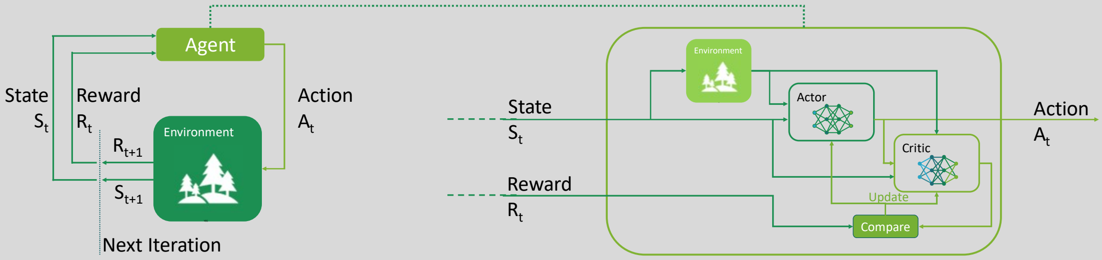
Reinforcement Learning in Differentiable Simulation for Heliostat Control

Safalya Pal[1], Max Pargmann*[1], Daniel Maldonado Quinto[1], Robert Pitz-Paal[1,2],

[1]German Aerospace Center (DLR), Institute of solar research; [2]Chair of Solar Technology, RWTH Aachen University
*max.pargmann@dlr.de

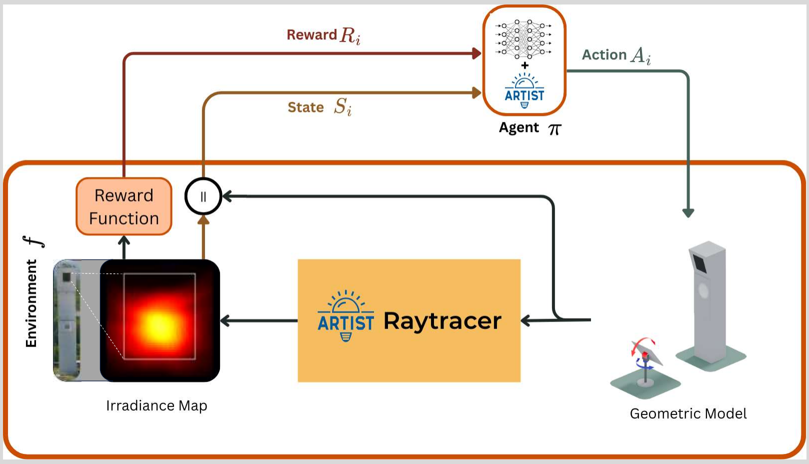
Data Availability

Reinforcement Learning (RL) (Sutton & Barto, 2018) is a branch of Machine Learning where an agent learns to take actions by interacting with the environment. The agent then recieves rewards or penalties, and the objective of the agent is to maximixe the total reward.



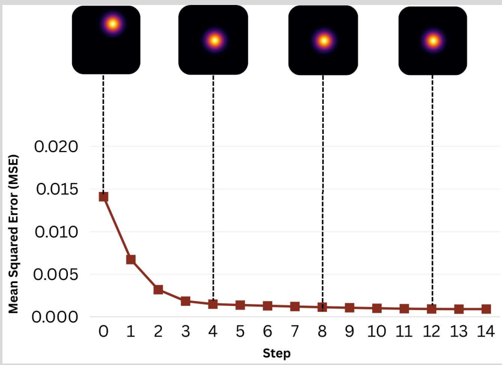
Training Environment at Solar Towers

Our method is based on the ARTIST differentiable simulator [2], which computes the gradients of the ray-tracing process with respect to heliostat configurations. This capability allows us to directly propagate error signals from the simulated flux distribution back to the motor positions, thereby enabling us to iteratively update our controlling agent's policy using these error signals.To handle the inherent complexities of high-dimensional control in motor position shifts (2 per heliostat), we adapt the Analytic Policy Gradient (APG) [3] framework to our task. Our differentiable simulator lets us compute the gradients of the reward function with respect to the policy parameters, these gradients then lets us update the policy network using stochastic gradient-decent or any of its variants.



In this work, we define two to different tasks. First to optimize the aim-point distribution on the receiver, second also predicting the best motor position of each heliostat to achieve this. Every run, random error alignment parameters are applied to every heliostat in the differentiable environment, unseen by the agent. This way the agent has to learn to adapt to various scenarios. The reward is the average difference between each heliostat's predicted flux center distance from the target center and the effective radius

Results



- Reflected irradiance spot on the receiver moves toward the center over the time-steps during evaluation.
- MSE decreases steadily over time-steps.
- Baseline model-free RL algorithms (SAC, PPO, TD3) failed to find motor positions produce valid irradiance maps in our environment and are hence excluded from comparison.

Conclusion & Outlook

- **Proof of Concept**
  - Demonstrated differentiable RL framework for heliostat control under simplified conditions.
  - Simulator & policy limited to:
    - Single heliostat, fixed location.
    - Policy specialized to training sun position.
- **Key Results**
  - Stable convergence of policy despite simplified setup.
  - System accounts for arbitrary alignment errors in heliostat model.
  - Shows potential for calibration-free, sensorless, closed-loop control.
  - Establishes baseline for comparison with future methods.
- **Future Work**
  - Scale to multiple heliostats, variable geometries, and dynamic sun positions.
  - Move toward realistic closed-loop control for entire heliostat fields.
  - Use simplified simulator for pre-training → fine-tune on high-fidelity simulators.
  - Bridge the gap from simulation to real-world deployment.

ARTIST: Differentiable Ray tracing Environment for irradiance prediction, heliostat calibration surface prediction and more

PAINT: Open source data set with 4 years of operation data (alignment, surface, weather and more data)

DLR