Progress Update: Impact of State Choice on RL Quadrotor Hover  
Project: Deep RL for Quadrotor Hover  
Date: 14 Aug 2025

### 1) Aim

Demonstrate reliable hover with RL and compare how two state designs—fast vs full—affect learning for DHP and SAC.

### 2) State Designs

* Fast states: [z, vz, roll, pitch, yaw, wx, wy, wz]  
  (altitude, vertical speed, attitude, angular rates)
* Full states: [x, y, z, vx, vy, vz, roll, pitch, yaw, wx, wy, wz]  
  (fast states + horizontal position/velocity)

### 3) Experiments (summary)

3.1 DHP (Dual Heuristic Programming)

* Fast states: Succeeded. Stable hover in plots and videos; small drift only.
* Full states: Failed. No convergence; unstable/erratic in videos.  
  Reason: adding slow/integrator-driven states (x, y, vx, vy) reduces the immediate control influence DHP relies on; delayed feedback undermines its gradient-based updates.

3.2 SAC (Soft Actor-Critic)

* Fast states: Failed. Couldn’t infer enough dynamics; no stable policy.
* Full states: Succeeded. Clear convergence and robust hover in videos.  
  Reason: richer observability lets SAC learn the state–action mapping, including the slower components.

### 4) Takeaways

* DHP is state-sensitive: it needs states that respond directly and quickly to control. Adding slow/integrated components hurts learning.
* SAC prefers full observability: the extra signals enable robust policies.
* Evidence matches plots/videos:
  + DHP + fast: converges, stable.
  + DHP + full: diverges/unstable.
  + SAC + fast: no learning.
  + SAC + full: converges, stable.

### 5) Documentation check

Review of repo notes, daily logs, and summaries confirms:

* Protocols, results, and lessons are well recorded.
* Plots/videos are generated per trial and archived.
* The rationale—fast for DHP, full for SAC—is consistently argued and observed across runs.

### 6) Next steps

* DHP: keep using fast, directly actuated states for hover; consider alternative formulations or reward shaping for harder tasks.
* SAC: standardize on the full state vector; extend to more complex tasks once hover is solid.
* General: continue rigorous logging (plots, videos); finalize hover before advancing to trajectories/navigation.

### 7) Conclusion

State representation is pivotal. DHP succeeds with fast, directly controllable states; SAC excels with the full state set. Quantitative plots and qualitative videos align with theory and prior notes.