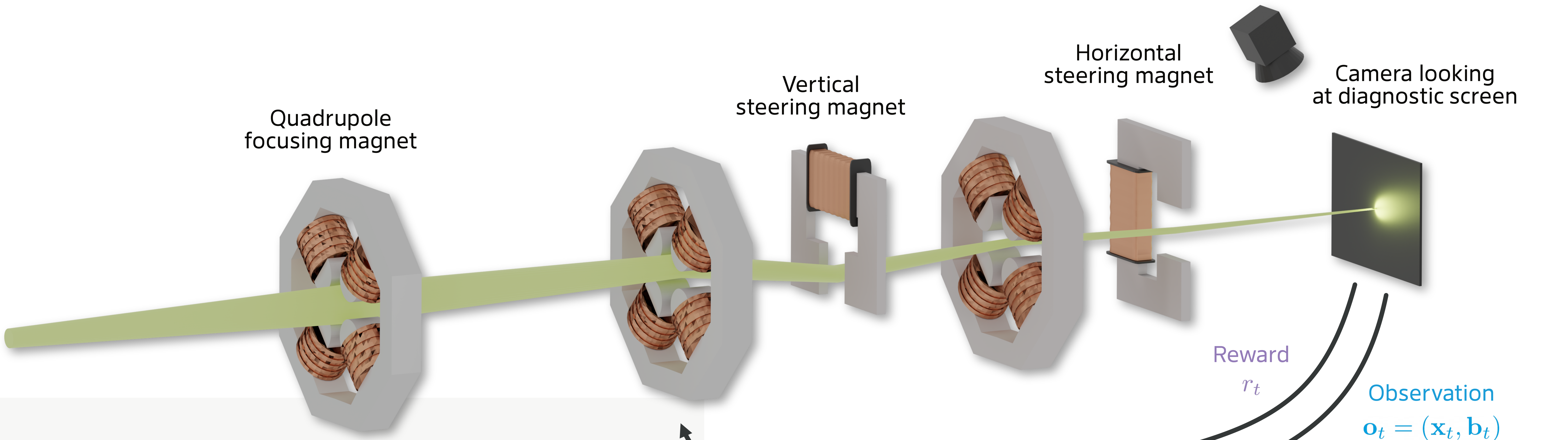
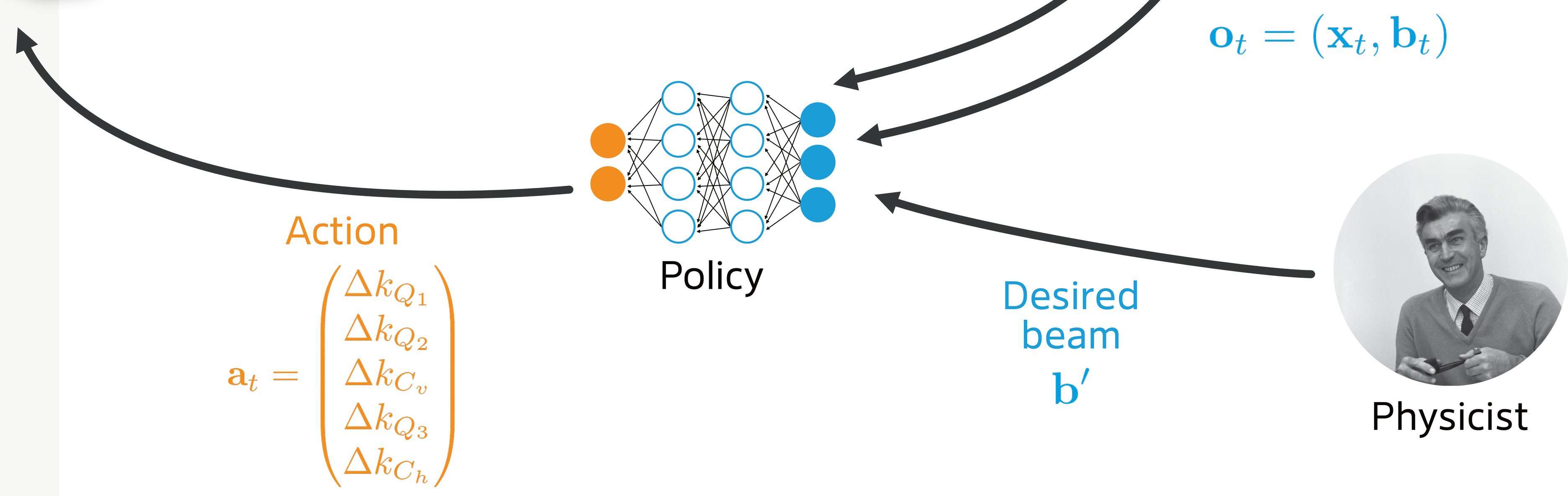
With deep **reinforcement learning**, autonomous agents can learn to **tune particle accelerators faster than human experts** and black-box optimisation algorithms.





