Dhirubhai Ambani
Institute of Information and Communication Technology

IE415: CONTROL OF AUTONOMOUS SYSTEMS

TUNING PID CONTROLLERS USING TD3-RL FOR NONLINEAR SYSTEMS

Name: Jeet Patel
ID: 202201216

Prof. Sujay Kadam

MOTIVATION

FOR CHOOSING THIS TOPIC

INTRODUCTION

- Brief overview of PID controllers and their significance in control systems.
  - Proportional: Responds to current error.
  - Integral: Addresses accumulated error over time.
  - Derivative: Predicts and compensates for future error changes.

- Challenges in tuning PID for nonlinear systems.
  - Nonlinearity
  - Dynamic Changes
  - Manual Tuning

- Motivation for using Reinforcement Learning (RL) and TD3 algorithm
  - Adaptability
  - Optimal Control
  - Automation

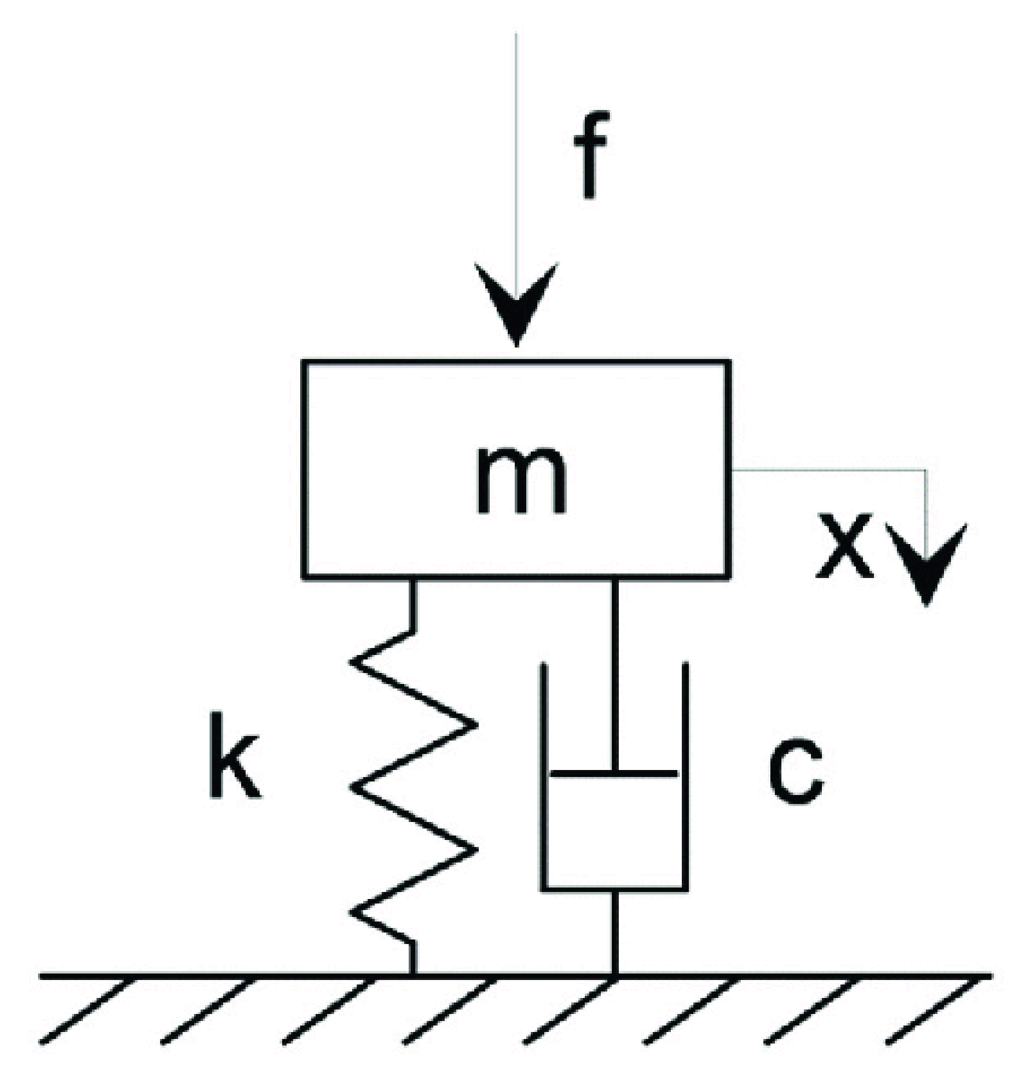
OBJECTIVE

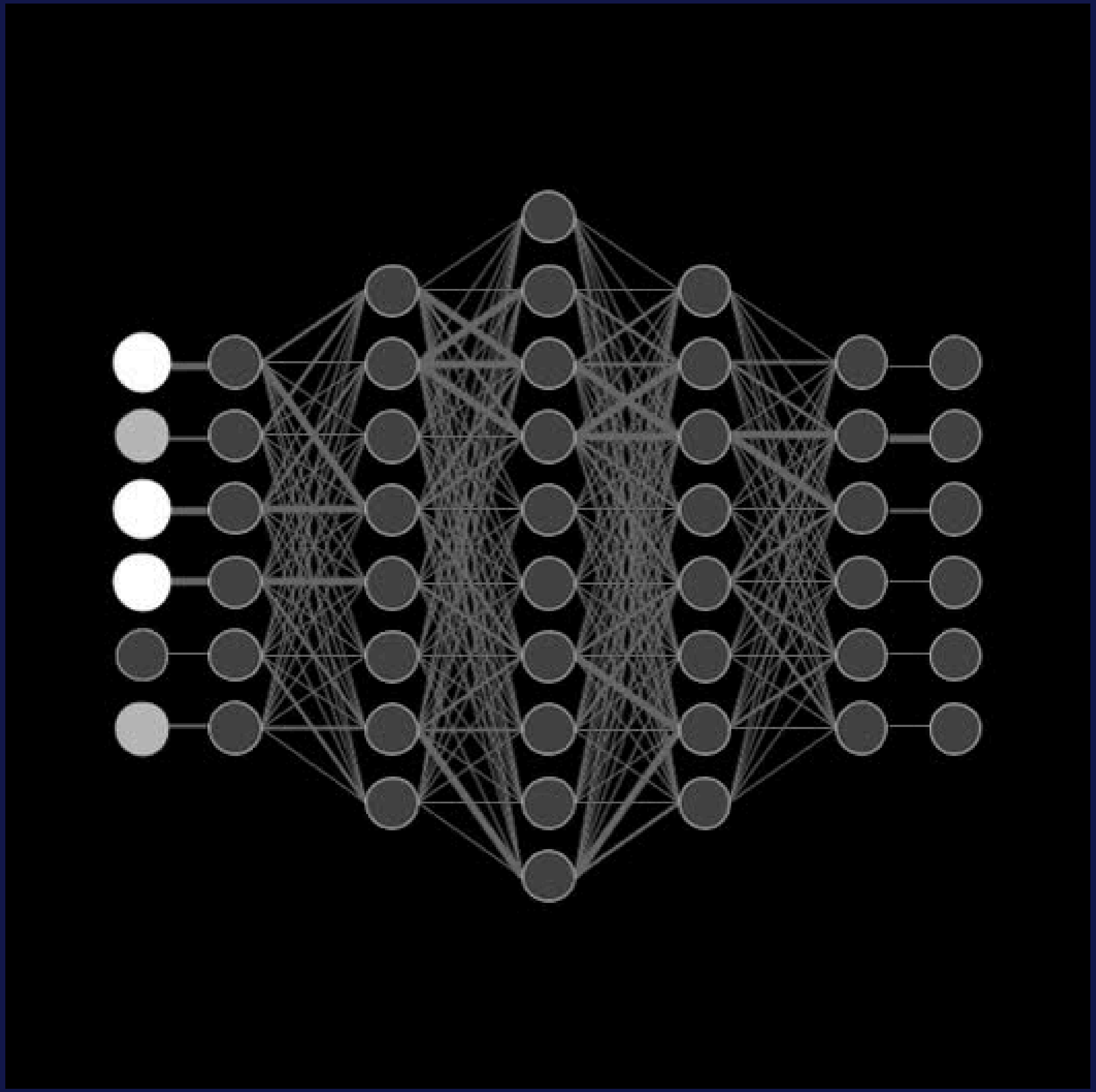
- Simulate and evaluate the performance of RL-tuned PID controllers.
- Compare RL-based tuning with conventional PID tuning.
- Apply the TD3 algorithm to train a PID controller for a mass-spring-damper system.

