Physics-Based Deep Neural Networks for Beam Dynamics in Charged Particle Accelerators

AMALEA Helmholtz Innovation Pool project

Andrei Ivanov
Hamburg, 17.12.2020

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

Overview

01 ML for High-Level Control in PETRAIV

02 Theory of the developed model

- From Taylor maps for ODEs to deep neural networks (TM-PNN)
- Regularization when learning dynamics with small datasets

03 Experiments on PETRAIII

- Orbit correction
- Tune recovering
- Optics measurement (data analysis in progress)

04 Other applications

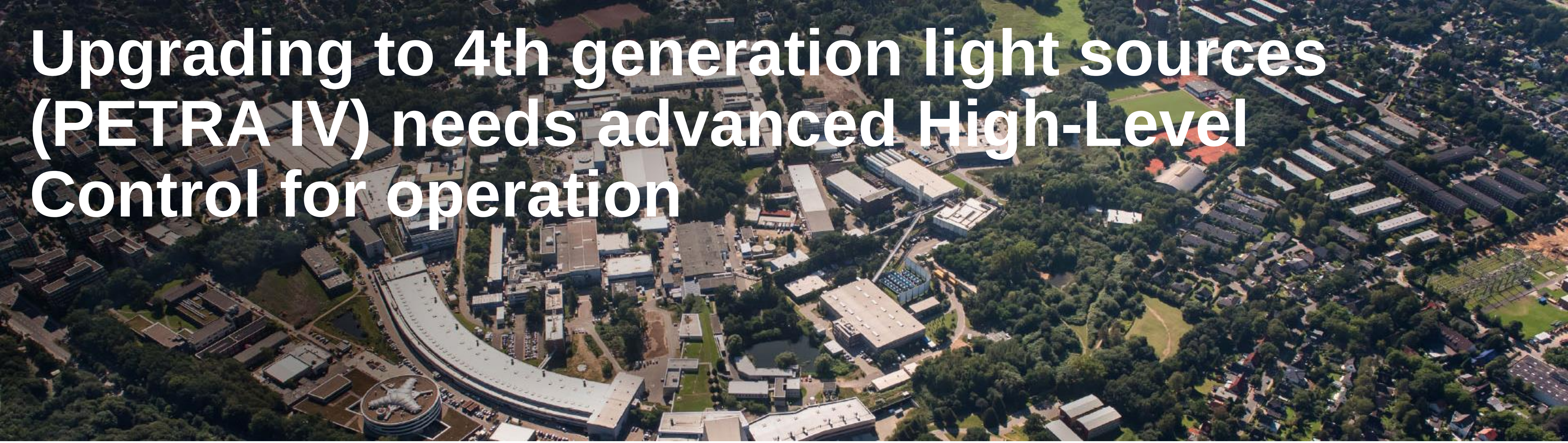
- RL for transmission problem
- UCI datasets (beyond accelerators)

04 Next steps

- Optimizations of the TM-PNN
- Optics measurement data processing
- Agent-based architecture for experiments and operation

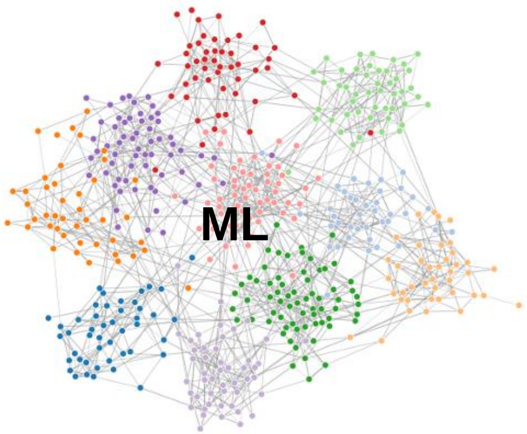
01 ML for High-Level Control in PETRAIV

Upgrading to 4th generation light sources (PETRA IV) needs advanced High-Level Control for operation



While PETRAIII is stable in operation and don’t require ML, the 4th generation rings (PETRAIV) faces certain difficulties:

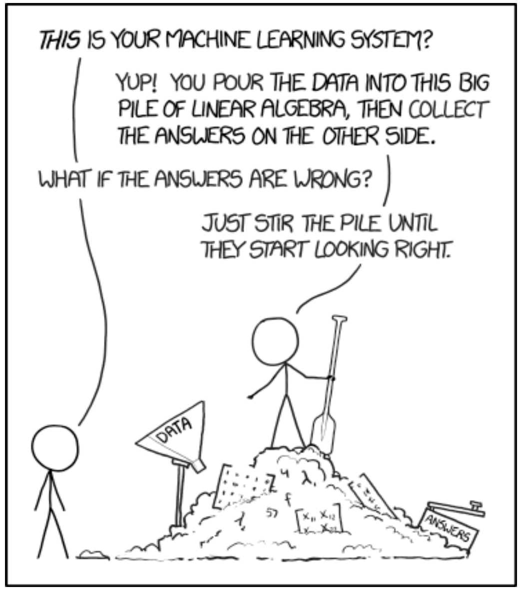
- The resolution gap between 1-10 nanometers
- Reliability demands grow (95% -> 99%)
- Machines are more sensitive with larger number of components
- High nonlinearities



Big Data is all about finding correlations

The model will be only as good or as bad as the data you have

- well-posed problem
- high performance computing
- ubiquitous data



https://xkcd.com/1838/

In the real-world, application of ML is more difficult than in research:

- learning on the real system from **limited samples**
- **high-dimensional** continuous state and action spaces.
- **safety constraints** that should never or at least rarely be violated
- tasks that may be **partially observable**, alternatively viewed as non-stationary or stochastic
- system operators who desire **explainable** policies and actions
- inference that must happen in **real-time** at the control frequency of the system

02 Theory of the developed model

- From Taylor maps for ODEs to deep neural networks
- Regularization when learning dynamics with small datasets

Novel approach for constructing deep neural networks for beam dynamics

with the following key features:

- accurate simulation of dynamics **without training**
- model **fine-tuning** with limited measurements

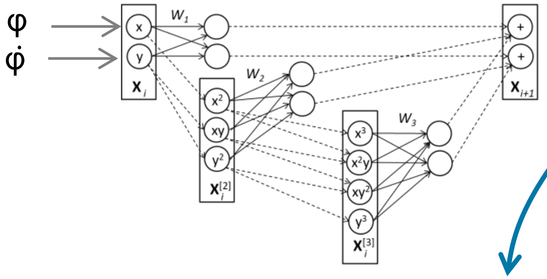
Novel approach for constructing deep neural networks for beam dynamics

with the following key features:

- accurate simulation of dynamics **without training**
- model **fine-tuning** with limited measurements

The key idea: If the dynamics of a system approximately follows a given differential equation, the Taylor mapping technique can be used to initialize the weights of a polynomial neural network

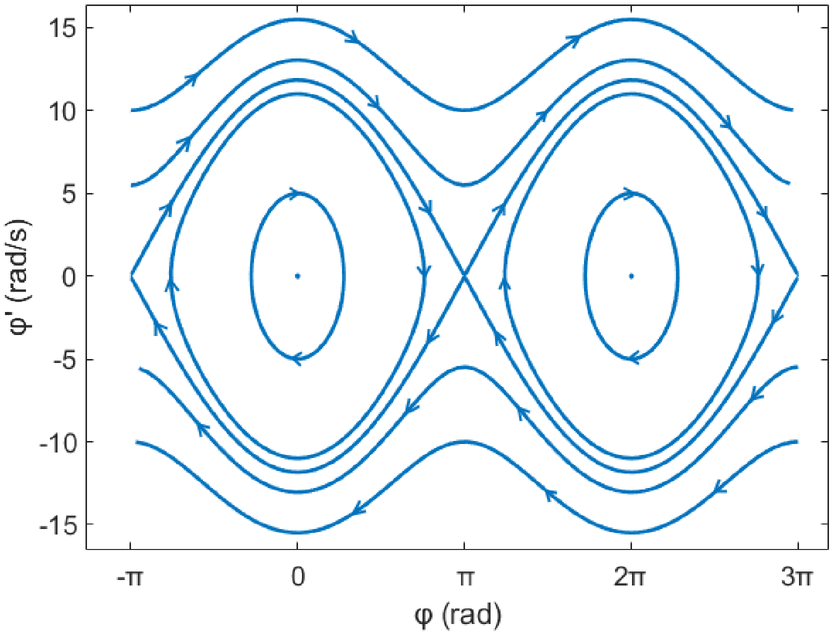
Pendulum oscillation: $\ddot{\varphi} = -\omega^2 \sin\varphi$



$$W_1 = \begin{pmatrix} 1 & 0.099 & -7.64E{-}06 & 0 \\ 0 & 1.000 & -1.54E{-}04 & 0 \end{pmatrix}$$