Learning-based Optimisation of Particle Accelerators Under Partial Observability Without Real-World Training.

Jan Kaiser, Oliver Stein and Annika Eichler

Motivation

- To this day, particle accelerators are largely tuned by hand.
- The tuning of accelerators often takes up hundreds of hours each year, which are not available to experiments.
- Fast and autonomous tuning methods increase the beam time available for research in a variety of fields and improves the reproducibility of results.
- Ultimately, autonomous tuning will let physicists formulate the experiment specifications without needing to know how to realise them.

The ARES Particle Accelerator

- *ARES* is an S-band radio frequency electron linear accelerator at DESY.
- The ARES *Experimental Area* (EA) is a section right in front of an experimental chamber where many of the experiments conducted at ARES place stringent requirements on the electron beam.
- Tuning the beam to desired parameters using the magnets in the EA as well as similar tasks in other locations are a recurring and time-consuming activity in the day-to-day operation of ARES.

Method

- Train a reinforcement learning agent using the *TD3* algorithm.
- Our agent observes the magnet settings in the EA, the beam's position and shape on a diagnostic screen at the end of the EA, and desired beam parameters set by the human operator. The agent is rewarded according to how much closer the new achieved beam parameters are compared to the previous ones.
- To solve the challenges that come with this real-world reinforcement learning task, we made a number of design decisions:
  - Choosing **changes to the magnet settings as actions** makes the agent robust to noise and unobserved parts of the environment.
  - **Training on a simulation** and then deploying the trained agent to the real world avoids the use of expensive beam time and risks to machine safety.
  - **Domain randomisation** in simulation allows the agent to transfer to the real world successfully despite known shortcomings of the simulation.
  - **Including magnet settings in the observation** provides the agent with a very short history to overcome partial observability.
  - Our **logarithmic reward** prevents vanishing rewards as the achieved beam parameters near the desired beam parameters and MAE deltas become orders of magnitude smaller.

