# (Some) Machine Learning Activities at DESY

Bianca Veglia, DESY TEMF Meeting,
Hamburg, June 1st 2023

HELMHOLTZ

DESY

# Accelerator Research
… and machine learning

‘Autonomous Accelerator’ with KIT

- First steps towards autonomous accelerator by bringing reinforcement learning to operation.
- Reinforcement learning involves measuring state values and adjusting control variables to determine their influence on each other, thus learning a control strategy that also takes into account its effects in the future. In the long run, this will completely replace manual intervention.
- Longitudinal bunch profile of two accelerators, ARES and FLUTE, located at DESY and KIT, respectively.
- These two test accelerators are dedicated to research and thus offer the unique opportunity of sufficient beam time to develop such a reinforcement learning algorithm.
- Using similar accelerators allows detailed research about the transferability of such algorithms and the resulting control agents.

*The ARES accelerator.*

KIT
Karlsruher Institut für Technologie

# Accelerator Research
... and machine learning

'HIR3X' with SLAC

- FELs generate femtosecond-duration X-ray pulses with a peak brightness more than a billion times higher than any previous source.
- As we transition from proof-of-principle experiments towards measurements of real systems by non-expert users, expansion of the experimental capabilities and reliability is necessary to be able to collect enormous datasets at high rates, over long periods of time.
- Specialized facilities: FLASH, the European XFEL, and the LCLS II.
- Optimizing all sub-systems of the source and instrumentation is the goal of this Helmholtz International Laboratory to achieve high-rate FEL measurements of complex systems.
- Applying Machine Learning to the operation of the accelerator and generation of X-ray pulses, as well as to the detection and analysis of X-ray signals.
- Aim at deploying robotic control of the delivery of samples to avoid interruptions and downtime, and to address challenges in the transport of high-power X-ray beams to experiments.

'IPC'

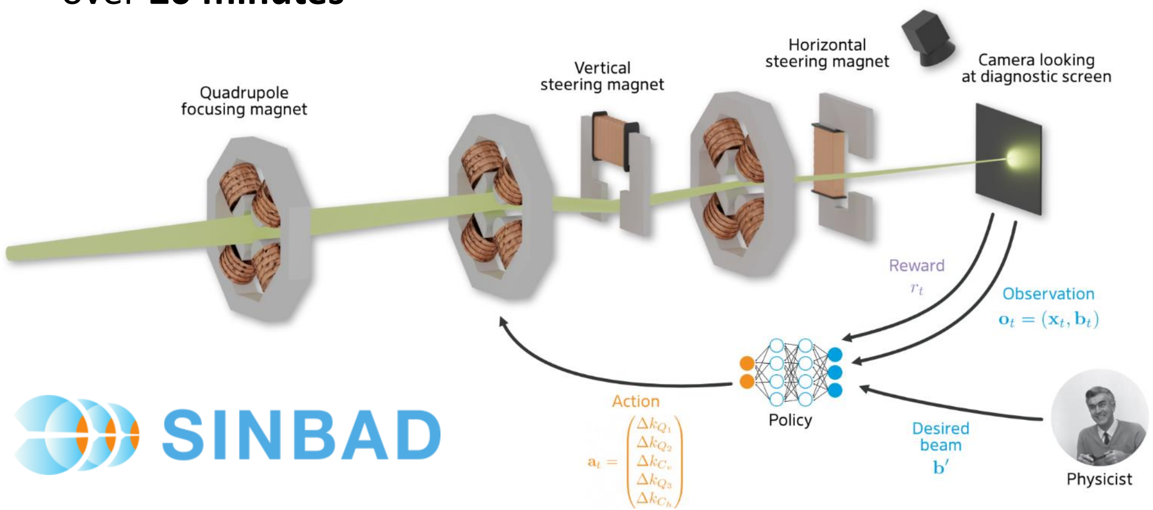
- Intelligent Process Controls (IPC) is a subgroup of MSK at DESY, pushing forward innovative research into autonomous accelerators using reinforcement learning and other cutting-edge optimization techniques.
- Engaged in developing advanced feedbacks and enhancing fault diagnosis and anomaly detection through machine learning.
- Aims to solve some of the most challenging problems facing particle accelerators today and in the future, including increasing their availability and developing autonomous accelerators.

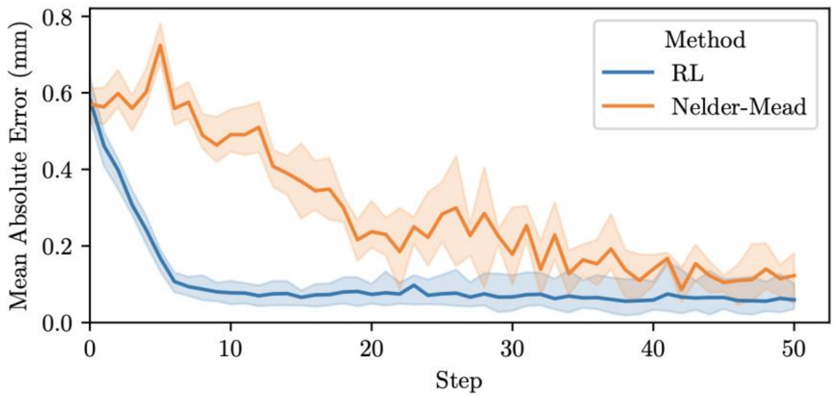
# Reinforcement Learning for Transverse Beam Tuning at ARES (Jan Kaiser)

