensified. We aim for the deviation penalty to have at least half of the weight to ensure tracking accuracy.

$$K_{Pc} = 0.5 \tanh(|P_{ccmd}(t_n) - P_c(t_n)|) + 0.5 \quad (12)$$

Finally the pressure control agent's reward function is as shown in Equation (13):

$$R_{Pc} = K_{Pc} R_{Pc\_deviation} + (1 - K_{Pc}) R_{Pc\_synchronization} + R_{Pc\_bonus} \quad (13)$$

#### 2.4.2. Thrust Distribution Agent's Reward

The thrust distribution agent controls the opening degree of each of the nozzles, with the opening degrees of these nozzles being unknown variables. The known constraints include the 5-axis thrust equation and the total nozzle opening degree command. The reward function design is shown in Equations (14) to (19).

$$R_{F_y} = -|F_{ycmd}(t_n) - F_y(t_n)| \quad (14)$$

$$R_{F_z} = -|F_{zcmd}(t_n) - F_z(t_n)| \quad (15)$$

$$R_{F_{roll}} = -|F_{rollcmd}(t_n) - F_{roll}(t_n)| \quad (16)$$

$$R_{F_{pitch}} = -|F_{pitchcmd}(t_n) - F_{pitch}(t_n)| \quad (17)$$

$$R_{F_{yaw}} = -|F_{yawcmd}(t_n) - F_{yaw\_target}(t_n)| \quad (18)$$

$$R_{val} = -|A_{tsumcmd}(t_n) - A_{tsum}(t_n)| \quad (19)$$

Under normal circumstances, pressure control is prioritized, making the reward item for the total valve opening degree the primary concern. A switch function with a threshold T is implemented; only when this reward item reaches T will other thrust deviation rewards be considered. The thrust distribution agent's reward is finally expressed in Equation (20).

$$R_F = \begin{cases} R_{val} + R_{F_y} + R_{F_z} + R_{F_{roll}} + R_{F_{pitch}} + R_{F_{yaw}}, & \text{if } R_{val} > T \\ R_{val}, & \text{if } R_{val} \leq T \end{cases} \quad (20)$$

#### 2.4.3. Opening Closed-loop Agent's Reward

The opening closed agent is designed to control the motor's displacement, ensuring it quickly approaches and stabilizes near the command, or accurately tracks a continuous curve.

- (1) Displacement Difference: Using the displacement difference as a reward component effectively penalizes significant deviations from the command during initial training, speeding up convergence.

$$R_{opening\_deviation} = -|A_{ticmd}(t_n) - A_{ti}(t_n)| \quad (21)$$

- (2) Velocity Term: To match the valve core's movement speed with the target curve's slope, a velocity term is added. This requires transforming the observation space, using the displacement difference between new and old states to represent the target curve's slope.

$$R_{opening\_synchronization} = -|(A_{ticmd}(t_n) - A_{ticmd}(t_{n-1})) - (A_{ti}(t_n) - A_{ti}(t_{n-1}))| \quad (22)$$

(3) Steady-State Reward: An additional reward is given when both displacement and speed rewards are satisfactory, promoting a state where both closely follow the target curve.

$$R_{opening\_bonus} = 1 \quad \text{if } R_{opening\_deviation} < \varepsilon_3 \quad \text{and} \quad R_{opening\_synchronization} < \varepsilon_4 \quad \text{else } 0 \quad (23)$$

(4) Dynamic Weighting: To ensure accuracy, the displacement difference should account for at least half of the weight in the reward function.

$$K_{opening} = 0.5 \tanh(|A_{ticmd}(t_n) - A_{ti}(t_n)|) + 0.5 \quad (24)$$

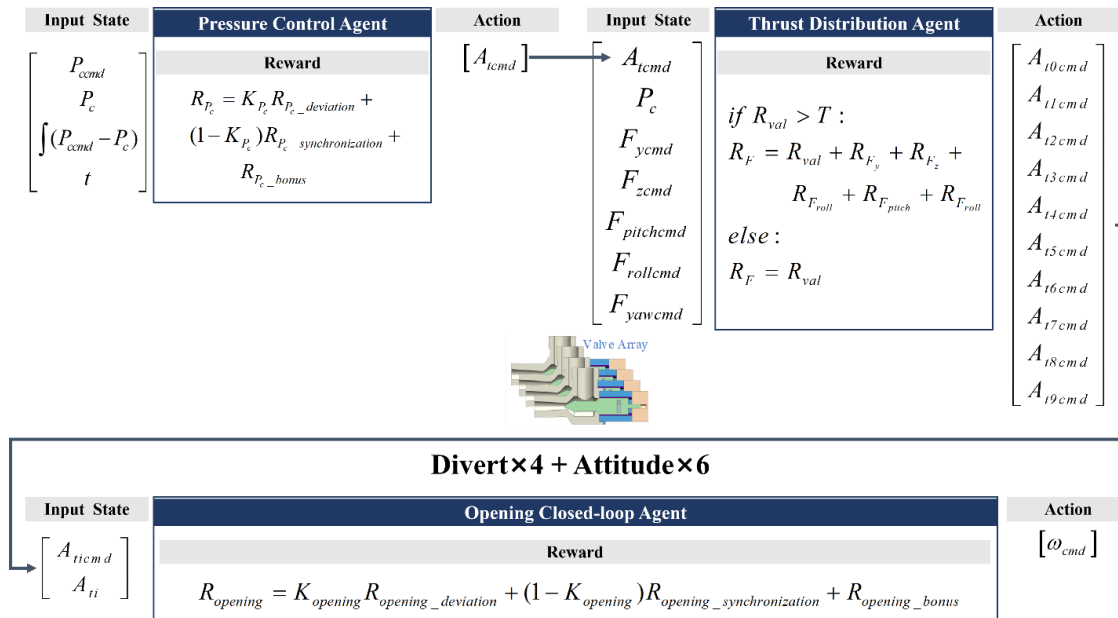
Finally the opening closed-loop agent's reward function is as shown in Equation (25):

