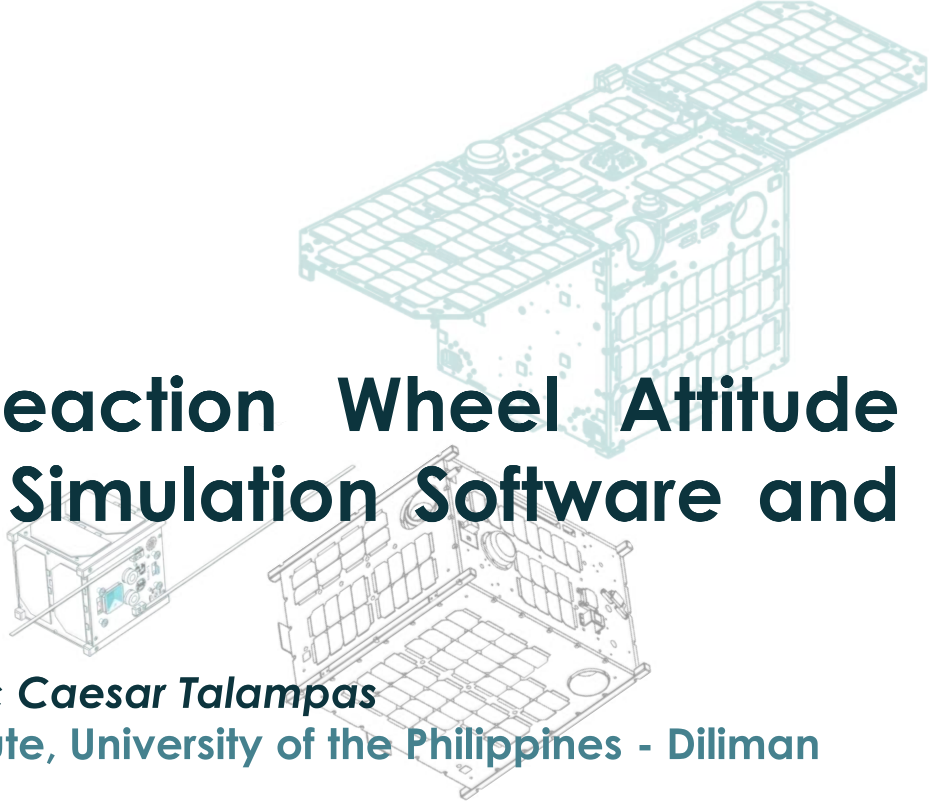
SSC21-WKIII-04

MATA-RL: Continuous Reaction Wheel Attitude Control using the MATA Simulation Software and Reinforcement Learning



Vanessa Tan, John Leur Labrador, and Marc Caesar Talampas
Electrical and Electronics Engineering Institute, University of the Philippines - Diliman

As earth observation satellites, Diwata microsatellites need to have a high degree of target pointing accuracy

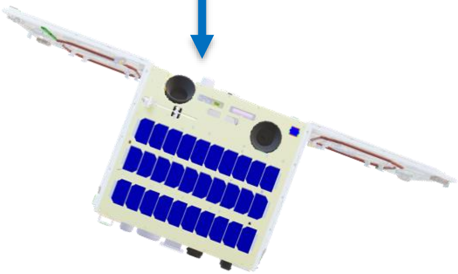
Current Status of Attitude Controllers

Current methods for attitude control have proven to be effective in stable environments. However, they are **prone to changes in control and mass parameters.**

[1] Wang Y., Ma Z., Yang Y., Wang Z., and Tang L. A new spacecraft attitude stabilization mechanism using deep reinforcement learning method. In 8TH European Conference for Aeronautics and Space Sciences (EUCASS), 2019.
[2] Su R., Wu F., and Zhao J. Deep reinforcement learning method based on ddpg with simulated annealing for satellite attitude control system. In 2019 Chinese Automation Congress (CAC), pages 390-395, 2019.

MATA-RL: Continuous Reaction Wheel Attitude Control using the MATA Simulation Software and Reinforcement Learning

State: Current Error Quaternion and Rotation Rates
Reward: Target Error Angle and Rotation Rate



Satellite Agent

Dynamics/Kinematics from MATA Simulator

Action Space: Reaction Wheel Speed

Main Contributions

Two deep reinforcement learning algorithms for continuous attitude control using the reaction wheel speed as action space

Development and utilization of MATA simulator for reinforcement learning environment

A comparison and analysis of attitude control performance between the RL algorithms and Diwata's PID control in different scenarios

Outline

- Spacecraft Kinematics and Dynamics
- Reinforcement Learning Algorithms
- Results and Case Studies
- Conclusion and Future Work

STAMINA4SPACE

Satellite Kinematics and Dynamics

$$\dot{\vec{\omega}} = \boldsymbol{I}^{-1}\left(\vec{T_c} + \vec{T_d} - (\vec{\omega} \times \boldsymbol{I}\vec{\omega}) - \left(\vec{\omega} \times \vec{h}_{rw}\right)\right)$$

- The dynamics equation for the satellite determines the angular acceleration from internal (control) and external torques (disturbance)

$$q(t) = \exp\left(\frac{1}{2}\Omega t\right)q(0)$$

- The kinematics equation for the satellite attitude uses quaternion expressions

Satellite Control

- The speed and mechanical alignment of each reaction wheel can be translated to the spacecraft's control torque

Satellite Control

$$\vec{T_c} = \vec{K_p}(er) + \vec{K_d}\frac{d}{dt}(er) + \vec{K_i}\int(er)dt$$

- PID control depends on the difference between the target and current attitude in addition to the satellite's rotation rate
- Gain values need to be "tuned" for best results

Reinforcement Learning

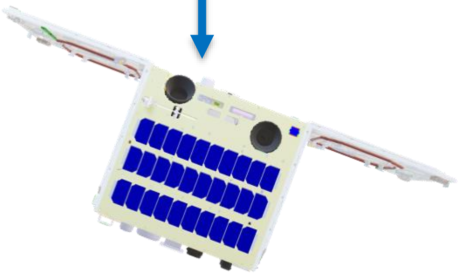
Source: https://images.app.goo.gl/Kj44uvBzWzMw1QzE9

- Agent – learner and decision maker
- Environment – where agent learns and decides what actions to perform
- Action – set of actions which agent can perform
- State – state of agent in the environment
- Reward – for each action selected by agent the environment provides a reward (usually a scalar value)

MATA-RL: Continuous Reaction Wheel Attitude Control using the MATA Simulation Software and Reinforcement Learning

$$s_t = \{\vec{q}_{error}, w\vec{b}i, w\vec{b}r\}$$

State: Current Error Quaternion and Rotation Rates
Reward: Target Error Angle and Rotation Rate