**Successfully deploying RL to an accelerator with zero-shot learning**

- Deploy a RL-trained optimisation algorithm to the **real-world** ARES accelerator with **zero-shot learning**
- **Faster optimisation** that alternative optimisation algorithms
- **Better final beam** than alternative optimisation algorithms
- **Autonomously** achieve tune in less than **5 minutes** that takes human operators over **20 minutes**

| 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

**Paper:** Kaiser, J., Stein, O., Eichler, A. **Learning-based Optimisation of Particle Accelerators Under Partial Observability Without Real-World Training**. In *Proceedings of the 39th International Conference on Machine Learning*. 2022.

# Cheetah - Simulation Faster than Ocelot

**Accelerating linear beam dynamics simulations for machine learning applications**

(Courtesy of Jan Keiser)

- Simple high-speed linear beam dynamics simulation package written in Python.
- Meant primarily to train reinforcement learning models or generate data for other machine learning applications
- Achieves speed-ups in multiple ways:
  - Trade accuracy for speed
  - PyTorch backend (including GPU acceleration)
  - Various smaller code optimisations
- Actively maintained, more features planned
  - GitHub: https://github.com/desy-ml/cheetah
  - Documentation: https://cheetah-accelerator.readthedocs.io

**Speed compared to other simulation codes**

| Simulation | Comment | Time (ms) |
|---|---|---|
| ASTRA | space charge | 609 000.00 |
| | no space charge | 234 000.00 |
| ELEGANT | | 300.00 |
| Ocelot | space charge | 24 600.00 |
| | no space charge | 247.00 |
| Cheetah | Particle Beam (CPU) | **1.33** |
| | Particle Beam (GPU) | **0.84** |
| | Parameter Beam | **0.27** |

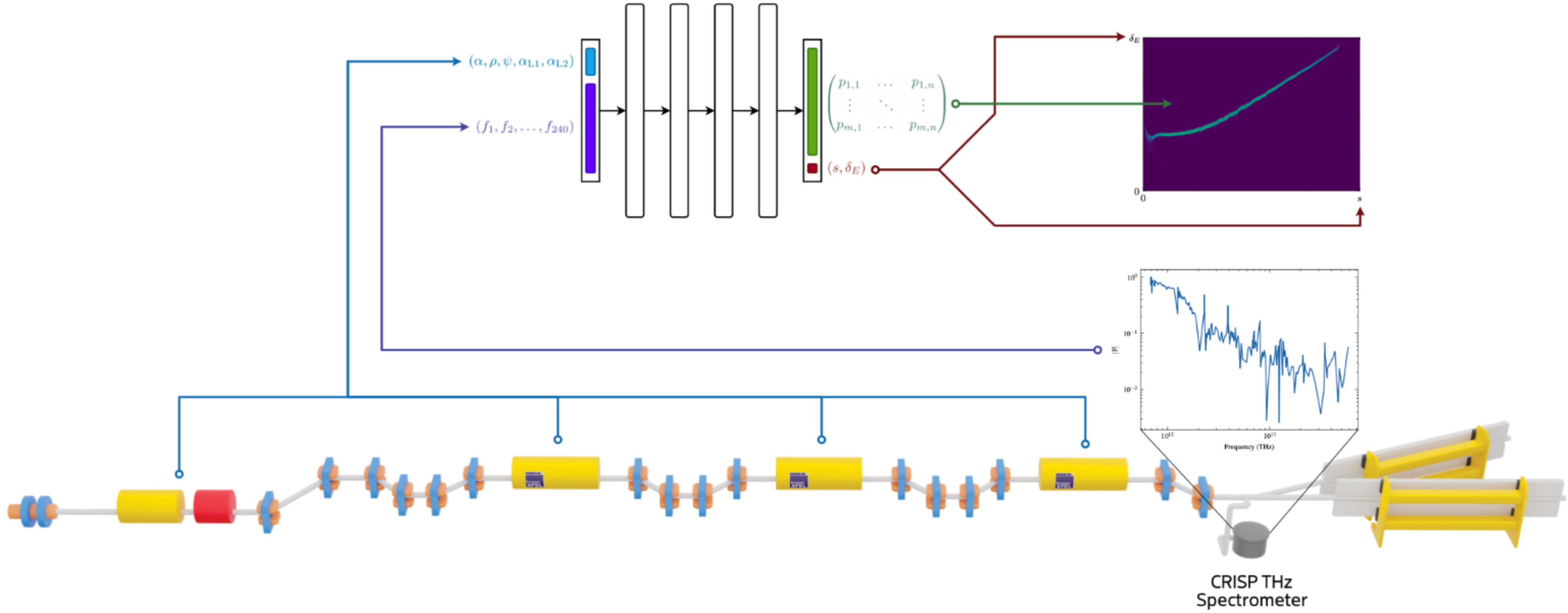
**Paper:** Stein, O., Kaiser, J., Eichler, A., Agapov, I. **Accelerating Linear Beam Dynamics Simulations for Machine Learning Applications**. In *Proceedings of the 13th International Particle Accelerator Conference*. 2022.