$$R_{opening} = K_{opening} R_{opening\_deviation} + (1 - K_{opening}) R_{opening\_synchronization} + R_{opening\_bonus} \quad (25)$$

### 3. Simulation and Results

The control process comprises three levels: pressure control, thrust distribution, and opening closed-loop, managed by the pressure control agent, thrust distribution agent, and opening closed-loop agents, respectively. These agents operate in a cascaded manner, as illustrated in Figure 7, which shows the control architecture of the SADPS.

- Pressure Control: The pressure control agent at the top level regulates the total opening of the nozzle based on the combustion chamber pressure command.
- Thrust Distribution: The thrust distribution agent at the middle level allocates the opening commands to each valve according to the thrust command.
- Opening Closed-loop: At the bottom level, ten opening closed-loop agents control their respective motors to adjust the opening of each valve.



**Figure 7.** Multi-agent control architecture.

The transmission process of the control signal is shown in Figure 8. The upper part represents the four divert control valves, and the lower part represents the six attitude control valves. The outer layer of the flowchart represents the pressure control loop, with the combustion chamber pressure command ( $P_{ccmd}$ ) and pressure ( $P_c$ ) as inputs, and the valve opening area command ( $A_{tsumcmd}$ ) as the output. The middle layer represents the thrust control loop, with the throat area command ( $A_{tsumcmd}$ ), thrust commands ( $F_{ycmd}$ ,  $F_{zcmd}$ ,  $F_{pitchcmd}$ ,  $F_{yawcmd}$ ,  $F_{rollcmd}$ ), and pressure ( $P_c$ ) as inputs, and the valve

openings as the output. Finally, the thrust vector synthesis of the thrusts output by each nozzle yields  $F_y$ ,  $F_z$ ,  $F_{pitch}$ ,  $F_{yaw}$ , and  $F_{roll}$ . As a result, the external system only needs to input the pressure command and the five-axis thrust command into the algorithm, which then outputs the five-axis thrusts, achieving closed-loop engine pressure control and thrust distribution. Due to confidentiality principles regarding the technology, the following simulation results are presented using normalized units.

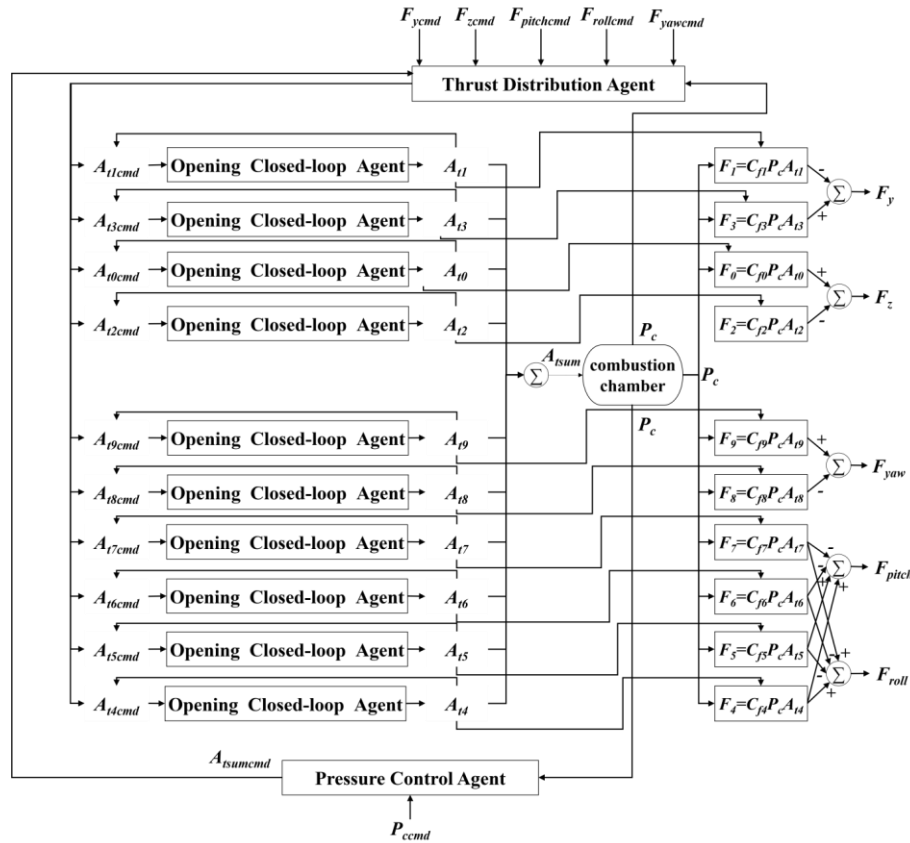


Figure 8. Control signal transmission flowchart.

