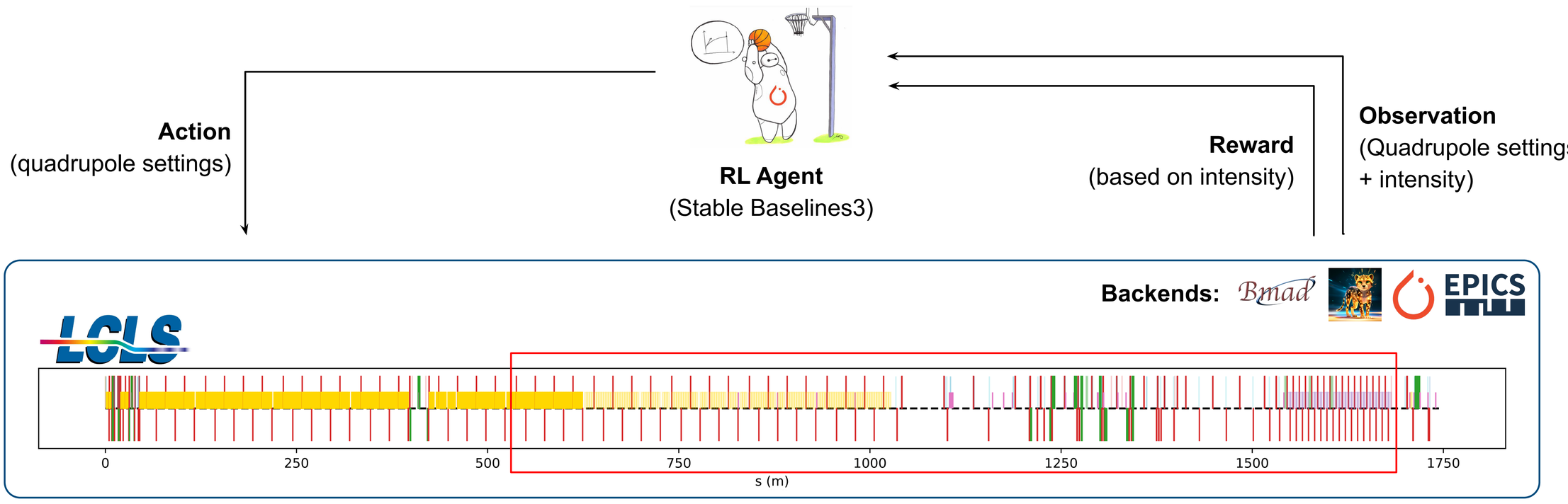
With **Cheetah** and **curriculum learning**, a reinforcement learning (RL) agent can learn to **tune 14 quadrupoles for FEL intensity**. **Gradient-based RL** enables **45x more sample-efficient** training.



Reinforcement Learning for Intensity Tuning at Large FEL Facilities.

J. Kaiser*, A. Eichler, A. Edelen, D. Ratner, M. Schram, K. Rajput

Motivation

- The FEL has to switch between energies and phase space shapes frequently, where the pulse intensity has to be optimised for each of these.
- Autonomous online tuning of the pulse intensity can facilitate faster fine tuning and switching of working points, ultimately increasing available experiment time and improving repeatability.
- Investigate tuning task of **14 quadrupoles to maximise FEL pulse intensity** at *LCLS*.
  - Starting point that can eventually be extended to 30+ quadrupoles, undulator taper, phase shifters and more actuators.
  - Solution should be able to transfer well to similar facilities, such as *European XFEL* and *FLASH*.

Gradient-free Reinforcement Learning-trained Optimisation (RLO)

- The relatively high dimensionality of the action space makes the task hard to solve. This is further amplified by the fact that the actuator space that leads to lasing is very narrow, resulting in sparse rewards. Solving this problem therefore requires a very large number of environment interactions.
- Existing simulations in *Bmad* etc. are too slow to collect the required number of samples in feasible time.
- We take various steps to improve sample efficiency and reduce the time it takes to collect each sample:
  - The ***Cheetah* simulation code** is purpose-built for fast sample collection in reinforcement learning and runs **multiple orders of magnitude faster**. This can be augmented with neural network surrogate models trained on real-world data.
  - **Curriculum learning** by increasing the **domain randomisation** ranges around the design values allows us to overcome the sparse reward setting.
- Successful training with Proximal Policy Optimisation (PPO) from *Stable Baselines3* takes 50 million samples and 1 day 16 hours of a HPC node.