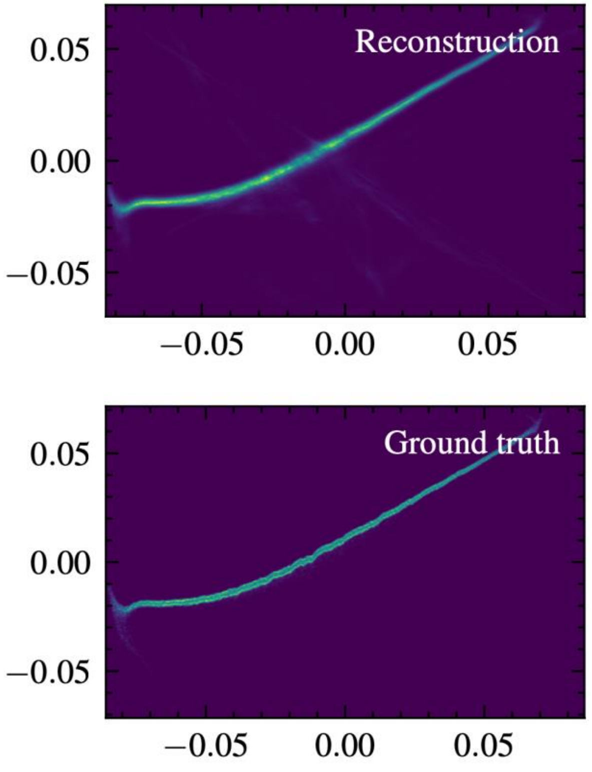
# Longitudinal Phase Space (LPS) Reconstruction at EuXFEL (Jan Kaiser)

## Neural networks for fast and non-destructive diagnostics

- Combine scalar (**RF settings**) and spectral (**CRISP THz spectrum**) information using a neural network
- Trained on data from a start-to-end beam dynamics simulation, **no need for real-world training data**
- Provides **adaptive resolution** of the LPS by predicting a bunch's extent in the LPS in addition to the LPS image.
- **Non-destructive** and very **fast** reconstruction

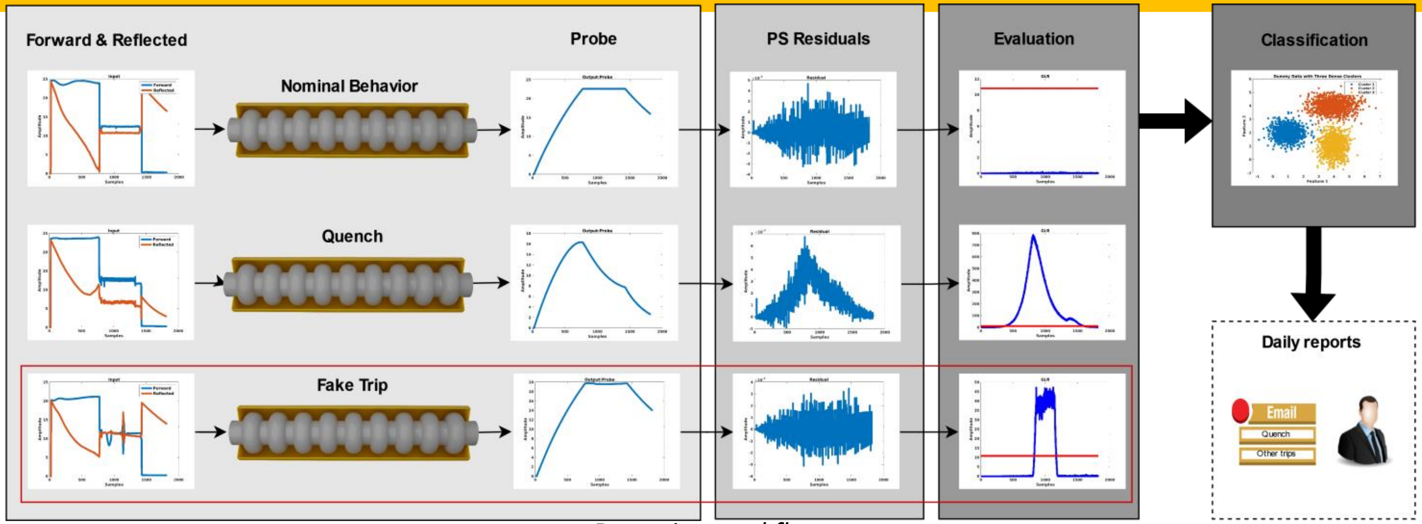
**Paper:** Kaiser, J., Eichler, A., Tomin, S., Zhu, Z. **Machine Learning for Combined Scalar and Spectral Longitudinal Phase Space Reconstruction**. In *Proceedings of the 14th International Particle Accelerator Conference*. 2023.