Satellite Agent

Dynamics/Kinematics from MATA Simulator

Action Space: Reaction Wheel Speed

$$a_t = \{RW1, RW2, RW3, RW4\}$$

STAMINA4SPACE

Reward Function

$$Q_{reward} = \exp[-0.1(\|\vec{q}_{target} - \vec{q}\|)]$$

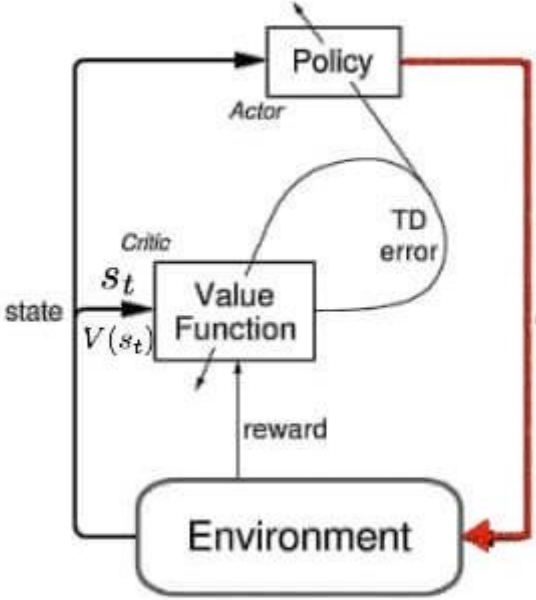
$$W_{reward} = \exp[-0.1(\|\vec{w}_{target} - \vec{w}\|)]$$

$$Reward_{total} = Q_{reward} * W_{reward}$$

Note: Additional +10 if $q_{error} < 0.1°$

Reinforcement Learning Algorithms

Actor-Critic



- Actor: decides which action to take
- Critic: tells the actor how good its action was and how it should adjust

(Figure from Sutton & Barto, 1998)

Reinforcement Learning Algorithms

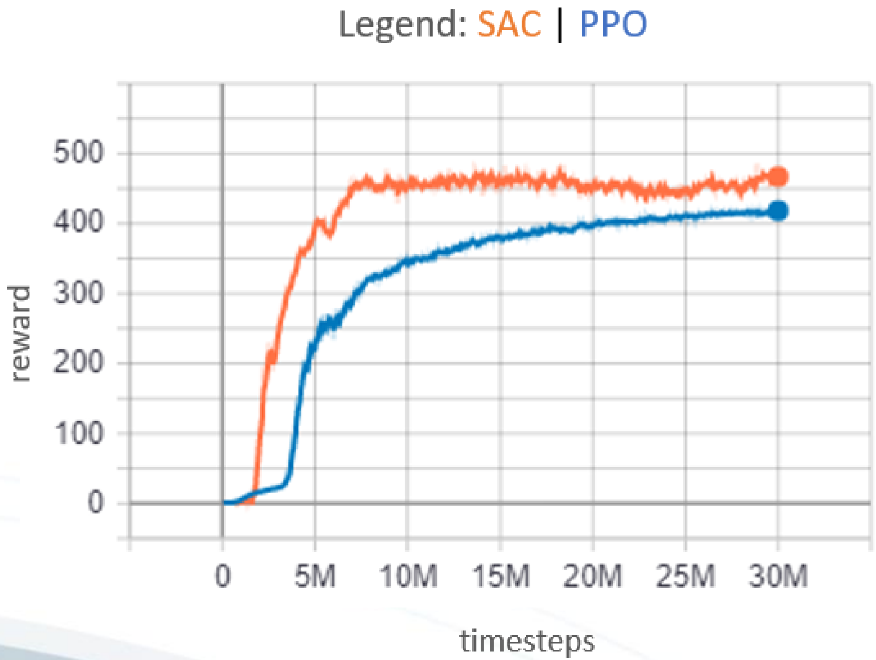
Proximal Policy Optimization (PPO)

- On-Policy
- Great performance for UAV attitude control
- Computational simplicity

Soft Actor Critic (SAC)

- Off-Policy
- Sample Efficient
- Can maximize the entropy of the policy

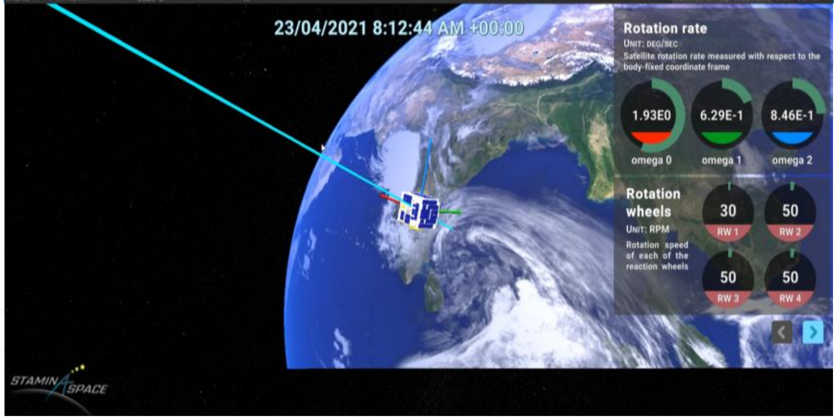
Training Results



- SAC achieved a higher cumulative reward (~450) than the PPO (~410)
- SAC reached convergence around 15M steps while the PPO needed 30M steps to achieve convergence

Training Results

Proximal Policy Optimization (PPO)

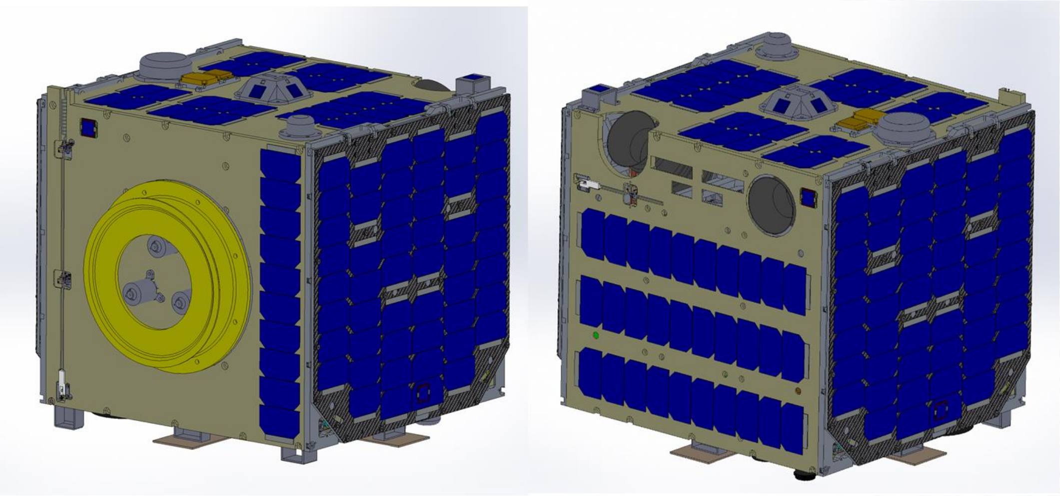


Soft Actor Critic (SAC)