SYSTEM DESCRIPTION

- Behavior:
  - Spring: Resists displacement proportionally (kx).
  - Damper: Opposes velocity (cx’) for stability.
  - Mass: Reacts to applied forces with acceleration (mx”).
- Purpose:
  - Models real-world systems like car suspensions or robotic joints.
  - Used to analyze dynamic responses to various inputs (e.g., step, sinusoidal).
- Control Objective:
  - Achieve stable performance with minimal overshoot and fast settling by controlling the force F.

$$mx'' + cx' + kx = F$$





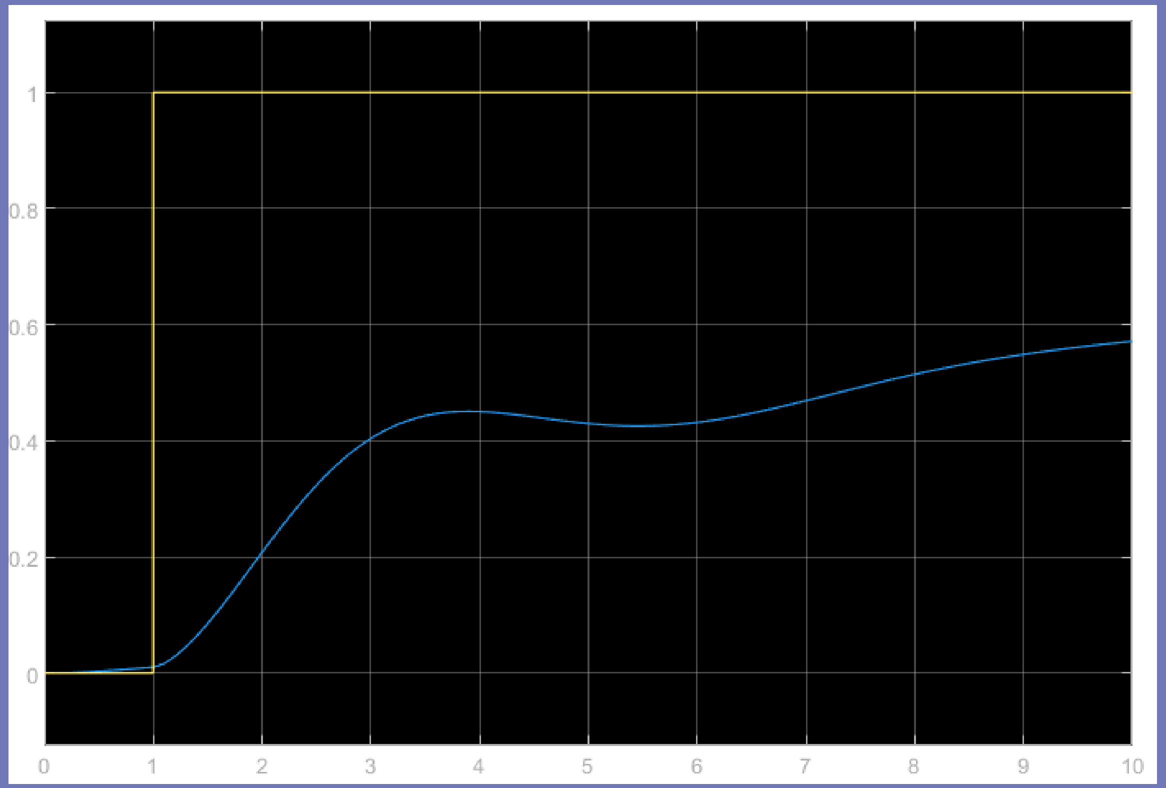
REINFORCEMENT LEARNING OVERVIEW

RL CONCEPTS

- Agent:
  - The decision-maker that learns to optimize control by interacting with the environment.
- Environment:
  - The system or process the agent interacts with (e.g., mass-spring-damper system).
- Policy:
  - Probability of selecting a possible action given a state and reward.
- State:
  - A representation of the system’s current condition (e.g., position, velocity).
- Action:
  - The control input applied by the agent to influence the environment (e.g., force).
- Reward:
  - Feedback signal guiding the agent’s learning, based on the desirability of its actions.

