



# Autonomous Observation Scheduling using Reinforcement Learning Under Operational and Visibility Constraints

Daniel Huterer Prats  
Ph.D. Student

Christopher Wheeler  
CTO  
Interactive Aptitude

Hanspeter Schaub  
*Distinguished Professor*

*RL for Space-to-Space Surveillance - AVS Meeting - October 16th, 2025*

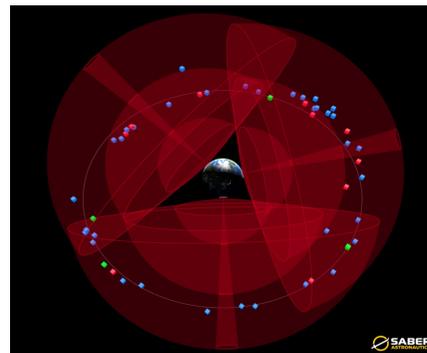


Ann and H. J. Smead Aerospace  
Engineering Sciences Department  
University of Colorado, **Boulder**

# Number of Space Objects is rapidly increasing...



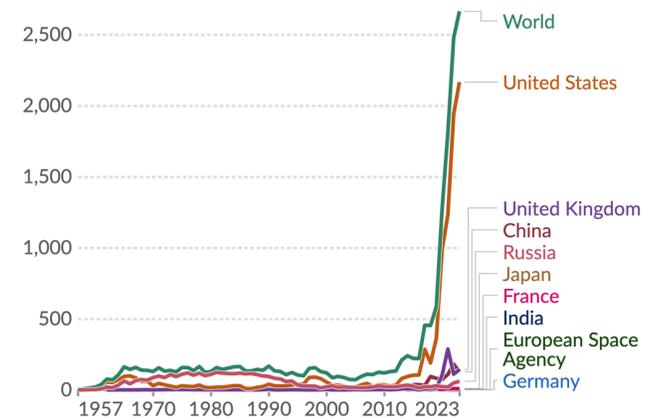
- Rapidly growing number of RSOs (Resident Space Objects) in orbit around Earth
  - >10,000 active and defunct satellites in orbit today (March 2023)
  - Tens of thousands expected next decade [1]
- Number of sensor to track targets is not keeping up
- Traditional approach using ground-based sensor has major limitations:
  - Weather, atmosphere, night-only, fixed Field of Regard (FoR)  
→ only ~25% operability in some cases [2]



## Annual number of objects launched into space

Our World in Data

This includes satellites, probes, landers, crewed spacecrafts, and space station flight elements launched into Earth orbit or beyond.



Data source: United Nations Office for Outer Space Affairs (2024)  
CC BY

# Need for tracking activities of SCs ...



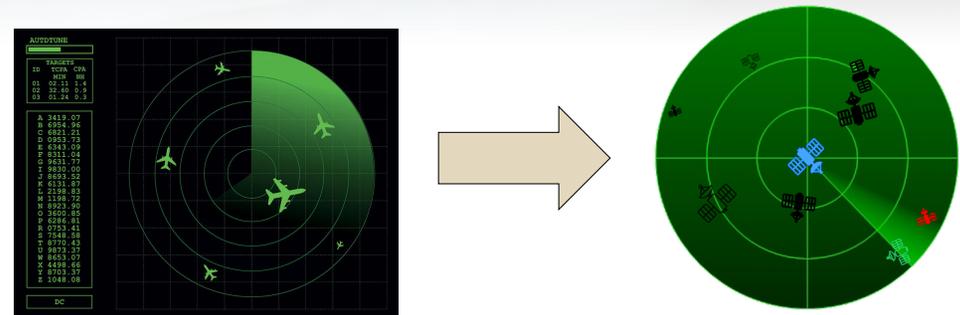
## Strategic Interests – Key Technologies

- Can we observe all the activity in particular region in space?

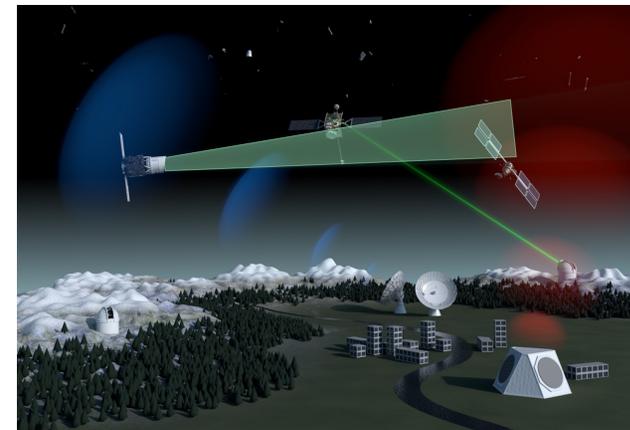
→ We want to be able to know what is going on above us...

→ Gain knowledge of states of targets (frequently)

- Need of space-based sensors to aid with SDA
  - Space-based sensors avoid atmospheric disturbances
  - Over an orbit they achieve global effective FoR



www.iadb.in/2023/07/30



Credit: ESA/Alan Baker

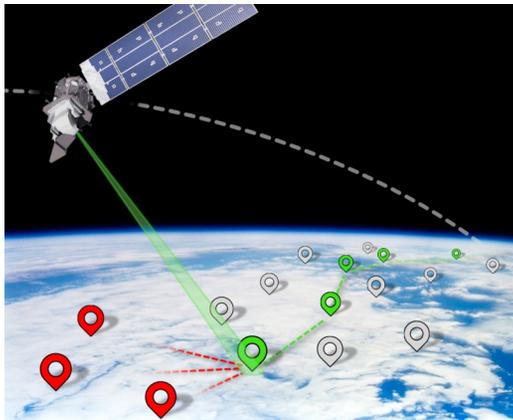
# Difficulty of creating an effective RSO-imaging schedule



## Agile Earth Observing Satellite (AEOS)

### → keeping catalog of ground-targets

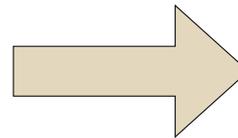
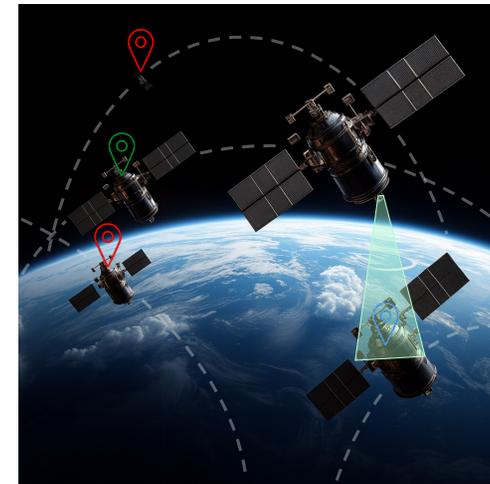
- Stationary/slow-moving targets
- Sequence of targets is predictable... (without considering clouds, etc.)
- All targets are 'below' the spacecraft and only smaller maneuvers need to be made to switch from one to another



## Space-to-Space Surveillance Satellite

### → keeping catalog of space-targets

- Fast- and slow-moving targets
- Sequence of targets is unpredictable!
- They can come from all direction
- Targets can be 'below' or above the spacecraft larger maneuvers (up to 180°) need to be made to switch from one to another



# Problem Statement: LEO-to-LEO imaging



- **Environment:**

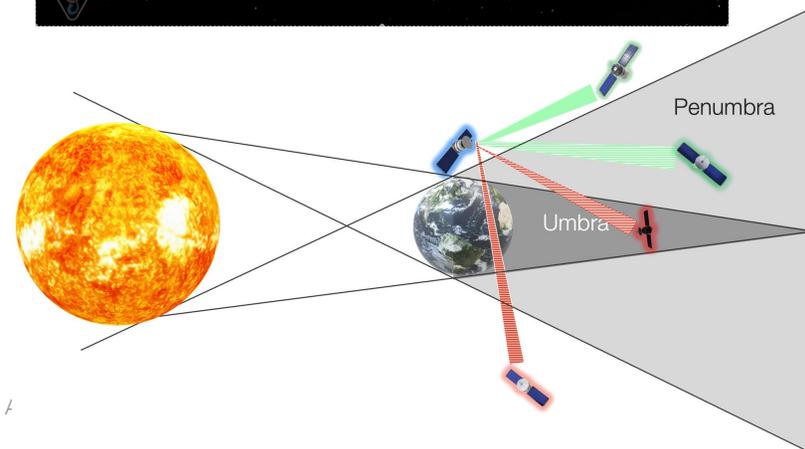
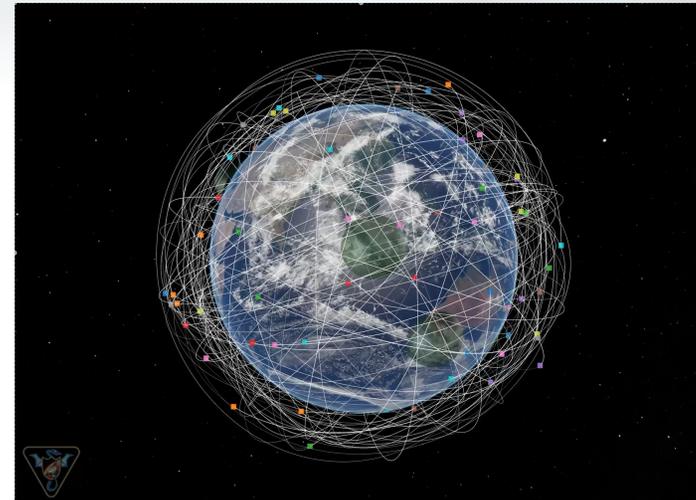
- Scanning Satellite & 100 RSOs: all in LEO and randomized orbital elements
- Episode duration of 150 imaging actions (~8 orbits)