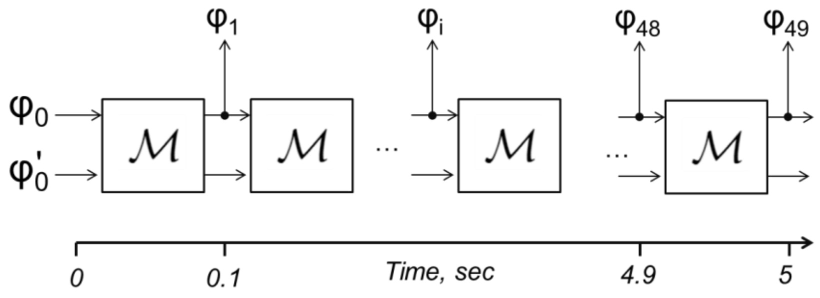
$$W_2 = \begin{pmatrix} 0\ 0\ 0\ 0\ 0\ 0\ -1.58E{-}05\ 0\ 6.11E{-}10\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ -4.80E{-}04\ 0\ 2.47E{-}08\ 0 \end{pmatrix}$$

$$W_3 = \begin{pmatrix} 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 2.46E{-}05\ 0\ -2.28E{-}09\ 0\ 7.61E{-}10\ 5.87E{-}14\ 0\ -1.95E{-}14\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 9.95E{-}04\ 0\ -1.15E{-}07\ 0\ 3.84E{-}08\ 3.56E{-}12\ 0\ -1.18E{-}12\ 0 \end{pmatrix}$$

