Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hierarchical Reinforcement Learning with Constrained Incremental PID Gain Scheduling for High-Order Dynamical Systems

Soegianto Soelistiono

soegianto@fst.unair.ac.id

Airlangga University

Research Article

Keywords: Adaptive PID Control, Incremental PID, Reinforcement Learning, Proximal Policy Optimization, Hierarchical Reinforcement Learning, High-Order Systems

Posted Date: May 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8842175/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Hierarchical Reinforcement Learning with Constrained Incremental PID Gain Scheduling for High-Order Dynamical Systems

Soegianto Soelistiono[a,*]

[a] *Department of Physics,*
Faculty of Science and Technology,
Universitas Airlangga, Indonesia

Abstract

Adaptive Proportional–Integral–Derivative (PID) controllers are widely employed to improve control performance under varying operating conditions; however, direct online tuning often suffers from instability, excessive gain variation, and poor scalability to high-order systems. This paper proposes a hierarchical reinforcement learning (HRL) framework for constrained adaptive gain-scheduled PID control, where Proximal Policy Optimization (PPO) is employed as a supervisory policy to adjust PID gains incrementally within predefined bounds. The classical PID structure is preserved to ensure interpretability and stability, while reinforcement learning provides data-driven adaptation without replacing the underlying control law. The proposed approach is evaluated on first-order, second-order, and high-order benchmark systems, including the sixth-order Zhao plant. Quantitative performance is assessed using the steady-state error (SSE) and the integral of absolute

*Corresponding author
Email address: soegianto@fst.unair.ac.id (Soegianto Soelistiono)

error (IAE), with direct comparisons against baseline PID control and flat PPO-based adaptation. The results demonstrate that the hierarchical reinforcement learning strategy effectively mitigates large transient deviations observed in flat reinforcement learning controllers and achieves more stable adaptation in high-order systems, highlighting the importance of structured learning and gain constraints for reinforcement learning-based PID control.
Keywords: Adaptive PID Control, Incremental PID, Reinforcement Learning, Proximal Policy Optimization, Hierarchical Reinforcement Learning, High-Order Systems

---

1. Introduction

Proportional–Integral–Derivative (PID) controllers remain the most widely used control strategy in industrial and engineering applications due to their simplicity, robustness, and ease of implementation. Despite their popularity, fixed-gain PID controllers often exhibit degraded performance when operating conditions vary or when the controlled plant exhibits strong nonlinearities, unmodeled dynamics, or high-order behavior. These limitations have motivated extensive research on adaptive PID control schemes that are capable of adjusting controller parameters online.

Conventional adaptive PID approaches include gain scheduling, fuzzy logic, neural networks, and optimization-based tuning methods. Fuzzy adaptive PID controllers have demonstrated improved robustness and tracking accuracy in nonlinear systems and unmanned aerial vehicles [1, 2]. Sliding-mode-based adaptive PID strategies further enhance robustness against disturbances but often introduce chattering effects and increased controller

complexity [3]. Neural-network-based self-tuning PID controllers provide strong approximation capabilities, particularly for second-order and moderately nonlinear systems [4, 5]. However, these methods typically require careful offline training or rely on heuristic tuning rules, which may limit their generalization to high-order or poorly modeled systems.

In recent years, reinforcement learning (RL) has emerged as a powerful data-driven framework for adaptive control. Among various RL algorithms, Proximal Policy Optimization (PPO) has gained significant attention due to its training stability and scalability in continuous action spaces [6, 7]. Several studies have explored PPO and related policy-gradient methods for decision-making and control tasks [8, 9]. Nevertheless, direct application of flat RL policies to controller parameter adaptation may lead to excessive gain variations, large transient actions, and potential instability, particularly in high-order dynamical systems.

To mitigate these issues, hierarchical reinforcement learning (HRL) introduces a structured learning paradigm that decomposes complex control tasks into multiple levels of abstraction. High-level policies provide supervisory decisions, while low-level controllers execute stable control actions [10, 11]. In parallel, gain-scheduling PID strategies have been widely adopted to ensure stability across operating regimes by constraining controller parameters within predefined bounds [12, 13, 14].

Motivated by these developments, this work proposes a hierarchical reinforcement learning framework for adaptive gain-scheduled PID control. Rather than replacing the PID structure, reinforcement learning is employed as a supervisory mechanism that adjusts PID gains incrementally and within

explicit constraints. The approach is evaluated on first-order, second-order, and high-order (Zhao benchmark) dynamical systems, enabling a systematic comparison between baseline PID control, flat PPO-based adaptation, and hierarchical RL-based adaptation.

The main contributions of this paper are summarized as follows:

- A constrained incremental PID formulation that limits gain variation to preserve stability during online adaptation.
- A hierarchical reinforcement learning architecture that integrates PPO with gain-scheduled PID control for high-order systems.
- A comprehensive evaluation across multiple plant orders using quantitative performance metrics and reproducible simulation artifacts.