- **Objective: maximize the fraction of successfully imaged targets/out of all the RSOs in the environment**

- **Control:** Image Target, Charge, Desat and Downlink

- **Constraints:**

1. Target visibility constraints
2. Maintain spacecraft safety (battery, RWs)
3. Restricted storage capacity



# Initial Condition Details



Table 2: Spacecraft and control parameters

Parameter	Value(s)
<i>Physical</i>	
Mass, Inertia	$m = 330\text{kg}$ , $[I_{xx}, I_{yy}, I_{zz}] = [82.1, 98.4, 121.0]\text{ kg m}^2$
Actuators	3× Reaction Wheels (orthogonal axes)
RW max torque	$u_{\text{max}} = 0.4\text{Nm}$ (per-axis)
RW speed limit	6000 RPM
Initial RW Speeds	±500 RPM
Initial Body Rates	$<0.0001\text{ rad s}^{-1}$ (random tumble)
Battery Capacity	500 Wh
Initial Stored Charge	25–50% of capacity
Solar panel size	1 m <sup>2</sup> (efficiency 20 %)
<i>Power Drains</i>	
Base Power Draw	10 W
Instrument Power Draw	30 W
Transmitter Power Draw	25 W
Thruster Power Draw	80 W
<i>Other Properties</i>	
Sensor-boresight-axis	Spacecraft z-axis
Desaturation Attitude	Sun-pointing
<i>Control (mrpFeedback)</i>	
Steering gains	$K = 7.0$ , $P = 35.0$

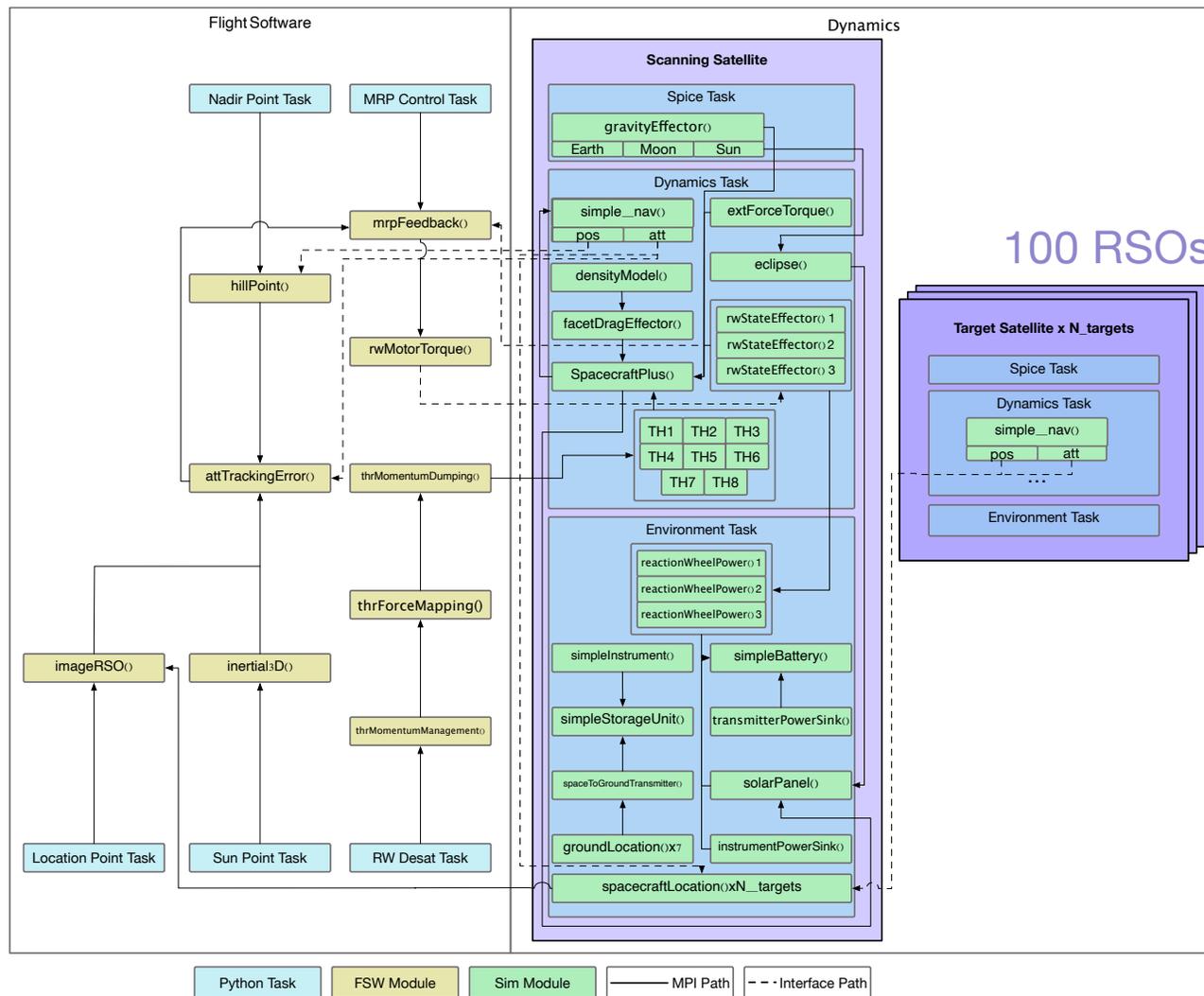
Table 1: Orbital Parameters for Scanning Satellite and Passive RSOs

Orbital Element	Scanning Satellite	Passive RSOs (N=100)
Semi-major axis ( $a$ )	6871 km <sup>a</sup>	6871 km to 8371 km <sup>b</sup>
Eccentricity ( $e$ )	0 (circular orbit)	[0.0, 0.02]
Inclination ( $i$ )	0° to 180°	0° to 180°
Right Ascension ( $\Omega$ )	0° to 360°	0° to 360°
Argument of Periapsis ( $\omega$ )	0° to 360°	0° to 360°
True Anomaly ( $f$ )	0° to 360°	0° to 360°

Groundstations



# Simulation Architecture



# BSK-RL: Environments for Satellite RL

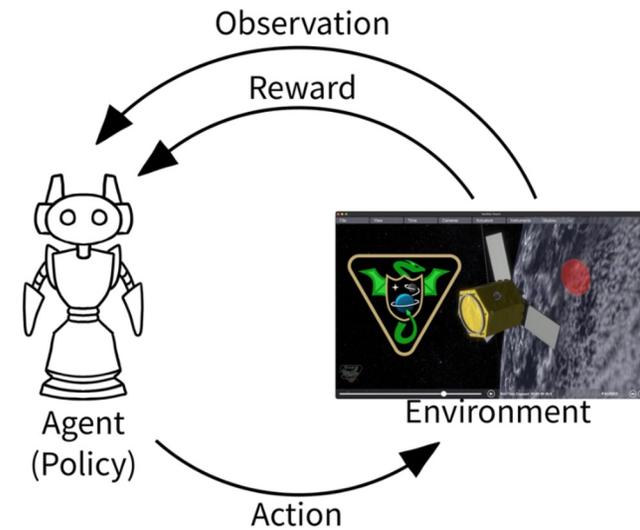


## • Inherits the best of Basilisk and Gymnasium

- Modular spacecraft and environments
- High-fidelity simulations
- Compatible with every major RL library (SB3, RLLib, etc.)



