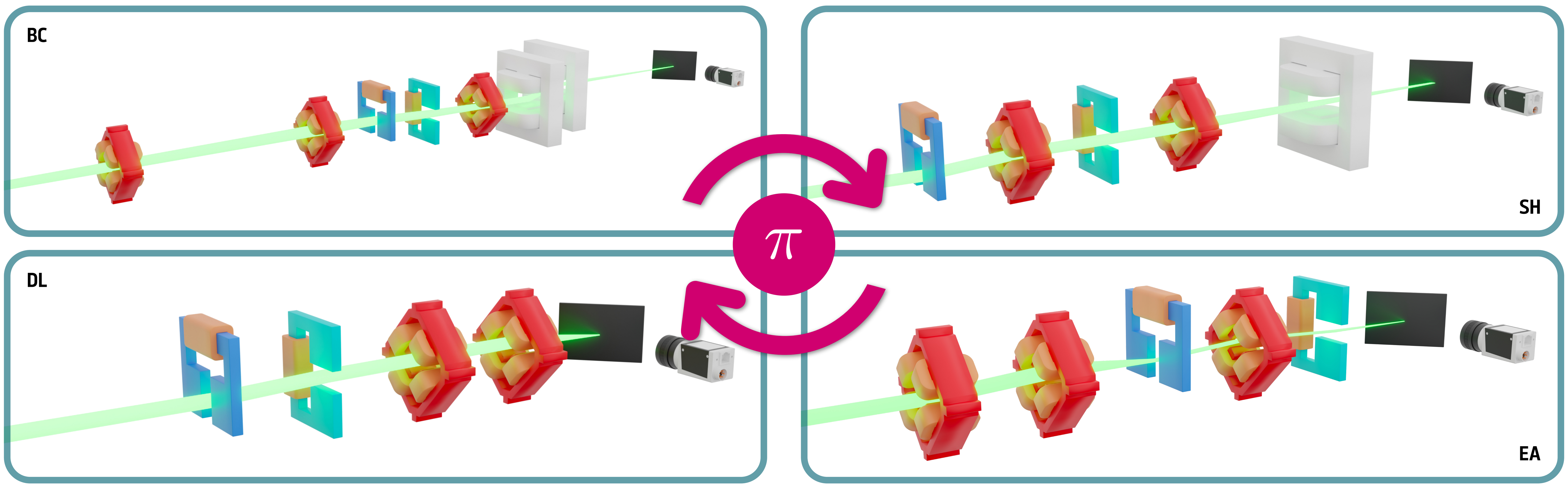
Lattice-agnostic policies for accelerator tuning can be trained with the help of **domain randomisation**. As a result, training effort can be significantly reduced by reusing policies on different lattices, because we now have **one policy to tune them all**.



Towards Lattice-Agnostic Reinforcement Learning Agents for Transverse Beam Tuning.

Jan Kaiser*, Chenran Xu, Annika Eichler, Andrea Santamaria Garcia

Motivation

- Reinforcement learning can solve accelerator tuning and control tasks that other methods are not powerful enough to solve, but training policies - even in simulation - is expensive, and trained policies are typically limited to solving one particular tuning task.
- In accelerator operations, there are many tuning and control tasks that are very similar in nature. These tasks often have the same inputs and objective, but differ in the exact device (e.g. lattice) they are performed on. With the current state of the art, one would need to train a separate policy on each of these task instances.

Case study tuning task

- We consider a transverse tuning task, where quadrupoles and dipole steerers are used to tune the transverse beam parameters on a diagnostic screen towards an arbitrary requested target.
- This task is ubiquitous across accelerator facilities. At the ARES facility at DESY alone, there are 4 locations with 4 different lattices, where this task is regularly performed. A similar section is also present at the FLUTE facility at KIT.
- Being based on PyTorch, all models built in Cheetah support automatic differentiation, making Cheetah differentiable.

Lattice-agnostic reinforcement learning

- We propose the use of **domain randomisation on the lattice** underlying the tuning task to generalise trained policies such that they work on any lattice. Two stages of domain randomisation are studied:
  - **Geometry randomisation:** A common lattice order of (Q, Q, Q, Cv, Ch, S) of 3 quadrupoles followed by two steerers and the screen is assumed. Domain randomisation is introduced by sampling random lengths for the drift sections in between them at the start of each training episode.
  - **Order randomisation:** The geometry-randomised approach is extended by randomly choosing the ordering of the three quadrupoles and two steerers at the start of each training episode.