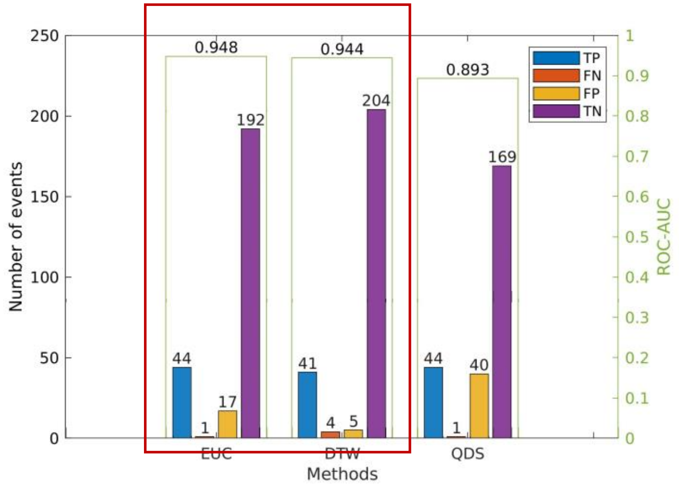
# Quench detection at the EuXFEL (Lynda Boukela)

- At XFEL, electron bunches are accelerated to high energies of up to 17.5 GeV
- The Linac uses hundreds of superconducting cavities
- **Anomalous events** disrupt the normal functioning of the accelerator, **quenches** are the most severe
- Currently the quenches are detected by monitoring the cavities quality factor
- With the ML-powered approach:
  - The residuals of the model are exploited
  - A statistical test is used to detect the anomalies
  - Quenches recognition is achieved through clustering




*Detection workflow*

- Experimental evaluation
- Signals from the second half of 2022 are used
- The ML approach outperforms the current system
- Potential hardware implementation in the future



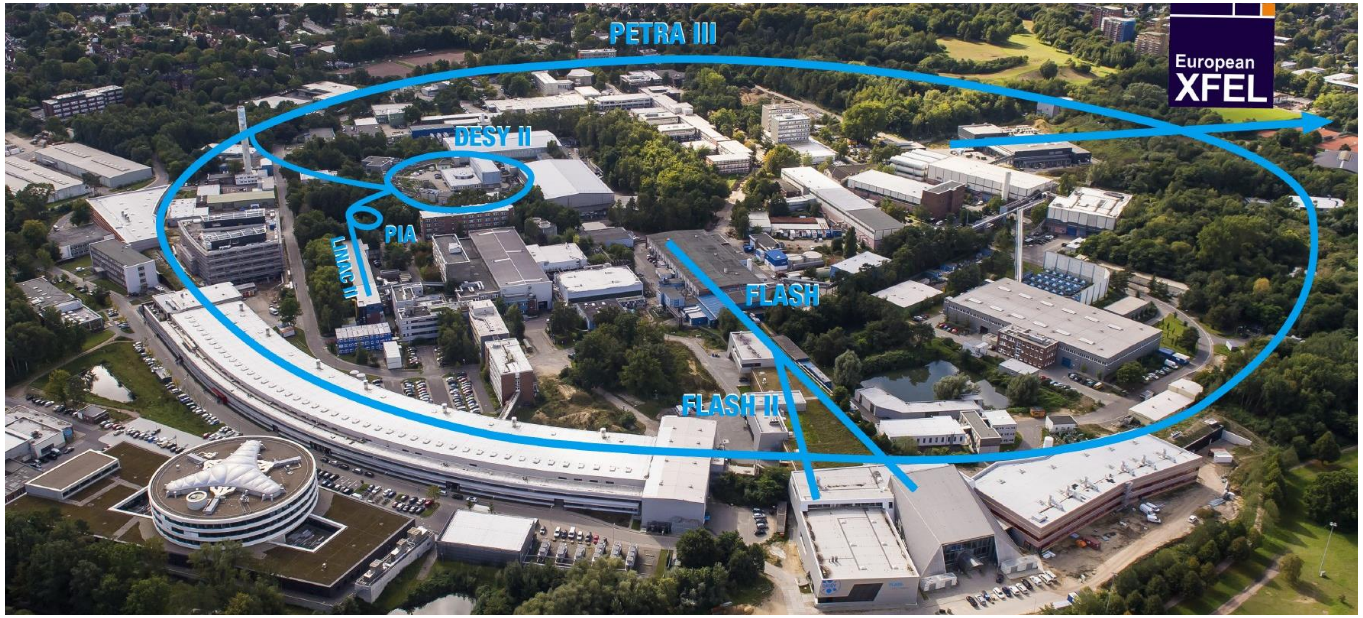

*Evaluation and comparison results*

# Neural Networks for ID Gap Orbit Distortion Compensation in PETRA III

- IDs gap variations affect the circulating electron beam orbit.
- Neural Networks can be used to predict the distortion.
- Four NNs architectures were developed and trained with PETRA III measurements to predict the beam orbit at any given operational ID gaps configuration.
- The models were optimised through hyperparameter sweeps to obtain high predictability.