**BSK-RL Documentation**  
[https://avslab.github.io/bsk\\_rl/](https://avslab.github.io/bsk_rl/)



# MDP Formulation



## • Observation Space:

Element	Description	Dim.
$s_{data}$	Fraction of onboard data storage used	1
$s_{batt}$	Normalized battery charge level	1
$s_{mom}$	Normalized wheel momentum	1
$\epsilon_i$	Elevation angles of visible target $i$	$N$
$\mathbf{r}_{BR,i}^{\mathcal{H}}$	Relative position vectors to target $i$ in Hill frame, $\mathcal{H}$	$3 \times N$
$\theta_i$	Angle between boresight and target $i$	$N$
$d_i$	Distance to target $i$	$N$
$s_i$	Illumination factor (shadowing value) for target $i$	$N$
$e_{start}, e_{end}$	Eclipse Normalized start/end times	2
$g_{open}, g_{close}$	Normalized ground stations window open/close times	2x5

## • Actions: 10x Image Target (LOS), Charge, Desat and Downlink

## • Reward Function:

- I.  $w_i$  - weight of the target
- II.  $s_i$  - illumination factor of the target
- III.  $e_{thresh}$  - illumination threshold
- IV.  $LOS_i$  - binary measure whether Line-of-Sight is present
- V.  $\angle(\hat{\mathbf{z}}_b, \hat{\mathbf{p}}_i)$  - angular pointing requirement

$$r(t) = w_i$$

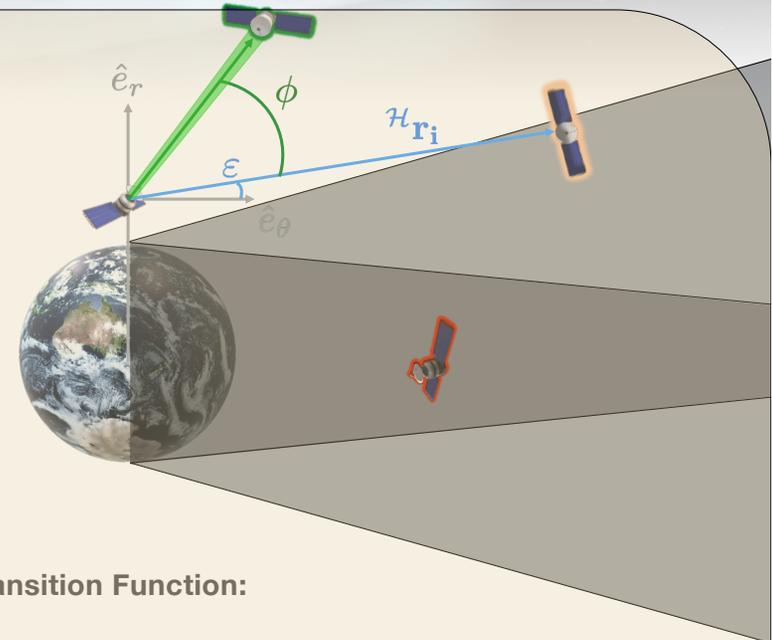
$$\text{s.t. } s_i > e_{thresh},$$

$$LOS_i,$$

$$\angle(\hat{\mathbf{z}}_b, \hat{\mathbf{p}}_i) \leq \text{MRP}(0.01)$$

## • Transition Function:

- Deterministic transitions from high-fidelity simulator
- Step duration  $\Delta t$  is given by chosen action duration
- Terminal States: empty battery; RW speeds exceeded



Throwback to Lorenzo's amazing analogy presentation



## ~~Parenting Styles And How To Raise Kids~~

## How To Have The Best Time Ever... As A Kid



Throwback to Lorenzo's amazing analogy presentation

## Problem Formulation

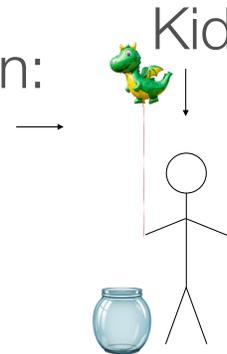
taking illuminated  
images of RSO

Spacecraft Resource  
Management



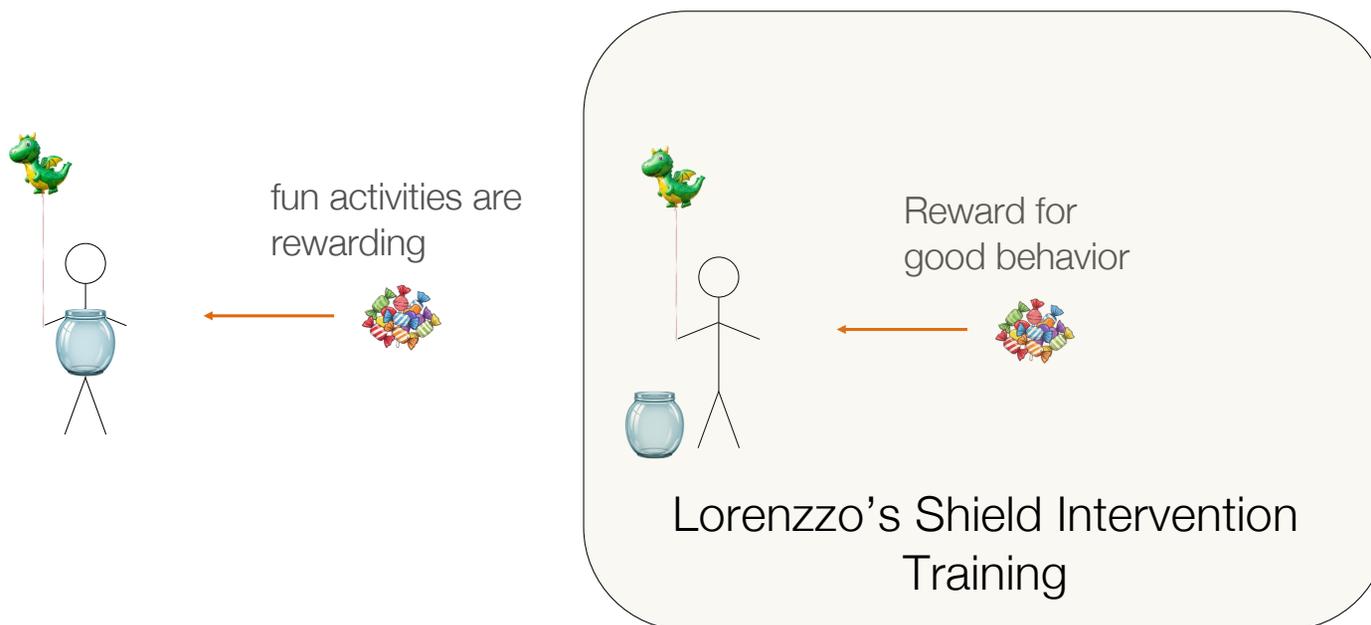
Fun/Good vs Necessary action:

- play with toys
- brush teeth
- make the bed
- tidy the room



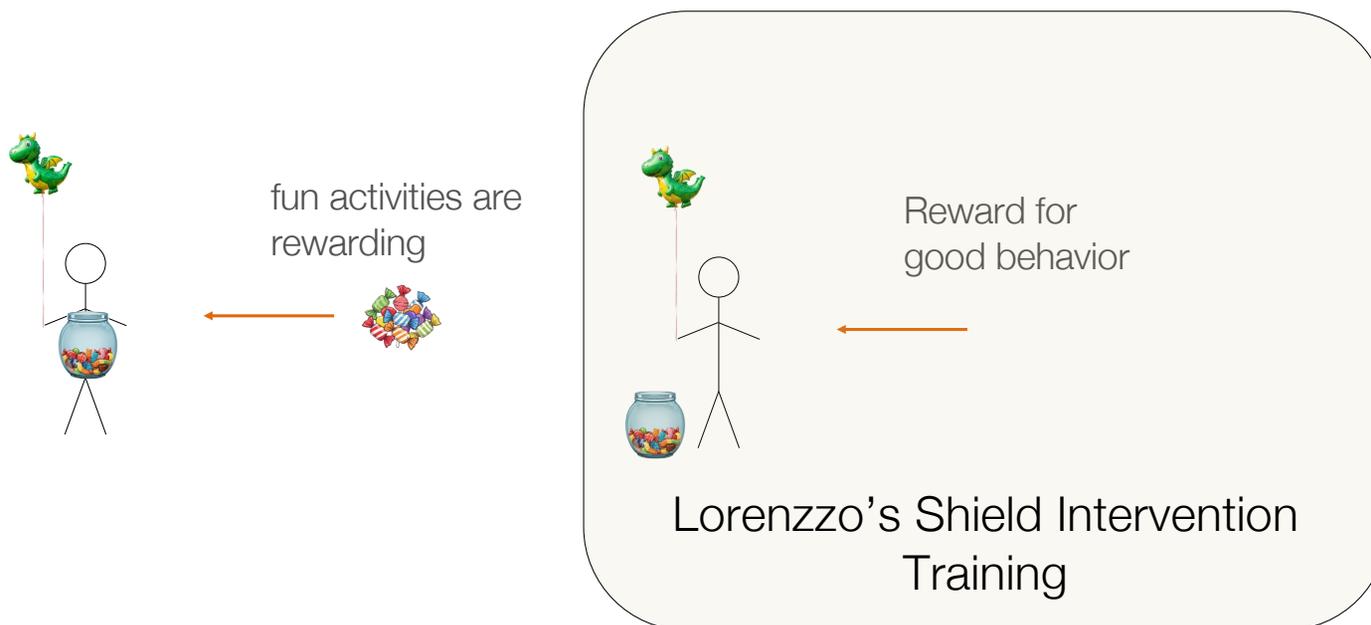
Throwback to Lorenzo's amazing analogy presentation

