Deep Reinforcement Learning for Spacecraft Attitude Control During Atmospheric Re-Entry

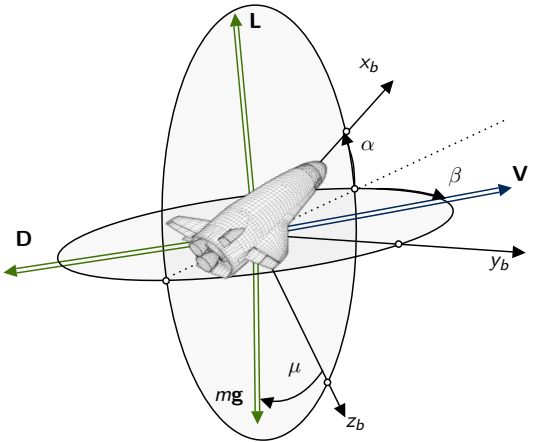
Alexander Fabisch[1], Melvin Laux[1], Mariela De Lucas Álvarez[1],
Edoardo Caroselli[2], Julian Theis[2]

[1]DFKI GmbH, Robotics Innovation Center: {first_name.last_name}@dfki.de
[2]Airbus Defence and Space GmbH: {first_name.last_name}@airbus.com

dfki ai Robotics Innovation Center

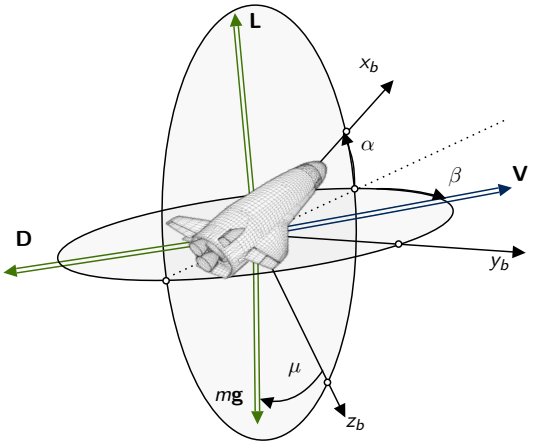
AIRBUS

Attitude Control During Hypersonic Re-Entry



- Altitude $93\,\mathrm{km} \rightarrow 10\,\mathrm{km}$
- Speed *Ma* $26.8 \rightarrow 0.5$
- Dynamic pressure $\bar{q}$: $50 \rightarrow 5825\,\mathrm{Pa}$

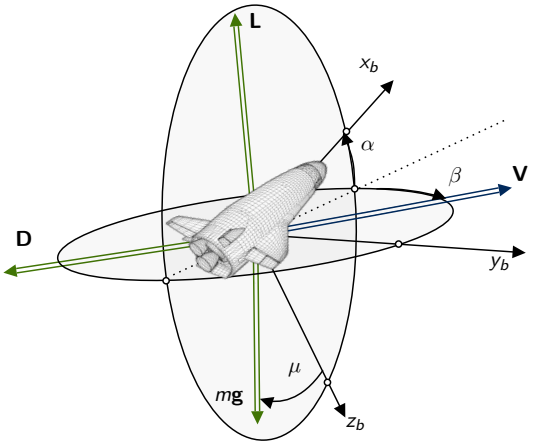
Attitude Control During Hypersonic Re-Entry



- Altitude $93\,\mathrm{km} \rightarrow 10\,\mathrm{km}$
- Speed *Ma* $26.8 \rightarrow 0.5$
- Dynamic pressure $\bar{q}$: $50 \rightarrow 5825\,\mathrm{Pa}$

- Guidance precomputes trajectory
- Track α_{cmd} and μ_{cmd}, hold $\beta \approx 0$
- Flaps: pitch, roll. Thrusters: yaw

Attitude Control During Hypersonic Re-Entry

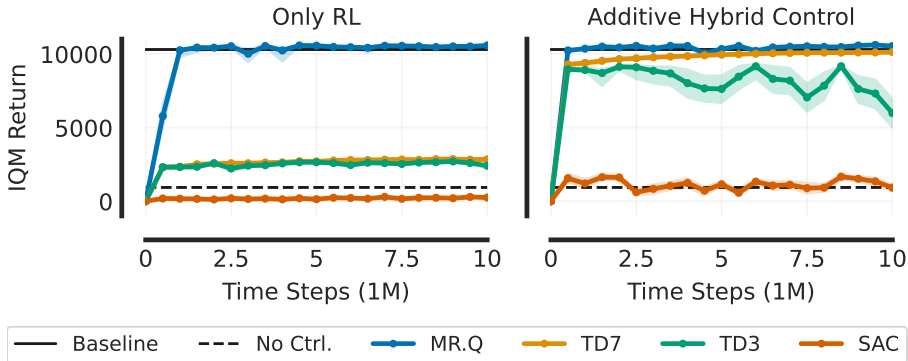


- Altitude $93\,\mathrm{km} \rightarrow 10\,\mathrm{km}$
- Speed *Ma* $26.8 \rightarrow 0.5$
- Dynamic pressure $\bar{q}$: $50 \rightarrow 5825\,\mathrm{Pa}$