STAMINA SPACE

Case Studies

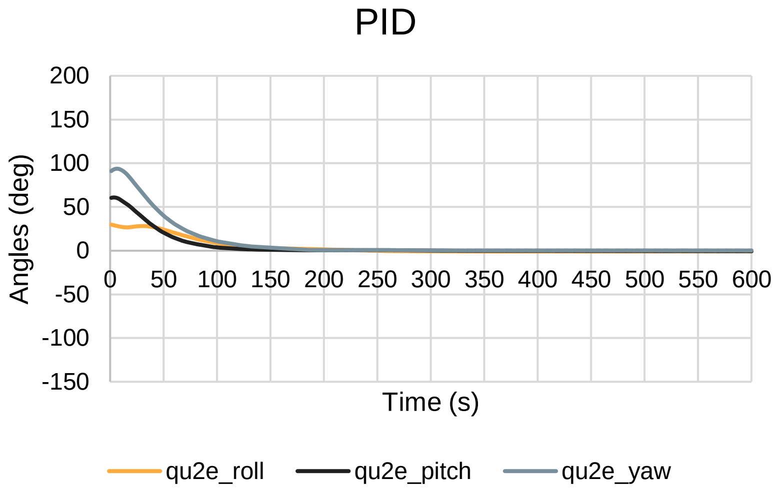
Scenario 1: Diwata 2

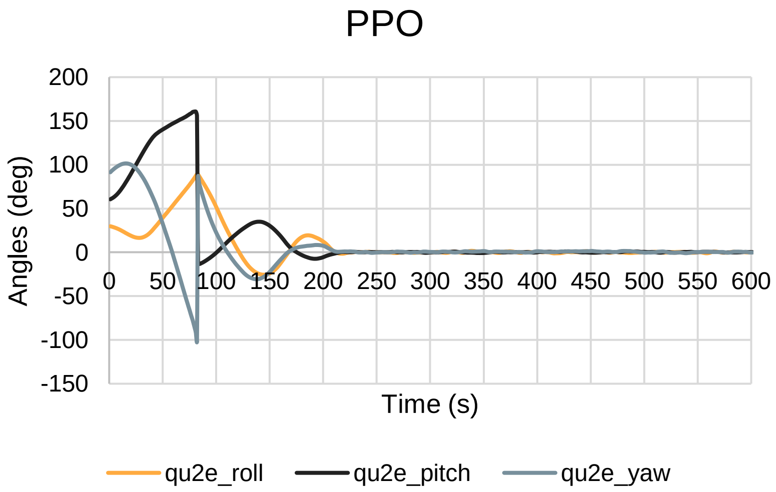


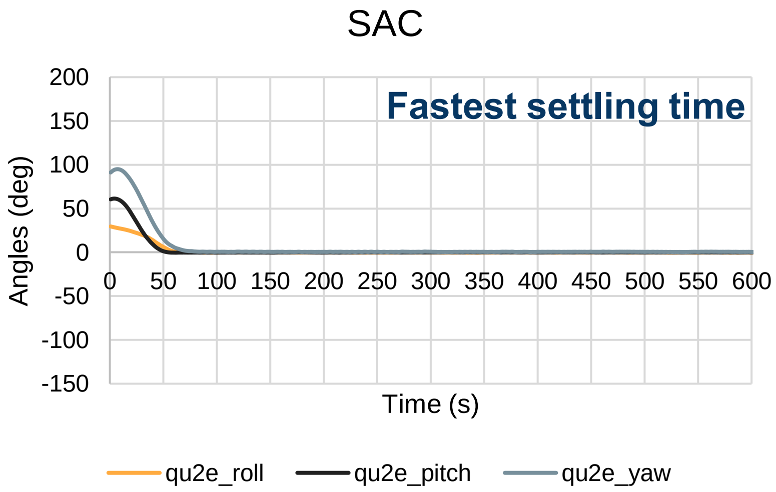
Diwata 2 Stowed Configuration (Baseline)

[1] PHL-Microsat. Diwata-2. https://phl-microsat.upd.edu.ph/diwata2.