## Problem Formulation



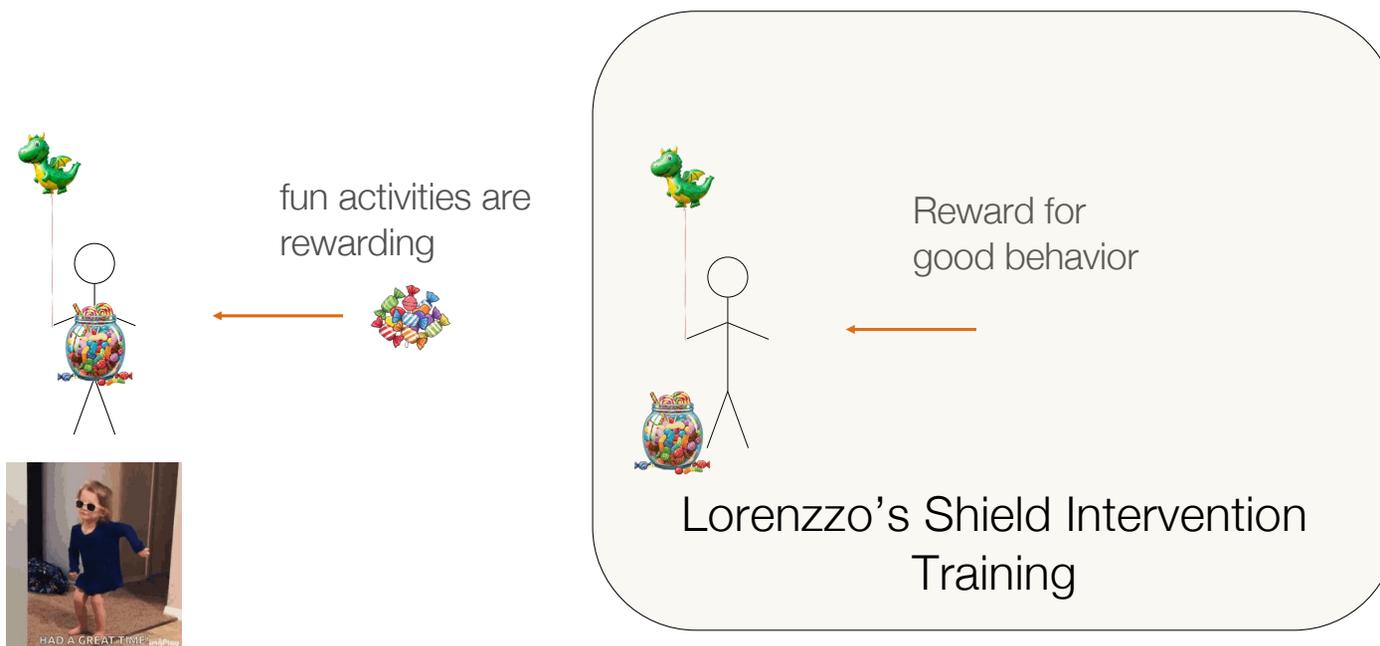
Throwback to Lorenzo's amazing analogy presentation

## Problem Formulation



Throwback to Lorenzo's amazing analogy presentation

## Problem Formulation

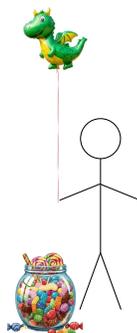


Throwback to Lorenzo's amazing analogy presentation

## Problem Formulation

If the agent only images  
other RSO, then batteries,  
RWs and data storage  
may be compromised...

You have spent too much  
time on fun activities and  
neglected some of the  
necessary ones...



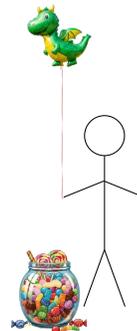
Your parents are about to  
get really mad if you keep  
playing with your toy, instead  
of going to bed!

Oh no...

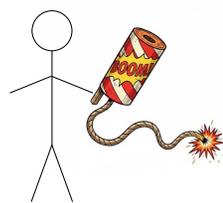


Throwback to Lorenzo's amazing analogy presentation

## Problem Formulation



Oh no...



How to have a great time as a kid, without the setting the house on fire?

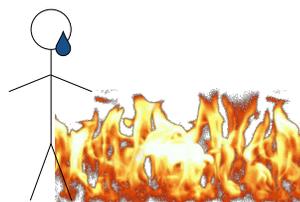
→ or getting a shoe thrown at you...

Throwback to Lorenzo's amazing analogy presentation

## Parenting Styles

Failure penalty

If



\*M. Stephenson, L. Q. Mantovani and H. Schaub, "Learning Policies for Autonomous Earth-Observing Satellite Scheduling over Semi-MDPs," *Journal of Aerospace Information Systems*. doi:[10.2514/1.1011649](https://doi.org/10.2514/1.1011649)

Throwback to Lorenzo's amazing analogy presentation

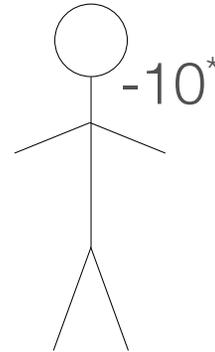
## Parenting Styles

Failure penalty

If



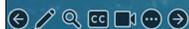
then



\*M. Stephenson, L. Q. Mantovani and H. Schaub, "Learning Policies for Autonomous Earth-Observing Satellite Scheduling over Semi-MDPs," *Journal of Aerospace Information Systems*. doi:[10.2514/1.1011649](https://doi.org/10.2514/1.1011649)



# Back to Space-to-Space Surveillance



# MDP Formulation



## • Observation Space:

Element	Description	Dim.
$s_{data}$	Fraction of onboard data storage used	1
$s_{batt}$	Normalized battery charge level	1
$s_{mom}$	Normalized wheel momentum	1
$\epsilon_i$	Elevation angles of visible target $i$	$N$
$\mathbf{r}_{BR,i}^{\mathcal{H}}$	Relative position vectors to target $i$ in Hill frame, $\mathcal{H}$	$3 \times N$
$\theta_i$	Angle between boresight and target $i$	$N$
$d_i$	Distance to target $i$	$N$
$s_i$	Illumination factor (shadowing value) for target $i$	$N$
$e_{start}, e_{end}$	Eclipse Normalized start/end times	2
$g_{open}, g_{close}$	Normalized ground stations window open/close times	$2 \times 5$

## • Actions: 10x Image Target (LOS), Charge, Desat and Downlink

### • Reward Function:

- I.  $w_i$  - weight of the target
- II.  $s_i$  - illumination factor of the target
- III.  $e_{thresh}$  - illumination threshold
- IV.  $LOS_i$  - binary measure whether Line-of-Sight is present
- V.  $\angle(\hat{\mathbf{z}}_b, \hat{\mathbf{p}}_i)$  - angular pointing requirement

$$r(t) = w_i$$

s.t.  $s_i > e_{thresh}$ ,

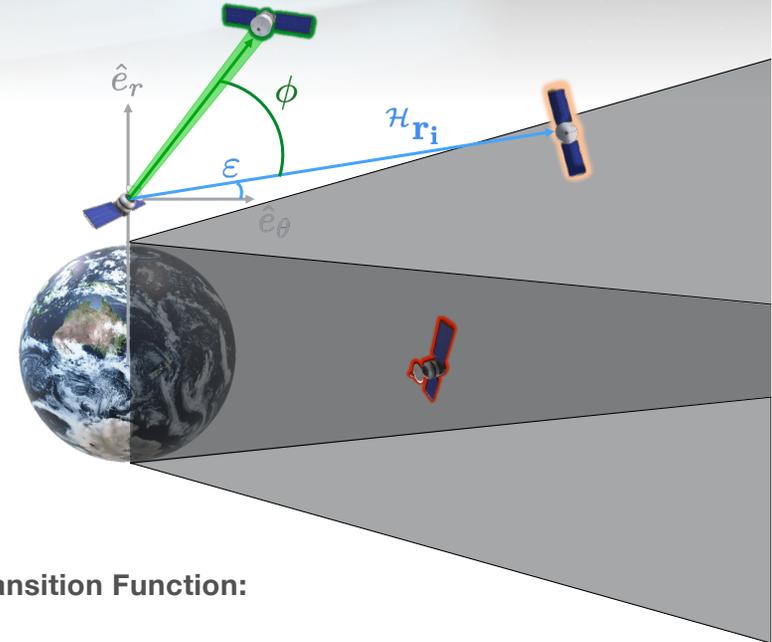
$$LOS_i,$$

$$\angle(\hat{\mathbf{z}}_b, \hat{\mathbf{p}}_i) \leq MRP(0.01)$$

### • Transition Function:

- Deterministic transitions from high-fidelity simulator
- Step duration  $\Delta t$  is given by chosen action duration
- Terminal States: empty battery; RW speeds exceeded