### 3.1. Performance of the Pressure Control Agent

The conditions are as follows: the initial combustion chamber pressure is set to 1 standard atmosphere, then 1 for 1 second, 0.22 for 1 second, and 1.33 for 1 second, continuously cycling 7 times. The control effect is shown in Figure 9.

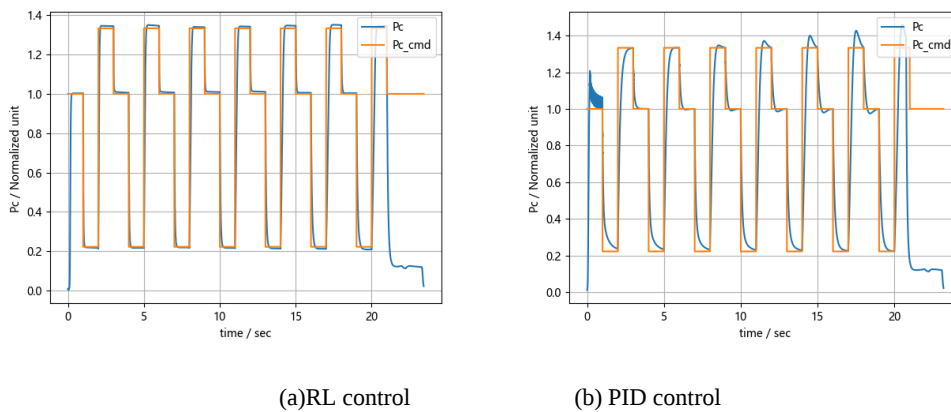


Figure 9. The control performance by pressure control agent.

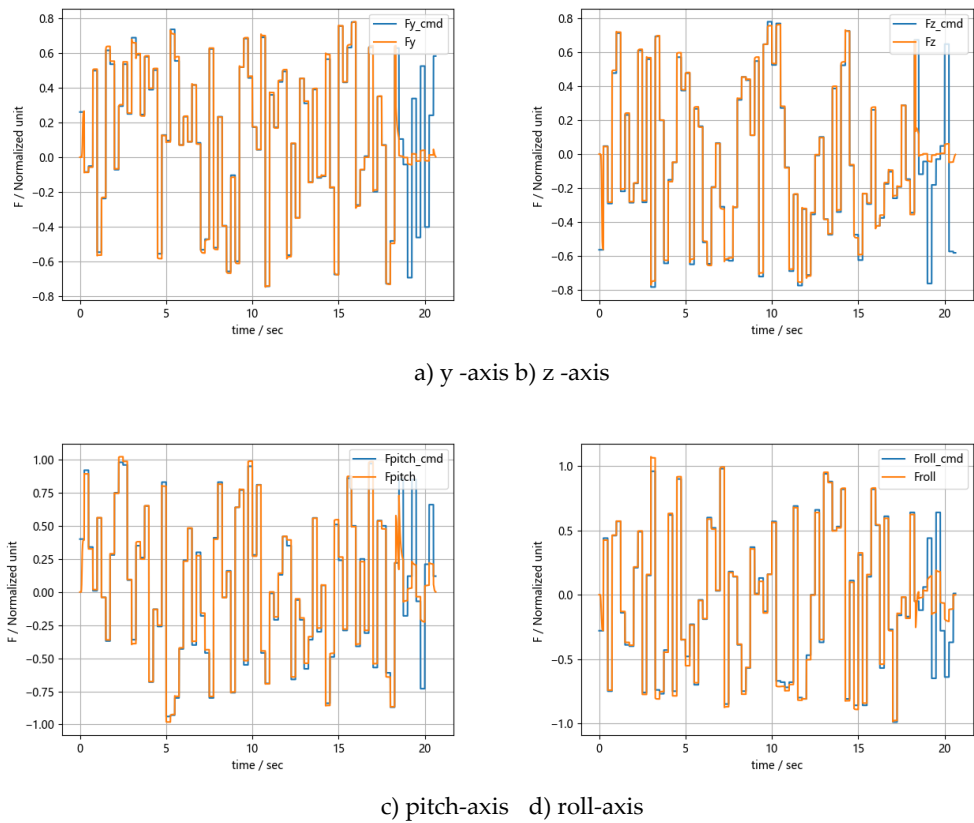
Figure 9 shows that closed-loop pressure regulation using reinforcement learning (RL) is significantly faster and more precise than with PID control. RL adapts its strategy over time as the engine's combustion chamber volume changes, while PID control uses fixed parameters and cannot quickly adjust to the time-varying system. Excluding the initial ignition phase, Table 1 shows the cumulative deviation and average deviation between the pressure command and the actual pressure.