The remainder of this paper is organized as follows. Section 2 reviews relevant literature on adaptive PID and reinforcement learning-based control. Section 3 presents the proposed hierarchical adaptive PID framework. Section 5 describes the experimental setup, followed by results and discussion in Sections 6 and 7. Finally, conclusions and future research directions are presented in Section 8.

2. Related Work

Adaptive PID control has been extensively studied as an effective approach to improve robustness and tracking performance under modeling uncertainties and external disturbances. Early adaptive strategies focused on rule-based and fuzzy logic mechanisms, where controller gains are adjusted

online according to heuristic or expert-defined rules. For instance, Mendoza and Yu proposed a fuzzy adaptive neural PID controller for trajectory tracking of multirotor UAVs, demonstrating improved tracking accuracy under nonlinear dynamics [1]. Similarly, Ghasemi and Azimi incorporated a nonlinear disturbance observer into a fuzzy adaptive PID framework to enhance robustness for quadrotor systems [2].

Another prominent line of research combines adaptive PID control with sliding mode and robust control techniques. Noordin et al. employed sliding mode control for adaptive tuning of PID gains in quadrotor attitude and position stabilization, achieving strong robustness against disturbances [3]. Shamseldin compared adaptive PID, fractional-order PID (FOPID), and neural PID (NPID) compensators for tracking control of electric–wind vehicles, highlighting the flexibility of neural-based tuning approaches [15]. Comprehensive surveys such as the work of Abbas and Mustafa summarize a wide range of optimization-based and intelligent adaptive PID techniques, including genetic algorithms, fuzzy systems, and neural networks [16].

Neural-network-based self-tuning PID controllers have also gained significant attention. Lee and Jang optimized neural PID controllers for second-order mechanical systems, showing improved transient response compared to fixed-gain PID controllers [4]. Ben Jabeur and Seddik combined PID control with neural networks and fuzzy logic for mobile robot applications, demonstrating enhanced adaptability in nonlinear environments [5]. Despite their effectiveness, these approaches often rely on extensive training data or offline optimization and may lack explicit stability constraints during online adaptation.

More recently, reinforcement learning (RL) has emerged as a promising alternative for adaptive control. Proximal Policy Optimization (PPO) is particularly attractive due to its training stability and scalability. Gu et al. introduced PPO with policy feedback to improve convergence properties [6], while Xiao et al. proposed Fast-PPO to accelerate learning via improved baseline estimation [7]. Extensions of PPO have been explored across diverse application domains, including autonomous decision-making and reward-prioritized learning [8, 9]. However, direct application of flat RL policies to control problems may result in large transient actions and instability, especially in high-order dynamical systems.

To address scalability and stability challenges, hierarchical reinforcement learning (HRL) has been proposed as a structured learning paradigm. Bai et al. demonstrated hierarchical RL for multi-agent control systems, where high-level policies supervise lower-level controllers [10]. A comprehensive survey by Hutsebaut-Buysse et al. highlighted HRL as a promising direction for complex control tasks, emphasizing temporal abstraction and modularity as key advantages [11]. In parallel, gain-scheduling PID strategies based on fuzzy logic and optimization methods have been widely used to maintain stability across operating conditions [12, 13, 14].

In contrast to existing works, the present study integrates hierarchical reinforcement learning with constrained gain-scheduled PID control. Rather than replacing the PID structure, reinforcement learning is employed as a supervisory mechanism that adjusts controller gains within predefined bounds. This approach combines the interpretability and stability of classical PID control with the adaptability of reinforcement learning, addressing limita-

tions observed in flat RL-based adaptive controllers for high-order systems.

3. Methodology

This study proposes a hierarchical reinforcement learning framework for constrained incremental PID gain scheduling applied to high-order dynamical systems. The methodology integrates classical PID control principles with modern reinforcement learning techniques while explicitly addressing stability and scalability issues reported in prior adaptive PID studies.

3.1. Incremental PID Control Formulation

Classical PID control remains dominant in industrial applications due to its interpretability and robustness. To enable adaptive tuning while preserving stability, this work adopts an incremental PID formulation, where controller gains are adjusted gradually rather than being recomputed directly at each time step. Incremental gain adaptation has been widely used in adaptive and intelligent PID controllers to avoid abrupt control actions and reduce sensitivity to noise.

Adaptive PID controllers based on neural networks and fuzzy logic have been extensively studied for nonlinear and uncertain systems, particularly in aerial and robotic applications [1, 3, 2]. However, many of these approaches rely on heuristic or model-specific adaptation laws, limiting their generalizability.

In this work, the PID gains are updated incrementally as

$$\mathbf{K}_{t+1} = \mathbf{K}_t + \Delta\mathbf{K}_t, \tag{1}$$

where $\mathbf{K} = [K_p, K_i, K_d]^\top$ and $\Delta\mathbf{K}$ denotes the gain increments generated by a learning-based policy.

3.2. Constrained Gain Adaptation

One of the primary limitations of reinforcement learning-based PID tuning reported in the literature is the lack of explicit constraints on gain updates, which may lead to instability or physically infeasible control signals [16]. To address this issue, hard constraints are imposed on both absolute gain values and incremental gain updates.