Results

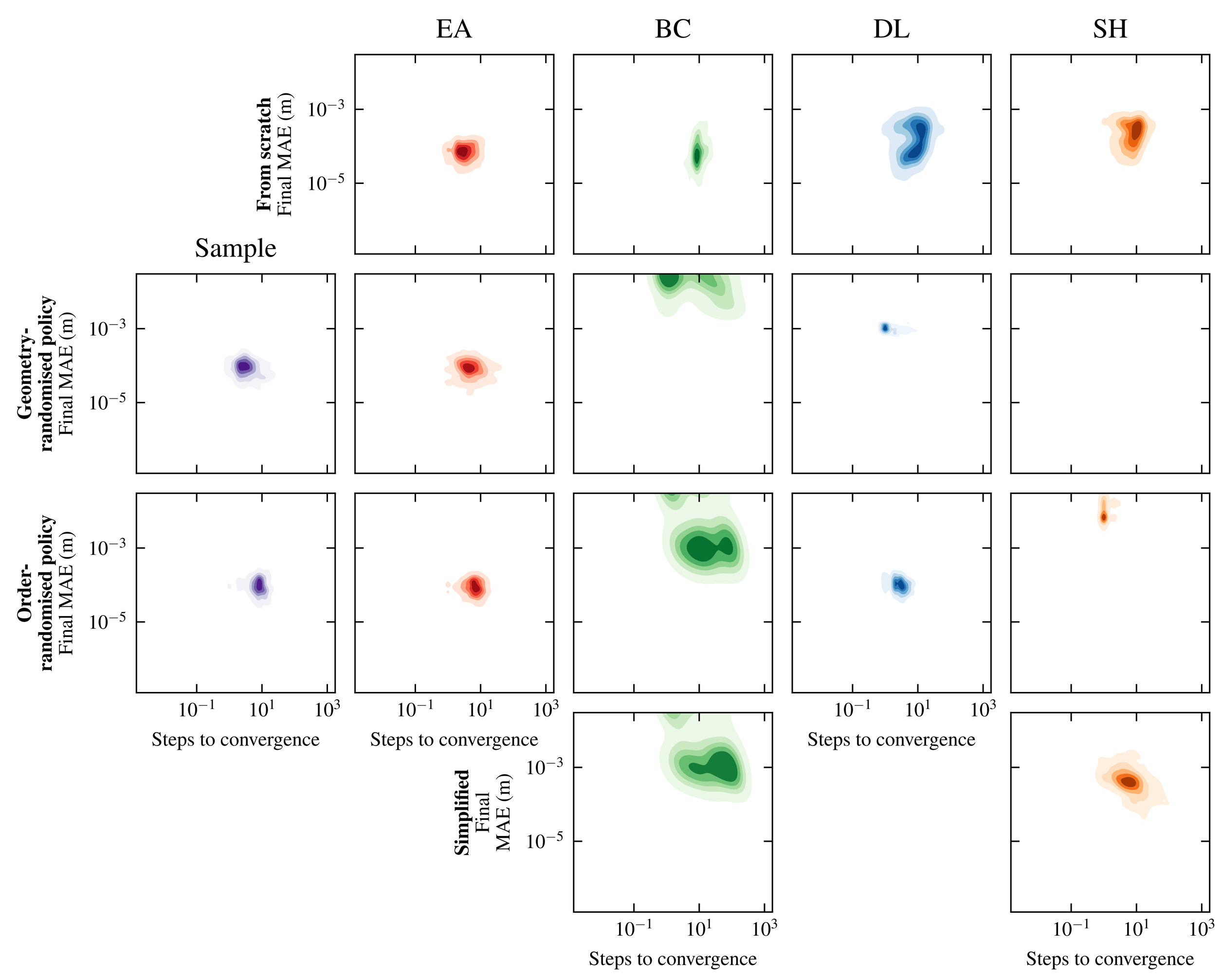
- Policies trained to be lattice-agnostic match and often even outperform specialised policies.
- There is some robustness to out-of-training samples when they are of the same nature, e.g. larger drift length, but less robustness when the violation is of a different nature, e.g. additional elements like bends or apertures being present in the beam line.