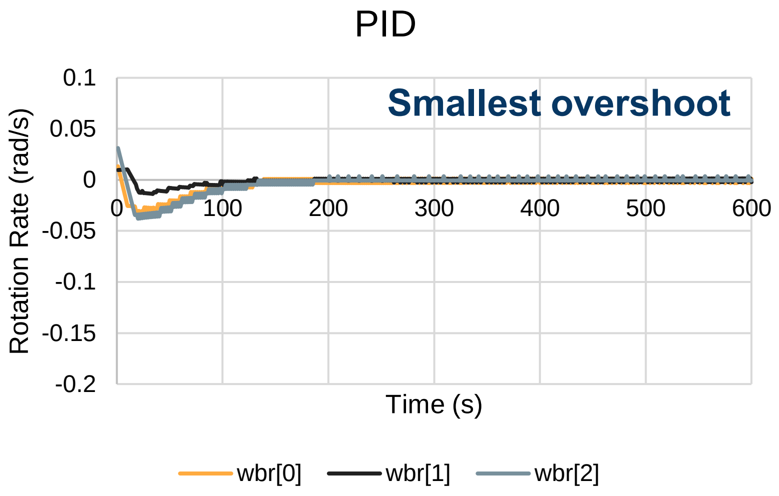
Control Performance for Attitude Angles

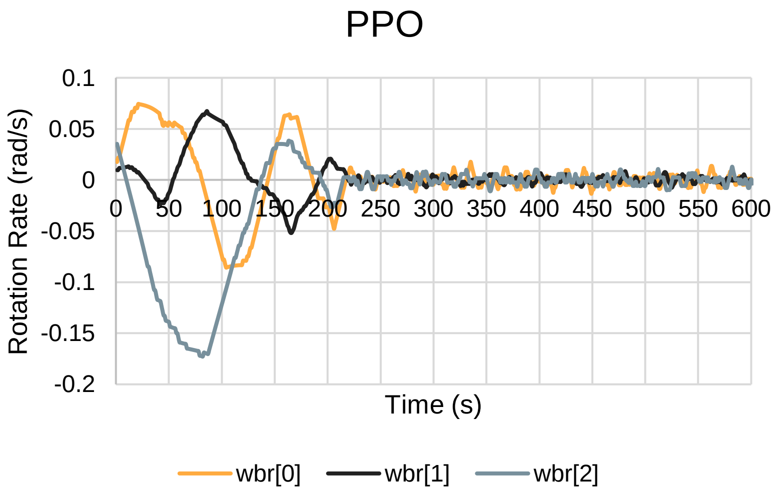


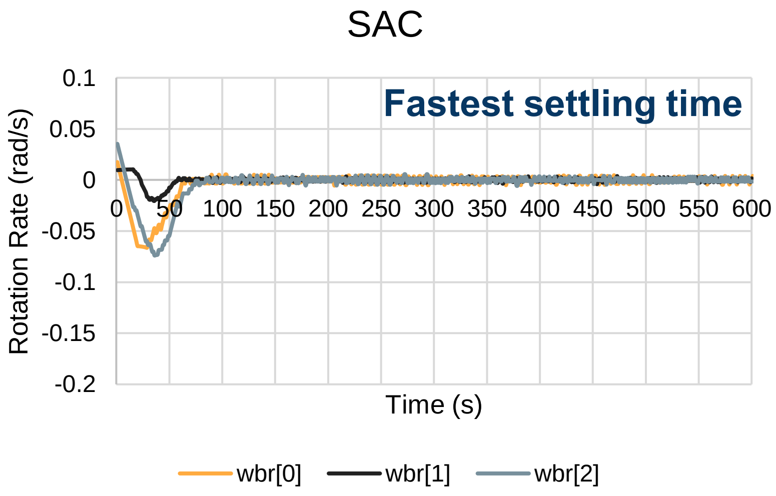




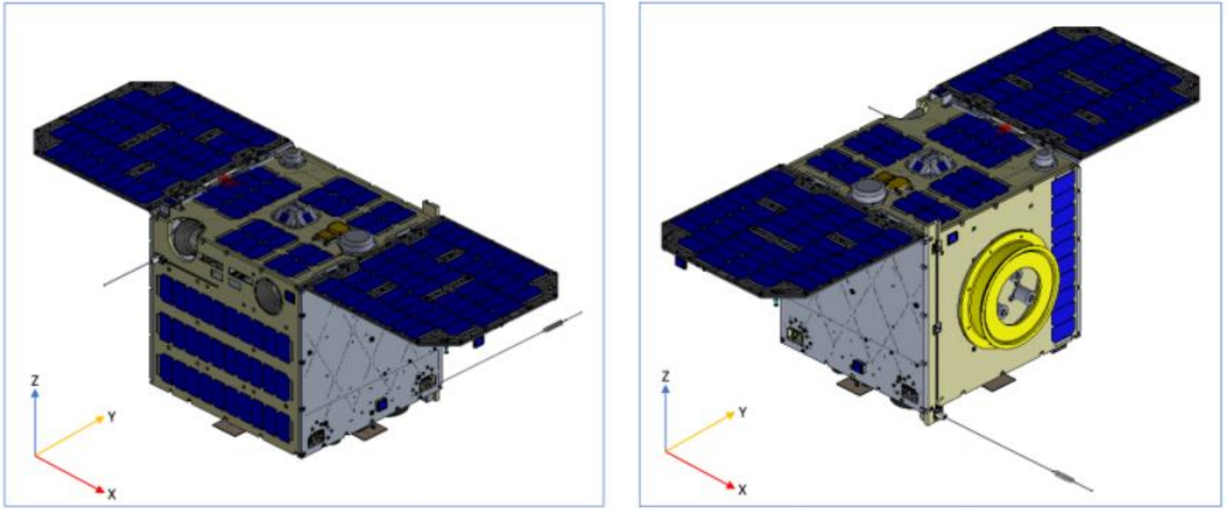
Control Performance for Rotation Rates







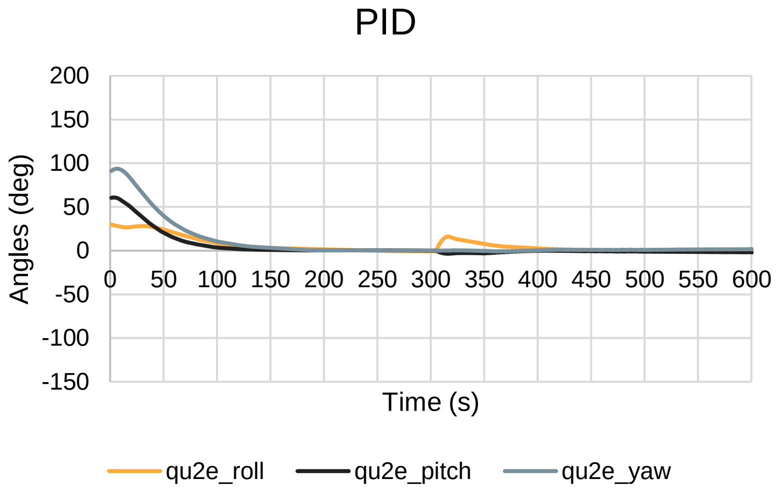
Scenario 2: Diwata 2 Deployed Configuration

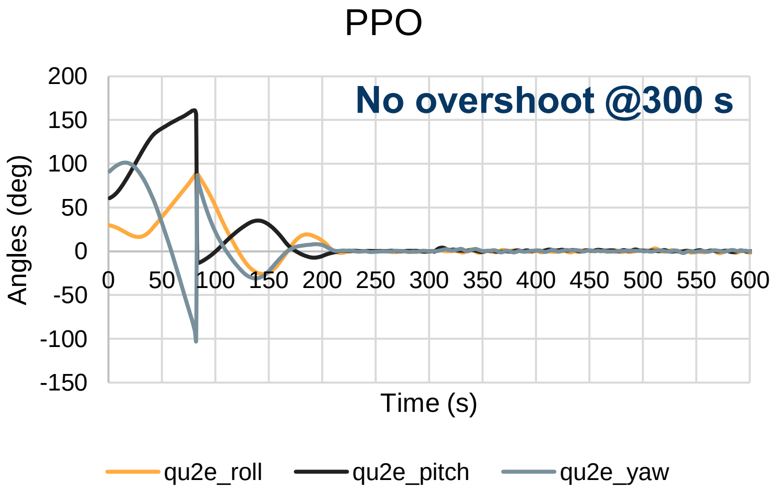


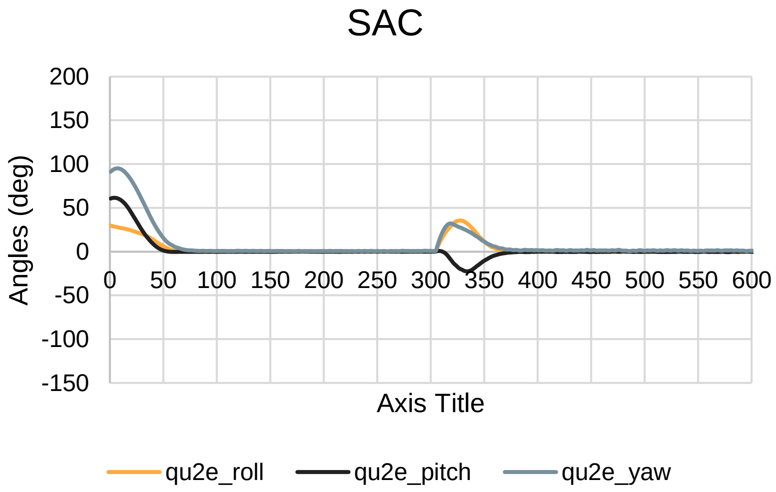
Diwata 2 with Deployed Solar Panels and Antenna (t = 300 s)

[1] PHL-Microsat. Diwata-2. https://phl-microsat.upd.edu.ph/diwata2.