- Guidance precomputes trajectory
- Track α_{cmd} and μ_{cmd}, hold $\beta \approx 0$
- Flaps: pitch, roll. Thrusters: yaw

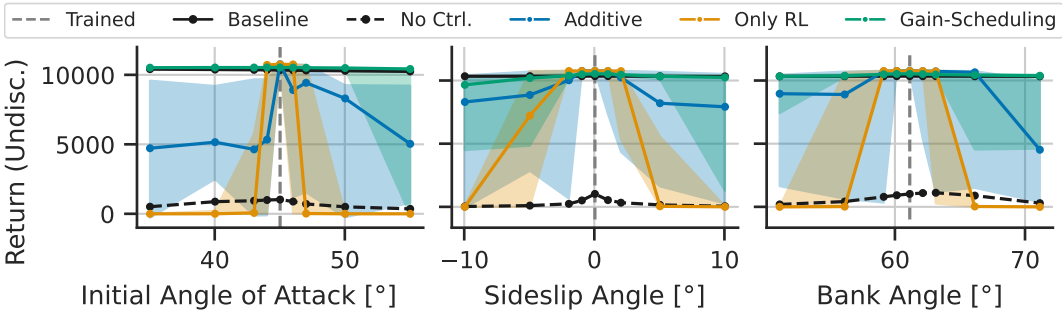
- Baseline: industry-standard PID
- Pure / additive / gain-scheduling RL

MR.Q Beats the Industry-Standard Controller



- MR.Q reaches and then surpasses the baseline, without task-specific tuning
- 10 seeds, IQM, 95 % bootstrap CI

. . . But It Does Not Generalize Out of Distribution



- **Pure RL collapses a few degrees away, hybrid controllers are more robust**
- Grey dashed line: the initial attitude seen during training
- Flip side: the baseline degrades drastically below $14\,\mathrm{rad/s}$ flap bandwidth (nominal 30)

Robustness in the Operational Envelope

Controller	Success rate	98-percentile error [deg] $\lvert e_\alpha \rvert$	$\lvert e_\beta \rvert$	$\lvert e_\mu \rvert$
Baseline (gain-scheduled PID)	79 %	2.96	0.57	2.35
Additive hybrid (PID + MR.Q)	93–94 %	0.59–0.83	0.65–0.85	0.87–1.19
Pure RL (MR.Q)	98–100 %	0.80–0.86	1.76–2.03	1.49–2.57

- **RL increases success rates by tracking angle of attack better**
- Dynamics randomization of mass, inertia tensor, and flap actuator bandwidth
- Range over 3 seeds. Best per column. Evaluated on 100 held-out contexts.

Key Insights

- MR.Q generalizes to every variation seen during training (e.g., exploring actions).
- It does not generalize out-of-distribution (OOD).
- OOD generalization requires hybrid control: policy + domain-specific controller.

Paper, code, and project page

Poster #49 • 15:00–18:00 • today

Backup: Reward, Observations, and Control Architectures

Reward:

$$r_{t'} = \underbrace{\max(0.001, e^{-w_c c(t)})}_{\text{attitude reward} > 0} - \text{control cost}$$

- $c(t)$: normalized squared attitude error
- > 0: no incentive to terminate early
- Control cost penalizes actuation

Observations: aerodynamic angles, rates, commands, and per angle the error, its integral, and its derivative

Control architectures:

- **Baseline:** PID with gain scheduling
- **Pure RL:** $a = \pi(\boldsymbol{o})$
- **Additive hybrid:** $a = a_{\mathrm{PID}}(\boldsymbol{o}) + \pi(\boldsymbol{o})$
- **GS hybrid:** $a = a_{\mathrm{PID}}(\boldsymbol{o}, \pi(\boldsymbol{o}))$, policy scales the PID gains

Algorithms (no task-specific tuning): SAC, TD3, TD7, MR.Q