ML and optimisation algorithms for CERN accelerators

Status and Results

V. Kain for the CERN ATS ML community forum

Motivation

→ **Automation and better modelling** for increased flexibility, reproducibility, availability and thus for overall better accelerator performance

Classical models/approaches not good at problems such as field prediction with hysteresis, multi-dimensional optimisation problems, …

…or are simply too time consuming.

→ **Complement** classical modelling with advanced algorithms

Types of algorithms / tasks

- ★ Numerical optimisers and optimal control

- ★ Machine Learning for
  - ★ Regression
  - ★ Computer vision
  - ★ Anomaly detection
  - ★ Trending and forecasts
  - ★ Reinforcement learning

Disclaimer:

Examples are from the **ML community forum**, albeit far from everything that was discussed.

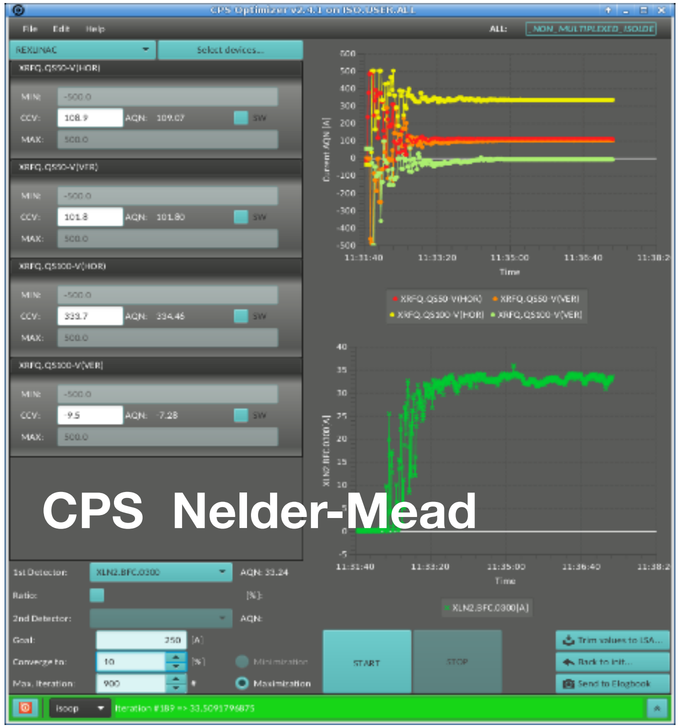
Numerical Optimisers pre-ML community

Was becoming standard … **ZS alignment, LEIR injection, ….**

Usually use derivative free algorithms due to lack of model.

★ Convex problems: Nelder-Mead, Powell, COBYLA

★ Non-convex problems: BOBYQA, Bayesian Optimisation, …



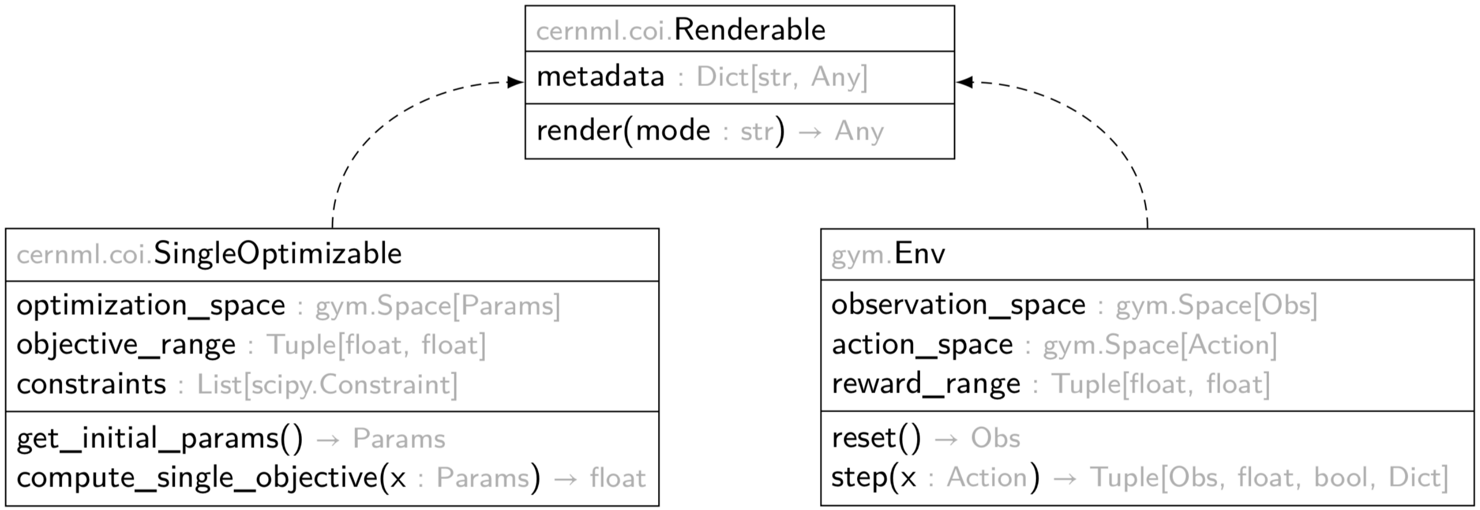
First generic optimiser (2018): Powell, COBYLA, ….