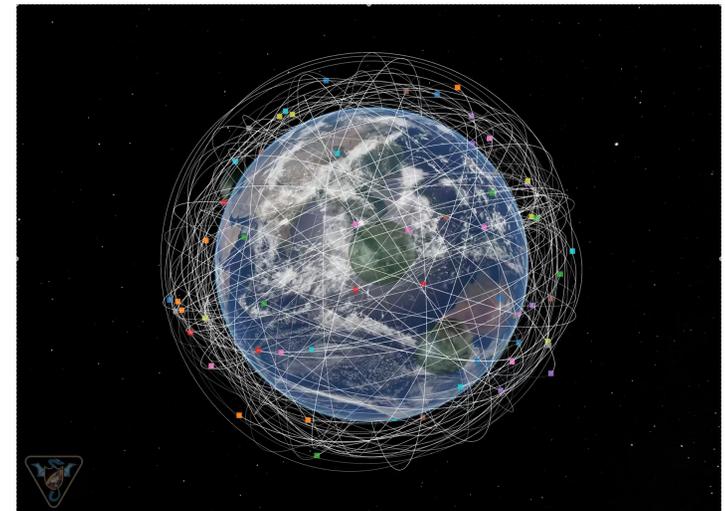
penalties



## What is 'optimality' for such a problem...



- **Finite time  $\neq$  full coverage:** Across 8 orbits, randomized inclinations/RAAN/anomalies can yield **no mutual visibility** for some RSOs within the episode.
- **LEO-to-LEO is geometry-limited:** Compared to GEO, **Earth occlusion and high relative rates** shorten access windows and shrink overlap in fields of regard.
- **Small instantaneous LOS set:** At any time, only a **small fraction** of the catalog is above the horizon and inside sensor FoR.
- **Illumination filter: Eclipse** (target or observer) removes additional opportunities; valid **sunlit** sightings are fewer than LOS sightings.





# Results

**Baseline Comparison  
Unlimited Resources**

# Unlimited Resources - Results



## Heuristic Policy:

- Uses full list of unimaged targets currently within LOS
- Chooses the one with the smallest angular pointing error (behavior to point to adjacent targets was observed in Siew et al [8])
- In comparisons with restricted resources, heuristic is fully reliant on the shield to maintain SC resources and manage storage

## Baseline Comparison\*:

- Isolating the imaging problem from the resource management
- Sample of 15,000 imaging actions for heuristic and RL-policy
- Performance is dominated by LOS geometry and target selection

RL-policy slightly outperforms heuristic

Metric	Heuristic	RL policy
Illuminated images (count)	89.54 ± 2.99	90.47 ± 2.95
Illumination fraction (%)	71.28 ± 2.90	88.97 ± 3.79

RL-policy actively chooses illuminated targets  
→ imaging behind/in-front when entering/leaving eclipse

\* Due to the randomized elements imaging all targets may not be physically possible (especially considering eclipse)



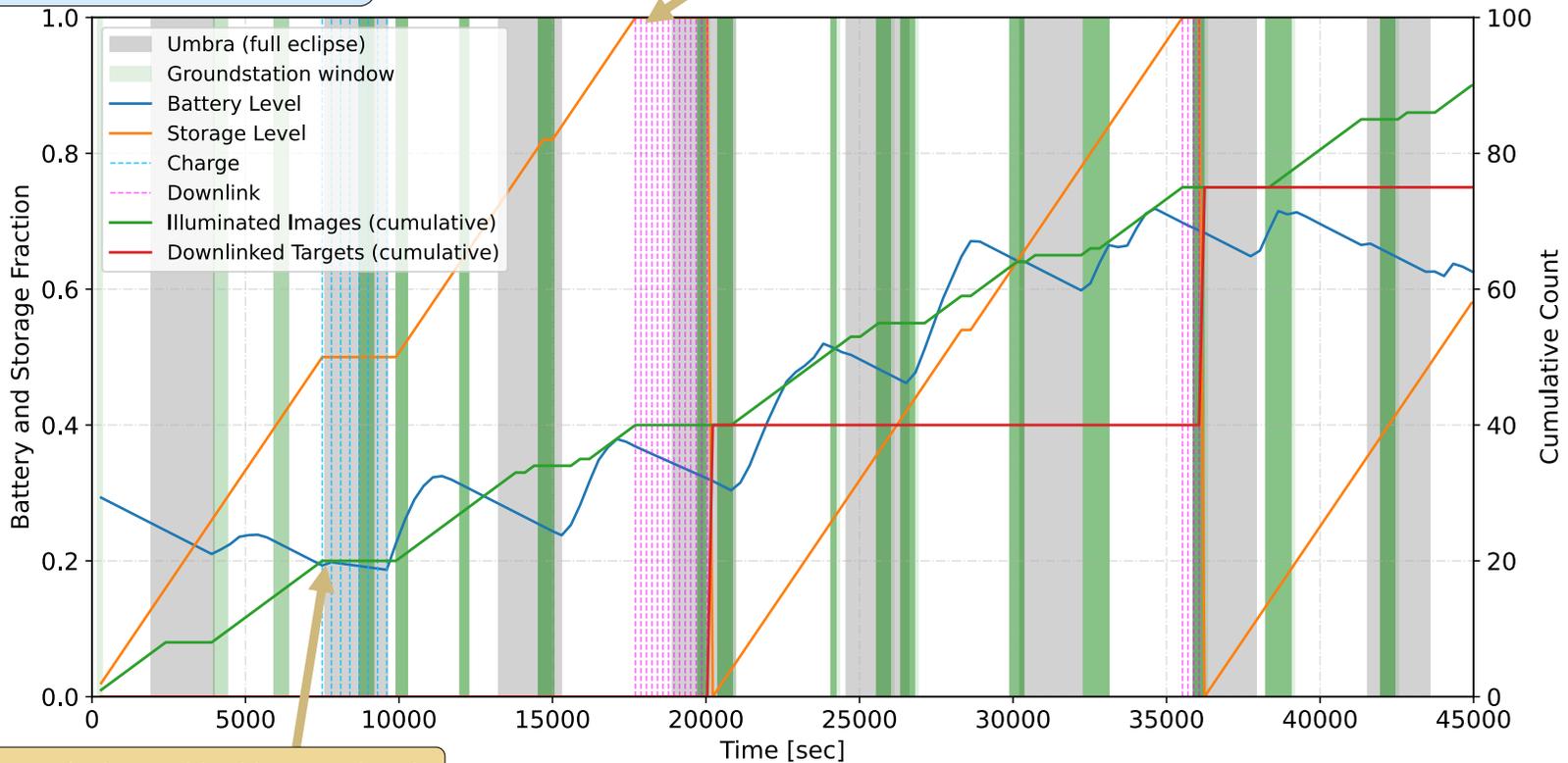
# Performance Comparison Under Restricted Resources

# Restricted Resources – Case Study (Heuristic)



Shield Intervention: storage >99%  
 =>downlink & battery <20% =>charge

Only downlinks when full (downlinking older- less valuable data)