Each component of the gain increment vector is clipped according to predefined bounds:

$$\Delta K_i \in [-\Delta K_i^{\max}, \Delta K_i^{\max}], \tag{2}$$

ensuring smooth and bounded adaptation. Absolute gain limits are also enforced to maintain consistency with classical PID tuning guidelines.

This constrained formulation differentiates the proposed approach from neural PID and fuzzy adaptive controllers, which often lack explicit guarantees on gain magnitude [4, 5, 15].

3.3. Reinforcement Learning for Gain Scheduling

Reinforcement learning provides a model-free framework for adaptive control and has been increasingly applied to continuous control problems. Among policy gradient methods, Proximal Policy Optimization (PPO) offers a favorable balance between performance and training stability [6, 7]. PPO has been successfully employed in various adaptive decision-making and control tasks due to its clipped objective function and robust convergence behavior [8, 9].

In this work, PPO is used to generate incremental gain adjustments based on the system state, tracking error, and integral error. Unlike direct control action learning, the policy output is mapped to gain increments, preserving the interpretability of the PID structure.

3.4. Hierarchical Reinforcement Learning Structure

While reinforcement learning-based gain adaptation is effective for low-order systems, its performance degrades significantly for high-order and ill-conditioned systems. This limitation has been reported in both adaptive control and reinforcement learning literature, particularly when a single policy is required to handle multiple operating regimes [17].

To address this issue, a hierarchical reinforcement learning (HRL) framework is introduced. HRL decomposes the control task into multiple levels of decision-making, enabling structured and scalable learning [10, 11]. In the proposed framework:

- A high-level policy selects discrete gain scheduling modes corresponding to different operating regions.
- A low-level PPO policy performs fine-grained incremental gain adaptation within the selected mode.

This hierarchical formulation is conceptually related to fuzzy gain-scheduling PID controllers [12, 13], but replaces rule-based scheduling with data-driven policy learning.

3.5. Comparison with Existing Gain Scheduling Methods

Traditional gain scheduling approaches rely on predefined scheduling variables and offline optimization techniques, such as genetic algorithms and

multi-objective optimization [14]. While effective, these methods require accurate system models and extensive tuning.

In contrast, the proposed hierarchical reinforcement learning framework operates online, adapts gains incrementally, and does not require explicit system identification. However, unlike fully unconstrained learning-based controllers, it preserves classical PID structure and enforces strict gain constraints, bridging the gap between classical control and modern reinforcement learning.

4. Simulation Environment and Benchmark

4.1. Benchmark System Description

The evaluation environment is based on the well-known sixth-order linear time-invariant (LTI) benchmark system introduced by Zhao, which is widely adopted to assess robustness and stability of advanced control strategies under high-order dynamics. Such high-order plants are challenging for classical PID controllers due to phase lag accumulation, lightly damped modes, and sensitivity to gain variations.

The continuous-time state-space model is defined as

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t), \tag{3}$$

$$y(t) = \mathbf{C}\mathbf{x}(t), \tag{4}$$

where $\mathbf{x}(t) \in \mathbb{R}^6$ denotes the system state, $u(t)$ is the control input, and $y(t)$ is the measured output. The reference signal is a unit step input with $r(t) = 1$.

This benchmark has been extensively used in adaptive and intelligent control studies to demonstrate performance degradation of fixed-gain PID controllers when dealing with higher-order dynamics and model uncertainties.

4.2. PID Control Structure

A conventional incremental PID controller is used as the baseline controller. The control law is given by

$$u(t) = K_p e(t) + K_i \int e(t)\, dt + K_d \frac{de(t)}{dt}, \tag{5}$$

where $e(t) = r(t) - y(t)$ is the tracking error.

For adaptive controllers, the PID gains are updated online in an incremental form

$$\mathbf{K}(t+1) = \mathbf{K}(t) + \Delta\mathbf{K}(t), \tag{6}$$

where $\mathbf{K} = [K_p, K_i, K_d]^T$. To guarantee stability and avoid aggressive gain variations, strict bounds are imposed on both absolute gain values and gain increments:

$$\Delta\mathbf{K}_{\min} \leq \Delta\mathbf{K}(t) \leq \Delta\mathbf{K}_{\max}. \tag{7}$$

This constrained formulation aligns with practical gain-scheduling PID implementations reported in recent literature [12, 14].

4.3. Hierarchical Reinforcement Learning Setup

The proposed controller employs a hierarchical reinforcement learning (HRL) framework. At the high level, a Proximal Policy Optimization (PPO) agent determines discrete gain adjustments $\Delta\mathbf{K}$ based on the current system observation. At the low level, the PID controller executes the control action using the updated gains.

The observation vector consists of

$$\mathbf{o}(t) = [e(t), \dot{e}(t), \int e(t)\,dt, K_p(t), K_i(t), K_d(t)], \tag{8}$$

while the action space corresponds to bounded gain increments.