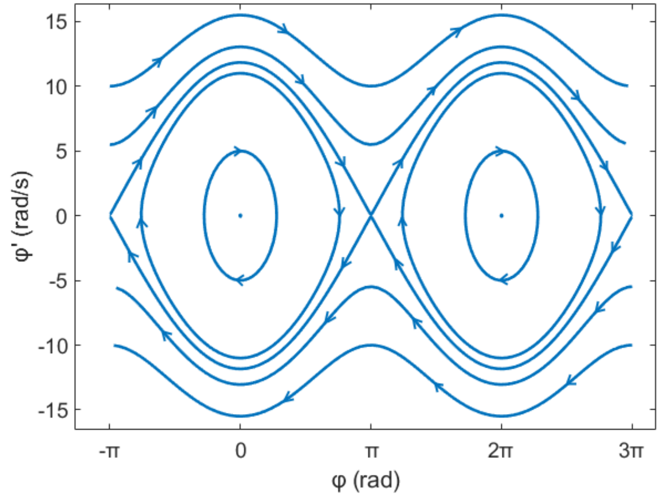


Novel approach for constructing deep neural networks for beam dynamics

fine-tuning of the NN with one or a few training samples

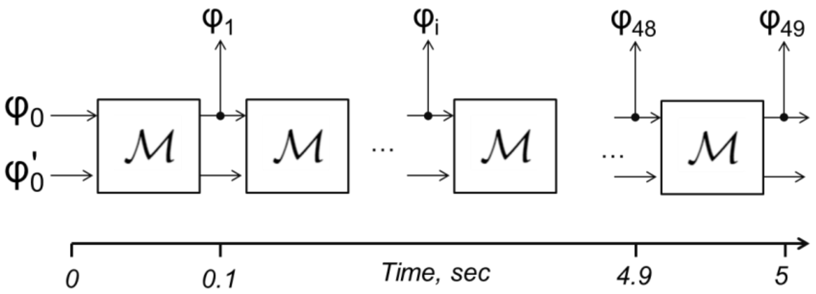


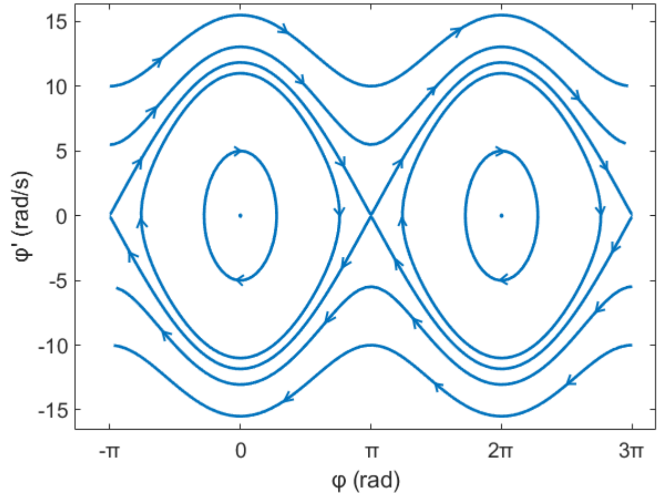
ideal pendulum



Novel approach for constructing deep neural networks for beam dynamics

fine-tuning of the NN with one or a few training samples

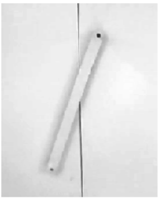


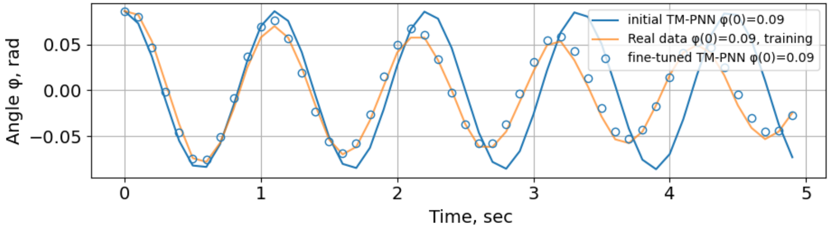


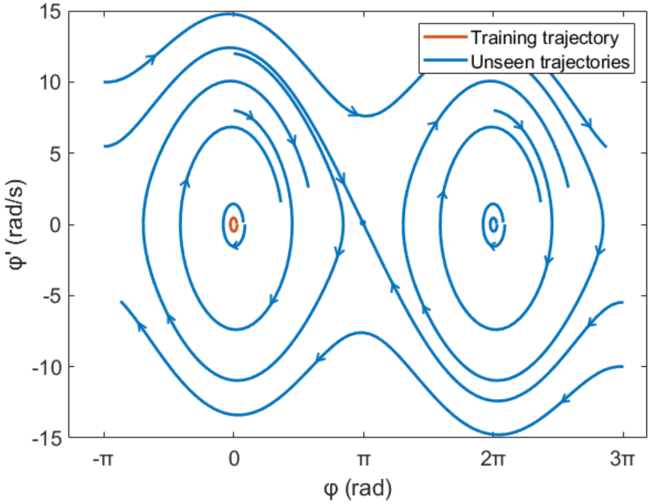
ideal pendulum

+

one measurement of dumped oscillations

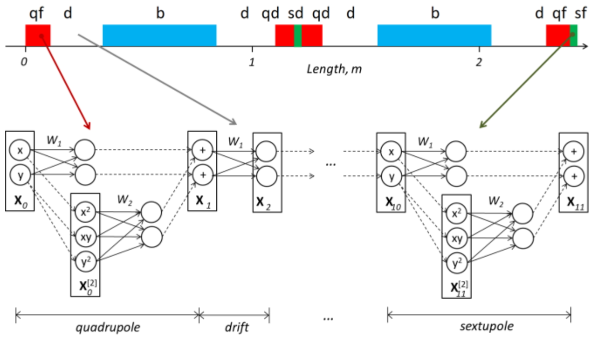


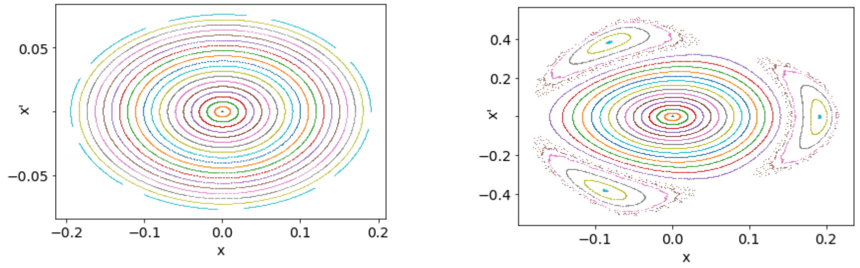




Translating lattices of the storage rings into deep NN

FODO: neural network with 12 layers represents **resonance**

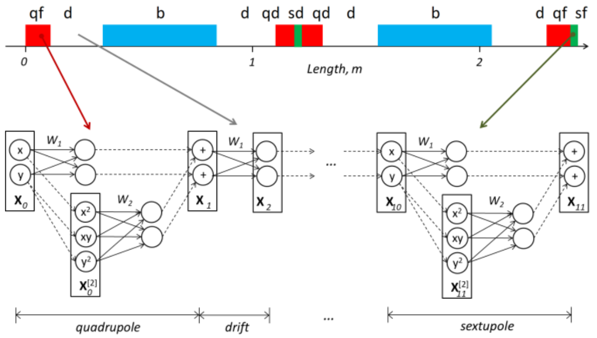


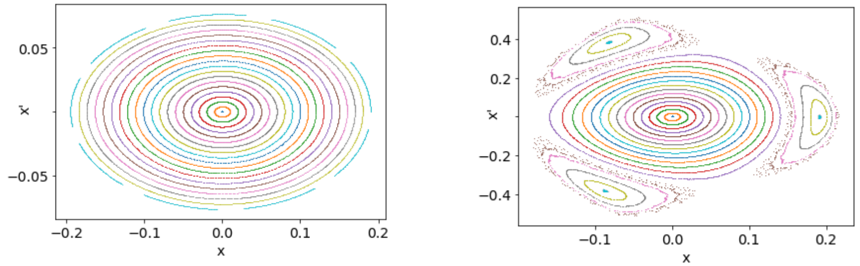


Initialized NN accurately represents the parametric dependency of dynamics on magnet strength, such as the appearance of a third-integer resonance

Translating lattices of the storage rings into deep NN

FODO: neural network with 12 layers represents **resonance**

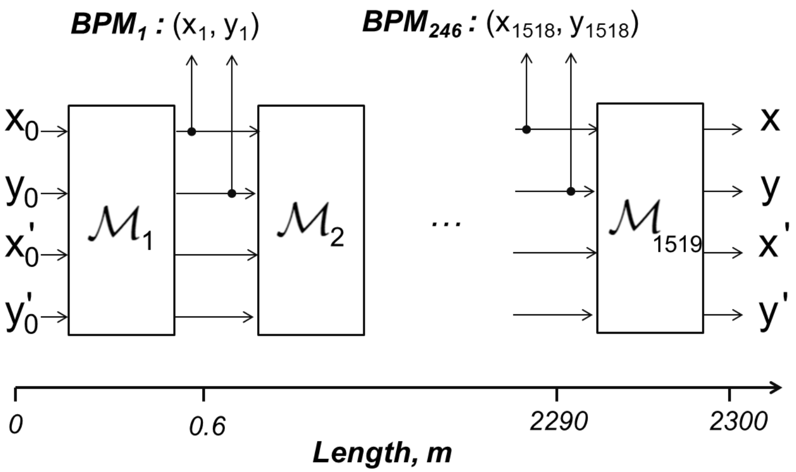


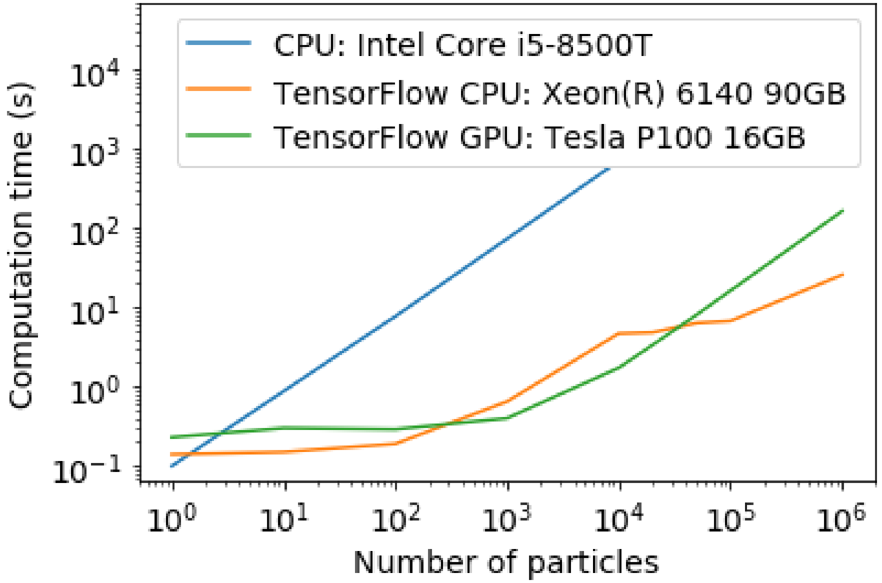


Initialized NN accurately represents the parametric dependency of dynamics on magnet strength, such as the appearance of a third-integer resonance

PETRAIII: deep neural network with **1519 layers** represents ideal lattice with fair accuracy

- 2,3 km length with **1519** magnets
- **210** horizontal and **194** vertical correctors
- **246** BMPS





Robustness to noise and physical consistency

Loss function for data-driven training contains a regularization term and avoid overfitting

$$Loss = \underbrace{\sum_{i=0}^{246} ||\mathbf{X}(0)_i^{\text{TM-PNN}} - \mathbf{X}(0)_i^{\text{BPM}}||}_{\text{depends on training data and weights}} + \lambda \underbrace{\sum_{j=0}^{1519} S(W_1^j, W_2^j)}_{\text{depends only on weights}},$$

Regularization aims to reduce the number of free parameters (weights) of the NN to ovoid overfitting. The traditional methods (L1-L2 norms) do not reflect physics and just try to reduce the absolute magnitude of the weights during training.

$$W_1 = \begin{pmatrix} 1 & 0.099 & -7.64\text{E}{-06} & 0 \\ 0 & 1.000 & -1.54\text{E}{-04} & 0 \end{pmatrix}$$

To handle problem of limited observations we implemented special regularization methods

Symplectic regularization

For Hamiltonian systems representing single-particle beam dynamics, the symplectic property can be used. The Hamiltonian structure of each layer is preserved for all new inputs which has a large impact on generalization.

$$W_1 = \begin{pmatrix} {w_1}^{11} & {w_1}^{12} \\ {w_1}^{21} & {w_1}^{22} \end{pmatrix}, \quad W_2 = \begin{pmatrix} {w_2}^{11} & {w_2}^{12} & {w_2}^{13} \\ {w_2}^{21} & {w_2}^{22} & {w_2}^{23} \end{pmatrix} \xrightarrow{\text{symplectic property}}$$

$$w_1^{11}w_1^{22} - w_1^{12}w_1^{21} - 1 = 0,$$
$$w_2^{11}w_2^{23} - w_2^{13}w_2^{21} = 0,$$
$$w_2^{12}w_2^{23} - w_2^{13}w_2^{22} = 0,$$

$$w_1^{11}w_2^{22} - w_1^{21}w_2^{12} + 2w_1^{22}w_2^{11} - 2w_1^{12}w_2^{21} = 0,$$
$$w_1^{22}w_2^{12} - w_1^{12}w_2^{22} + 2w_1^{11}w_2^{23} - 2w_1^{21}w_2^{13} = 0.$$