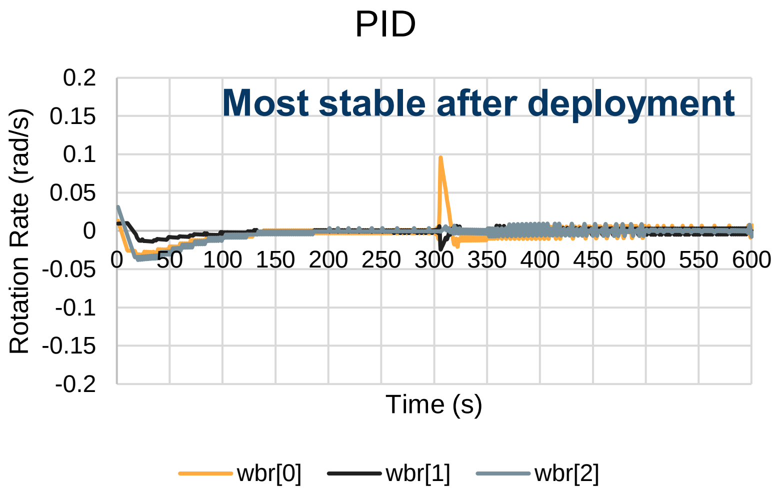
Control Performance for Attitude Angles

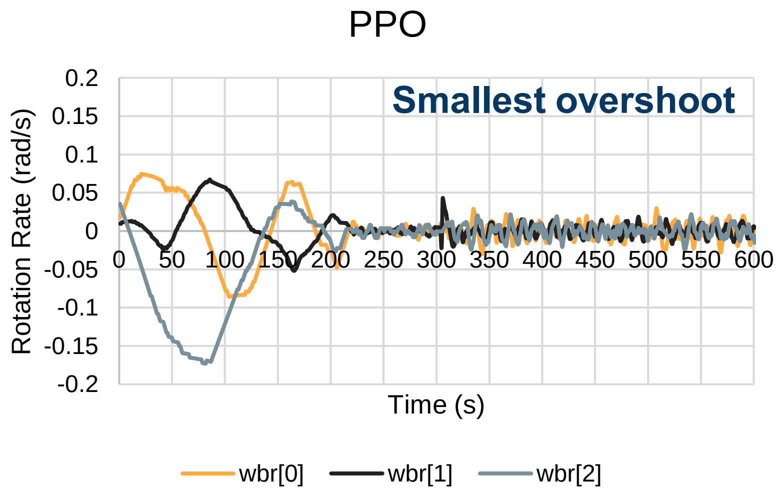


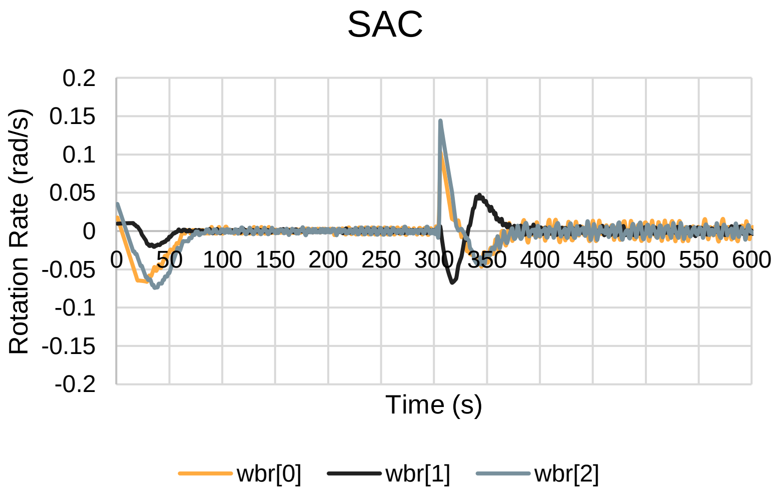




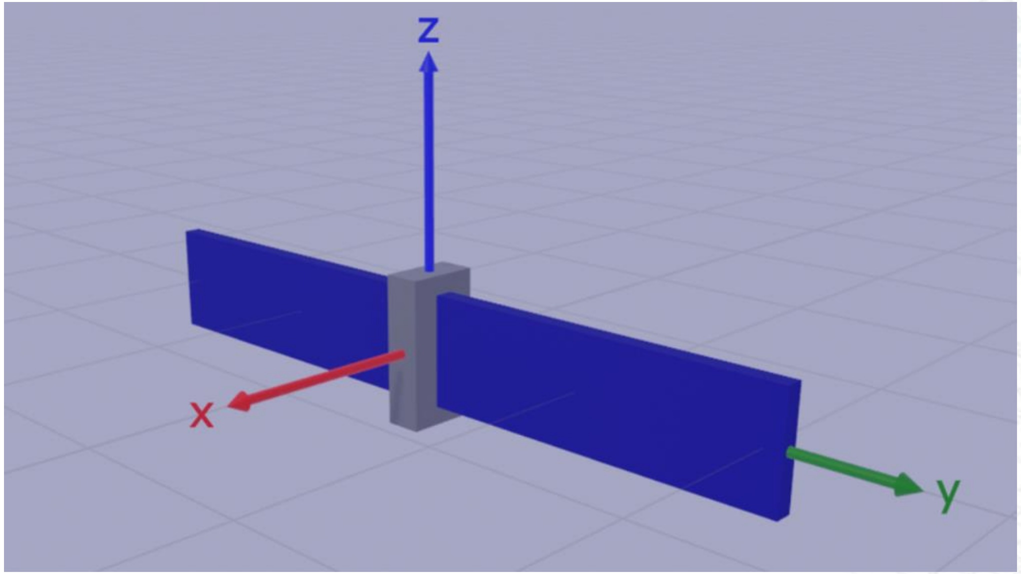
Control Performance for Rotation Rates







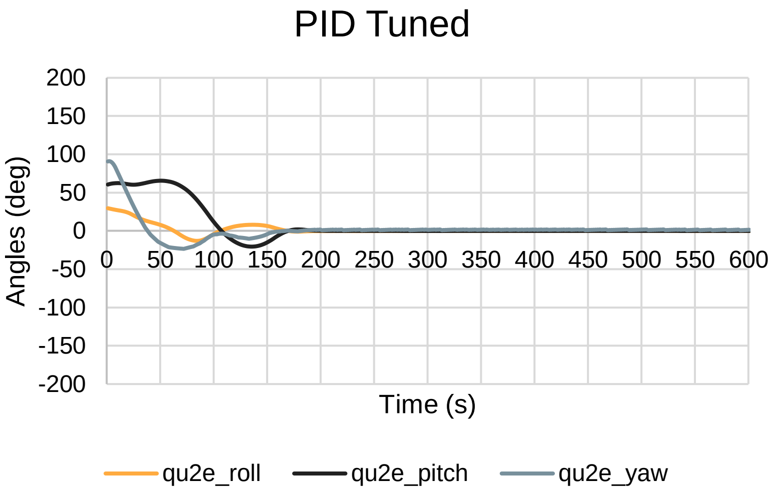
Scenario 3: LM 50

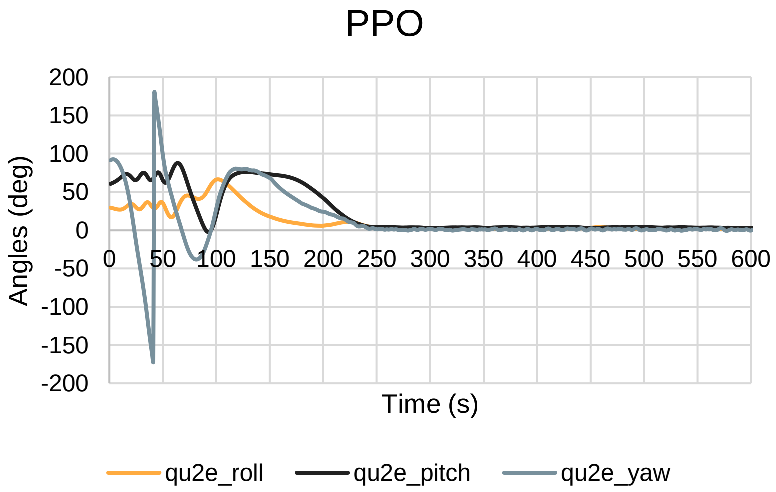


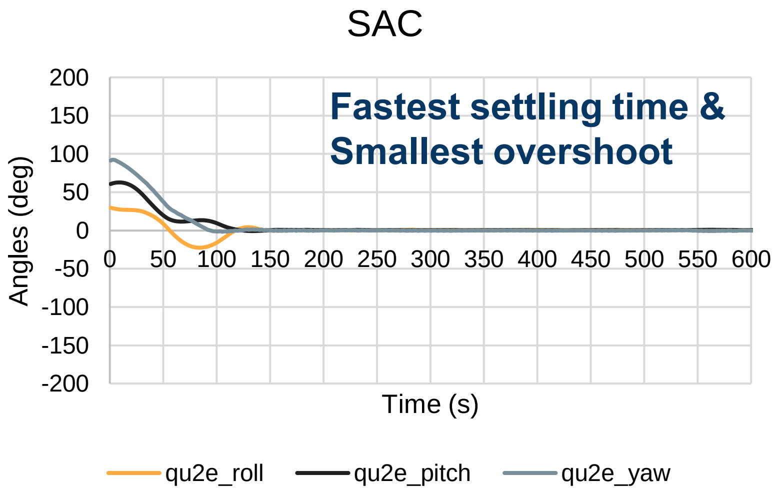
LM 50 (Different Flight Heritage and Mass with Diwata 2)