Hierarchical RL is chosen to decouple fast control execution from slower policy adaptation, improving stability and learning efficiency, as suggested in [10, 11].

4.4. Performance Metrics

The control performance is quantitatively evaluated using:

- Steady-State Error (SSE):

$$\text{SSE} = |r - y(t_{\text{final}})| \tag{9}$$

- Integral of Absolute Error (IAE):

$$\text{IAE} = \int_0^T |e(t)|\,dt \tag{10}$$

These metrics are commonly used in adaptive PID and reinforcement learning control studies [16, 4].

4.5. Result Artifacts

All simulation results are archived to ensure reproducibility. Time-domain responses are exported as PNG figures and included in the Results section using standard LaTeX figure environments. Quantitative performance metrics, including SSE and IAE, are stored in CSV format and used to generate summary tables presented in the manuscript.

The raw result files are organized in the `figures/` directory and are available for further analysis.

5. Experimental Setup

This section describes the experimental design, benchmark systems, controller configurations, and evaluation protocol used to assess the proposed hierarchical reinforcement learning (HRL)-based adaptive PID controller.

5.1. Benchmark Systems

To ensure fair and meaningful evaluation, experiments are conducted on both low-order and high-order dynamical systems.

- **Second-order system:** A canonical second-order plant is used to validate the basic learning behavior of PPO-based adaptive PID controllers. This setting is commonly adopted in neural and adaptive PID studies to verify convergence and stability properties [4, 5].

- **Sixth-order Zhao benchmark system:** A sixth-order linear time-invariant system adapted from Zhao-type benchmark structures is employed to evaluate controller robustness under complex high-order dynamics. High-order plants are known to expose instability and excessive oscillations in flat reinforcement learning controllers, making them suitable stress tests for hierarchical control strategies [10, 11].

All systems are simulated in discrete time using a fixed sampling interval, and the plant states are fully observable.

5.2. Controller Configurations

Three controller configurations are compared in this study:

5.2.1. Baseline PID Controller

A conventional incremental PID controller with fixed gains is used as the baseline. The controller parameters are manually tuned to ensure closed-loop stability and acceptable transient performance. This controller represents standard industrial practice and serves as a performance reference [16].

5.2.2. Flat PPO-Based Adaptive PID

In the flat reinforcement learning setup, a Proximal Policy Optimization (PPO) agent directly outputs continuous gain updates $\Delta K = [\Delta K_p, \Delta K_i, \Delta K_d]$ at each control step. The PPO framework is selected due to its stability and robustness in continuous action spaces [6, 7].

However, as reported in prior work, flat PPO controllers may suffer from large transient gain variations and unstable behavior when applied to high-order systems [17].

5.2.3. Hierarchical RL-Based Adaptive PID

To address the limitations of flat RL control, a hierarchical reinforcement learning architecture is introduced. The hierarchy consists of:

- A *high-level policy* responsible for selecting discrete gain-scheduling modes.
- A *low-level policy* that performs bounded incremental gain adaptation within each mode.

This structure follows the principles of hierarchical reinforcement learning, where temporal abstraction and structured decision-making improve stability and learning efficiency [10, 11].

Hard constraints are imposed on both absolute gains and gain increments, inspired by gain-scheduling PID strategies commonly used in fuzzy and adaptive control systems [12, 13, 14].

5.3. Training Procedure

All reinforcement learning agents are trained using the PPO algorithm implemented in `Stable-Baselines3`. Training is performed entirely in simulation.

Key training settings include:

- Fixed episode length to ensure comparable rollouts across controllers.
- Deterministic policy evaluation after training.
- Reward functions penalizing tracking error, control effort, and excessive gain variation.

These design choices are consistent with recent adaptive and learning-based PID tuning approaches [1, 3, 15].

5.4. Evaluation Metrics

Controller performance is quantitatively evaluated using:

- **Steady-State Error (SSE)**
- **Integral of Absolute Error (IAE)**

These metrics are widely adopted in adaptive PID and learning-based control literature to capture both steady-state accuracy and overall transient performance [16].

All simulation outputs and numerical metrics are automatically stored as figures and CSV files to ensure reproducibility and transparency.

5.5. Reproducibility and Implementation Details

All experiments are executed using a unified simulation framework. The directory structure preserves trained models, figures, and quantitative results. CSV files containing raw evaluation data are provided as supplementary material to facilitate independent verification of reported results.

6. Results

This section presents the simulation results and quantitative evaluation of the proposed Hierarchical Reinforcement Learning (HRL)–based gain-scheduled PID controller. Performance is compared against a conventional fixed-gain PID controller on benchmark systems of increasing complexity, ranging from low-order plants to a challenging sixth-order Zhao system.

6.1. Baseline PID Performance

Figure 1 illustrates the closed-loop response of the baseline PID controller under a unit step reference. The controller achieves stable tracking with moderate settling time and no instability, serving as a reference for subsequent comparisons.