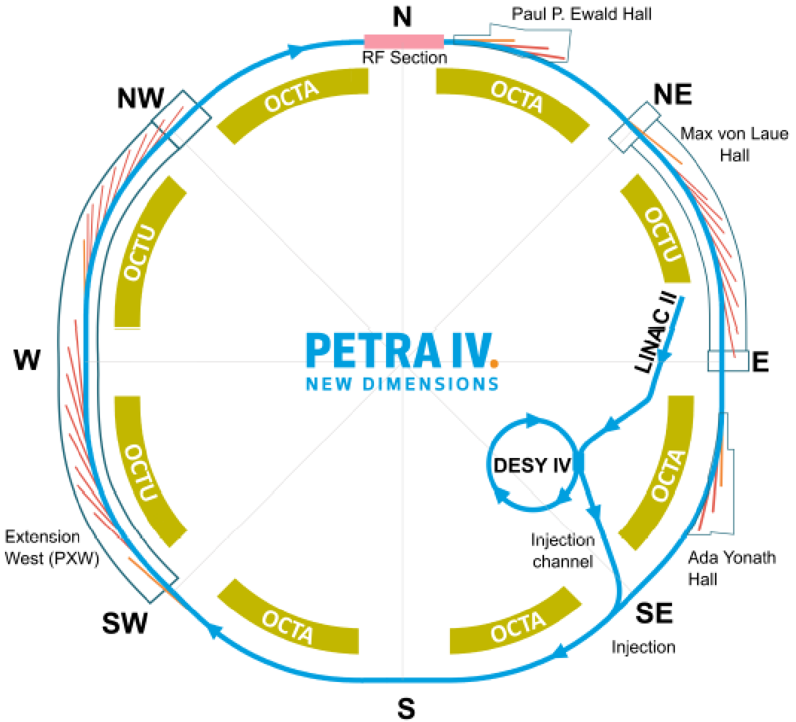
- The predicted displacements can be used in local or global correction scheme.

# PETRA III to PETRA IV.




- High brilliance 3rd Generation Synchrotron Radiation Source.
- Extremely low emittances: (hor. / vert.) 1.3 / 0.012 nm.
- 25 beamlines.

Each year, more than 2000 users are performing measurements at the PETRA III beamlines.

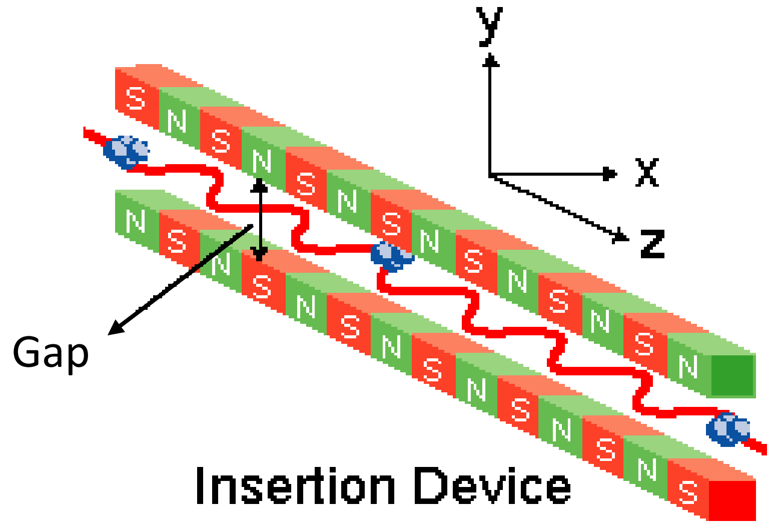



The storage ring will feed up to **30 undulator** insertions (photon beam can be further split to allow more experimental stations).

Ultra low emittances in the region of 10 pm.

# IDs induce an orbit distortion which varies with the gap size

- The magnetic fields of IDs introduce perturbations to the circulating electron beam and hence affect the linear and nonlinear beam dynamics of the electron beam in the storage ring.
- Often users adjust the spectrum from undulators by changing undulator gap size. It's important to keep the orbit constant during these field changes to not disrupt other users.



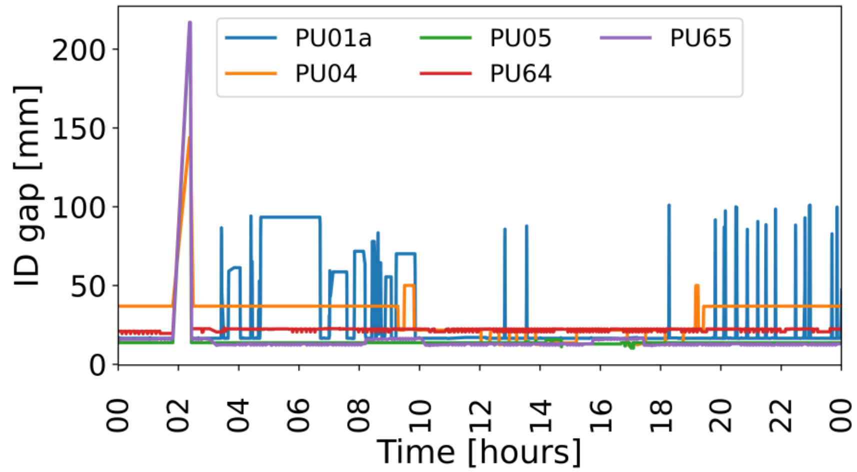

The field integrals determine the primary effect of the undulator on the electron beam orbit.