Generic Optimisation Framework

Developed ***Framework*** to facilitate the application of numerical optimisation.

Inspired by the approaches in Reinforcement Learning → split between ***optimisation problem*** and ***algorithm*** with predefined interface in between.

→ OpenAI Gym environments



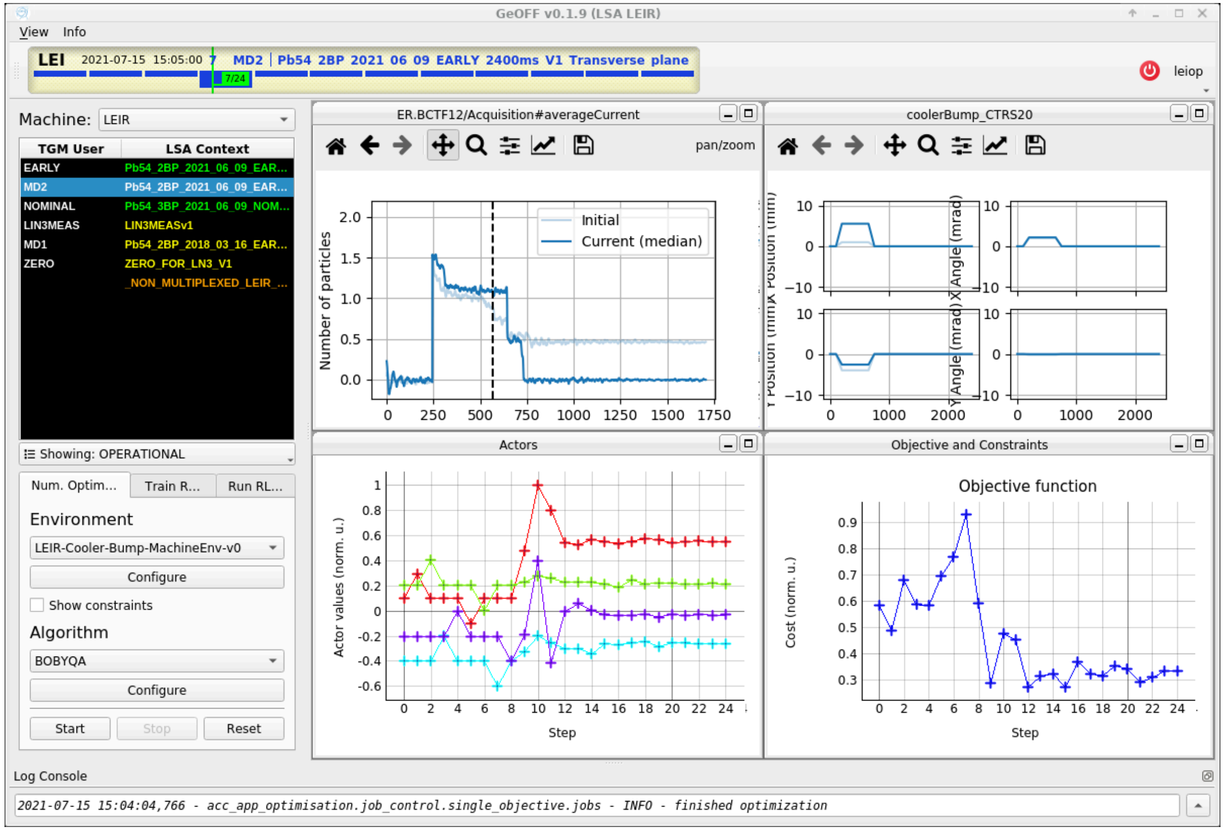
Currently (only) single objective problems.