TD3 ALGORITHM

Twin Delayed Deep Deterministic Policy Gradient



- DDPG shortcomings: i. Overestimation bias
  ii. Numerical instability

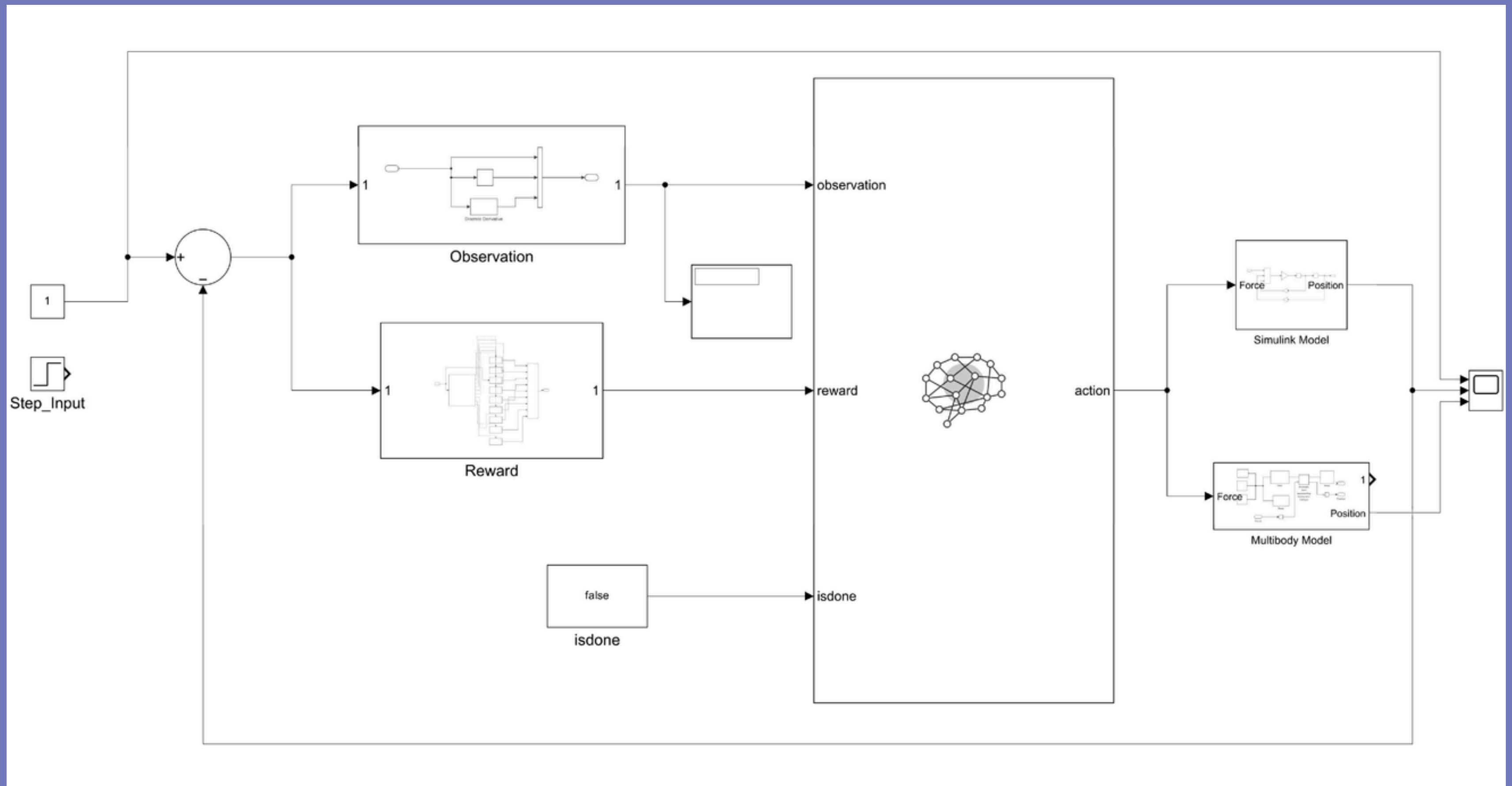
- Replay Buffer

- To reduce Overestimation bias in: DDPG TD3 uses three techniques
  1. Clipped double Q-learning
  2. Target policy smoothing
  3. Delayed policy and target updates

SIMULATION SETUP

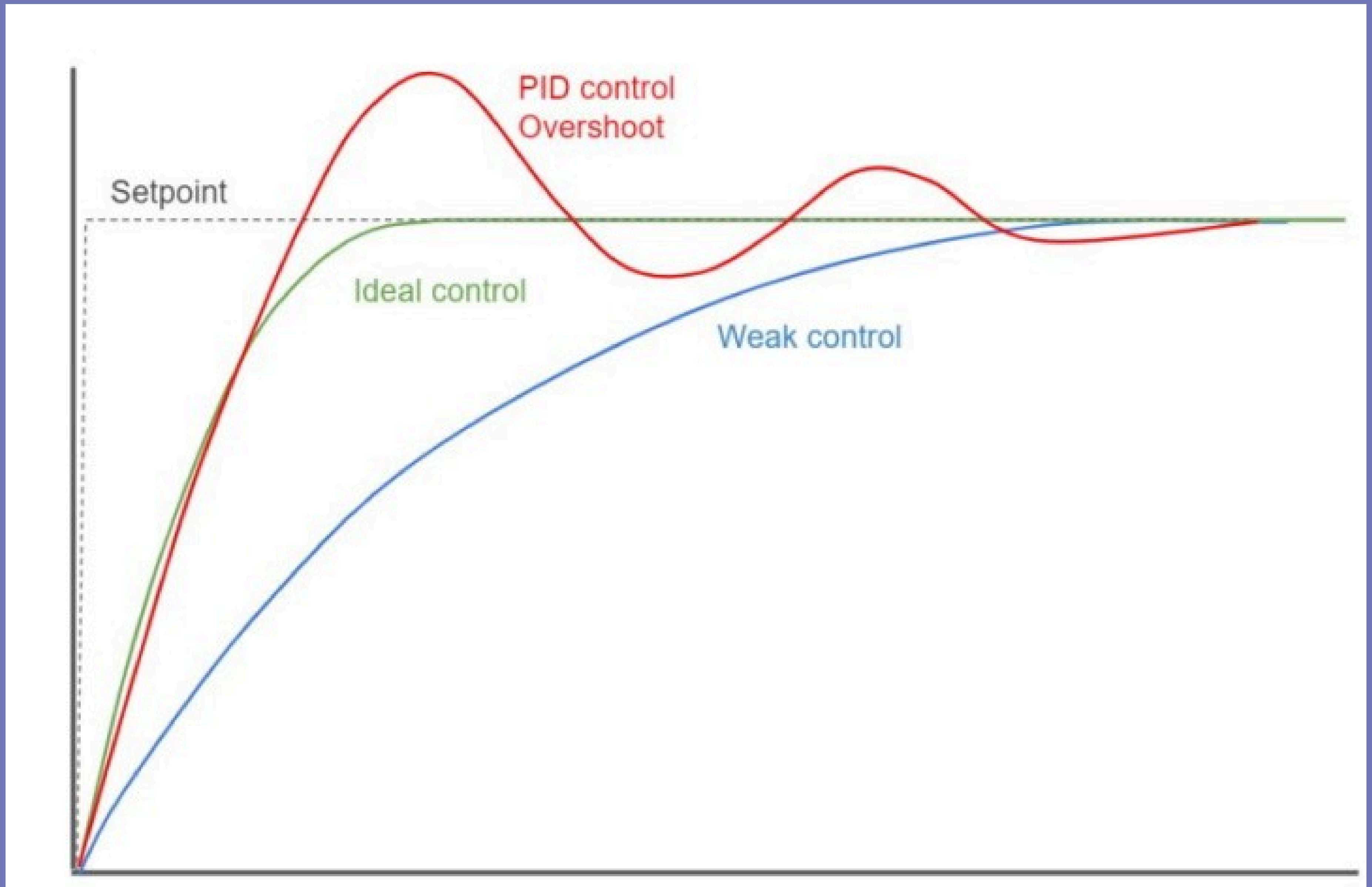
- Simulink model components:
  - Observation block: Outputs P, I, D.
  - Reward block: Implements the designed reward function.
  - RL Agent: Uses TD3 to adjust KP, KI, KD.