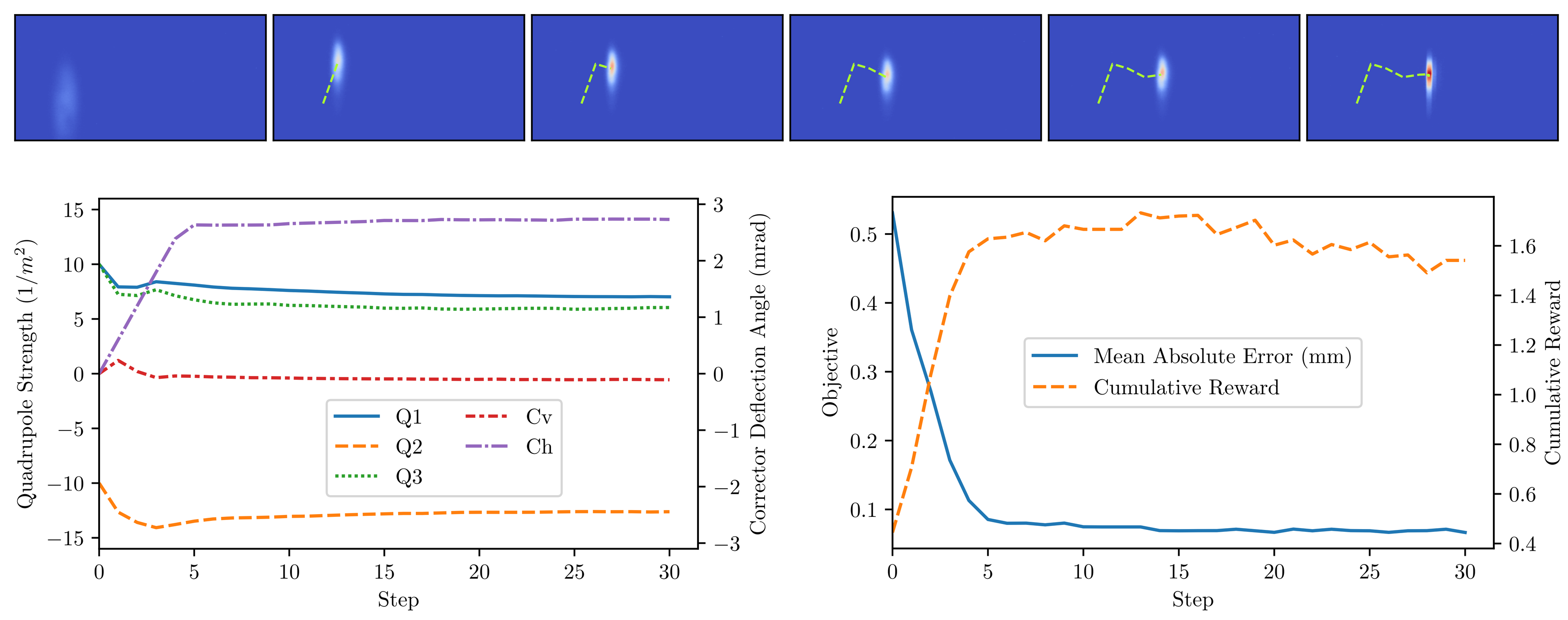
Results

- The reinforcement learning agent gets closer to the desired beam than all but one black-box optimisation algorithm we tested.
- Our agent converges around 10x faster than the only competitive algorithm.
- Despite never having seen the real accelerator in training, the agent continues to outperform other algorithms there as well.
- The agent achieves a beam competitive with that achieved by expert human operators, but does so faster and more consistently.

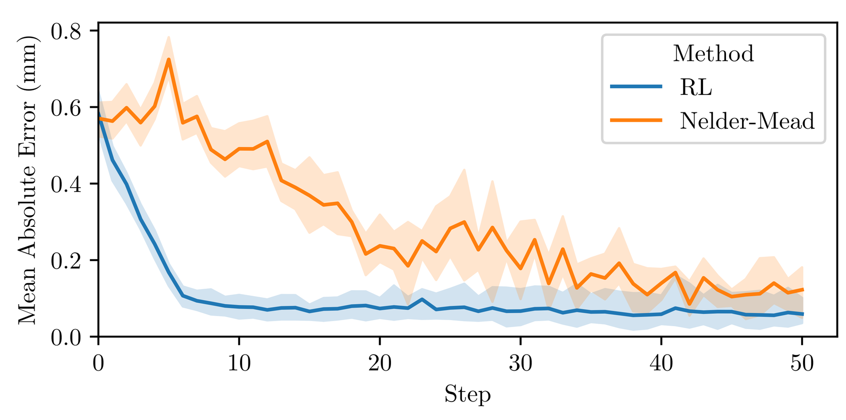
Scan me to download the full paper!

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

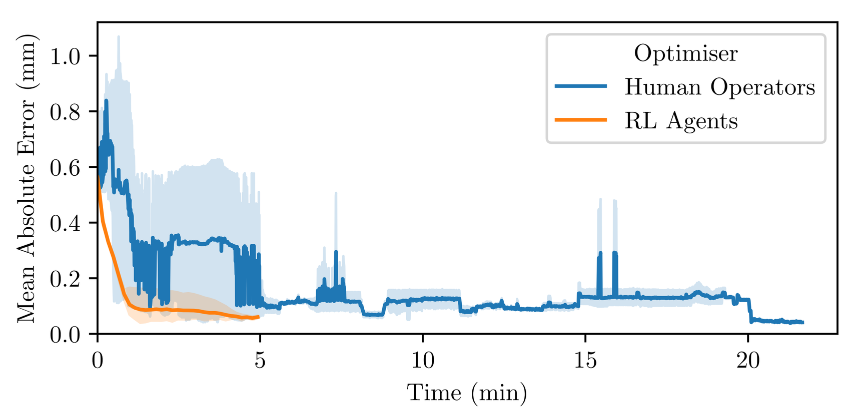
Example RL optimisation run



RL agents vs. Nelder-Mead optimisation



RL agents vs. human experts



Reward formulation

$$R(\boldsymbol{s}_t, \boldsymbol{a}_t) = \begin{cases} \hat{R}(\boldsymbol{s}_t, \boldsymbol{a}_t) & \text{if } \hat{R}(\boldsymbol{s}_t, \boldsymbol{a}_t) > 0 \\ 2 \cdot \hat{R}(\boldsymbol{s}_t, \boldsymbol{a}_t) & \text{otherwise}. \end{cases}$$

$$\hat{R}(\boldsymbol{s}_t, \boldsymbol{a}_t) = O(\boldsymbol{x}_t) - O(\boldsymbol{x}_{t+1})$$

$$O(\boldsymbol{x}_t) = \ln \sum_{p \in \boldsymbol{b}_t, p' \in \boldsymbol{b}'} w_p \left| p - p' \right|$$

RL agents vs. other algorithms

Algorithm	MAE Median (mm)	Convergence Median (Steps)
Do Nothing	1.122	0
Zero	0.588	1
FDF	0.699	1
Random	0.267	101
Powell	0.259	119
COBYLA	0.105	34
Nelder-Mead	0.007	112
Bayesian	0.081	101
Ours	0.008	7
Ours (Machine)	0.036	12

SINBAD HIR³X HELMHOLTZ DESY.