QUBO-based regularization (quadratic unconstrained binary optimization)

Since physical systems that are described by ODEs often lead to sparse weights, this should be preserved during training:

$$W_2 = \begin{pmatrix} \boxed{0\ 0\ 0\ 0\ 0\ 0} & -1.58\text{E}{-05} & \boxed{0} & 6.11\text{E}{-10} & \boxed{0} \\ 0\ 0\ 0\ 0\ 0\ 0 & -4.80\text{E}{-04} & 0 & 2.47\text{E}{-08} & 0 \end{pmatrix} \xrightarrow{\text{QUBO problem}}$$

Problem: fit weights with data and maximize number of zero elements

Solution: combinatorial problem that can be solved with Quantum Annealers

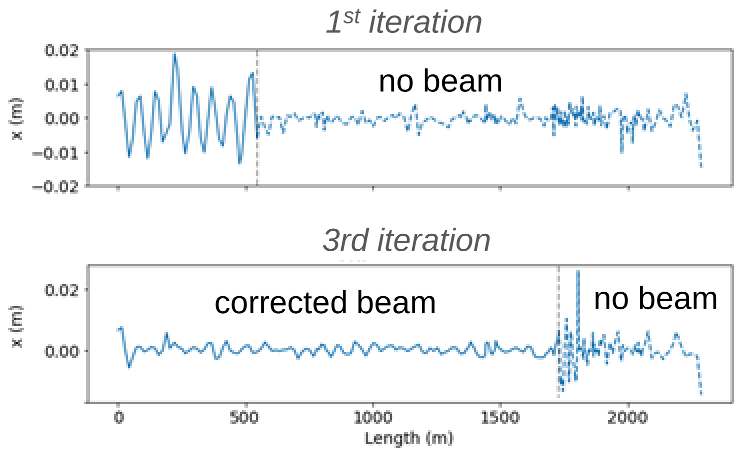
02 Experiments on PETRAIII

- Orbit correction
- Tune recovering
- Optics measurement (data analysis in progress)

One-shot learning of PETRAIII in experiments

Beam threading

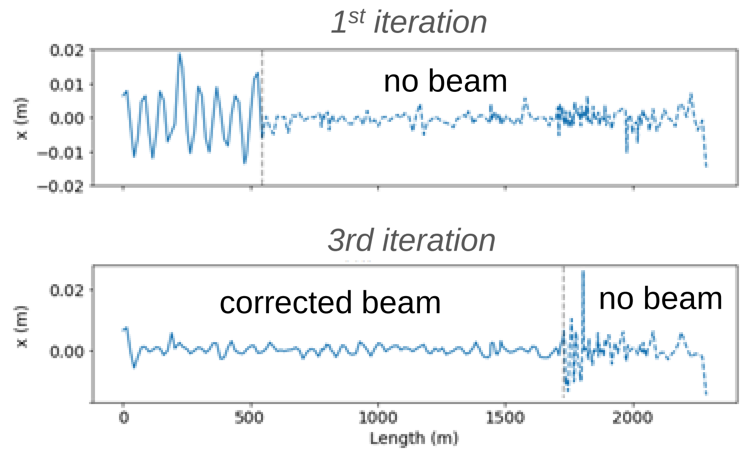
1. All corrector magnets are switched off
2. Beam is able to travel through only a part of the ring
3. **Neural Network predicts** an optimal control policy for beam propagation



One-shot learning of PETRAIII in experiments

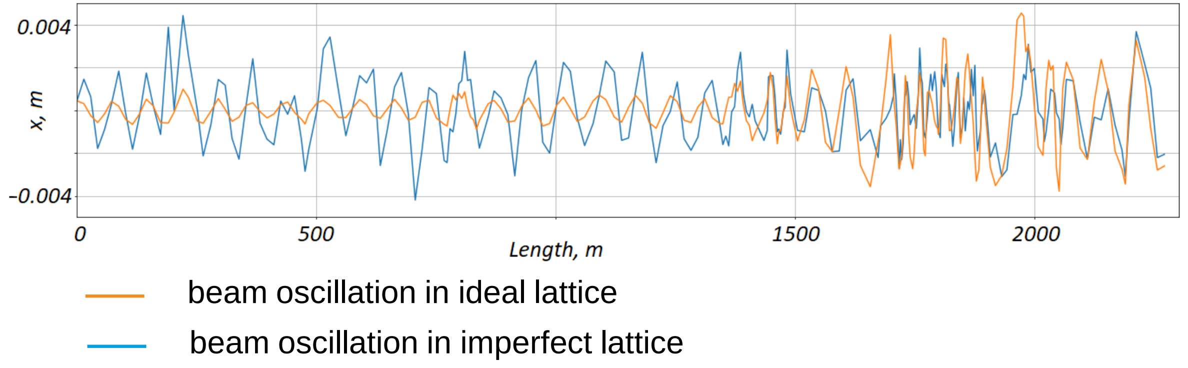
Beam threading

1. All corrector magnets are switched off
2. Beam is able to travel through only a part of the ring
3. **Neural Network predicts** an optimal control policy for beam propagation

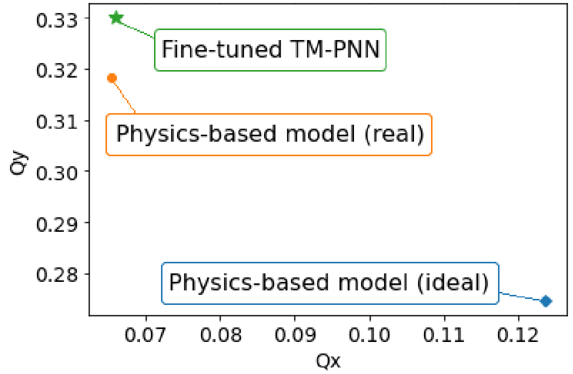


Tune recovering

1. Tune is the main multi-turn frequency of beam oscillation in the storage ring
2. The affected magnets cause the tune change from the designed values.
3. **Neural Network** is trained with only a **single-turn measurement** and estimates tunes with 95% accuracy.