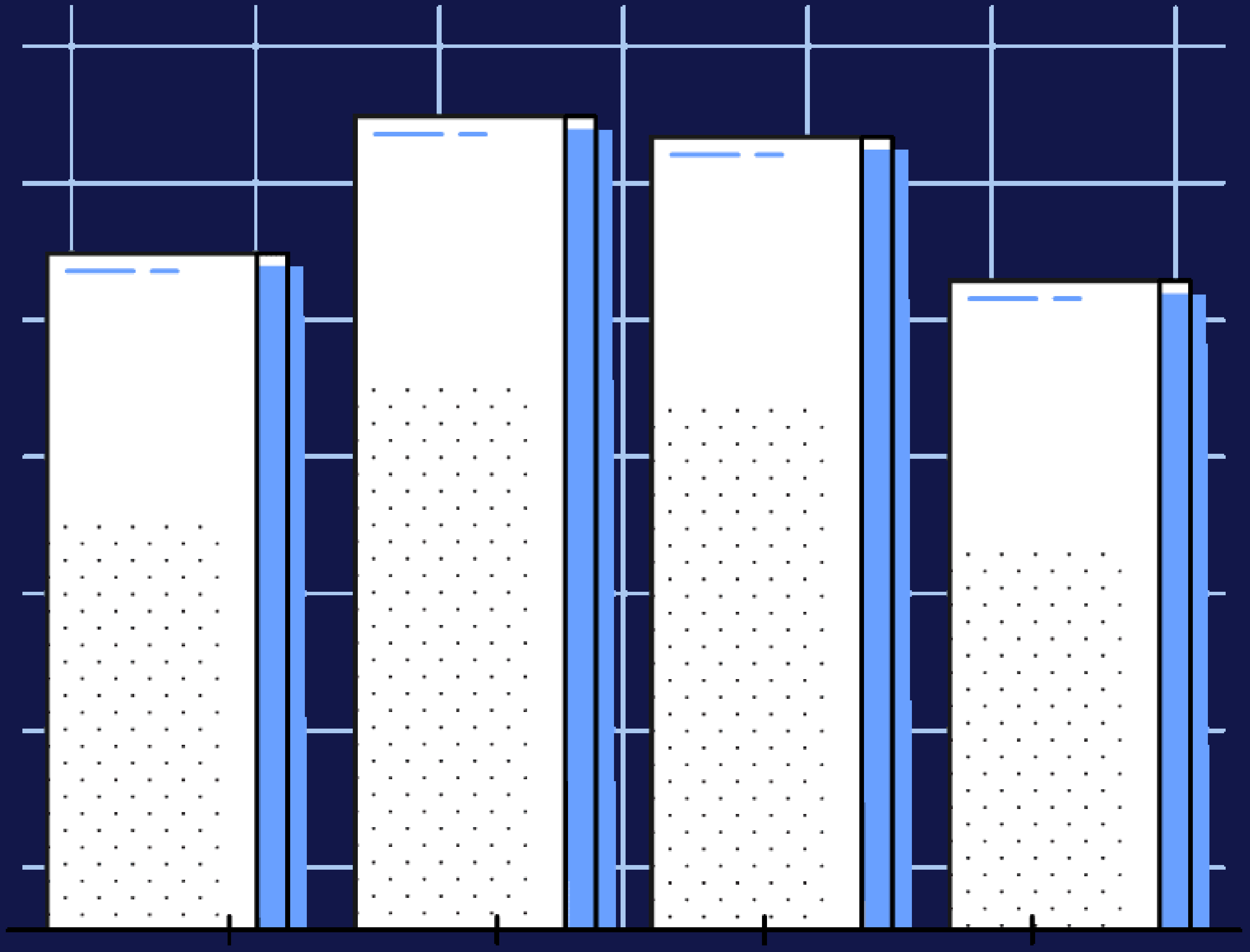
- Training parameters:
  - Episodes: 500
  - Mini-batch size: 128
  - Discount factor: 0.97
  - Learning rates for actor and critic.



REWARD FUNCTION DESIGN

- Reward components:
  - Penalty for overshooting: -2000.
  - Penalty for weak control: -100
  - Reward for staying within target: +3.
  - Penalty for undesirable ranges.
  - Specific rewards for fine-tuned output ranges.
  - Penalty for large deviations.