Generic Optimisation Framework

Libraries, support and GUI for testing offline and in control room.

Fully integrated with CERN settings management, role-base access, acc-py

→ also allows to optimise functions in cycling machines



Generic Optimisation Framework

Many examples now from many machines:

- ★ Electro-static septum alignment in the SPS
- ★ SPS Spill correction of 50 Hz harmonics
- ★ Crystal alignment
- ★ LEIR injection efficiency optimisation
- ★ LEIR extracted trajectory optimisation
- ★ LEIR e-cooler bump optimisation
- ★ SPS tune function optimisation
- ★ LINAC4 chopping efficiency optimisation
- ★ ISOLDE ion sources, transmission,…
- ★ …

Regression

Artificial Neural Networks (ANNs) are the “ideal" **universal function approximators**.

Fits without having to “necessarily" know what the real function can be parameterised with.

(At the precise sometimes of generalisation due to over-fitting.)

Useful for many purposes of course. But also:

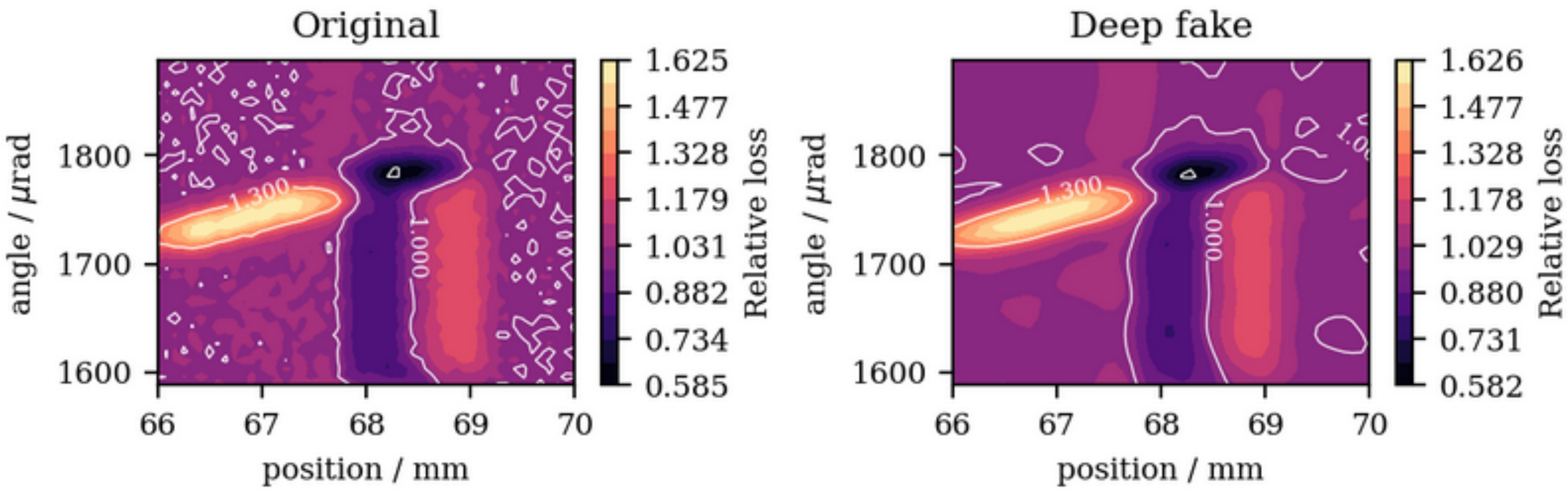
→ fit the dynamics/models. **Surrogate models**.