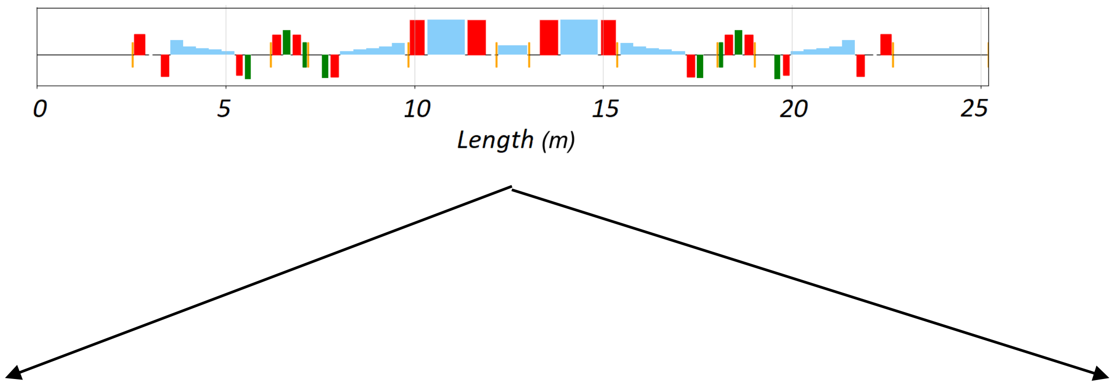


training NN with a single-turn measurements

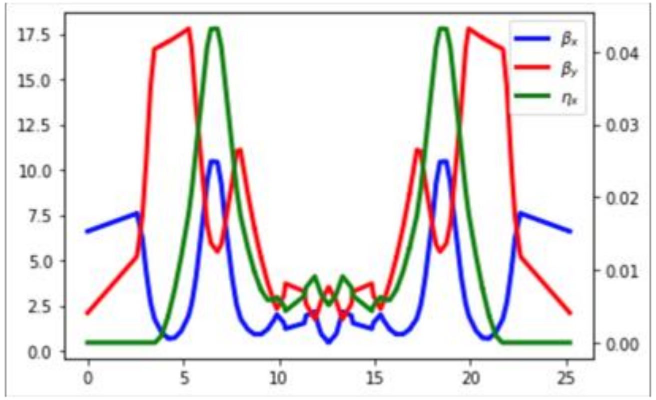


95% accuracy of the multi-turn prediction

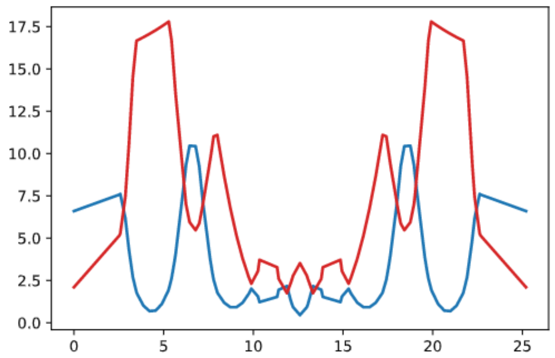
Optics measurement



Elegant



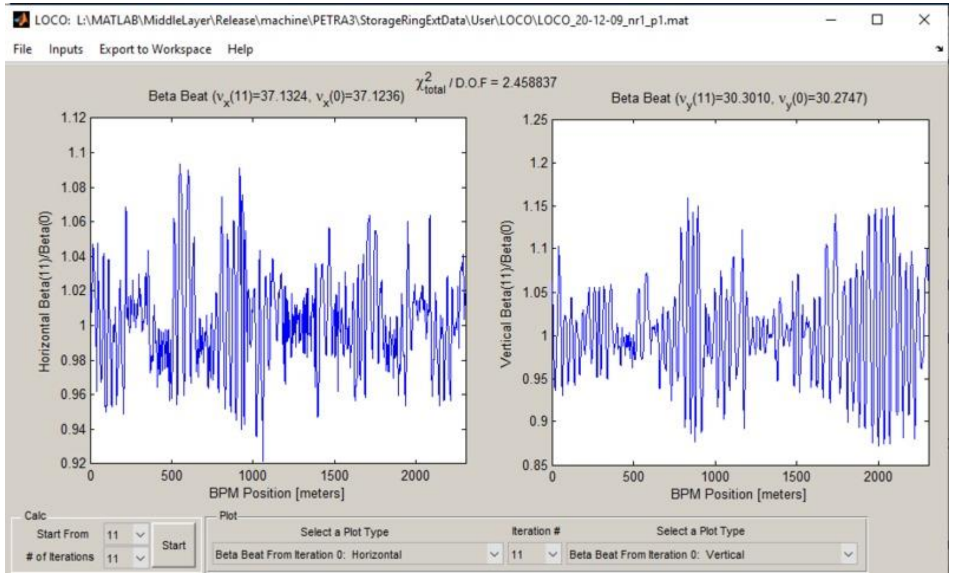
NN in TensorFlow



Optics measurement on PETRA III

LOCO as a benchmark

Linear response matrix
measurement for closed optics



Fine-tuning of the proposed NN architecture

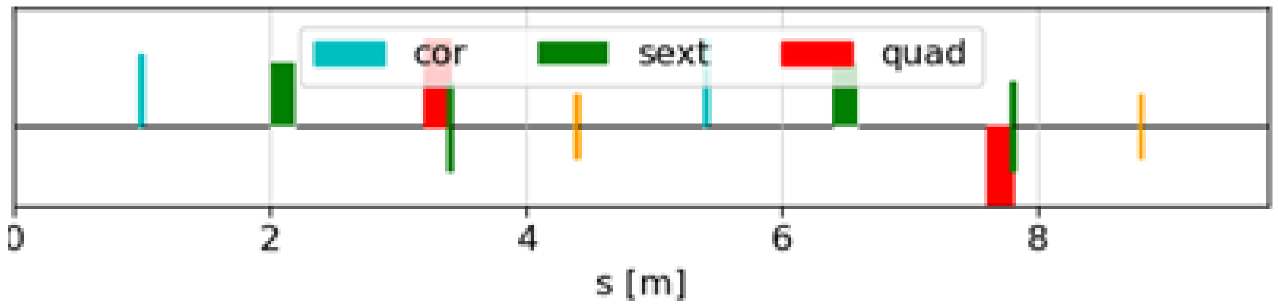
Measurements on December 9, 2020 are not analyzed yet.

04 Other applications

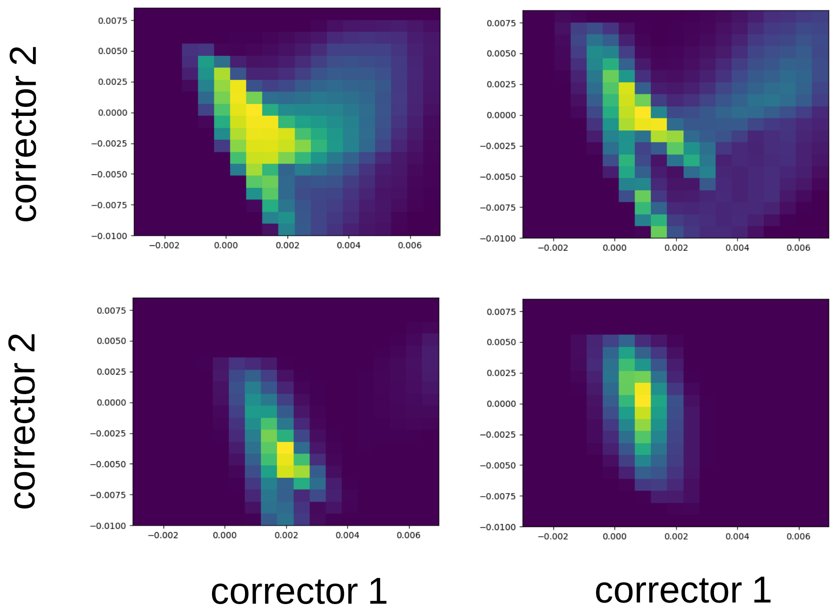
- RL for transmission problem
- UCI datasets (beyond accelerators)

Problem formulation

beam transmission: 2 actuators (correctors), 1 objective, sextupoles and apertures