**Table 1.** Statistical results of pressure control performance.

| Model | RL-PPO               |                   |                             | PID                  |                   |                             |
|-------|----------------------|-------------------|-----------------------------|----------------------|-------------------|-----------------------------|
| Item  | Cumulative Deviation | Average Deviation | Average Regulation Time, ms | Cumulative Deviation | Average Deviation | Average Regulation Time, ms |
| Value | 304.75               | 0.043             | 244                         | 352.50               | 0.051             | 827                         |

3.2. Performance of the Thrust Distribution Agent

The conditions are as follows: the pressure is set to 1. The divert control thrust commands( $F_{ycmd}$ ,  $F_{zcmd}$ ) undergo random step changes with an upper limit of 0.8, and the attitude control thrust commands( $F_{pitchcmd}$ ,  $F_{yawcmd}$ ,  $F_{rollcmd}$ ) undergo random step changes with an upper limit of 0.1. The results are shown in Figure 10.



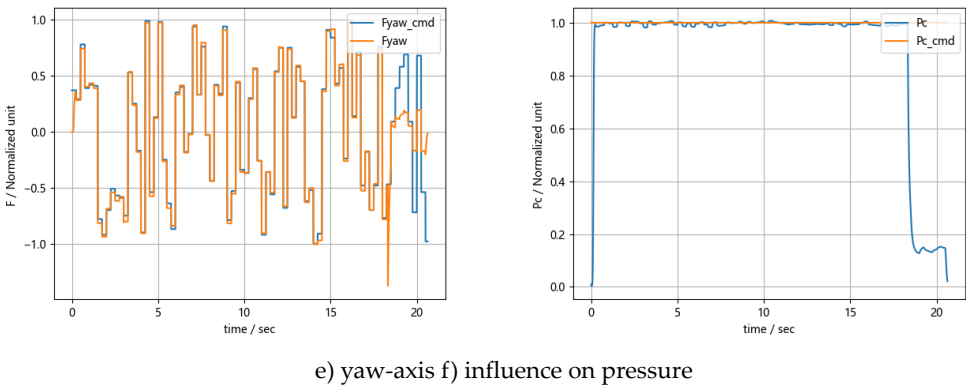


Figure 10. The control performance by thrust distribution agent.

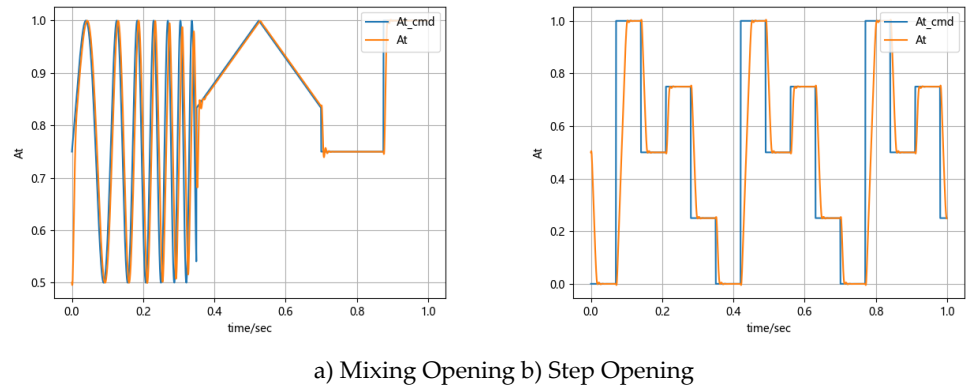
Figure 10 shows that the error for orbital control thrust( $F_y, F_z$ ) ranges from 2% to 5%, while the error for attitude control thrust( $F_{pitch}, F_{yaw}, F_{roll}$ ) ranges from 5% to 10%. Pressure fluctuations due to thrust changes are within  $\pm 2\%$ . Excluding initial ignition and propellant depletion phases, the cumulative and step deviations between thrust command and actual thrust are shown in Table 2.

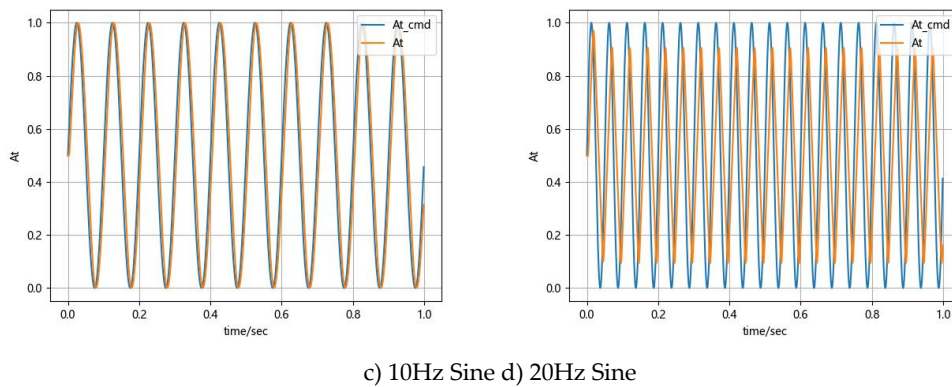
Table 2. Statistical results of thrust distribute performance.

| Item                   | Divert Control Thrust | Attitude Control Thrust | Impact on Pressure |
|------------------------|-----------------------|-------------------------|--------------------|
| Cumulative Deviation   | 252.8                 | 732.3                   | 30.1               |
| Step Average Deviation | 0.036                 | 0.011                   | 0.004              |

3.3. Performance of the Opening Closed-Loop Age

The control effectiveness of the nozzle opening closed-loop agent was tested by recording displacement tracking, as shown in Figure 11. The closed-loop control showed an error of less than 1% at the step stable position and demonstrated excellent response time and anti-overshoot characteristics.