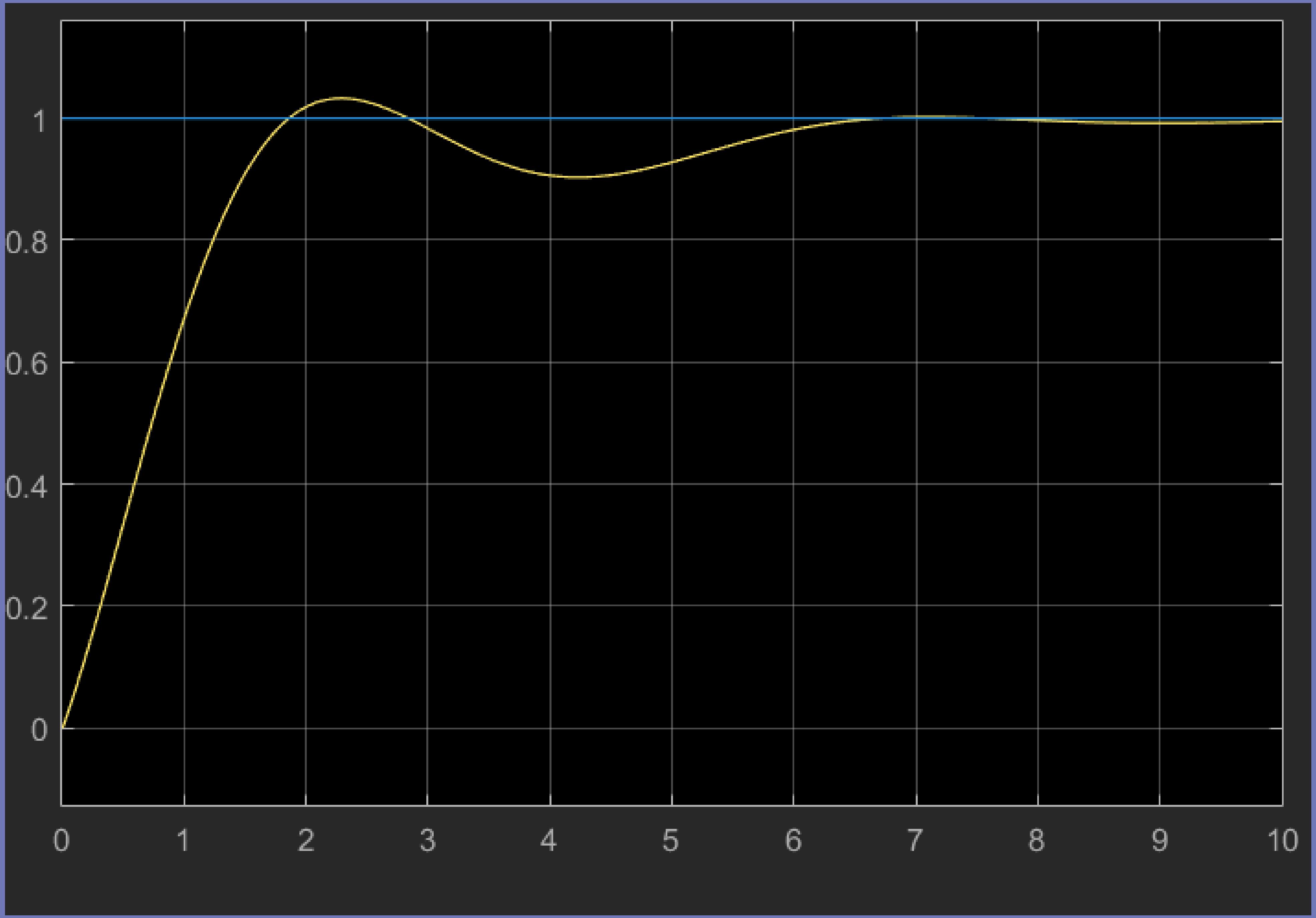


RESULTS

I: Conventional PID Tuning

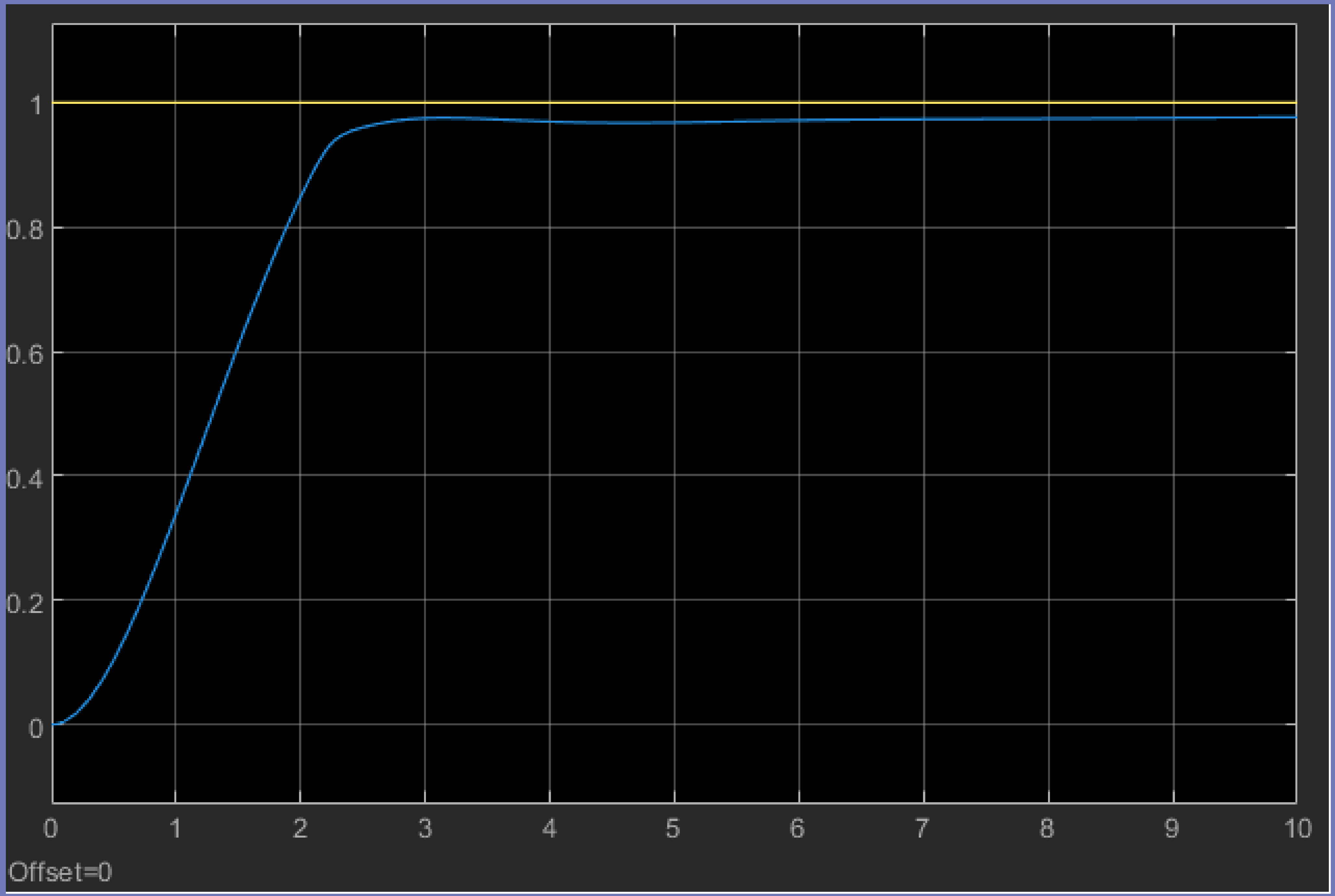
II: RL-Tuned PID

I: CONVENTIONAL PID TUNING



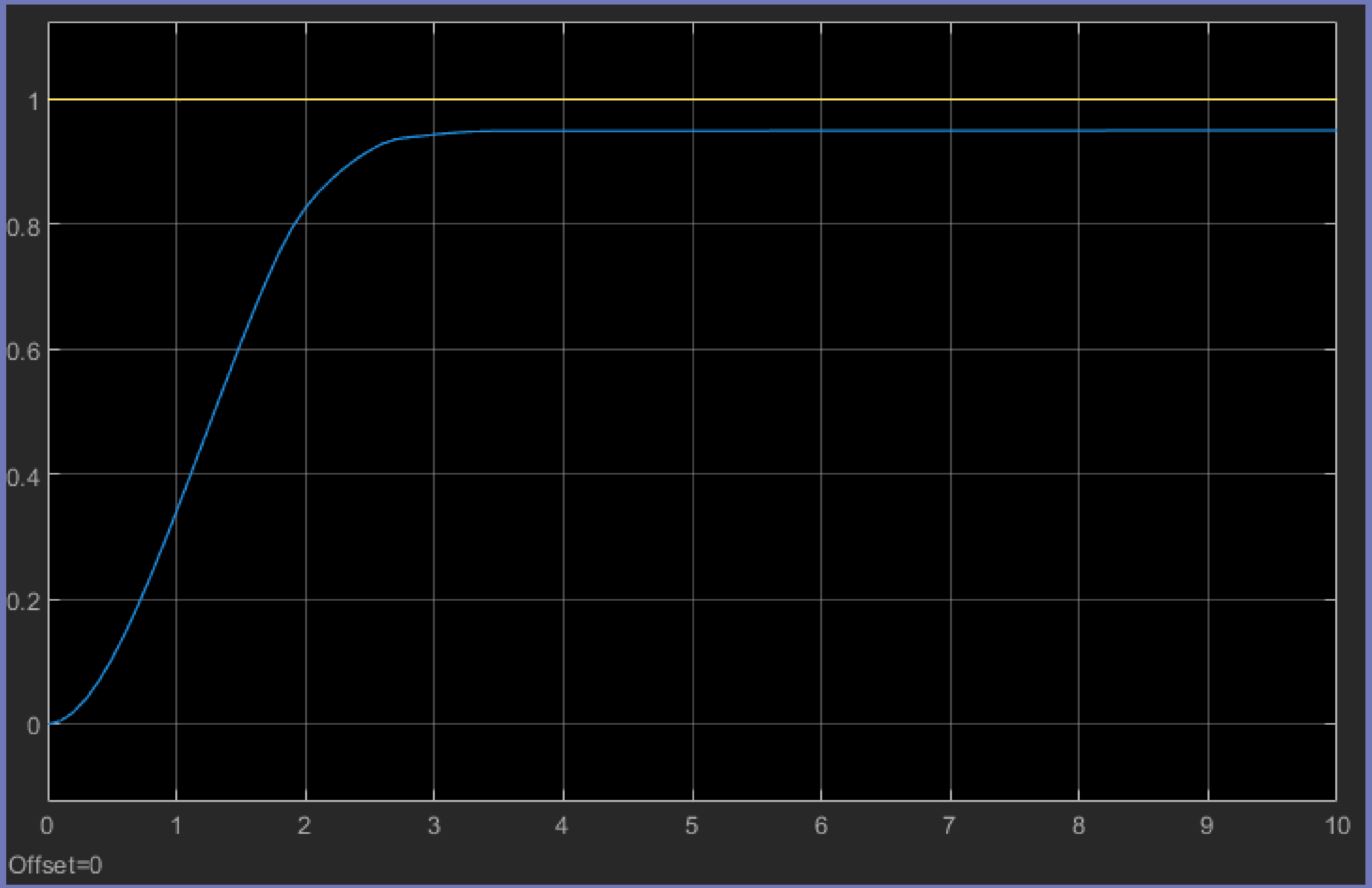
II: RL-TUNED PID

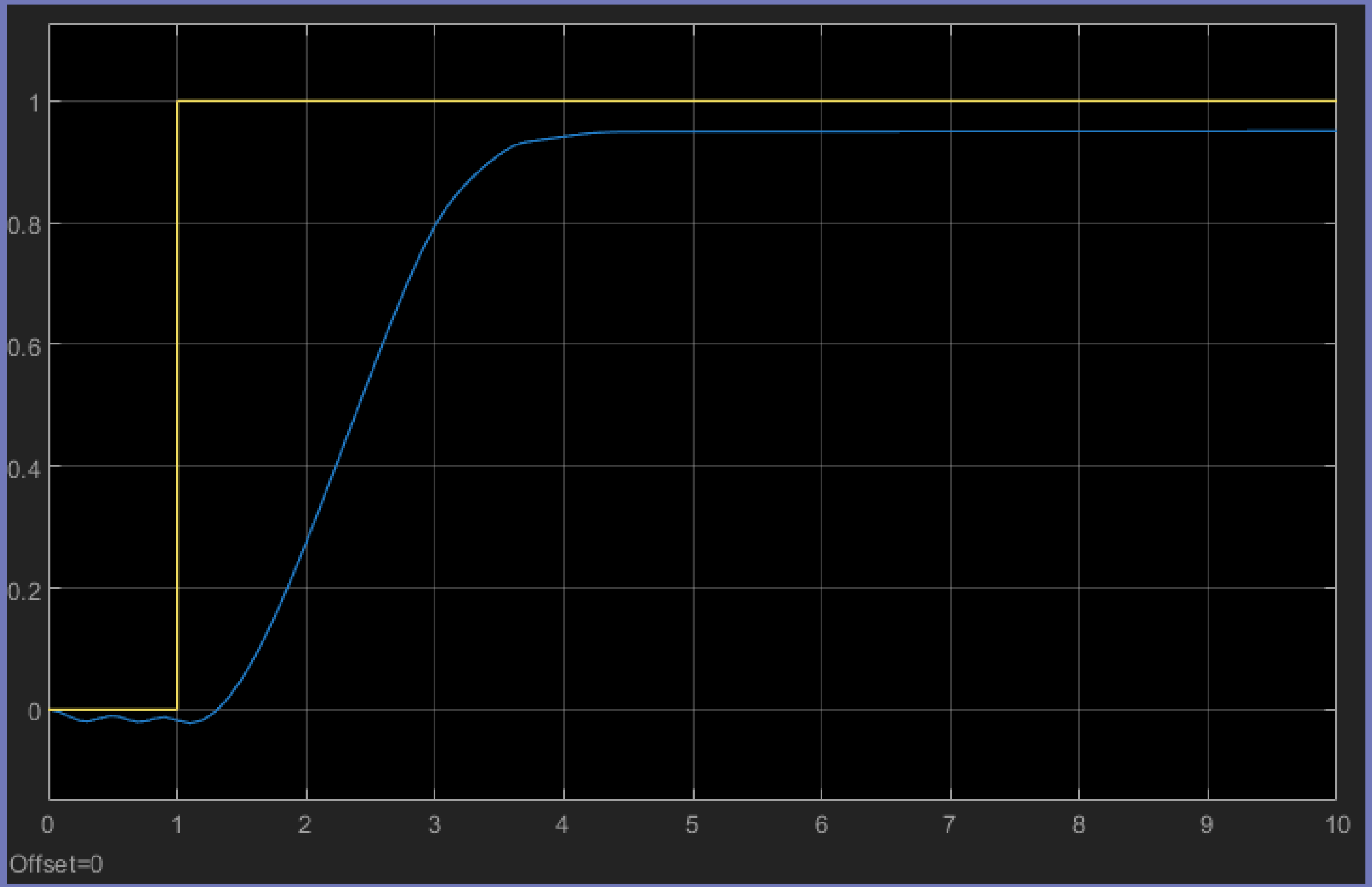
Trail: 18