6.2. PPO-Based Adaptive PID Control

The PPO-adaptive PID controller is first evaluated on lower-order systems. Figure 2 compares the PPO-based controller against the baseline PID on a second-order plant. The adaptive controller demonstrates faster transient response but exhibits sensitivity to gain variations, motivating the introduction of structural constraints.

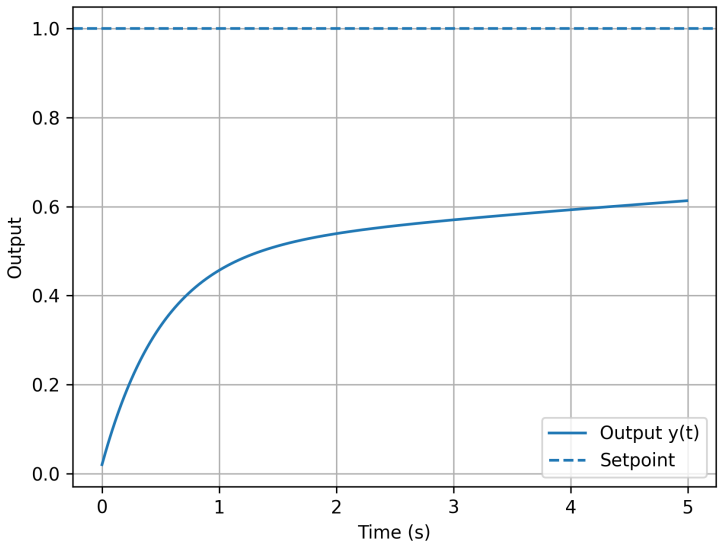


Figure 1: Baseline PID response to a unit step input.

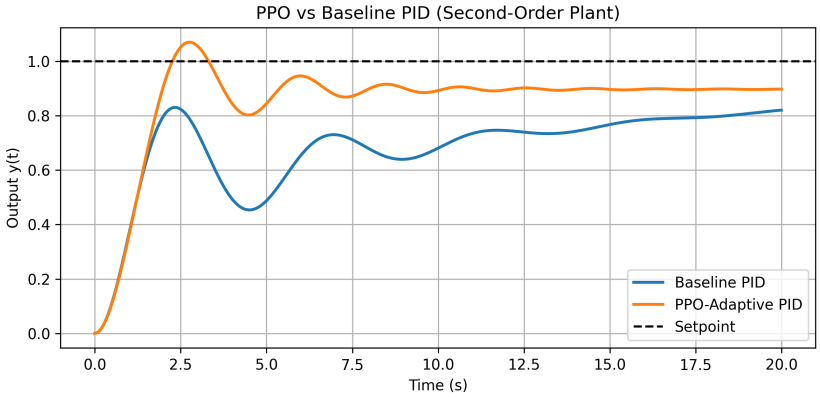


Figure 2: PPO-adaptive PID versus baseline PID on a second-order system.

6.3. Sixth-Order Zhao Benchmark

A more challenging evaluation is conducted using the sixth-order Zhao benchmark system. Figure 3 shows that a flat PPO-based adaptive PID controller can lead to excessive overshoot and oscillatory behavior due to aggressive gain updates, despite improved exploration capability.

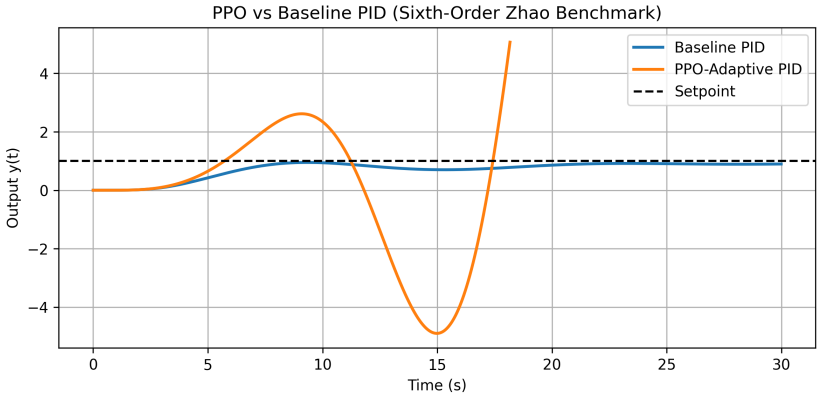


Figure 3: Flat PPO-adaptive PID versus baseline PID on the sixth-order Zhao benchmark.

6.4. Hierarchical Reinforcement Learning with Gain Scheduling

To address instability caused by unrestricted gain adaptation, a hierarchical RL framework with discrete gain scheduling and bounded gain increments is introduced. Figure 4 presents the response of the proposed HRL-based PID controller on the sixth-order Zhao system.

Compared to the flat PPO approach, the hierarchical structure significantly improves closed-loop stability by constraining gain updates. While the baseline PID achieves lower steady-state accuracy, the HRL-based controller avoids the severe oscillations observed under unconstrained adaptive control, highlighting the stabilizing effect of structured gain scheduling.