**Figure 11.** The control performance by opening closed-loop agent.

Adjustment time, amplitude, and phase lag under different valve opening commands were tested, as shown in Table 3 and Table 4.

**Table 3.** Control performance for step valve opening commands.

| Command Type | Adjustment Time ( $t_{90}$ ) /ms |
|--------------|----------------------------------|
| 0~1          | 37                               |
| 0~0.75       | 29                               |
| 0~0.5        | 22                               |
| 0~0.25       | 22                               |
| 1~0          | 37                               |
| 0.75~0       | 29                               |
| 0.5~0        | 26                               |
| 0.25~0       | 25                               |

**Table 4.** Control performance for sinusoidal valve opening command.

| Frequency | Amplitude Attenuation | Phase Lag |
|-----------|-----------------------|-----------|
| 5hz       | 0.016-0.992           | 18°       |
| 10hz      | 0.036-0.936           | 36°       |
| 15hz      | 0.052-0.968           | 75.6°     |
| 20hz      | 0.188-0.828           | 144°      |

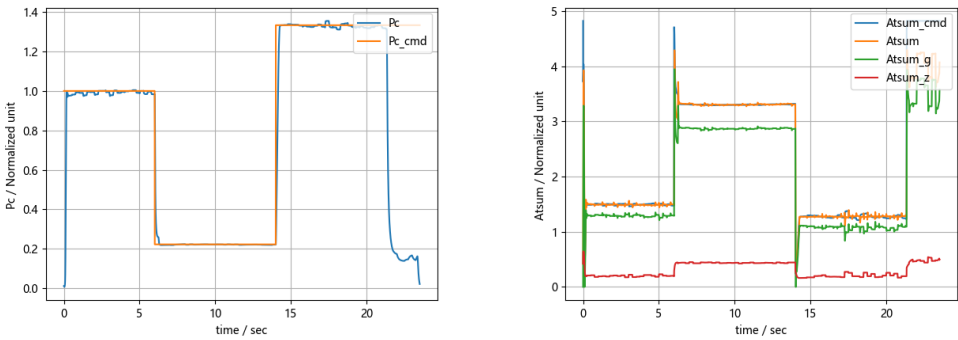
In step curve tracking, the reinforcement learning algorithm controlled overshoot effectively, keeping it below 1% in all scenarios. The full-stroke step response time was  $\leq 40$  ms. For continuous curve tracking, the full-stroke 20 Hz amplitude attenuation was  $\leq 40\%$ , with a phase lag of 144°.

### 3.4. Performance of the Muti-Agents

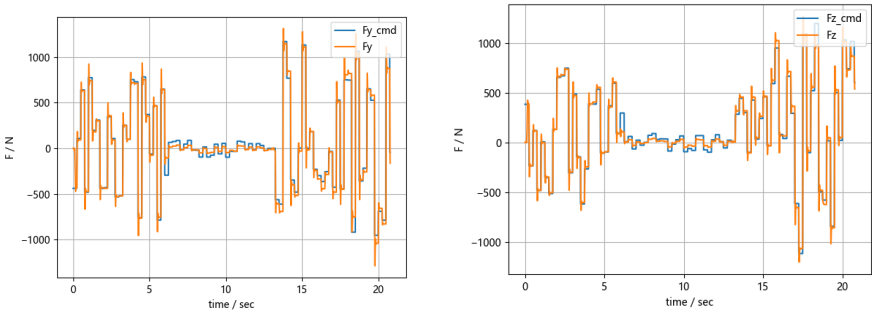
The trained model of the aforementioned multi-agent system was tested under the typical conditions shown in Figure 2, tracking the pressure and thrust commands. The conditions were designed as follows:

- a. Initial 6 seconds:
  - Pressure:1
  - Divert control thrust command ( $F_{ycmd}$ ,  $F_{zcmd}$ ): Random steps within the upper limit of 0.8
  - Attitude control thrust ( $F_{pitchcmd}$ ,  $F_{yawcmd}$ ,  $F_{rollcmd}$ ): Random steps within the upper limit of 1
- b. Next 8 seconds:
  - Pressure: 0.22
  - Divert control thrust command ( $F_{ycmd}$ ,  $F_{zcmd}$ ): Random steps within the upper limit of 0.1
  - Attitude control thrust ( $F_{pitchcmd}$ ,  $F_{yawcmd}$ ,  $F_{rollcmd}$ ): Random steps within the upper limit of 0.3