***Corresponding authors:** jan.kaiser@desy.de | chenran.xu@kit.edu

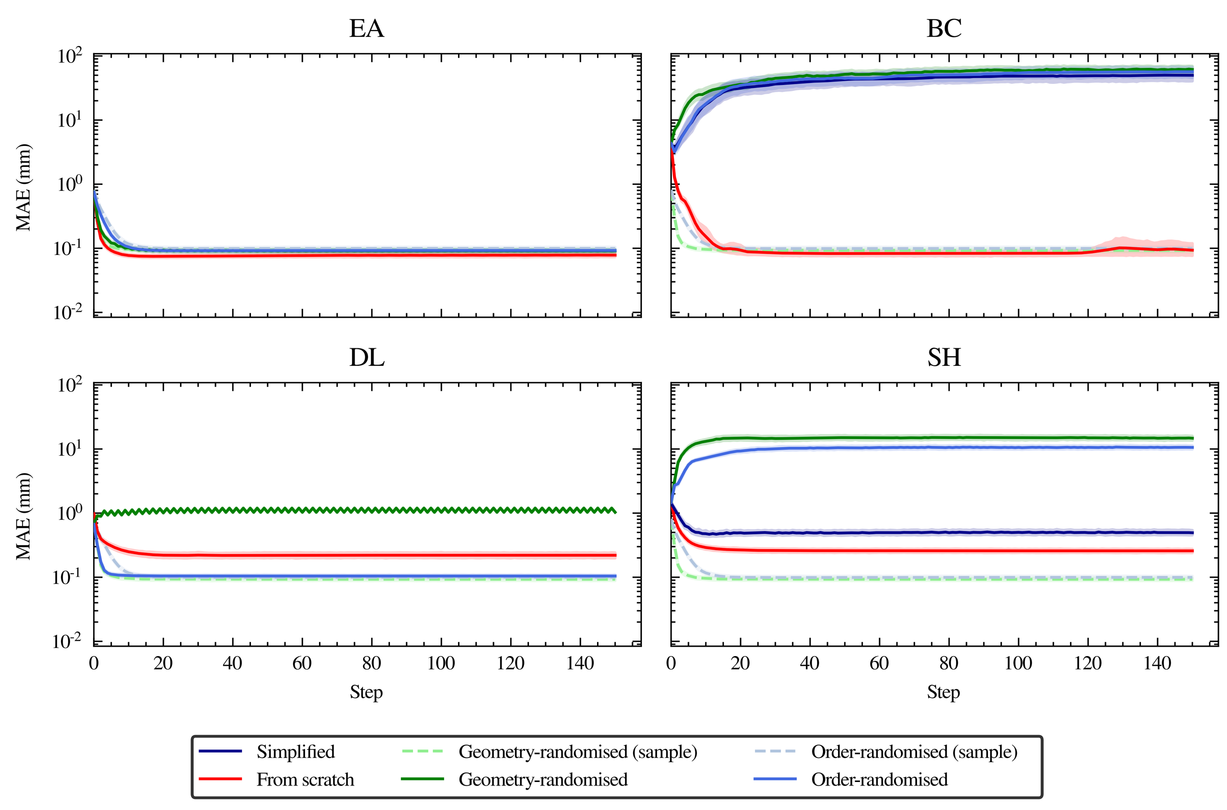
Scan me to download this poster!

Deutsches Elektronen-Synchrotron DESY
Karlsruhe Institute of Technology KIT
Research Centres of the Helmholtz Association

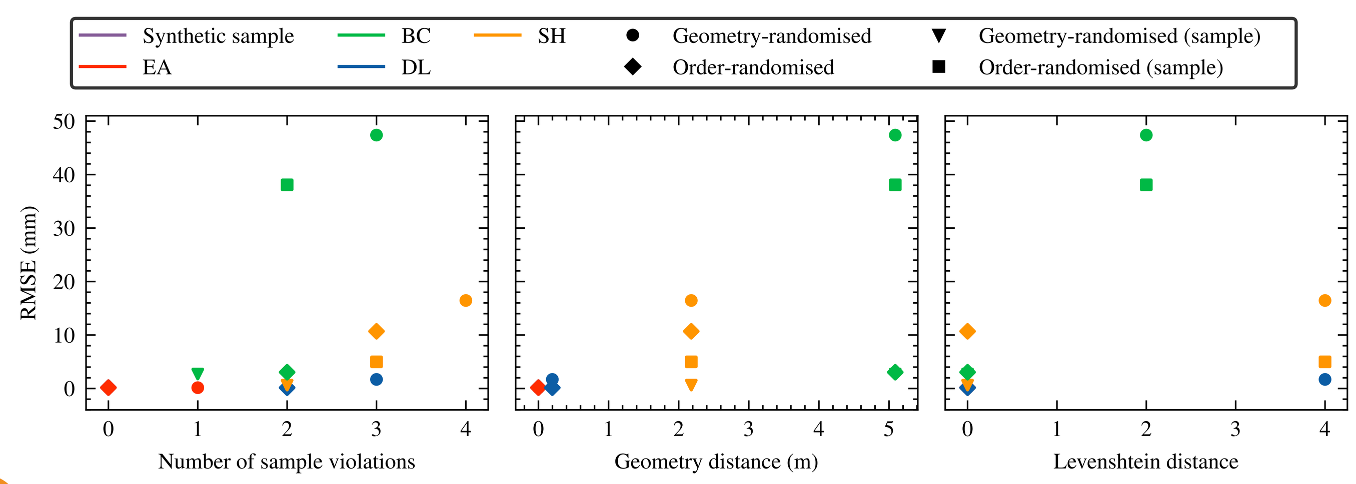
Lattices for geometry-randomised training and real test instances



Objective convergence by different policies on different tasks



Correlation between out-of-distribution metrics and test-time performance



SINBAD HIR³X HELMHOLTZAI RL4AA Karlsruher Institut für Technologie DESY.