[1] Elkins J., Sood R., and Rumpf C. Autonomous spacecraft attitude control using deep reinforcement learning. In 71st International Astronautical Congress, October 2020.
[2] Elkins J., Sood R., and Rumpf C. Adaptive continuous control of spacecraft attitude using deep reinforcement learning. In 2020 AAS/AIAA Astrodynamics Specialist Conference, August 2020.

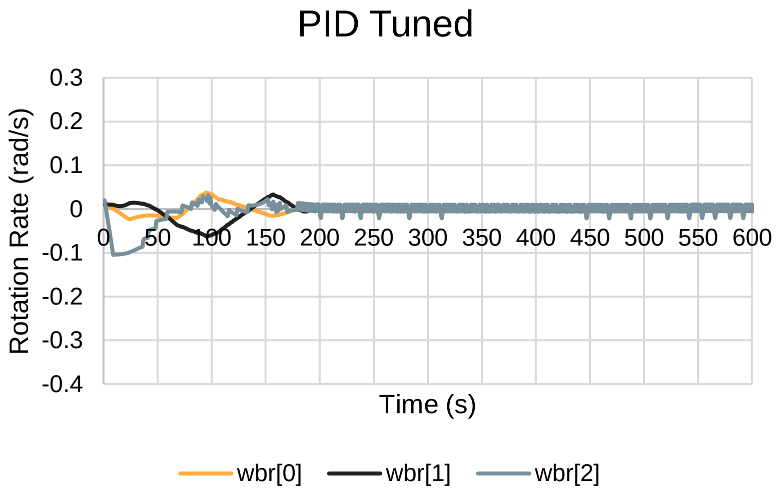
Control Performance for Attitude Angles

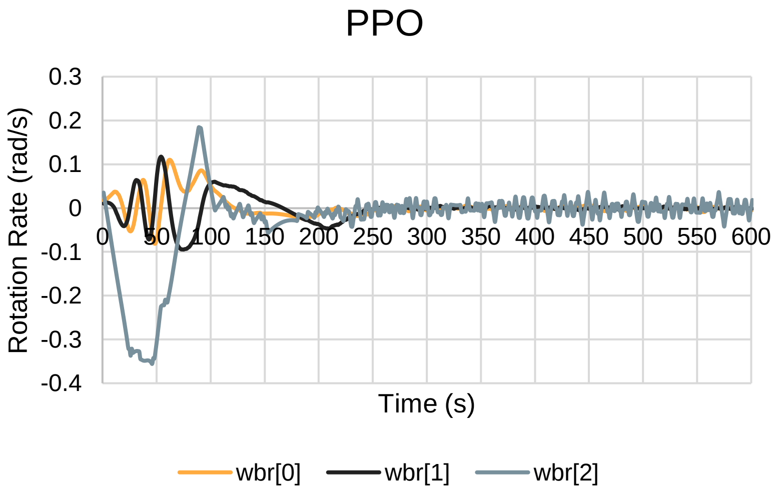


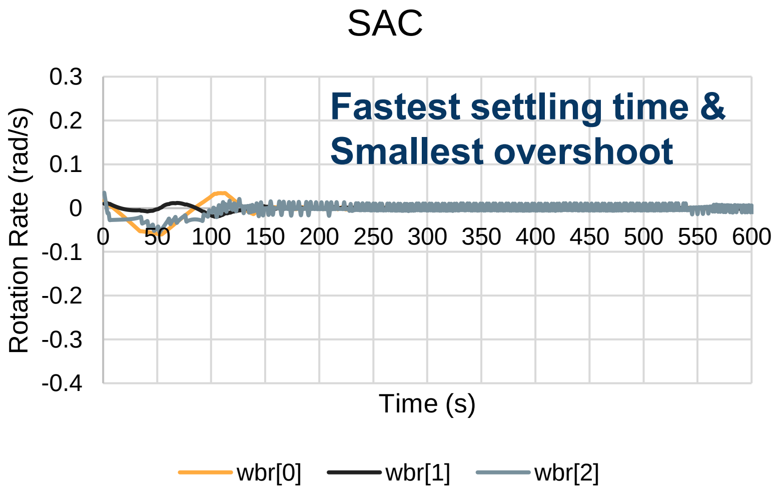




Control Performance for Rotation Rates



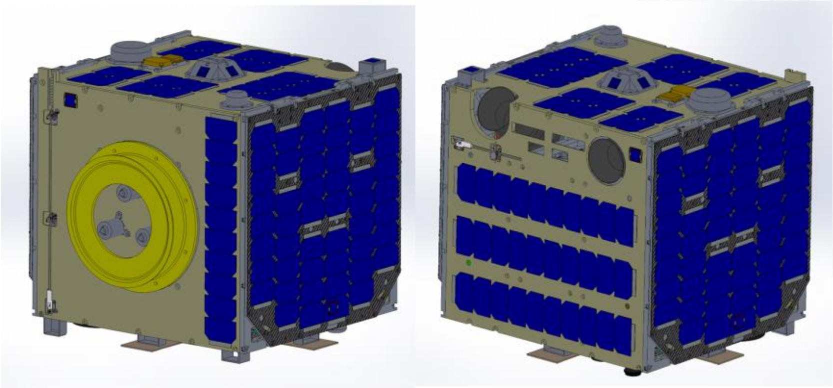




Case Studies Summary

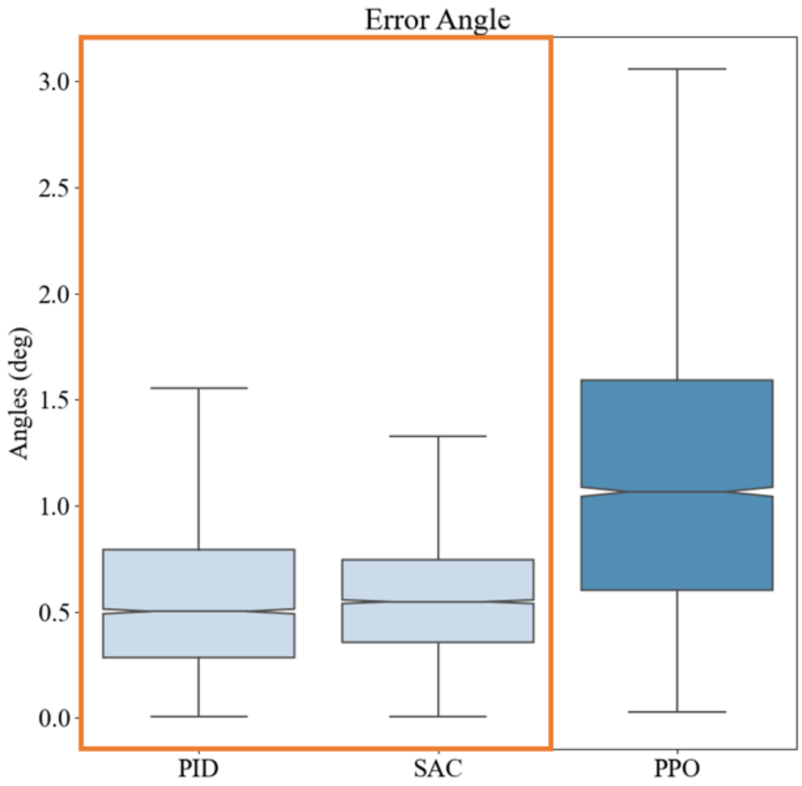
- SAC is the fastest attitude controller when no sudden disturbances occur
- SAC is also comparable with the PID controller in terms of the stability and overshoot metrics
- PPO has the worst performing metrics, however, it is the most resilient to sudden disturbances