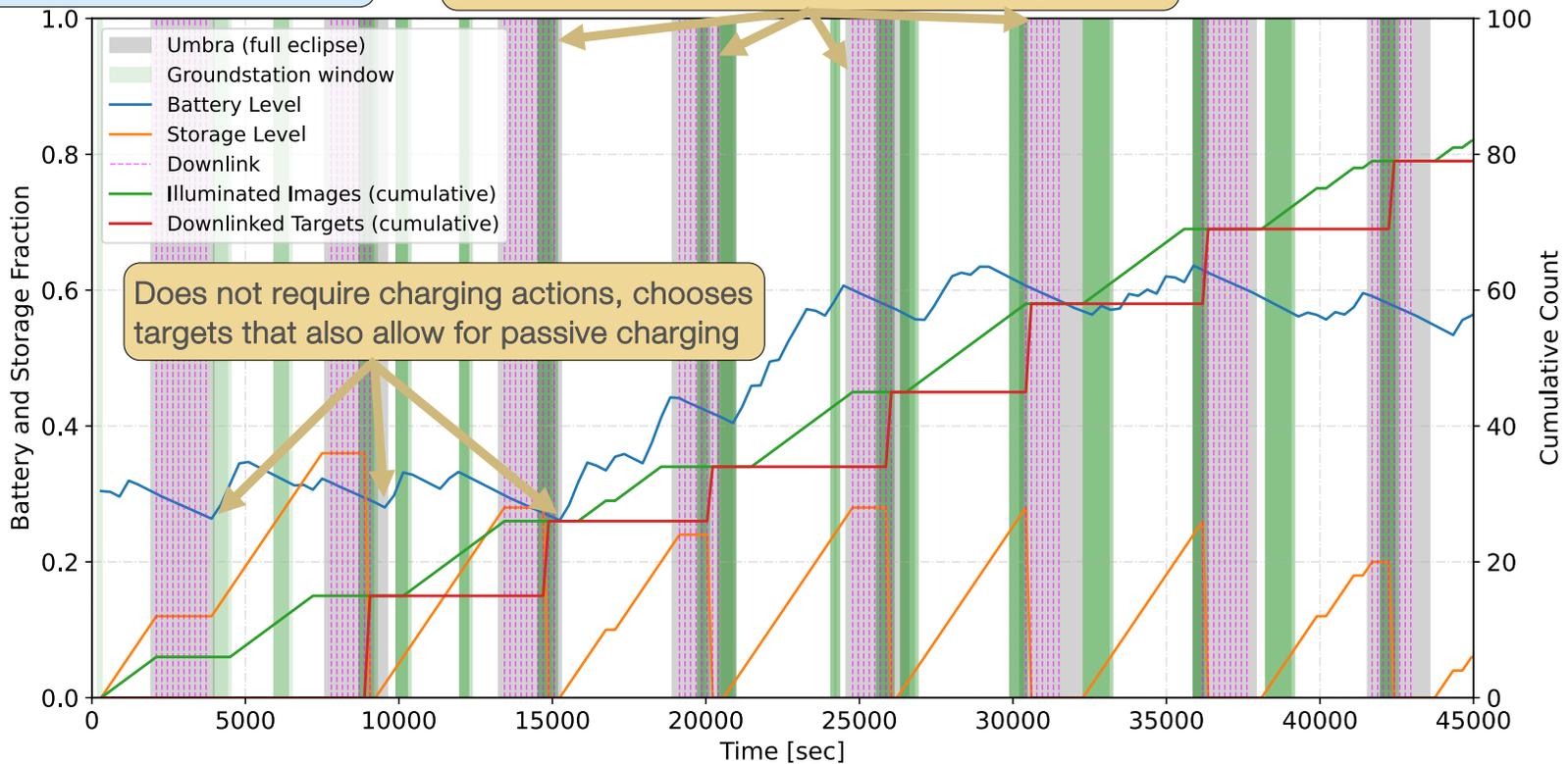
Drops below critical battery levels

# Restricted Resources – Case Study (RL-policy)



Shield Intervention: storage >99%  
 =>downlink & battery <20% =>charge

Learns to use eclipse-periods, when most targets within LOS are also dark, to downlink frequently

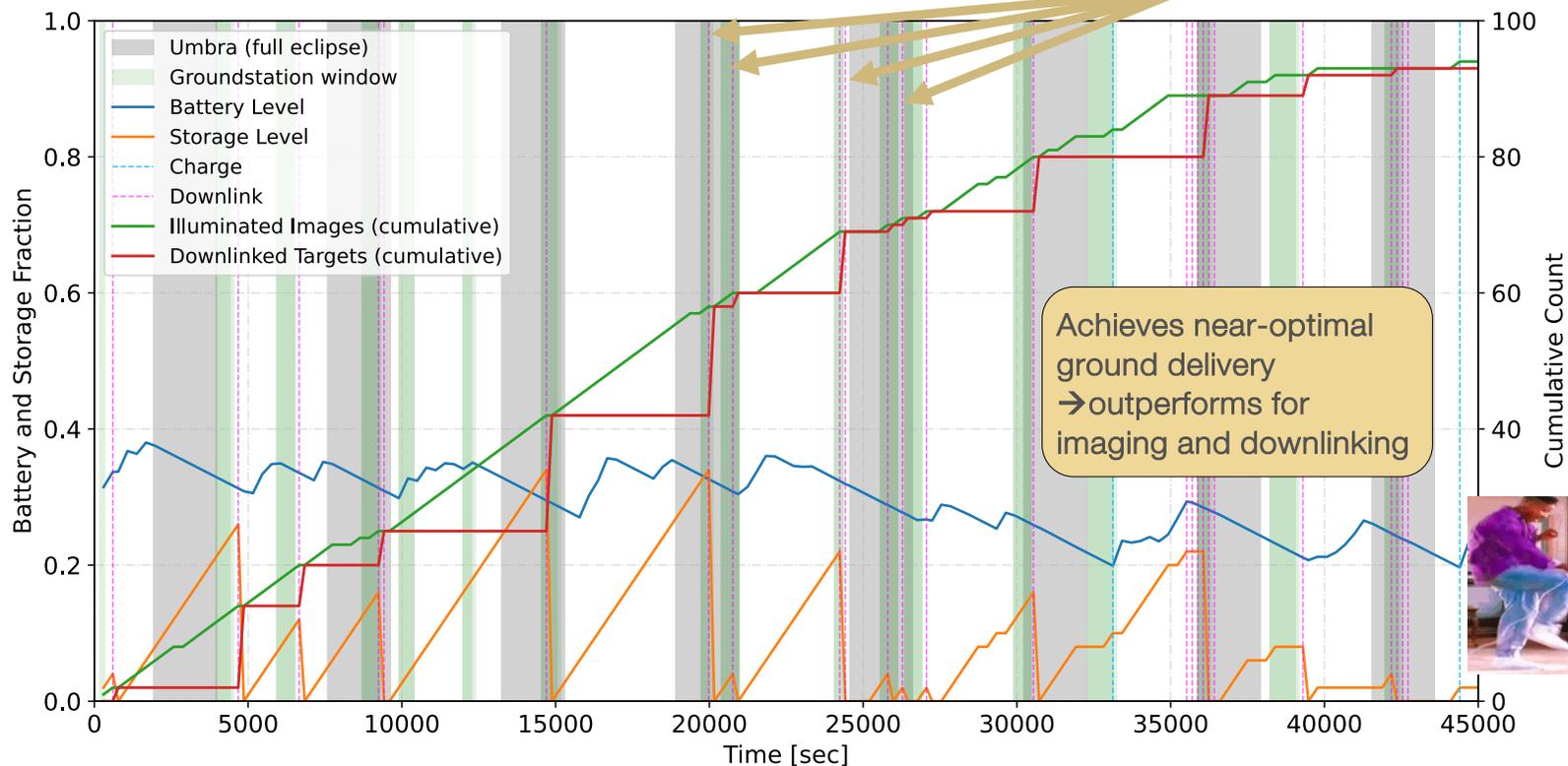


# Varying the Reward Policy – With Improved Training



Using increased training durations and larger batch sizes

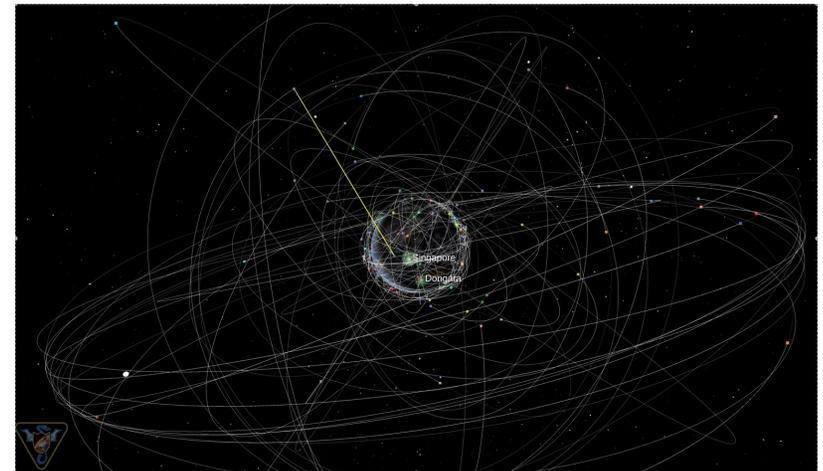
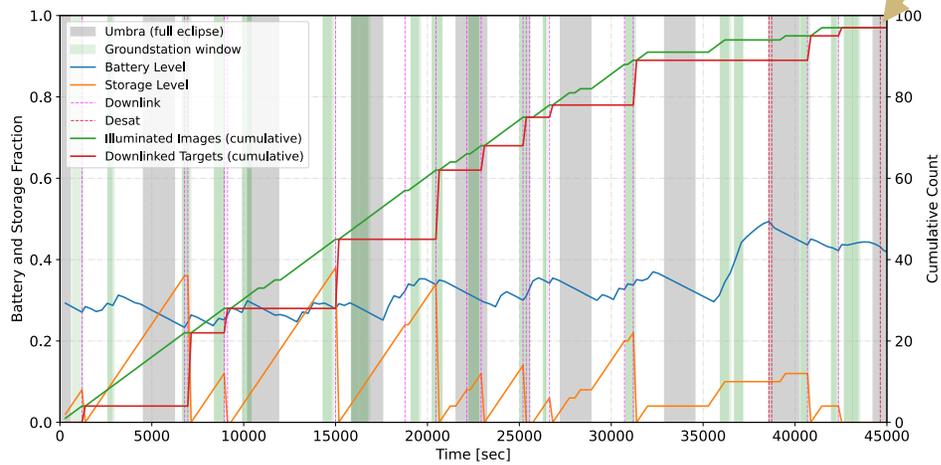
Learns to downlink when in range of ground-station



# Testing Mixed Orbit Regimes – with LEO-trained Policy



Achieved 95 out of 98 RSOs that were ever visible



The same policy also performs well in a mixed regime environment!