$$I_{1x} \equiv \int_{z_0}^{z_0+L} B_x(z_1)\,dz_1 \qquad x'_{exit} = -\frac{q}{\gamma m v_z} I_{1y}$$

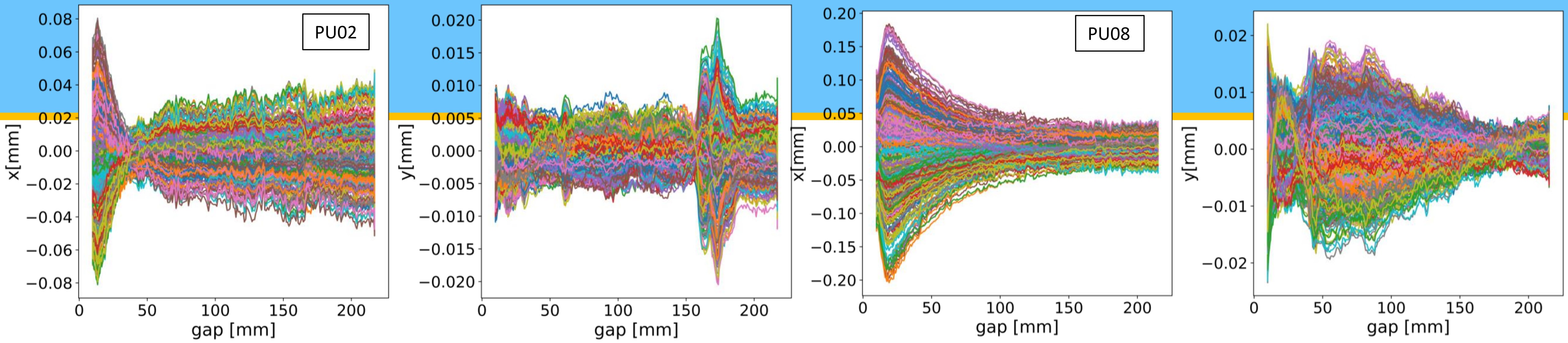
$$I_{1y} \equiv \int_{z_0}^{z_0+L} B_y(z_1)\,dz_1 \qquad y'_{exit} = \frac{q}{\gamma m v_z} I_{1x}$$

$$I_{2x} \equiv \int_{z_0}^{z_0+L} \int_{z_0}^{z_2} B_x(z_1)\,dz_1\,dz_2 \qquad x_{exit} = -\frac{q}{\gamma m v_z} I_{2y}$$

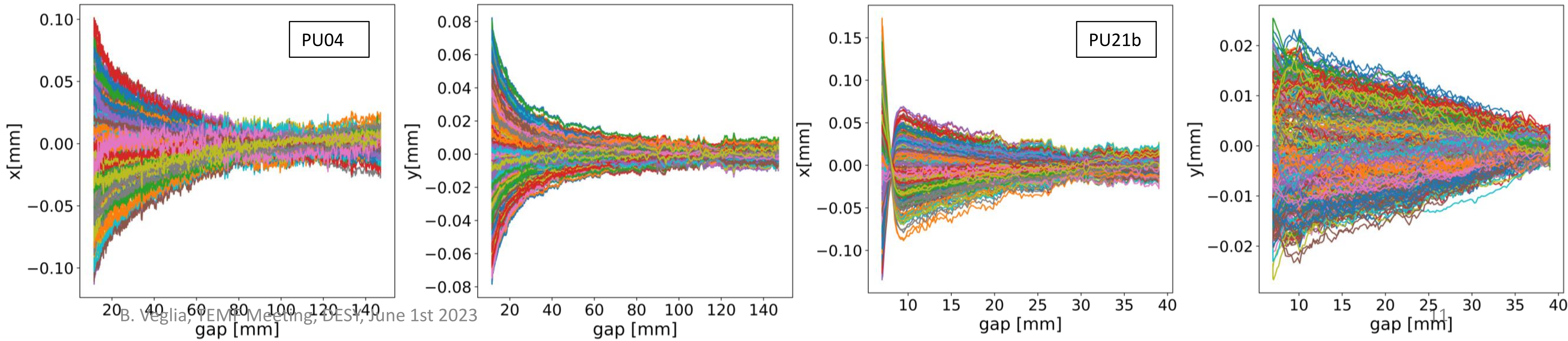
$$I_{2y} \equiv \int_{z_0}^{z_0+L} \int_{z_0}^{z_2} B_y(z_1)\,dz_1\,dz_2 \qquad y_{exit} = \frac{q}{\gamma m v_z} I_{2x}$$