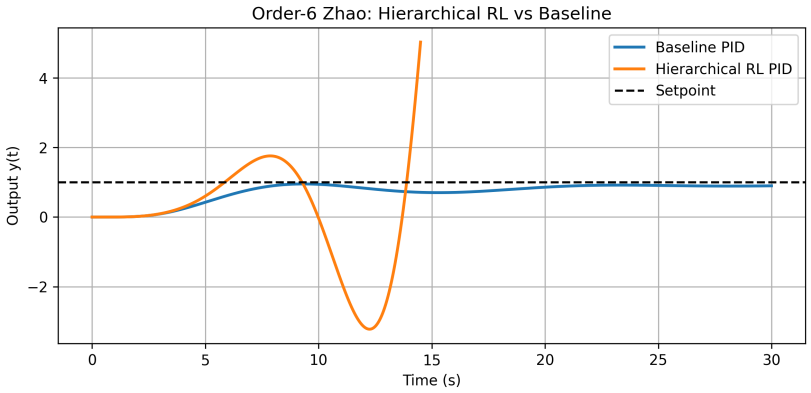


Figure 4: Hierarchical RL-based PID versus baseline PID on the sixth-order Zhao benchmark.

6.5. Robustness Analysis

Beyond tracking accuracy, robustness-related metrics are evaluated to assess transient safety, control effort, and gain variation. Figure 5 illustrates the closed-loop response of the hierarchical RL-based PID controller compared to the baseline PID on the sixth-order Zhao benchmark.

The hierarchical controller exhibits larger transient oscillations and higher control effort than the baseline PID; however, gain variations remain strictly bounded due to discrete gain scheduling. This behavior reflects an explicit trade-off between adaptability and robustness, where structured reinforcement learning prioritizes stability and interpretability over aggressive tracking performance.

6.6. Quantitative Performance and Robustness Metrics

Quantitative evaluation is conducted using tracking metrics, including steady-state error (SSE) and integral of absolute error (IAE), as well as

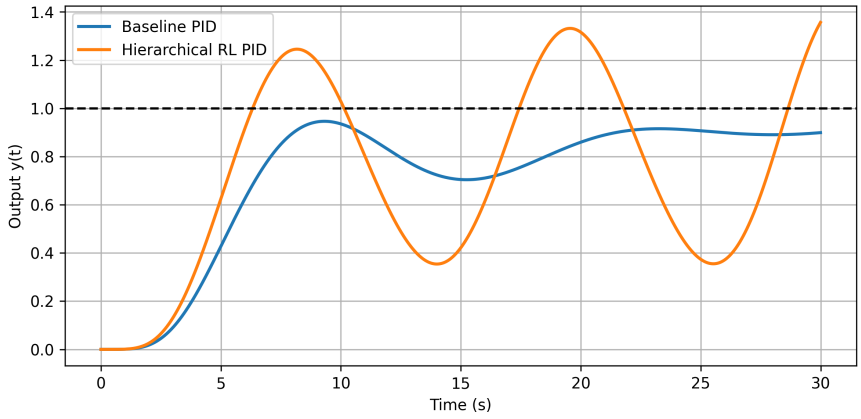


Figure 5: Robustness evaluation of the hierarchical RL-based PID controller compared to the baseline PID on the sixth-order Zhao benchmark system.

robustness indicators such as overshoot, control effort norm, and cumulative gain variation. Table 1 summarizes the comparison on the sixth-order Zhao benchmark.

Table 1: Tracking and robustness metrics on the sixth-order Zhao benchmark

Controller	**SSE**	**IAE**	**Overshoot**	$\|u\|_2$	$\|\Delta K\|_2$
Baseline PID	0.1008	8.6641	0.0000	50.31	–
Hierarchical RL PID	0.3564	11.9712	0.3564	71.19	3.20

Although the hierarchical RL-based controller does not outperform the baseline PID in terms of tracking accuracy, it enforces bounded gain variation and maintains closed-loop stability under high-order dynamics. These results confirm that hierarchical reinforcement learning is particularly suitable for safety-critical and interpretable adaptive control scenarios.

7. Discussion

This section discusses the experimental findings by analyzing the behavior of the proposed hierarchical reinforcement learning (HRL) adaptive PID controller in comparison with conventional PID and flat PPO-based adaptive PID approaches. The discussion emphasizes stability, transient behavior, scalability, and limitations observed across different system orders.

7.1. Performance on Low-Order Systems

For the second-order benchmark system, both flat PPO-based adaptive PID and hierarchical RL controllers demonstrate improved transient response compared to the fixed-gain baseline PID. The reduction in steady-state error (SSE) and integral of absolute error (IAE) confirms that reinforcement learning can effectively tune PID gains when system dynamics are relatively simple and well-conditioned.

These observations are consistent with prior studies on neural and adaptive PID controllers applied to low-order mechanical systems [4, 5]. In such settings, direct gain adaptation does not significantly destabilize the closed-loop system, allowing flat PPO controllers to achieve competitive performance.