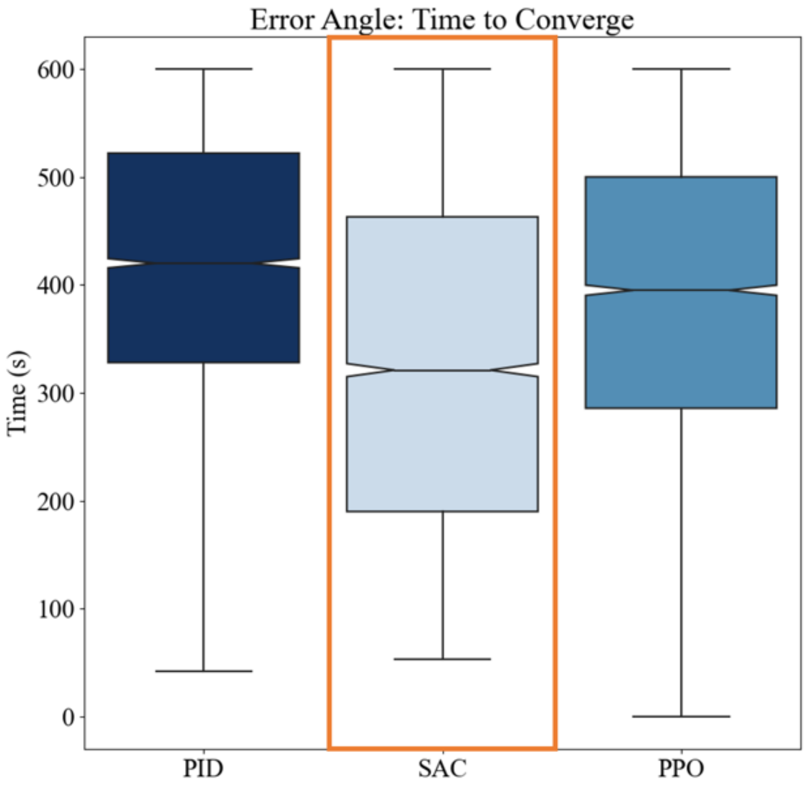
Overall Evaluation



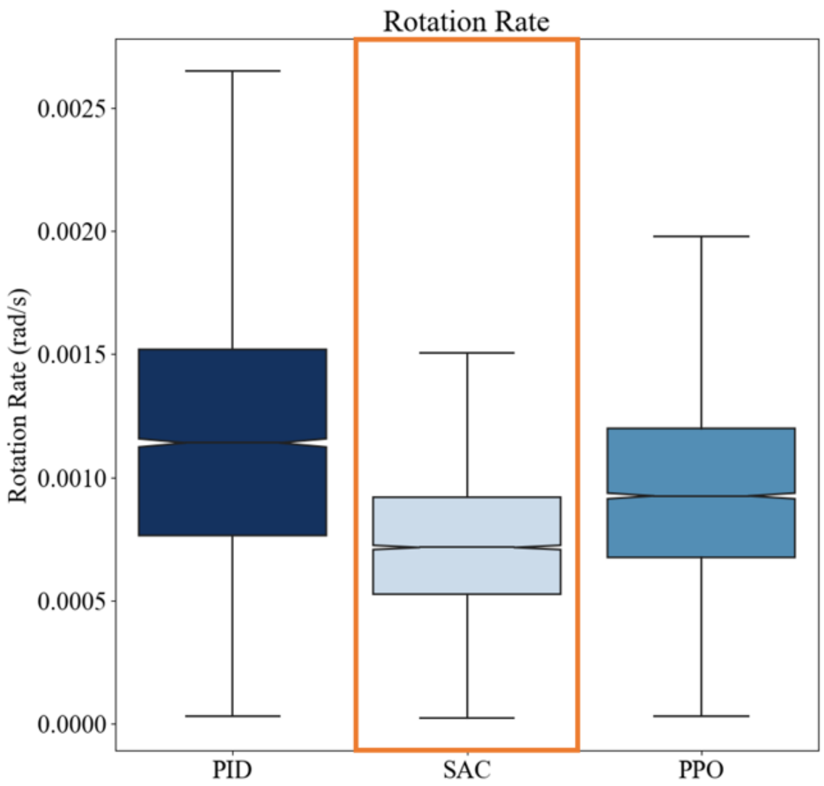
- For the overall evaluation, the Diwata 2 stowed configuration was utilized
- The initial state and target parameters were randomized for each episode
- Evaluation for 5000 episodes