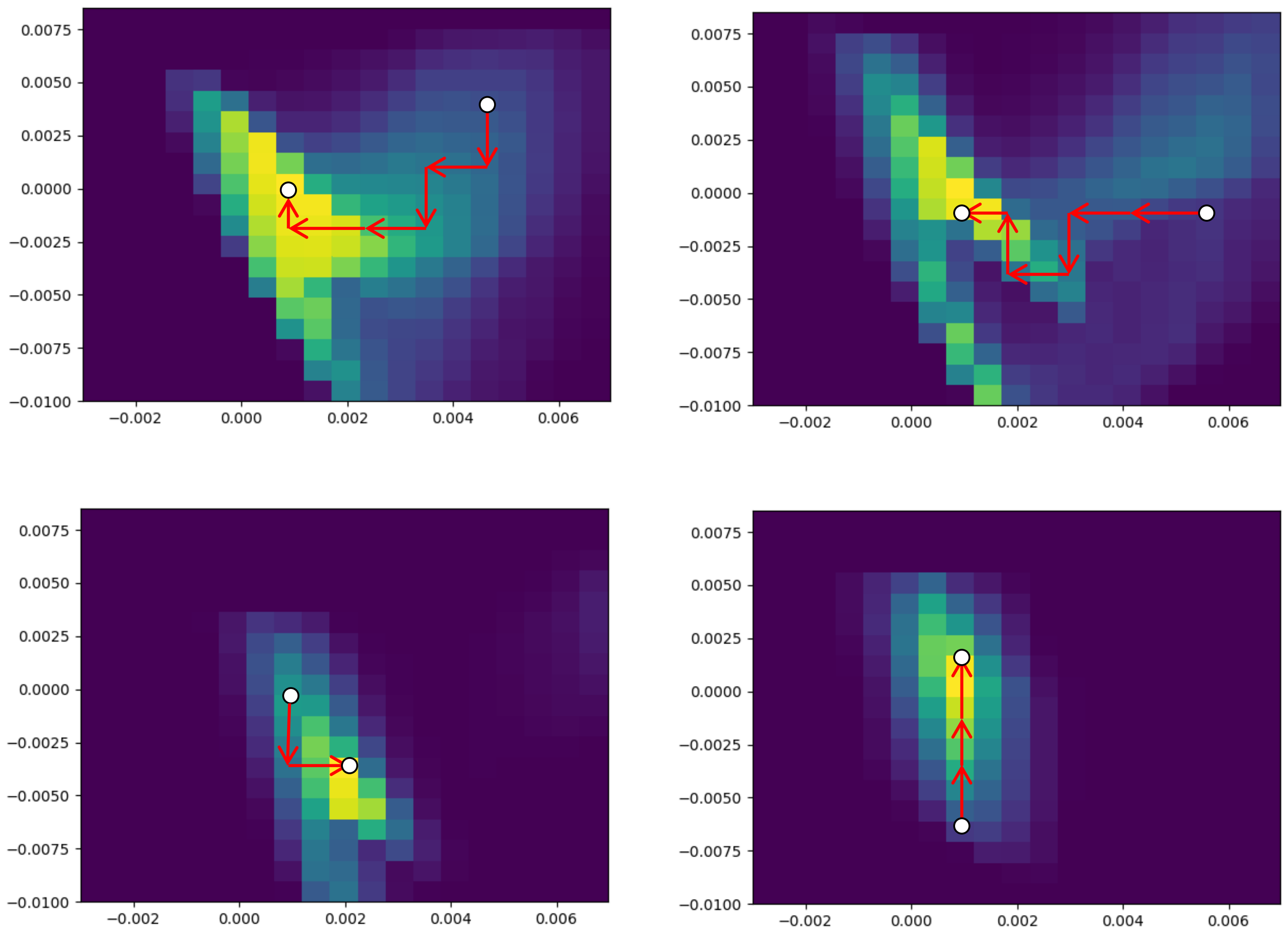


nonlinear response concerning the random misalignments of magnets



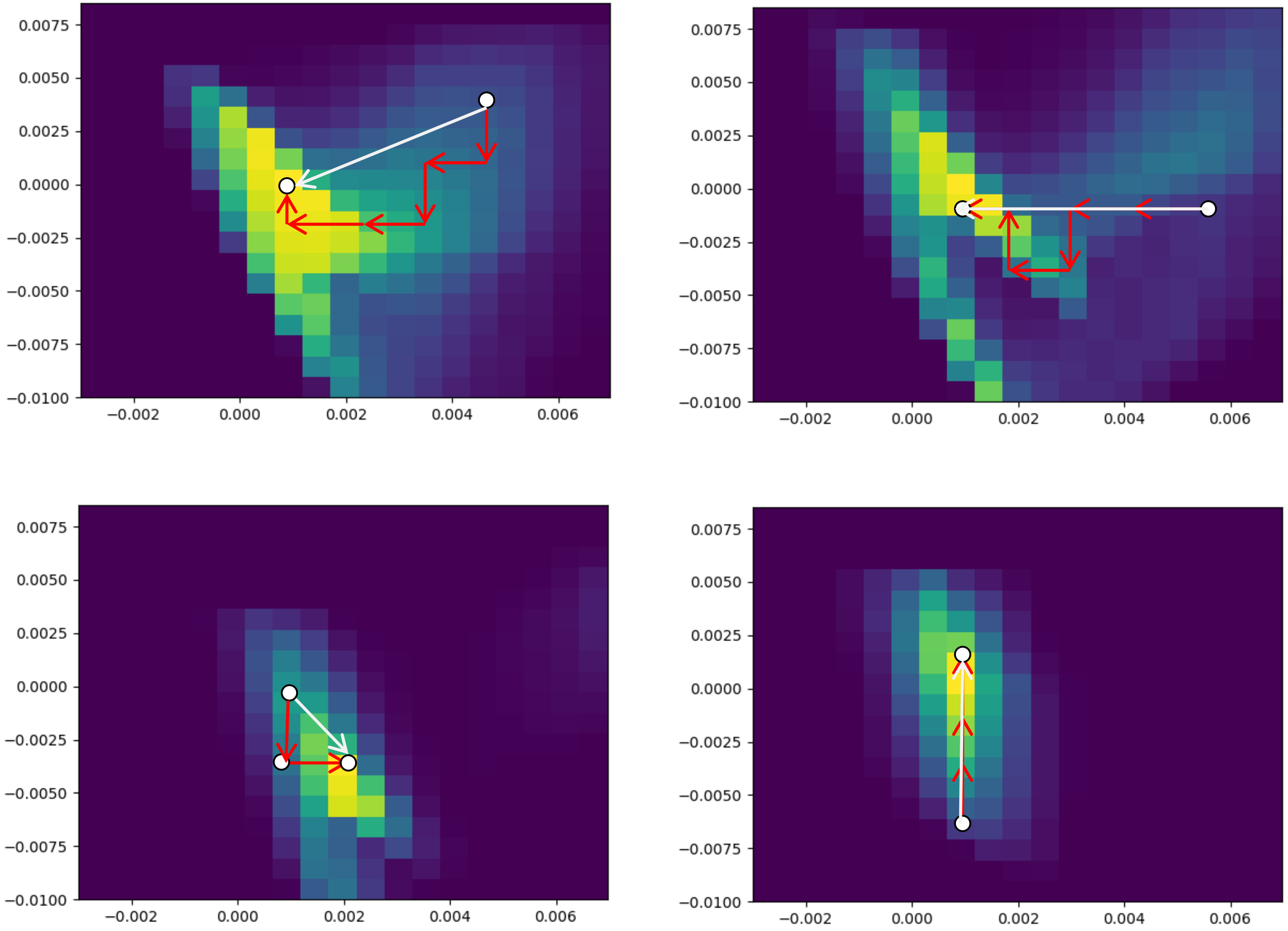
Numerical optimization

using traditional optimizers one can iteratively find out optimal corrector’s values