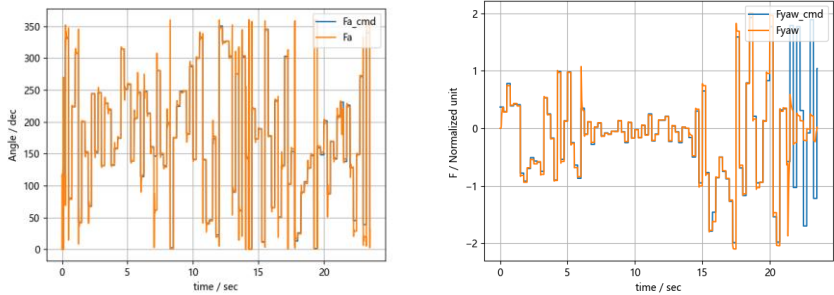
- c. Final 6 seconds:
- Pressure: 1.33
  - Divert control thrust command ( $F_{ycmd}$ ,  $F_{zcmd}$ ): Random steps within the upper limit of 1.2
  - Attitude control thrust ( $F_{pitchcmd}$ ,  $F_{yawcmd}$ ,  $F_{rollcmd}$ ): Random steps within the upper limit of 2
- The control time step was set to 2.5 ms. The control effect of the multi-agent system is shown in Figure 12.



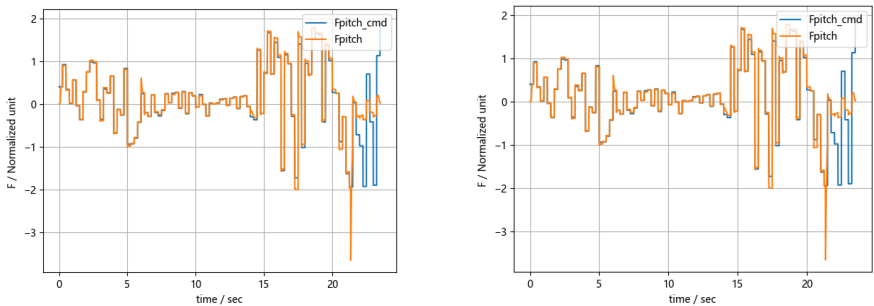
a) Pressure b) Nozzle opening



c) y-axis thrust d) z-axis thrust



e) Angle f) yaw-axis thrust



g) pitch-axis thrust h) roll-axis thrust

**Figure 12.** The performance of the joint agents.

A statistical analysis was conducted on the joint control effects of multiple intelligent agents. Under these conditions, the hierarchical multi-agent system achieved complete closed-loop control of the solid attitude and orbital control engine, including combustion chamber pressure, thrust, and valve opening.

In terms of error:

- The impact of thrust distribution on pressure closed-loop was within 5%.
- The divert control thrust error did not exceed 5%.
- The attitude control thrust error did not exceed 10%.

In terms of adjustment time:

- The pressure closed-loop adjustment time was within 300 ms.
- The thrust adjustment time was within 100 ms.
- The valve opening adjustment time was within 40 ms.

#### 4. Discussions and Conclusions

The solid attitude and divert propulsion system (SADPS) is characterized by its time-variant, strongly coupled, highly nonlinear, and complex nature. To tackle these challenges, we integrated reinforcement learning (RL) for control system design, adaptive parameter adjustment, and continuous data-driven learning. This study provides several key findings and insights into the effective management of pressure and thrust through a unified RL control architecture.

##### 4.1. Integrated Control Architecture and System Dynamics

Our approach involved designing a comprehensive control architecture to manage the closed-loop regulation of pressure, thrust, and valve positions. The system's dynamics, characterized by nonlinearity and time variance, show that operational duration notably influences the dynamic response of pressure regulation. The RL-based control system adapts to these dynamic changes more effectively than traditional PID control, significantly improving response speed and reducing regulation time.

##### 4.2. Impact of Coupling Effects on Thrust Optimization

The system's design includes shared combustion chambers among multiple valves, causing fluctuations in the total exhaust area. This configuration poses challenges in optimizing thrust distribution across multiple axes, considering constraints imposed by thrust equations for various directions. The coupling effects lead to acceptable pressure fluctuations, demonstrating the system's robustness in handling complex interactions within the propulsion system.

##### 4.3. Enhancing Control Through Reinforcement Learning and Adaptive Strategies

Integrating RL techniques with thrust deviation rewards optimized thrust control. RL enables adaptive adjustment based on real-time engine operating conditions, significantly improving pressure regulation effectiveness compared to traditional methods. The introduction of a thrust deviation reward in RL reduced errors in divert and attitude thrust, showcasing the potential of RL in handling under-constrained optimization problems inherent in thrust distribution.

##### 4.4. Development of Realistic Training Environments and State Space Modeling

A critical aspect of our methodology was creating a highly realistic RL training environment through meticulous state space identification. This ensured accurate initial training of neural networks and subsequent fine-tuning on the actual system. Incorporating engine operating time into the RL observation space significantly improved pressure regulation time compared to fixed-parameter PID control.

#### 4.5. Optimizing Training Efficiency and Technical Enhancements

We implemented technical optimizations such as simplifying neural network architectures, using learning rate decay strategies, and optimizing activation functions. These enhancements collectively contributed to a marked improvement in the overall performance and robustness of the control system. The RL-based multi-agent architecture outperformed traditional PID control by enhancing response speed in pressure closed-loop regulation, achieving under-constrained thrust distribution, and enabling high-speed control with minimal overshoot in valve opening closed-loop scenarios.