Gradient-based Approach Using Cheetah

- **Cheetah natively supports automatic differentiation**, meaning we can do gradient-based reinforcement learning using the true policy gradient.
- Implementation of gradient-based policy optimisation using Cheetah and *PyTorch Lightning* reaches the same reward threshold in **45x fewer samples**.

Outlook

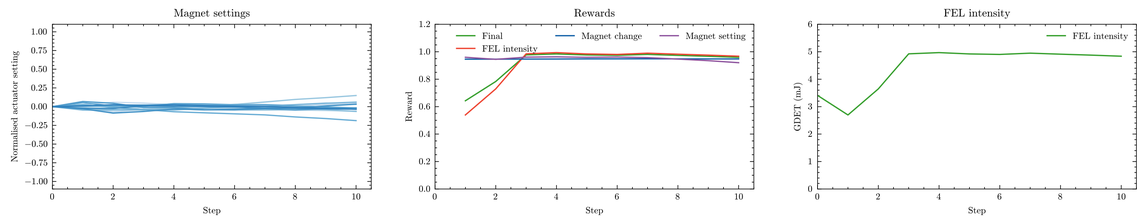
- Analyse agreement of Cheetah model to neural network surrogates trained on real-world data to facilitate zero-shot transfer to real machine or curriculum learning on different backends (Cheetah to NN surrogate to real machine).
- Compare RLO to Bayesian optimisation with neural network priors on the same task.

***Corresponding author:** jan.kaiser@desy.de

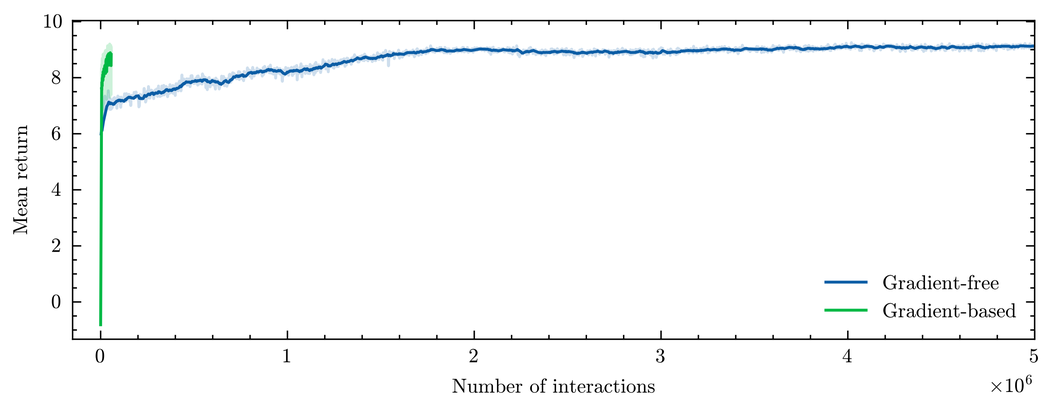
Scan me to download the poster!

Deutsches Elektronen-Synchrotron DESY
A Research Centre of the Helmholtz Association

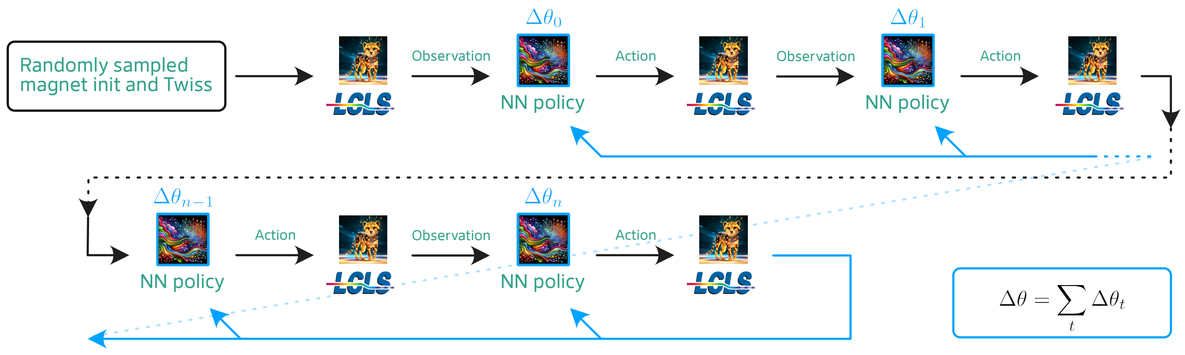
Example FEL tuning run by a trained policy



Mean return curves of gradient-free vs. gradient-based training



Gradient-based RL rollout graph



Areal views of LCLS and European XFEL

HIR³X · Jefferson Lab · SLAC NATIONAL ACCELERATOR LABORATORY · DESY.