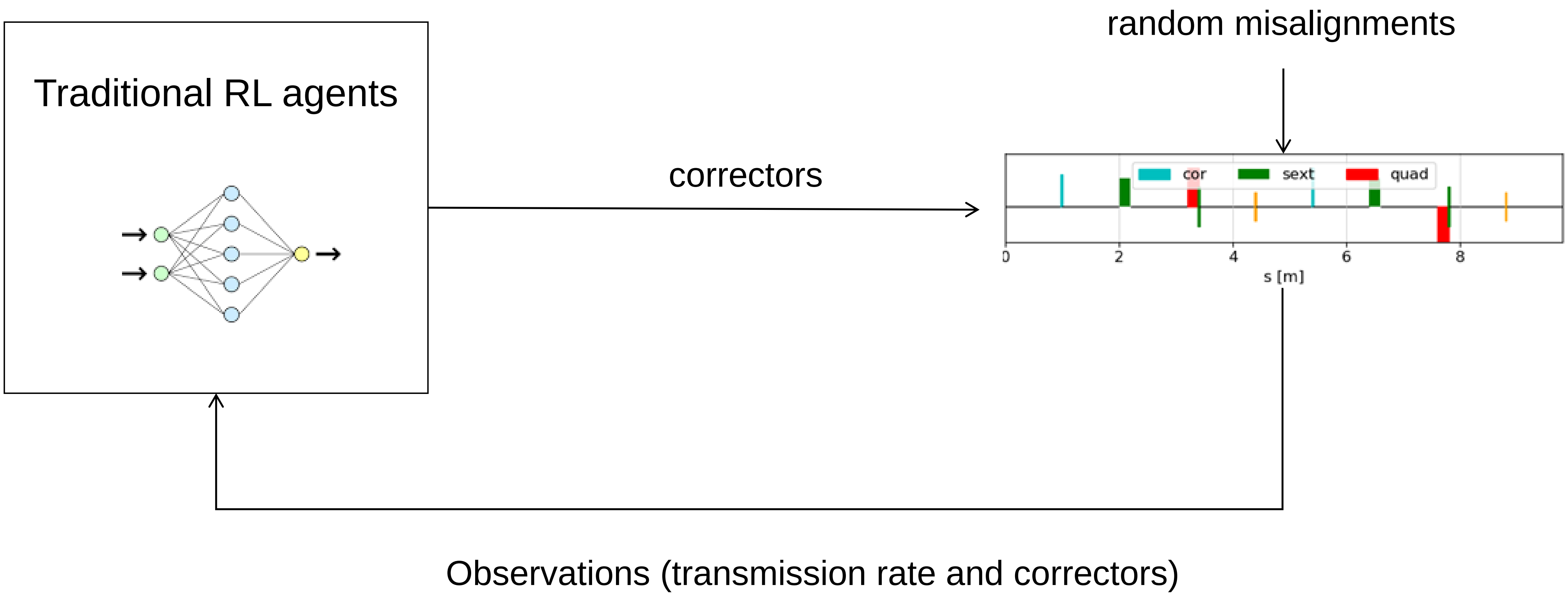
Numerical optimization

using traditional optimizers one can iteratively find out optimal corrector’s values



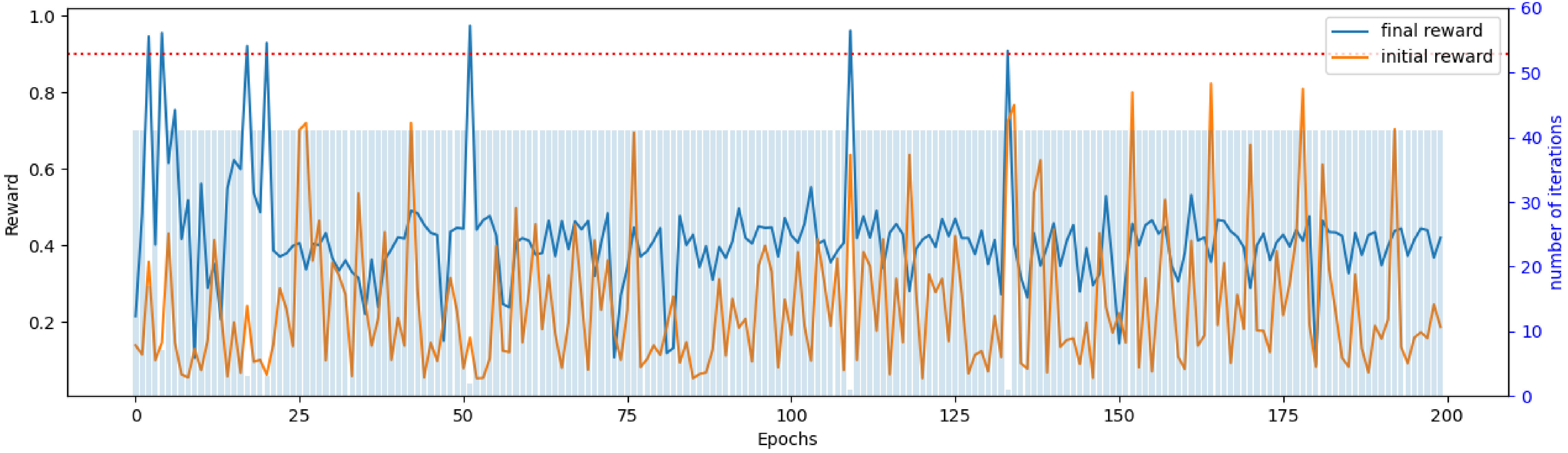
RL for control

NN is trained with historical data and learns an optimal policy



RL for control

It is hard to achieve meaningful results with black-box models

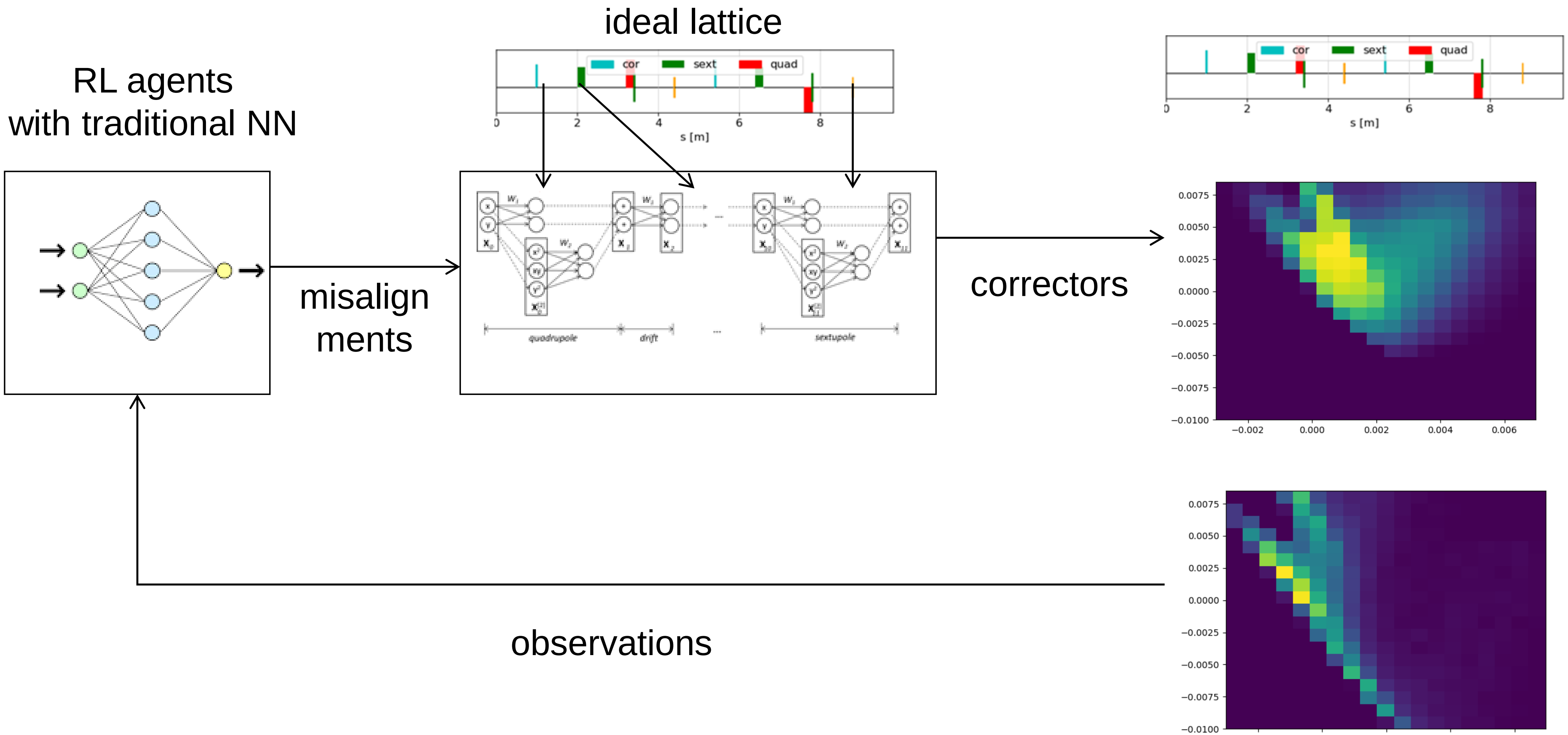


During each epoch NN is trained with simulated data for the given random misalignments and tries to maximize initial state (orange line). After max. 40 iterations the procedure begins again for new random misalignments.