Example: Surrogate model beam loss vs crystal position and angle

Vanilla feed-forward networks. Works very well…

- Simulations take up to 1 minute for one sample, NN basically instantaneous

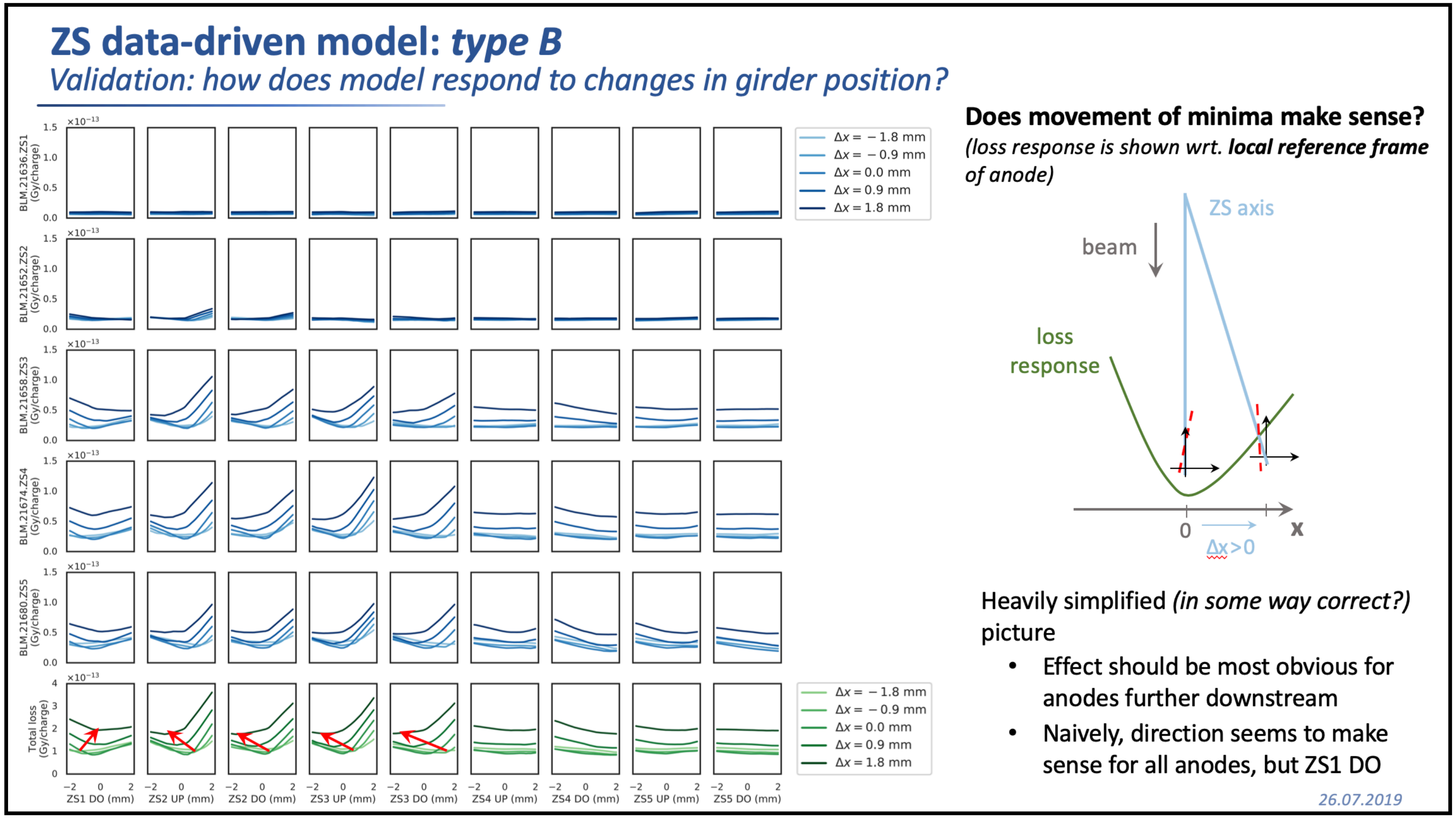
$$L_{ZS_{i,j,k}} = \mathcal{M}(\theta_i, x_j, ZS_k) \qquad \mathcal{H}$$