# Conclusions



**RL-policies can handle the dynamics of LEO-to-LEO imaging, while also managing spacecraft resources**

- RL was successfully used for space-space satellite tasking
- Comparable or improved performance over heuristic policies on different metrics

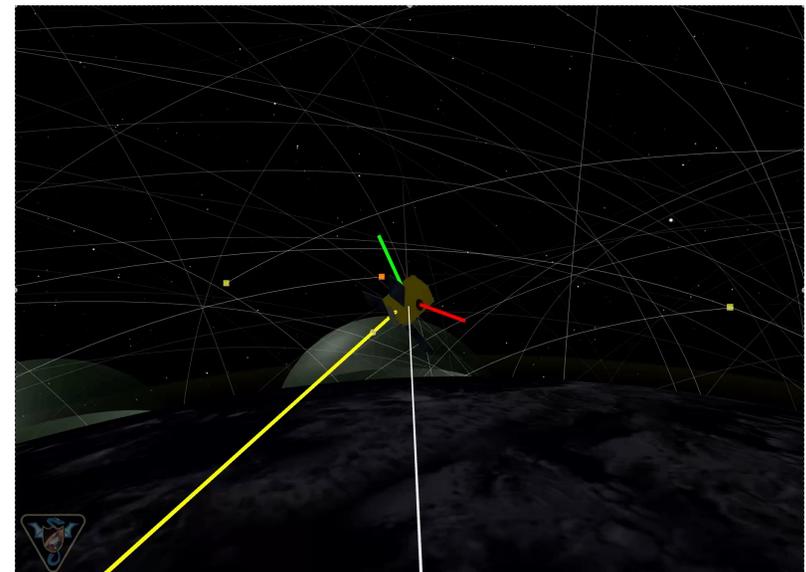
**→ Space-based sensor tasking using RL is a promising approach to help manage the growing SDA challenge as the number of RSOs is increasing**

• **Ongoing Work:**

- Analyze tradeoff between imaging versus downlink rewarding
- Expansion to larger number of RSOs (also in MEO and GEO)

• **Future Expansions:**

- Reimage targets → to maintain knowledge of RSO catalog (include uncertainties)
- Relax fixed-imaging-time constraint
- Expand to Multi-agent scenario → various scanning satellites working in tandem





# Questions?

## Contact

[daniel.hutereprats@Colorado.edu](mailto:daniel.hutereprats@Colorado.edu)  
<https://www.linkedin.com/in/dhutereprats/>



BSK-RL Documentation  
[https://avslab.github.io/bsk\\_rl/](https://avslab.github.io/bsk_rl/)

# Conclusions



**RL-policies can handle the dynamics of LEO-to-LEO imaging, while also managing spacecraft resources**

- RL was successfully used for space-space satellite tasking
- Comparable or improved performance over heuristic policies on different metrics

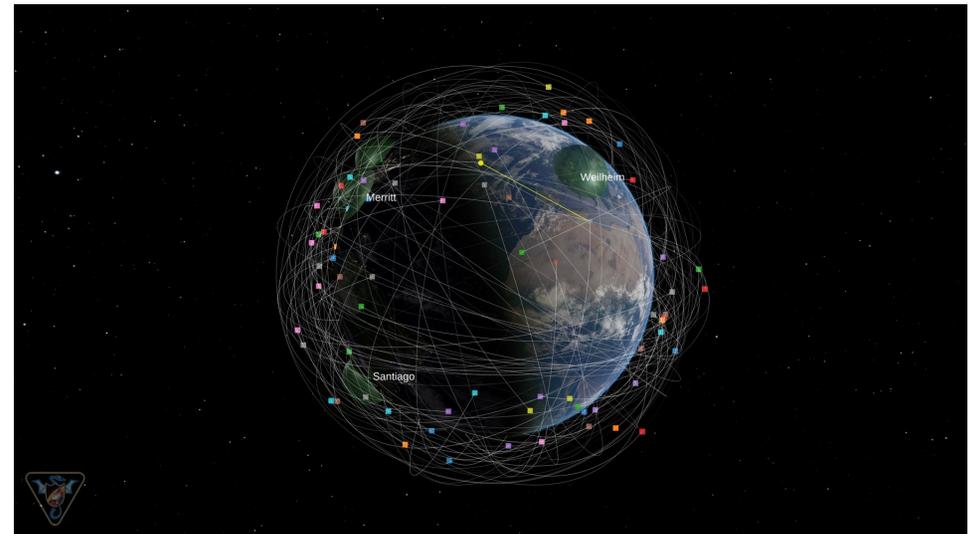
**→ Space-based sensor tasking using RL is a promising approach to help manage the growing SDA challenge as the number of RSOs is increasing**

**• Ongoing Work:**

- Analyze tradeoff between imaging versus downlink rewarding
- Expansion to larger number of RSOs (also in MEO and GEO)

**• Future Expansions:**

- Reimage targets → to maintain knowledge of RSO catalog (include uncertainties)
- Relax fixed-imaging-time constraint
- Expand to Multi-agent scenario → scanning agents collaborating



# References



- [1] Andrew Williams, Olivier Hainaut, Angel Otarola, Gie Han Tan, Andrew Biggs, Neil Phillips, and Giuliana Rotola. A report to eso council on the impact of satellite constellations. Technical report, European Southern Observatory (ESO), 2021.
- [2] Mark R Ackermann, Colonel Rex R Kiziah, Peter C Zimmer, J T McGraw, John T McGraw, J T McGraw, and David D Cox. A Systematic Examination Of Ground-Based And Space-Based Approaches To Optical Detection And Tracking Of Satellites.
- [3] Daniel Jang, Peng Mun Siew, David Gondelach, and Richard Linares. Space situational awareness tasking for narrow field of view sensors: A deep reinforcement learning approach. 71st International Astronautical Congress. International Astronautical Federation, the International Academy of Astronautics, and the International Institute of Space Law.
- [4] Richard Linares and Roberto Furfaro. An Autonomous Sensor Tasking Approach for Large Scale Space Object Cataloging.
- [5] Peng Mun Siew, Tory Smith, Ravi Ponmalai, and Richard Linares. Scalable Multi-Agent Sensor Tasking Using Deep Reinforcement Learning.
- [6] Richard Linares and Roberto Furfaro. Dynamic Sensor Tasking for Space Situational Awareness via Reinforcement Learning.
- [7] Benedict Oakes, Jason F Ralph, and Jordi Barr. Deep Reinforcement Learning Applications to Space Situational Awareness Scenarios.
- [8] Peng Mun Siew, Daniel Jang, Thomas G. Roberts, and Richard Linares. Space-Based Sensor Tasking Using Deep Reinforcement Learning. 69(6):1855–1892.
- [9] L. Q. Mantovani, Y. Nagano, and H. Schaub. Reinforcement learning for satellite autonomy under different cloud coverage probability observations. AAS/AIAA Astrodynamics Conference, 2023.
- [10] M. Stephenson and H. Schaub. Reinforcement learning for earth-observing satellite autonomy with event-based task intervals. AAS/AIAA Space Flight Mechanics Conference, 2024.
- [11] Mark Stephenson, Daniel Huterer Prats, and Hanspeter Schaub. Autonomous satellite inspection in low earth orbit with optimization-based safety guarantees. In International Workshop on Planning Scheduling for Space, Toulouse, France, April 28–30 2025.