RL for control enhanced by physics-based NN

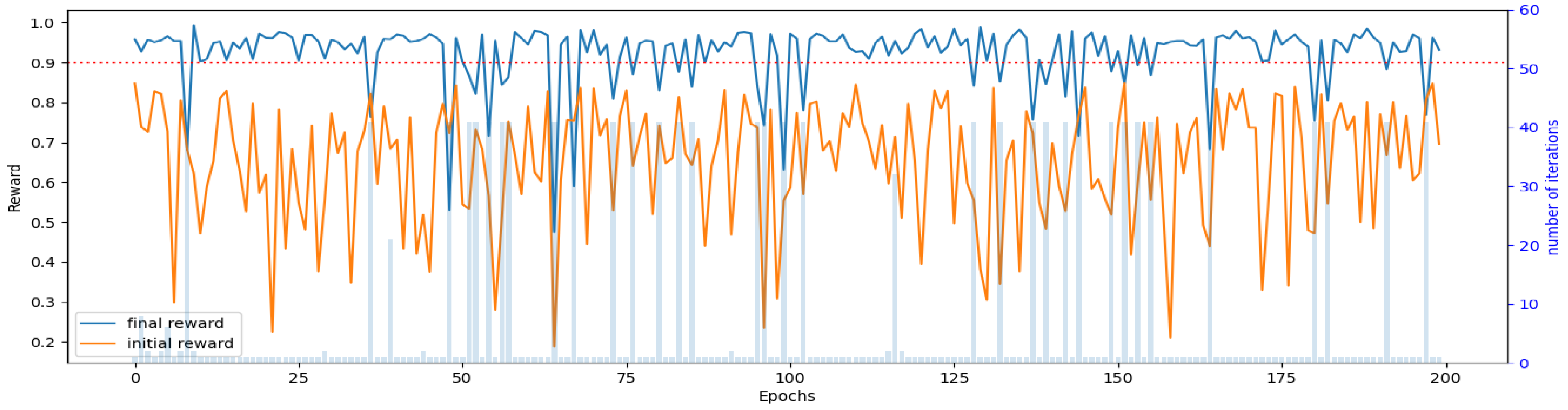
Incorporate a priory knowledge in form of a trainable NN

real lattice with random misalignments



RL agent recovers misalignments distribution from data and provides an optimal strategy

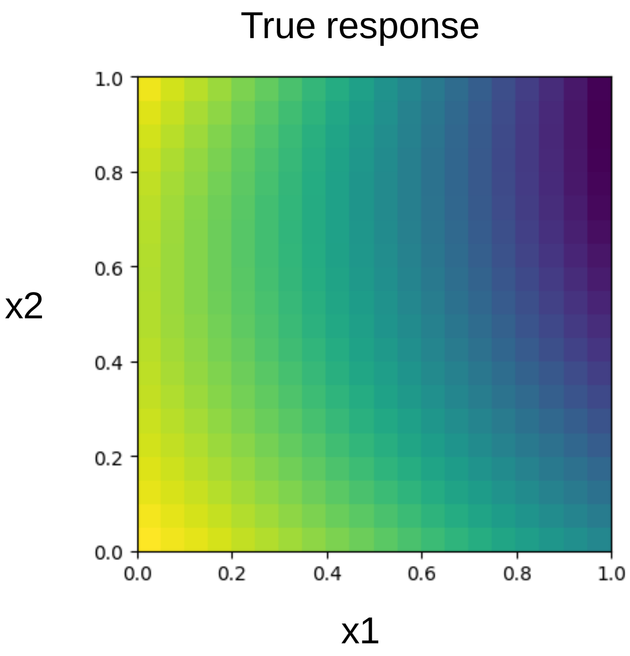
Similar to a traditional optimizer that utilizes knowledge from historical data and uses adaptive steps during objective maximization

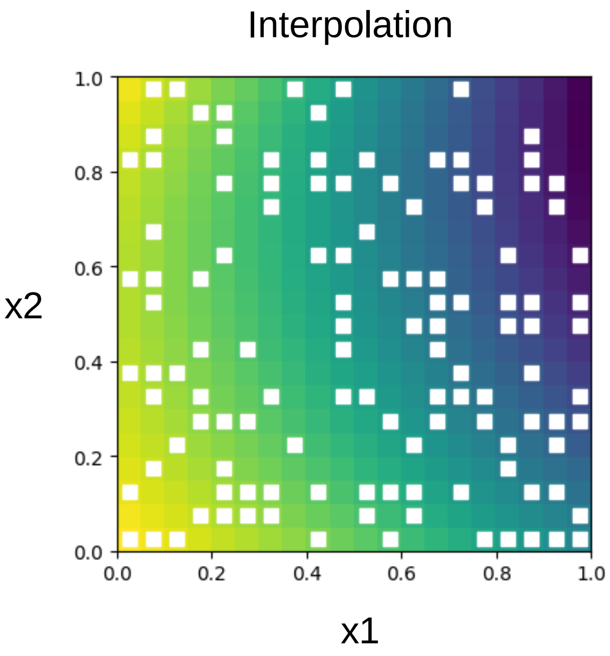


UCI benchmarks

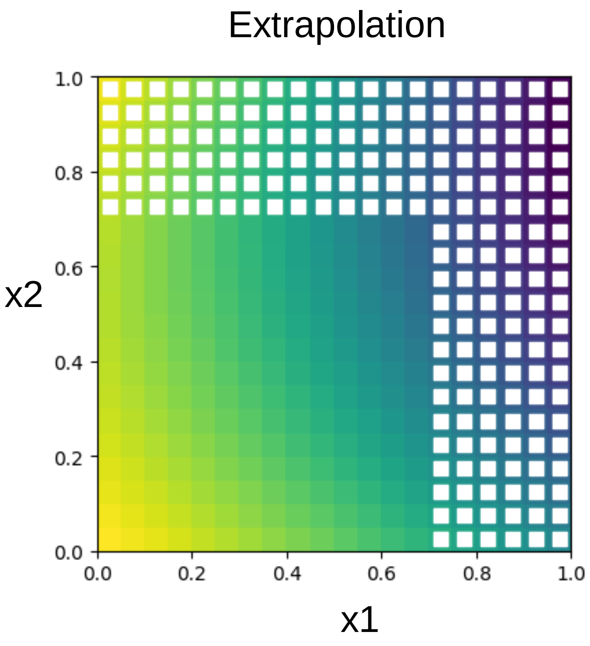
If dataset generated by a physical system then developed model can be applied for a general-purpose regression problem without a prior knowledge about ODEs

- **Airfoil Self-Noise Data Set**: NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel.
- **Yacht Hydrodynamics Data Set**: Delft data set, used to predict the hydrodynamic performance of sailing yachts from dimensions and velocity.





Prediction for data from the middle of training data



Prediction outside of the range of training data

UCI benchmarks

If dataset generated by a physical system then developed model can be applied for a general-purpose regression problem without a prior knowledge about ODEs

- **Airfoil Self-Noise Data Set**: NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel.
- **Yacht Hydrodynamics Data Set**: Delft data set, used to predict the hydrodynamic performance of sailing yachts from dimensions and velocity.

Performance of models on test sets for Yacht data:

	Interpolation			**Extrapolation**		
	MSE	MAE	R2	MSE	MAE	R2
Linear Regression	0.16	0.12	0.25	0.47	0.48	<0
Polynomial Regr.	0.08	0.05	0.82	0.34	0.33	<0
SVR	0.09	0.07	0.80	0.28	0.28	<0
XGB	0.02	0.01	0.99	0.27	0.28	<0
DNN	0.01	0.01	0.99	0.11	0.10	<0
TM-PNN	0.01	0.01	0.99	0.02	0.02	0.84

04 Next steps

- Optimizations of the TM-PNN
- Optics measurement data processing
- Agent-based architecture for experiments and operation

Results

European Conference
on Artificial Intelligence

Physical Review AB

Next steps

01 Optimizations of the TM-PNN

02 Optics measurement data processing

03 Agent-based architecture for experiments and operation

Thank you

Contact

DESY. Deutsches Elektronen-Synchrotron

www.desy.de

Andrei Ivanov
andrei.ivanov@desy.de