#### 4.6. Limitations and Future Work

Despite the significant improvements, the proposed control method has certain limitations. One major limitation is the dependence of agent training on the accuracy of the training environment. The simulated engine model within the training environment may not fully reflect the real operating conditions of the engine, and conducting real engine training is both costly and hazardous. Additionally, the control effect of the thrust distribution agent is challenging to optimize, often failing to achieve perfect coordination between pressure and thrust. The reward function struggles to balance these two aspects effectively.

To overcome these limitations, future work will focus on developing a semi-physical simulation system capable of realistically mimicking SADPS. This system aims to provide a reusable and realistic training environment for the algorithm. For thrust distribution, we plan to introduce global constraints into the RL framework, such as minimizing propellant consumption, maximizing total impulse, ensuring safe control, minimizing valve core oscillations, and maintaining constant openings for orbital and attitude control. These constraints will help the RL agents converge more effectively, achieving better coordination between pressure and thrust.

### 5. Conclusions

This study showcases the effectiveness of using reinforcement learning (RL) for pressure control and multi-agent systems for thrust and valve management in solid attitude and orbital control engines. The RL-based pressure control agent significantly outperforms traditional PID control, offering faster and more precise regulation. The thrust distribution agent maintains high accuracy and stability in controlling thrust forces with minimal impact on pressure, highlighting the importance of closed-loop control for valve response times. The opening closed-loop agent excels in tracking both step and continuous commands, ensuring low overshoot and rapid response times.

Under typical operating conditions, the hierarchical multi-agent system provides comprehensive closed-loop control, achieving desired performance metrics with minimal errors and swift adjustments. These findings underscore the potential of advanced control strategies in enhancing the performance and reliability of propulsion systems. However, addressing the limitations related to training environment accuracy and thrust coordination remains a priority for future research. Developing a semi-physical simulation system and integrating global constraints into the RL framework will be critical steps toward achieving more robust and reliable control in SADPS.

**Acknowledgments:** This work was supported in part by the Foundation Strengthening Plan Project and the National Natural Science Foundation of China under Grant 61973210. The authors would like to extend their sincere gratitude to the Shanghai Space Propulsion Technology Research Institute, China, for their invaluable technical support, especially to Xin Duan, Weihua Zhou and Wenguo Hou. Additionally, heartfelt appreciation is extended to the technical support provided by Shanghai Jiao Tong University, China, particularly to Haiyu Wang and Zi Fang.

**Declaration of Competing Interest:** The authors declare that they have no conflicts of interest to this work. All authors have read and approved this version of the article. Neither the entire paper nor any part of its contents has been published or has been accepted elsewhere.