II: RL-TUNED PID

Trail: 23

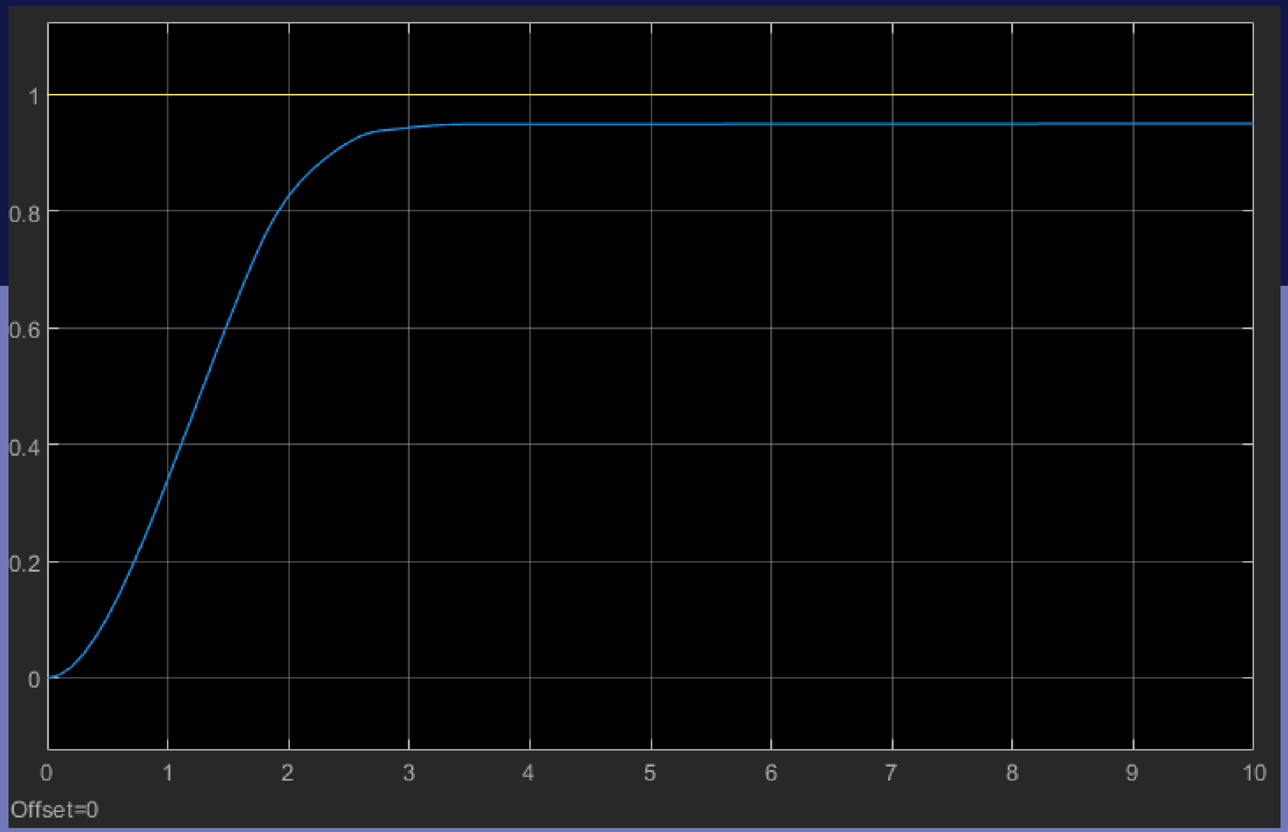


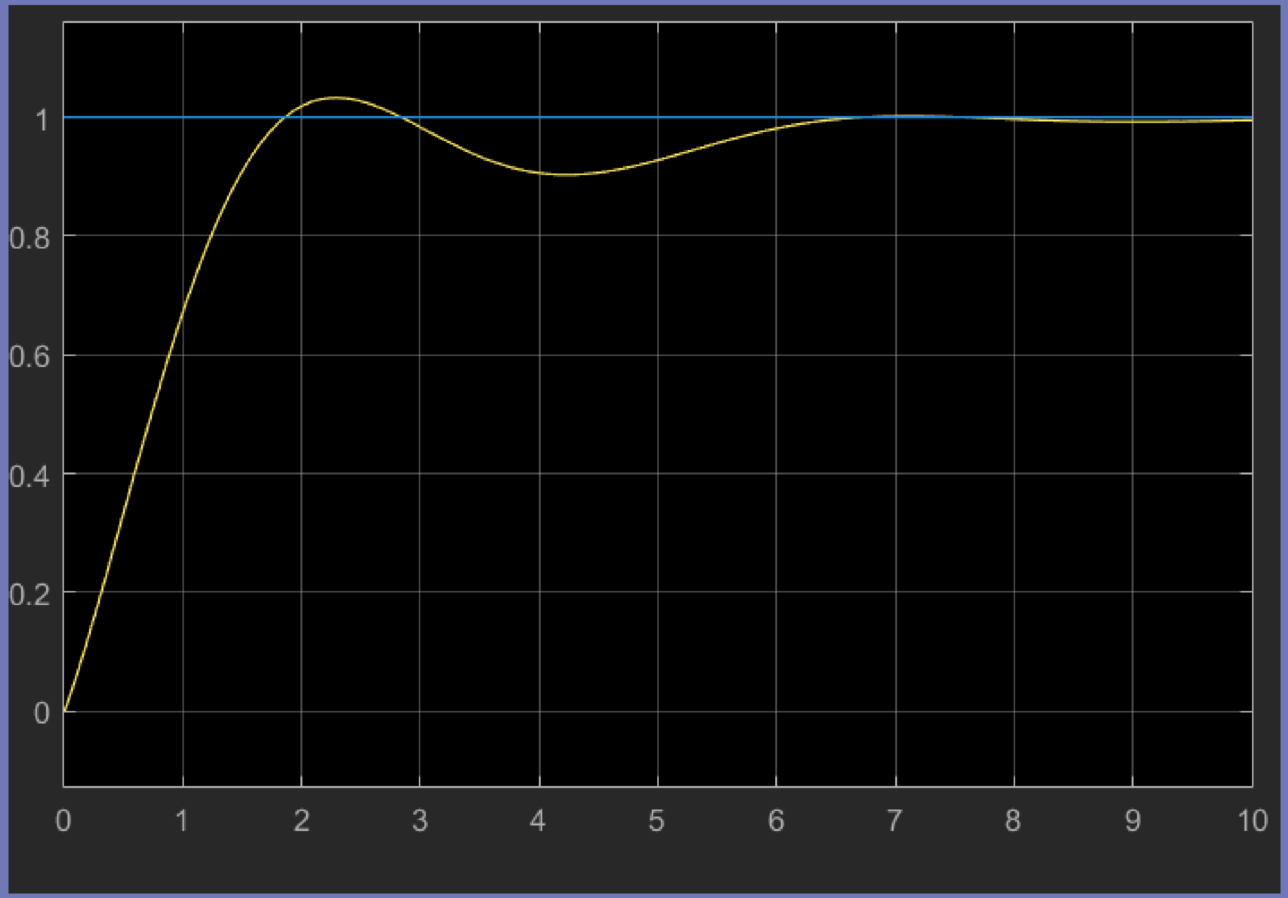


COMPARISON

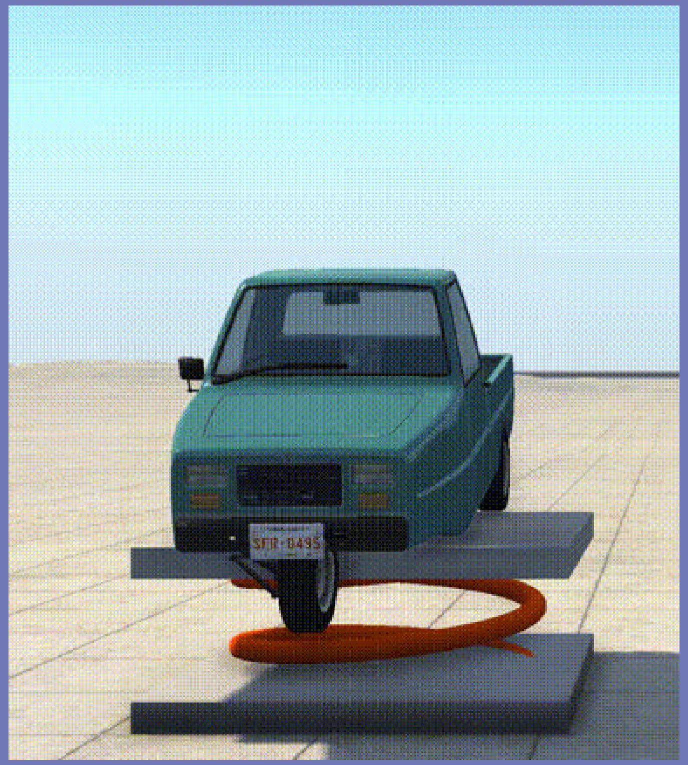
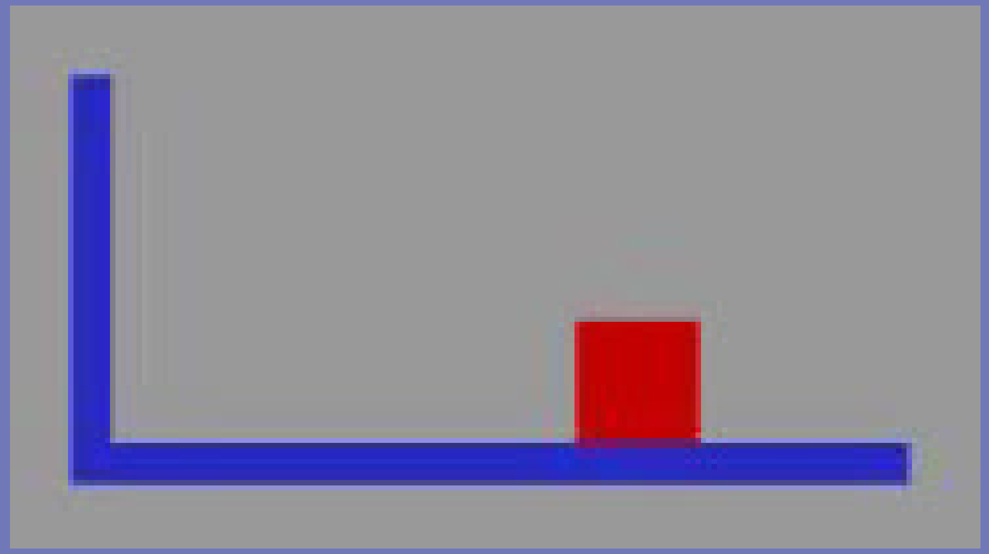
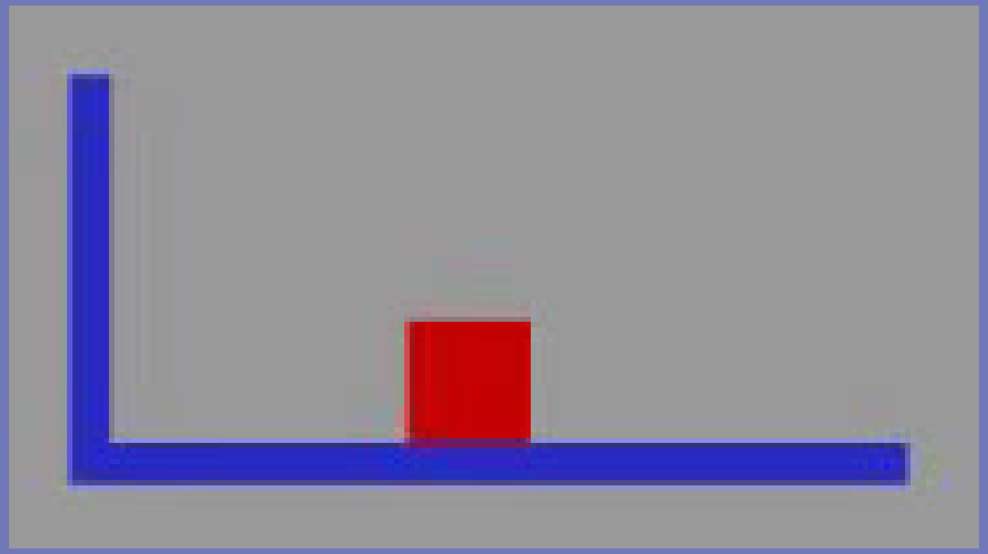
- Comparison Points:
  - Overshoot
  - Settling time
  - Adaptability

- TD3's Advantages in Dynamic Environments:
  - Reduced Overestimation Bias
  - Stable Policy Updates
  - Smooth Control





COMPARISON

	Conventional PID Controller	RL-Tunned PID Controller
Example		
Simscape Model		

CONCLUSION

REFERENCES

- “Tuning of PID Controllers Using Reinforcement Learning for Nonlinear Systems Control” - Gheorghe Bujgoi and Dorin Sendrescu
- Simscape Model From Matlab’s Community File Exchange - Link
- MATLAB and Simulink documentation
- Reinforcement Learning Onramp

Thank You