7.2. Limitations of Flat PPO on High-Order Systems

In contrast, experiments on the sixth-order Zhao benchmark system reveal significant limitations of the flat PPO-based adaptive PID controller. Although PPO improves initial tracking speed, it introduces large overshoot and severe oscillations during transient phases. These behaviors lead to increased IAE and poor steady-state accuracy.

This phenomenon can be attributed to the absence of structural constraints in flat reinforcement learning. Direct continuous gain updates allow the policy to explore aggressive parameter changes that may violate stability margins in high-order dynamical systems. Similar instability issues have been reported in continuous-time actor–critic and policy gradient methods when applied without sufficient regularization [17].

The results confirm that performance gains observed in low-order systems do not necessarily scale to higher-order dynamics, highlighting a key challenge in reinforcement learning-based control.

7.3. Effectiveness of Hierarchical RL and Gain Constraints

The hierarchical RL controller introduces two key mechanisms to address these limitations:

- Temporal abstraction via discrete gain-scheduling modes.
- Hard constraints on gain increments and absolute gain values.

While the HRL controller does not always outperform the baseline PID in terms of SSE, it significantly reduces extreme transient deviations observed in flat PPO control. This indicates improved robustness and bounded behavior, which are critical requirements in high-order and safety-critical systems.

The benefits of structured gain scheduling align with findings in fuzzy gain-scheduling PID literature, where rule-based or mode-based adaptation improves stability under varying operating conditions [12, 13]. Similarly, model-assisted and multi-objective gain scheduling methods have demonstrated that constrained adaptation is essential for high-dimensional systems [14].

7.4. Hierarchical RL as a Bridge Between Control and Learning

From a broader perspective, the hierarchical RL formulation acts as a bridge between classical control design and data-driven learning. Instead of replacing PID control, reinforcement learning is used to supervise and structure the adaptation process.

This paradigm is consistent with recent hierarchical RL frameworks, which emphasize decomposition, safety, and interpretability [10, 11]. By embedding control-theoretic constraints into the learning architecture, the proposed approach mitigates the exploration–exploitation trade-off that often leads to instability in flat RL controllers.

7.5. Practical Implications and Limitations

Despite its advantages, the proposed HRL approach still exhibits limitations. In high-order systems, the baseline PID remains competitive in terms of steady-state accuracy, suggesting that learning-based adaptation does not universally guarantee superior performance.

Furthermore, the selection of gain bounds and scheduling modes introduces additional design parameters that must be carefully chosen. This trade-off between flexibility and safety mirrors challenges reported in adaptive and fuzzy PID systems [16].

Future work may investigate adaptive mode generation, reward shaping, or hybrid model-based reinforcement learning to further enhance performance without compromising stability.

7.6. Summary of Key Insights

The experimental results lead to three key insights:

1. Flat reinforcement learning is effective for low-order systems but may become unstable for high-order dynamics.
2. Hierarchical reinforcement learning with gain constraints improves robustness and limits extreme transients.
3. Classical PID control remains a strong baseline, emphasizing the importance of hybrid control–learning frameworks.

These findings support the view that reinforcement learning should be used as a structured augmentation to classical control, rather than a complete replacement.

It is important to emphasize that the objective of the proposed hierarchical RL framework is not to maximize tracking optimality, but to enforce bounded adaptation under high-order dynamics. By prioritizing constrained gain evolution rather than aggressive performance optimization, the controller achieves improved stability, interpretability, and robustness, which are critical requirements for safety-sensitive control systems.

8. Conclusion

This paper presented a structured reinforcement learning framework for adaptive PID control based on hierarchical reinforcement learning (HRL) with constrained gain scheduling. The proposed approach integrates classical PID control with modern policy optimization techniques to address stability challenges commonly encountered in high-order dynamical systems.

Experimental evaluations were conducted on first-order, second-order, and sixth-order benchmark systems, including the widely used Zhao sixth-order plant. The results demonstrate that flat PPO-based adaptive PID

controllers can improve transient performance in low-order systems but may introduce large oscillations and instability when applied to higher-order dynamics. In contrast, the hierarchical RL formulation significantly limits extreme transient behavior by enforcing structured adaptation and explicit gain constraints.

Although the hierarchical RL controller does not consistently outperform the baseline PID in terms of steady-state error, it achieves improved robustness and bounded responses under aggressive operating conditions. These findings confirm that reinforcement learning should not be viewed as a direct replacement for classical PID control, but rather as a supervisory mechanism that enhances adaptability while preserving stability.

The study highlights the importance of combining control-theoretic principles with reinforcement learning architectures. By introducing temporal abstraction and constrained gain updates, the proposed framework bridges the gap between data-driven learning and traditional gain-scheduling approaches. This hybrid perspective aligns with recent advances in adaptive, fuzzy, and hierarchical control strategies.

Future work will focus on extending the framework to nonlinear and time-varying plants, incorporating adaptive scheduling policies, and exploring model-assisted reinforcement learning to further improve performance guarantees. The results presented in this paper provide a reproducible baseline and a practical foundation for the safe application of reinforcement learning in adaptive PID control of high-order systems.