# Closed orbit distortion measurements

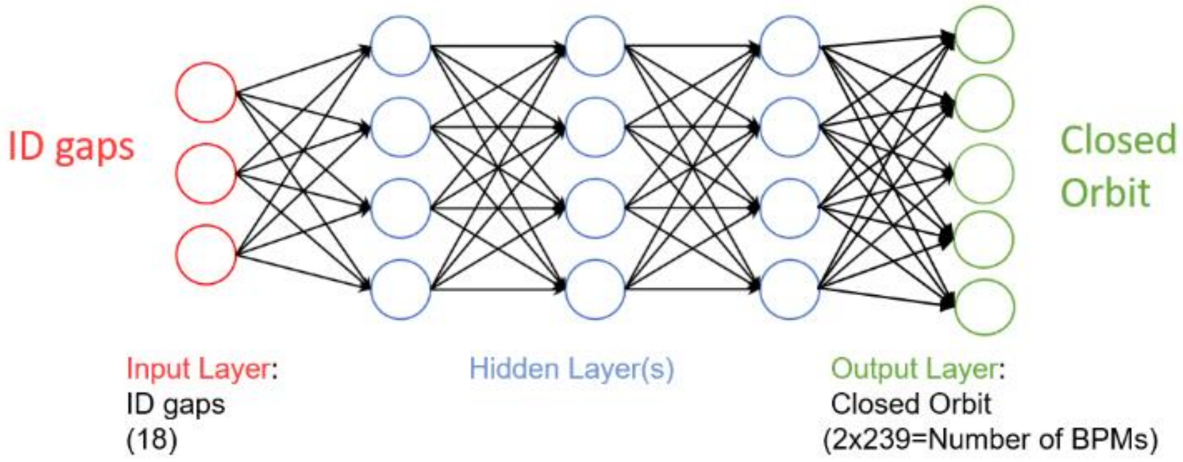



Measurements of the horizontal and vertical orbit were taken varying the gap size in their operation range for 18 IDs in PETRA III. Each colour represents a BPM along the ring.



# Building the NNs

- The NN takes as input a vector containing the ID gap sizes and gives as output the predicted orbit at the location of each BPM.
- The different model are trained on 80% of the measurements took last July and validated vs the remaining 20%




Keras

The models are trained using the back propagation method employing the Adam optimizer for 200 epochs (10-50 minutes).

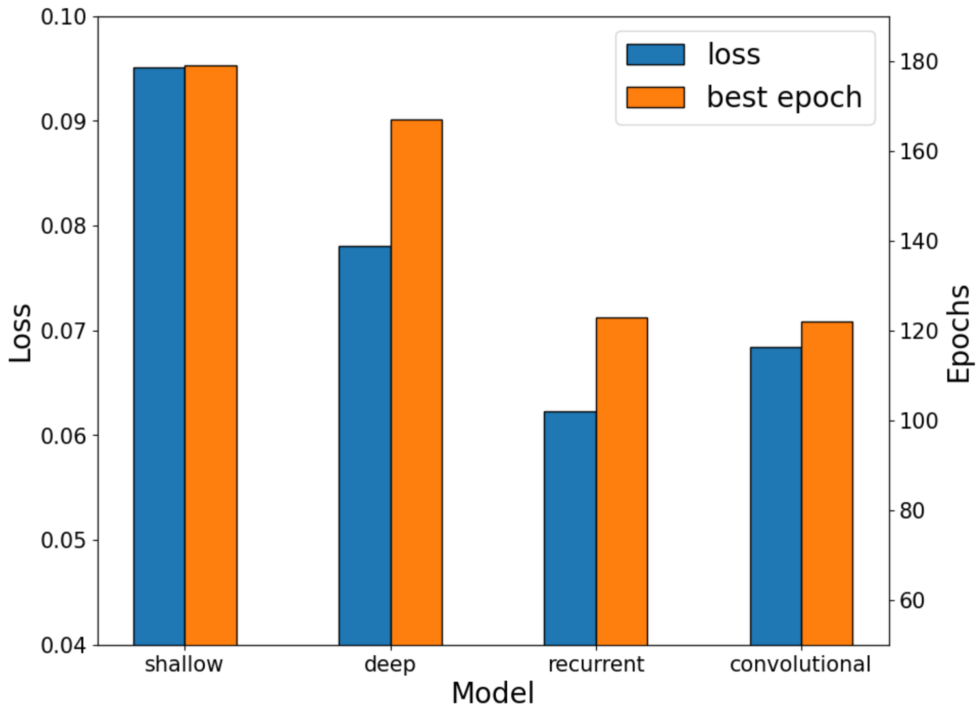
Hyperparmeter sweeps performed with

Weights & Biases

Why using machine learning on this specific problem?

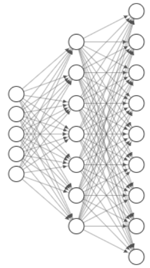
Because of its flexibility and ability to model also highly nonlinear processes.
Measurements for look-up tables are time consuming and need to be updated regularly.

# Comparing NN architectures

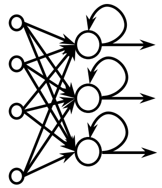



Four different architectures studied:

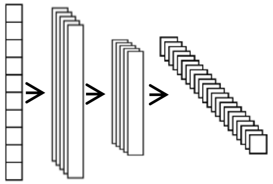
> Feed Forward NN (FFNN): the simplest architecture, information is always fed forward and there are no loops. Two considered here:
> - Shallow: only one hidden layer.
> - Deep: 3 hidden layers.



> Recurrent NN (RNN): connections between nodes can create a cycle feeding the output of a layer back to itself.