## References

1. Liu L, Liang G. Optimization Selection of Regulated Pressurization System Schemes for Liquid Attitude and Divert Propulsion Systems. *Procedia Engineering* 2015;99:1247–51. <https://doi.org/10.1016/j.proeng.2014.12.655>.
2. Kim S, Huh H. Recent Progress in R&D and Prospect of Divert and Attitude Control System(DACS). *Journal of the Korean Society of Propulsion Engineers* 2012;16:62–72. <https://doi.org/10.6108/KSPE.2012.16.6.062>.
3. Napier J, Garmy V. Controllable Solid Propulsion for Launch Vehicle and Spacecraft Application. 57th International Astronautical Congress, Valencia, Spain: American Institute of Aeronautics and Astronautics; 2006. <https://doi.org/10.2514/6.IAC-06-C4.2.04>.
4. Fang Z, Liang G. Transient simulation of a differential piston warm gas self-pressurization system for liquid attitude and divert propulsion system. *Chinese Journal of Aeronautics* 2018;31:698–709. <https://doi.org/10.1016/j.cja.2018.02.001>.
5. Ha D-S, Kim HJ. Dynamic characteristic modeling and simulation of an aerospike-shaped pintle nozzle for variable thrust of a solid rocket motor. *Acta Astronautica* 2022;201:364–75. <https://doi.org/10.1016/j.actaastro.2022.09.031>.
6. Dai C, Qiang H, Wang X. Research on Divert and Attitude Control System Technology of Ballistic Missile Midcourse Maneuver Penetration Warhead. *FCIS* 2023;3:97–109. <https://doi.org/10.54097/fcis.v3i1.6342>.
7. Davis C, Gerards A. Variable Thrust Solid Propulsion Control Using Labview. 39th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Huntsville, Alabama: American Institute of Aeronautics and Astronautics; 2003. <https://doi.org/10.2514/6.2003-5241>.
8. Pérez-Roca S, Marzat J, Piet-Lahanier H, Langlois N, Farago F, Galeotta M, et al. A survey of automatic control methods for liquid-propellant rocket engines. *Progress in Aerospace Sciences* 2019;107:63–84. <https://doi.org/10.1016/j.paerosci.2019.03.002>.
9. Zhu M, Li Y, Chen X, Mi J, Li W, Xue H. Design of nonlinear control method for pressure of pintle solid rocket motor. *Acta Astronautica* 2024;215:296–307. <https://doi.org/10.1016/j.actaastro.2023.12.008>.
10. Bergmans J, Di Salvo R. Solid Rocket Closed-Loop Pressure Control. 39th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Huntsville, Alabama: American Institute of Aeronautics and Astronautics; 2003. <https://doi.org/10.2514/6.2003-4968>.
11. Joner S, Quinquis I. Control of an Exoatmospheric Kill Vehicle with a Solid Propulsion Attitude Control System. AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, Colorado: American Institute of Aeronautics and Astronautics; 2006. <https://doi.org/10.2514/6.2006-6572>.
12. Waxenegger-Wilfing G, Sengupta U, Martin J, Armbruster W, Hardi J, Juniper M, et al. Early detection of thermoacoustic instabilities in a cryogenic rocket thrust chamber using combustion noise features and machine learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 2021;31:063128. <https://doi.org/10.1063/5.0038817>.
13. Wang M, Xie Y. Design of the optimal thruster combinations table for the real time control allocation of spacecraft thrusters. Proceedings of the 48th IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference, Shanghai, China: IEEE; 2009, p. 5063–8. <https://doi.org/10.1109/CDC.2009.5400664>.
14. Martel F. OPTIMAL 6 AXIS COMMAND OF A SPACE VEHICLE WITH A PRECOMPUTED THRUSTER SELECTION CATALOGUE TABLE n.d.
15. Ankersen F, Wu S-F, Aleshin A, Vankov A, Volochinov V. Optimization of Spacecraft Thruster Management Function. *Journal of Guidance, Control, and Dynamics* 2005;28:1283–90. <https://doi.org/10.2514/1.11531>.
16. Lee W, Eun Y, Bang H, Lee H. Efficient Thrust Distribution with Adaptive Pressure Control for Multinozzle Solid Propulsion System. *Journal of Propulsion and Power* 2013;29:1410–9. <https://doi.org/10.2514/1.B34645>.
17. Lee H, Bang H. Efficient Thrust Management Algorithm for Variable Thrust Solid Propulsion System with Multi-Nozzles. *Journal of Spacecraft and Rockets* 2020;57:328–45. <https://doi.org/10.2514/1.A34550>.
18. Lim Y, Lee W, Bang H, Lee H. Thrust distribution for attitude control in a variable thrust propulsion system with four ACS nozzles. *Advances in Space Research* 2017;59:1848–60. <https://doi.org/10.1016/j.asr.2017.01.002>.
19. Javaid M, Haleem A, Singh RP, Suman R. Artificial Intelligence Applications for Industry 4.0: A Literature-Based Study. *J Ind Intg Mgmt* 2022;07:83–111. <https://doi.org/10.1142/S2424862221300040>.
20. Nian R, Liu J, Huang B. A review On reinforcement learning: Introduction and applications in industrial process control. *Computers & Chemical Engineering* 2020;139:106886. <https://doi.org/10.1016/j.compchemeng.2020.106886>.
21. [21] Xu, Weili, Xuerong Yang, and Gefei Shi. 2024. "The Maintenance of Orbital States in a Floating Partial Space Elevator Using the Reinforcement Learning Method." *Aerospace* 11(10): 855. doi:10.3390/aerospace11100855.

22. [22] Skarka, Wojciech, and Rukhseena Ashfaq. 2024. "Hybrid Machine Learning and Reinforcement Learning Framework for Adaptive UAV Obstacle Avoidance." *Aerospace* 11(11): 870. doi:10.3390/aerospace11110870.
23. Pham, Tan-Hanh, Godwyll Aikins, Tri Truong, and Kim-Doang Nguyen. 2024. "Adaptive Compensation for Robotic Joint Failures Using Partially Observable Reinforcement Learning." *Algorithms* 17(10): 436. doi:10.3390/a17100436.
24. Wang, Xing, Hong Yi, Jia Xu, Chuanyi Xu, and Lifei Song. 2024. "PID Controller Based on Improved DDPG for Trajectory Tracking Control of USV." *Journal of Marine Science and Engineering* 12(10): 1771. doi:10.3390/jmse12101771.
25. Song A, Wang N, Li J, Ma B, Chen X. Transient flow characteristics and performance of a solid rocket motor with a pintle valve. *Chinese Journal of Aeronautics* 2020;33:3189–205. <https://doi.org/10.1016/j.cja.2020.04.023>.
26. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal Policy Optimization Algorithms 2017. <https://doi.org/10.48550/ARXIV.1707.06347>.
27. Schulman J, Moritz P, Levine S, Jordan M, Abbeel P. High-Dimensional Continuous Control Using Generalized Advantage Estimation 2015. <https://doi.org/10.48550/ARXIV.1506.02438>.
28. Engstrom L, Ilyas A, Santurkar S, Tsipras D, Janoos F, Rudolph L, et al. Implementation Matters in Deep Policy Gradients: A Case Study on PPO and TRPO 2020. <https://doi.org/10.48550/ARXIV.2005.12729>.
29. X.S. Wu, J. Chen, D. Wang. Principles of Solid Rocket Motors. Beijing: Weapon Industry Press; 2015.
30. Wang H, Fu Z, Yan J, Hua Z, Guan Y. Modeling and Simulation of High Frequency and High Precision Multi Axis Cooperative Control Based on Reinforcement Learning. 2023 3rd International Conference on Computer, Control and Robotics (ICCCR), Shanghai, China: IEEE; 2023, p. 286–90. <https://doi.org/10.1109/ICCCR56747.2023.10193913>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.