# Simulation Set Up Details

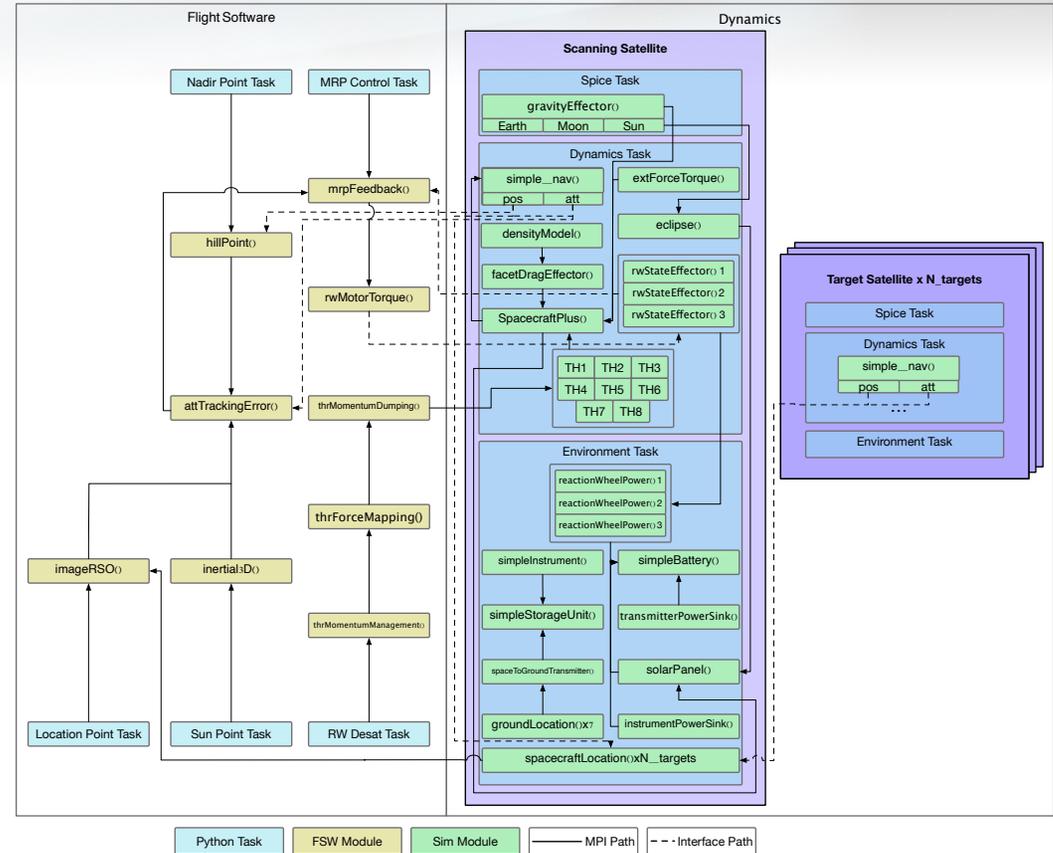


Table 1: Orbital Parameters for Scanning Satellite and Passive RSOs

Orbital Element	Scanning Satellite	Passive RSOs (N=100)
Semi-major axis ( $a$ )	6871 km <sup>a</sup>	6871 km to 8371 km <sup>b</sup>
Eccentricity ( $e$ )	0 (circular orbit)	[0.0, 0.02]
Inclination ( $i$ )	0° to 180°	0° to 180°
Right Ascension ( $\Omega$ )	0° to 360°	0° to 360°
Argument of Periapsis ( $\omega$ )	0° to 360°	0° to 360°
True Anomaly ( $f$ )	0° to 360°	0° to 360°

Table 2: Spacecraft and control parameters

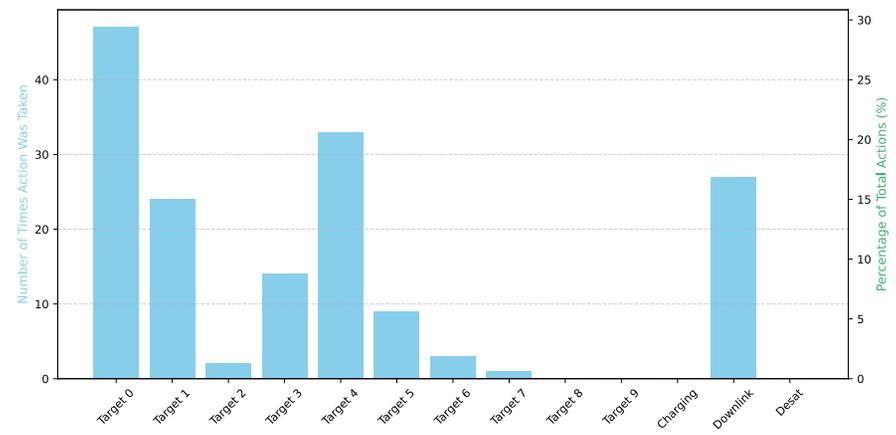
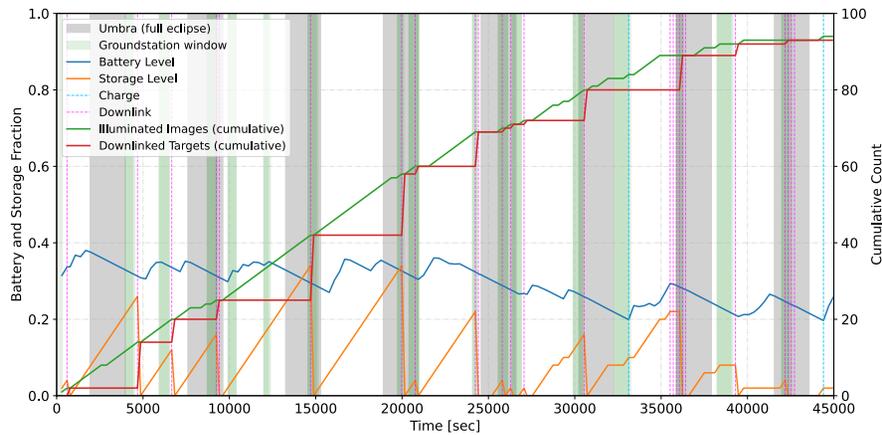
Parameter	Value(s)
<i>Physical</i>	
Mass, Inertia	$m = 330$ kg, $[I_{xx}, I_{yy}, I_{zz}] = [82.1, 98.4, 121.0]$ kg m <sup>2</sup>
Actuators	3× Reaction Wheels (orthogonal axes)
RW max torque	$u_{max} = 0.4$ N m (per-axis)
RW speed limit	6000 RPM
Initial RW Speeds	±500 RPM
Initial Body Rates	<0.0001 rad s <sup>-1</sup> (random tumble)
Battery Capacity	500 Wh
Initial Stored Charge	25–50% of capacity
Solar panel size	1 m <sup>2</sup> (efficiency 20%)
<i>Power Drains</i>	
Base Power Draw	10 W
Instrument Power Draw	30 W
Transmitter Power Draw	25 W
Thruster Power Draw	80 W
<i>Other Properties</i>	
Sensor-boresight-axis	Spacecraft z-axis
Desaturation Attitude	Sun-pointing
<i>Control (mrpFeedback)</i>	
Steering gains	$K = 7.0$ , $P = 35.0$



# Restricted Resources – Case Study (future paper)



Using increased training durations and larger batch sizes



# Full MDP Formulation



## • Observation Space:

Element	Description	Dim.
$s_{\text{data}}$	Fraction of onboard data storage used	1
$s_{\text{batt}}$	Normalized battery charge level	1
$s_{\text{mom}}$	Normalized wheel momentum	1
$\epsilon_i$	Elevation angles of visible target $i$	$N$
$\mathbf{r}_{BR,i}^{\mathcal{H}}$	Relative position vectors to target $i$ in Hill frame, $\mathcal{H}$	$3 \times N$
$\theta_i$	Angle between boresight and target $i$	$N$
$d_i$	Distance to target $i$	$N$
$s_i$	Illumination factor (shadowing value) for target $i$	$N$
$e_{\text{start}}, e_{\text{end}}$	Eclipse Normalized start/end times	2
$\mathbf{g}_{\text{open}}, \mathbf{g}_{\text{close}}$	Normalized ground stations window open/close times	$2 \times 5$

## • Actions: 10x Image Target (LOS), Charge, Desat and Downlink

## • Reward Function:

- I.  $w_i$  - weight of the target
- II.  $s_i$  - illumination factor of the target
- III.  $e_{\text{thresh}}$  - illumination threshold
- IV.  $\text{LOS}_i$  - binary measure whether Line-of-Sight is present
- V.  $\angle(\hat{\mathbf{z}}_b, \hat{\mathbf{p}}_i)$  - angular pointing requirement

$$r(t) = w_i$$

$$\text{s.t. } s_i > e_{\text{thresh}},$$

$$\text{LOS}_i,$$

$$\angle(\hat{\mathbf{z}}_b, \hat{\mathbf{p}}_i) \leq \text{MRP}(0.01)$$

## • Transition Function:

- Deterministic transitions from high-fidelity simulator
- Step duration  $\overline{\Delta t}$  is given by chosen action duration
- Terminal States: empty battery; RW speeds exceeded

## • Training:

Table 9: RL Training Parameters

Name	Value
Learning rate	$1 \times 10^{-6}$
Discount factor ( $\gamma$ )	0.9997
Gradient clip	1.0
PPO clip parameter ( $\epsilon$ )	0.15
Training batch size	3200

## • Shield Implementation (during testing):

- If storage-fraction > 99%  $\rightarrow$  task downlink action
- If rw-speed-fraction > 90%  $\rightarrow$  task desaturation action
- If battery fraction < 20%  $\rightarrow$  task charging action

# Illumination setup and angles used for observations