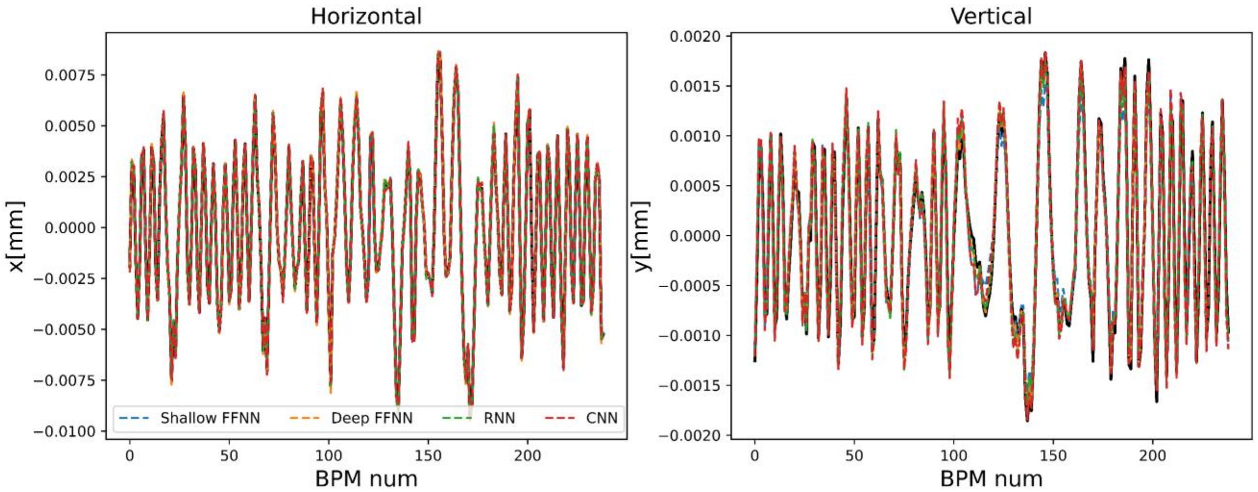


> 1D Convolutional NN (CNN): the convolution layer systematically apply learned filters to input in order to extract features. It applies a kernel (a matrix, in this case 1D) of weights which are multiplied with the input to extract relevant features.



The convolutionl and recurrent structure outperform the fully connected NN in a reasonable amount of epochs.

# From prediction to correction

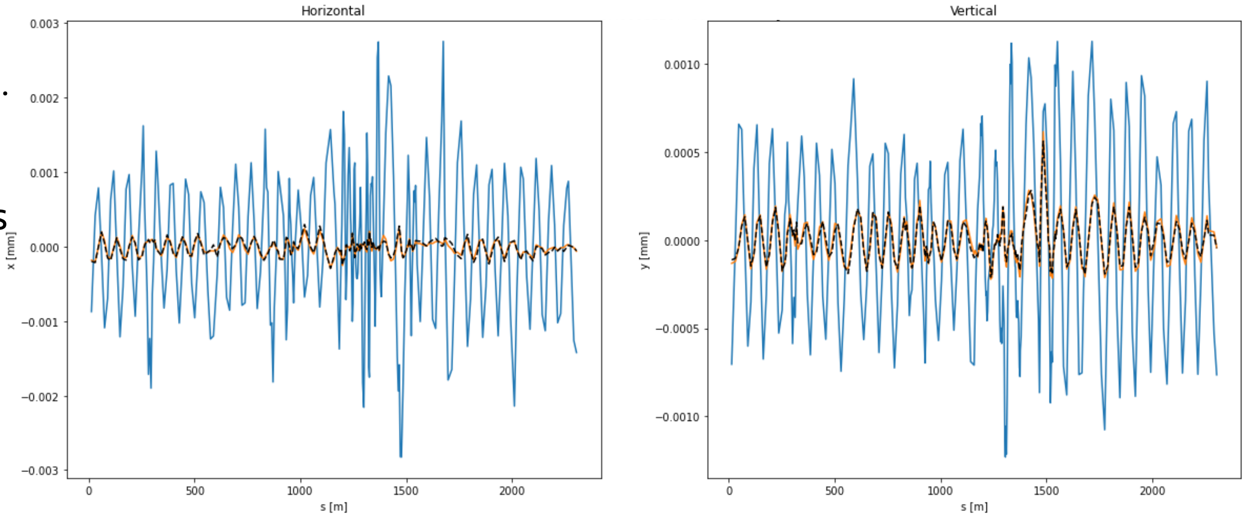



- With the models it is possible to scan 18 IDs through the entire operational parameter space and evaluate the expected transverse position.
- The orbit distortion predicted by the trained models can then be used to calculate the required strength in the corrector magnets.

Global correction: through the ORM it is possible to compute the kick at the correctors along the ring necessary to counteract the distortion.
Local correction: use the trained NN to map the ID movements to the current of the compensation coils at each end of the undulators.

\> Tests of the implementation of the related compensation schemes planned in the next future.

A similar approach could also be considered to counteract the perturbation introduced by ID gap variations to the betatron coupling and the vertical dispersion that could impact the extremely low emittances of PETRA IV .



# Thank you

**Contact**

DESY. Deutsches
Elektronen-Synchrotron

www.desy.de

Bianca Veglia

MPY

bianca.veglia@desy.de