

st²⁹, rt

RL Optimizer (PPO)

Neural Network Policy π_{θ}

Input: State st²⁹

- Position/velocity errors
- Current hyperparameters

Output: Action at¹⁷

- Q weights (12D)
- R weights (4D)
- Horizon N (1D)

$\theta = [Q, R, N]$ ¹⁷

MPC Controller (CasADi/IPOPT)

Optimization Problem:

$$\min J = \sum (x'Qx + u'Ru)$$

Subject to:

- Dynamics: $\dot{x} = f(x, u)$
- Control limits
- State bounds

Output u⁴

[Thrust, wroll, wpitch, wyaw]

u⁴

UAV Environment (PyBullet)

Physics Simulation (PyBullet)

Dynamics: $\dot{x} = f(x, u, \theta_{\text{drone}})$

State x¹²:

- Position (3D), Velocity (3D)
- Euler angles (3D), Angular rates (3D)

Reward: $r = -||\text{epos}||^2 - ||\text{evel}||^2 - \lambda ||u||^2$