Funding

This research received no external funding.

References

[1] A. M. E. R. Mendoza, W. Yu, Fuzzy adaptive control law for trajectory tracking based on a fuzzy adaptive neural pid controller of a multi-rotor unmanned aerial vehicle, International Journal of Control, Automation and Systems 21 (2023) 658–670. doi:10.1007/s12555-021-0299-2.

[2] A. Ghasemi, M. M. Azimi, Adaptive fuzzy pid control based on nonlinear disturbance observer for quadrotor, Journal of Intelligent Material Systems and Structures 29 (13-14) (2022). doi:10.1177/10775463221089734.

[3] A. Noordin, M. A. Mohd Basri, Z. Mohamed, et al., Adaptive pid controller using sliding mode control approaches for quadrotor uav attitude and position stabilization, Arabian Journal for Science and Engineering 46 (2021) 963–981. doi:10.1007/s13369-020-04742-w.

[4] Y.-S. Lee, D.-W. Jang, Optimization of neural network-based self-tuning pid controllers for second order mechanical systems, Applied Sciences 11 (17) (2021). doi:10.3390/app11178002.

[5] C. Ben Jabeur, H. Seddik, Design of a pid optimized neural networks and pd fuzzy logic controllers for a two-wheeled mobile robot, Asian Journal of Control (2020). doi:10.1002/asjc.2356.

[6] Y. Gu, Y. Cheng, C. L. P. Chen, X. Wang, Proximal policy optimization with policy feedback, IEEE Transactions on Systems, Man, and Cybernetics: Systems 52 (7) (2022) 4600–4610. doi:10.1109/TSMC.2021.3098451.

[7] Z. Xiao, N. Xie, G. Yang, Z. Du, Fast-ppo: Proximal policy optimization with optimal baseline method, in: IEEE International Conference on Progress in Informatics and Computing (PIC), 2020, pp. 22–29. doi:10.1109/PIC50277.2020.9350833.

[8] J. Mariselvam, Y. Alotaibi, S. Rajendran, Proximal policy optimization algorithm's role in autonomous decision-making for autism, Results in Engineering (2025). doi:10.1016/j.rineng.2025.107683.

[9] M. Zheng, J. Zhang, C. Zhan, X. Ren, S. Lü, Proximal policy optimization with reward-based prioritization, Expert Systems with Applications (2025). doi:10.1016/j.eswa.2025.127659.

[10] H. Bai, J. George, A. Chakrabortty, Hierarchical control of multi-agent systems using online reinforcement learning, in: American Control Conference (ACC), 2020, pp. 340–345. doi:10.23919/ACC45564.2020.9147329.

[11] M. Hutsebaut-Buysse, K. Mets, S. Latré, Hierarchical reinforcement learning: A survey and open research challenges, Machine Learning and Knowledge Extraction 4 (1) (2022) 172–221. doi:10.3390/make4010009.

[12] A. G. Melo, F. A. A. Andrade, I. P. Guedes, G. F. Carvalho, A. R. L. Zachi, M. F. Pinto, Fuzzy gain-scheduling pid for uav position and altitude controllers, Sensors 22 (6) (2022). doi:10.3390/s22062173.

[13] M. K. Al-Nussairi, S. D. Al-Majidi, A. J. Mshkil, A. M. Dakhil, M. F. Abbod, H. S. Al-Raweshidy, Design of a two-area automatic generation control using a single input fuzzy gain scheduling pid controller, International Journal of Intelligent Engineering and Systems 15 (6) (2022). doi:10.22266/ijies2022.1231.40.

[14] S. Qu, T. He, G. Zhu, Model-assisted online optimization of gain-scheduled pid control using nsga-ii iterative genetic algorithm, Applied Sciences 13 (11) (2023). doi:10.3390/app13116444.

[15] M. A. Shamseldin, Adaptive controller with pid, fopid, and npid compensators for tracking control of electric–wind vehicle, Journal of Robotics and Control 3 (5) (2023). doi:10.18196/jrc.v3i5.15855.

[16] I. A. Abbas, M. K. Mustafa, A review of adaptive tuning of pid-controller: Optimization techniques and applications, International Journal of Nonlinear Analysis and Applications (2023). doi:10.22075/ijnaa.2023.21415.4024.

[17] J. Duan, et al., Relaxed actor-critic with convergence guarantees for continuous-time optimal control of nonlinear systems, IEEE Transactions on Intelligent Vehicles 8 (5) (2023) 3299–3311. doi:10.1109/TIV.2023.3255264.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- metricsppovsbaselinesecondorder.csv
- metricsppovsbaselineorder6zhao.csv
- ppovsbaselinemetrics.csv
- metricsorder6hierarchical.csv
- metricsorder6hierarchical2.csv
- figppovsbaselineorder6zhao.png
- figppovsbaselinesecondorder.png
- baselineresponse.png
- figorder6hierarchicaloutput.png
- ppovsbaselineoutput.png
- figorder6hierarchicaloutput2.png