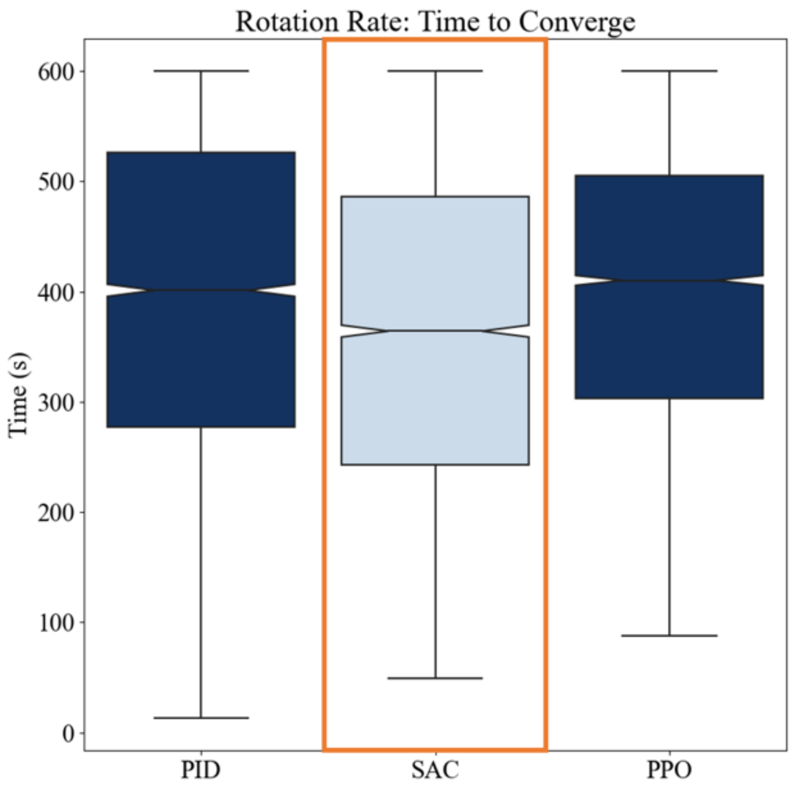
Overall Results



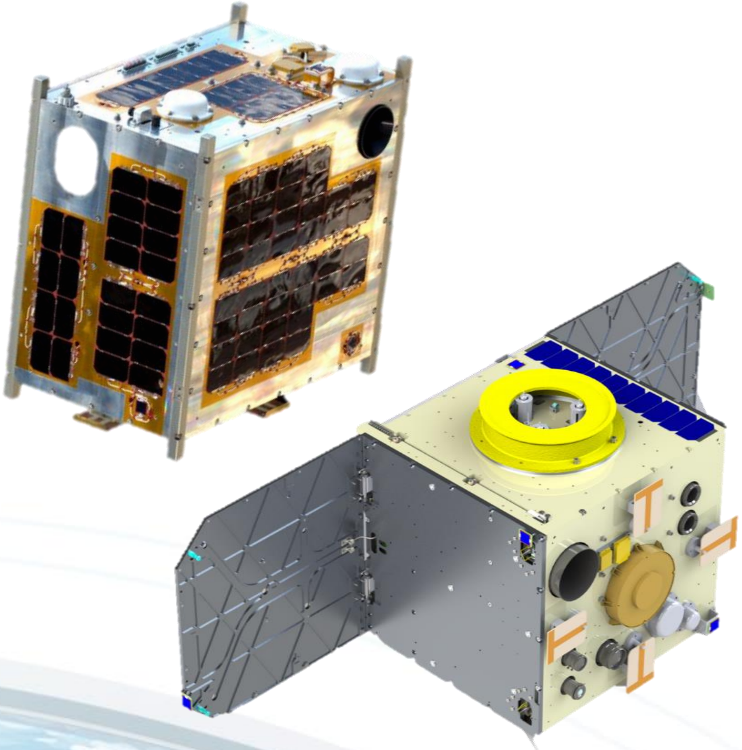


Overall Results



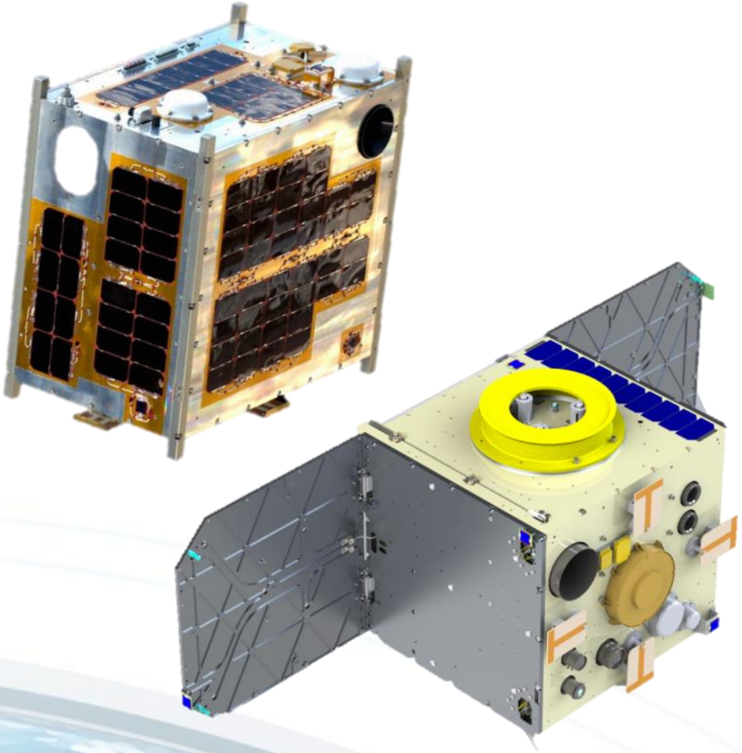


Conclusion



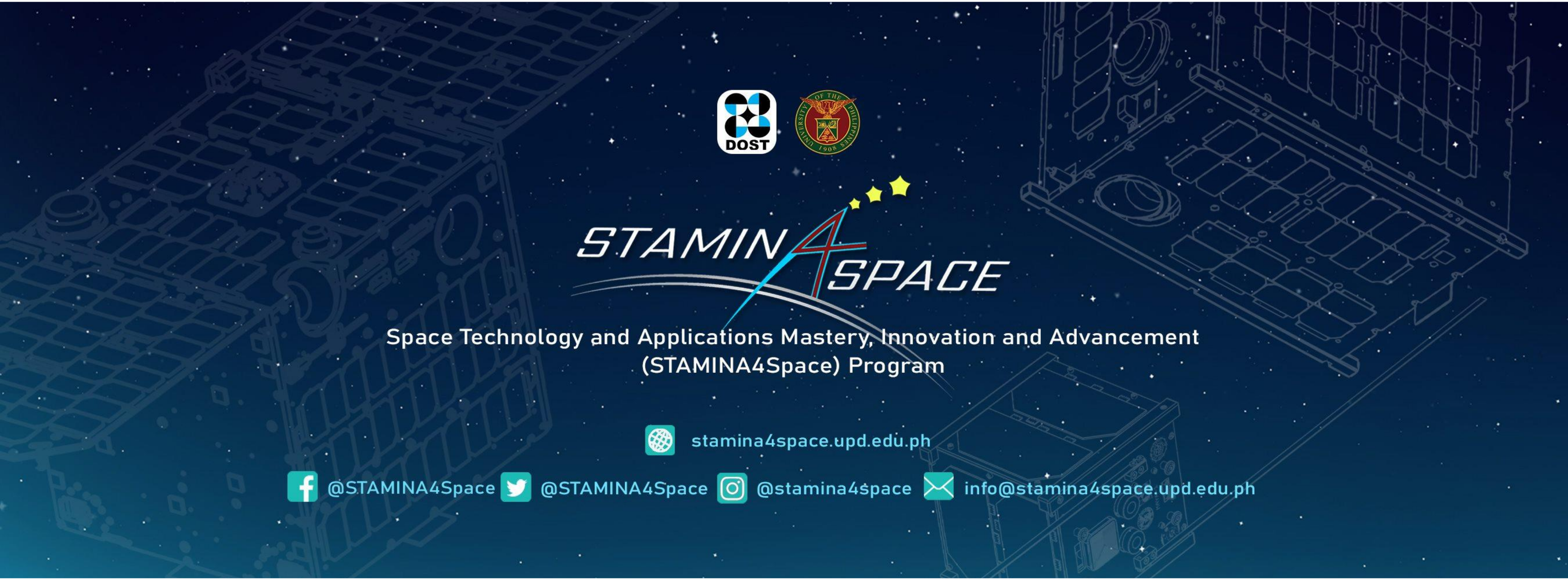
- If the priority of the satellite is to be robust in sudden disturbances, PPO is the best algorithm
- For fast attitude target, the best RL algorithm is the SAC. It is also the most comparable algorithm with the PID controller in terms of stability
- No need to re-tune RL algorithms to get a good response

Future Work



- RL algorithms with the combined features of PPO and SAC can be explored for future work
- Exploration of RL algorithms without reward engineering
- Investigate how to implement and test RL algorithms in an engineering model

Thank You!



vanessa.tan@eee.upd.edu.ph