F.Velotti, B. Goddard

Example: Surrogate model beam loss vs electro-static (ES) anode positions + girder

Vanilla feed-forward network. Works very well…

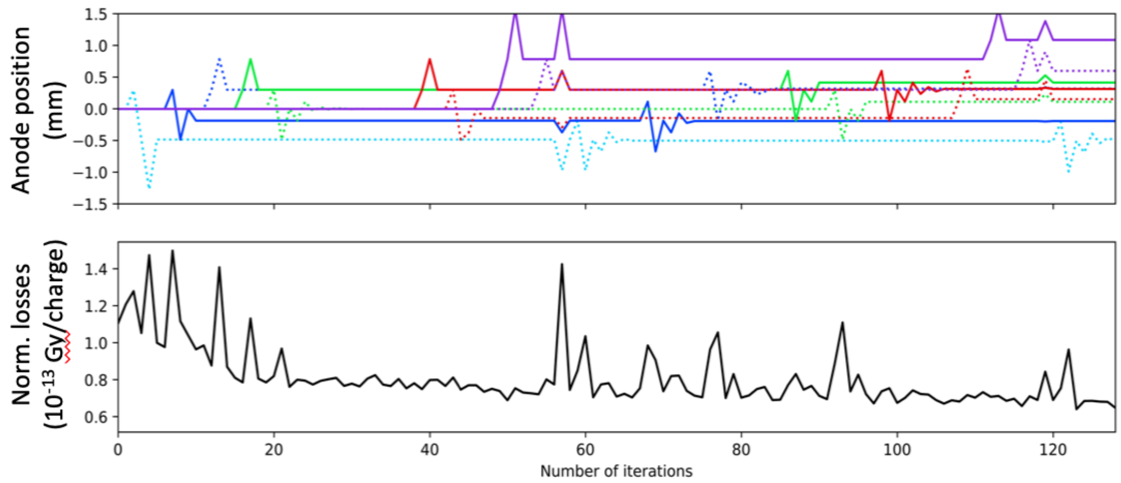


M. Schenk et al

Example: Surrogate model beam loss vs electro-static (ES) anode positions + girder

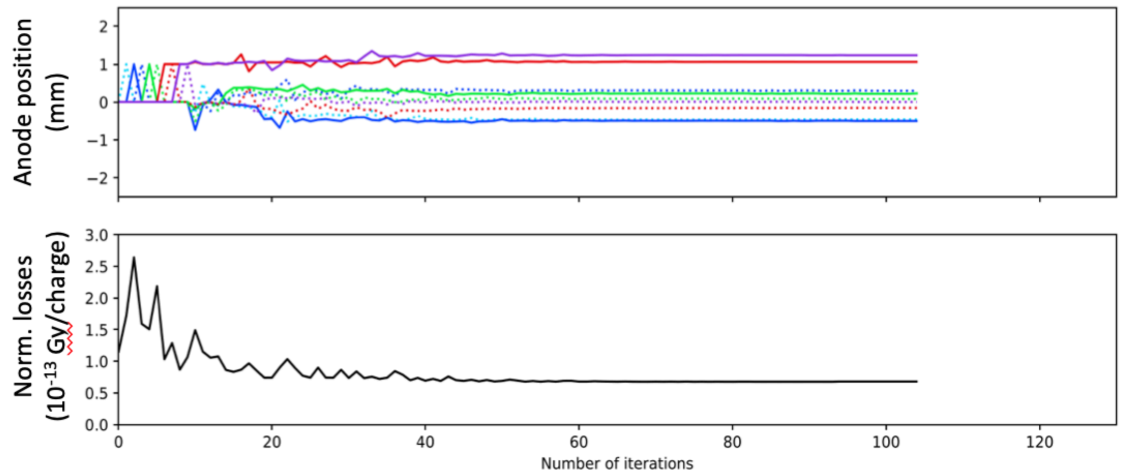
Compare optimisation algorithms with surrogate model

Powell method



Was used in 2018 reduced alignment time from 8 h to ~ 45 minutes

COBYLA



Candidate for 2021

ES alignment in 10-15 minutes?

Result electro-static septum alignment 2021

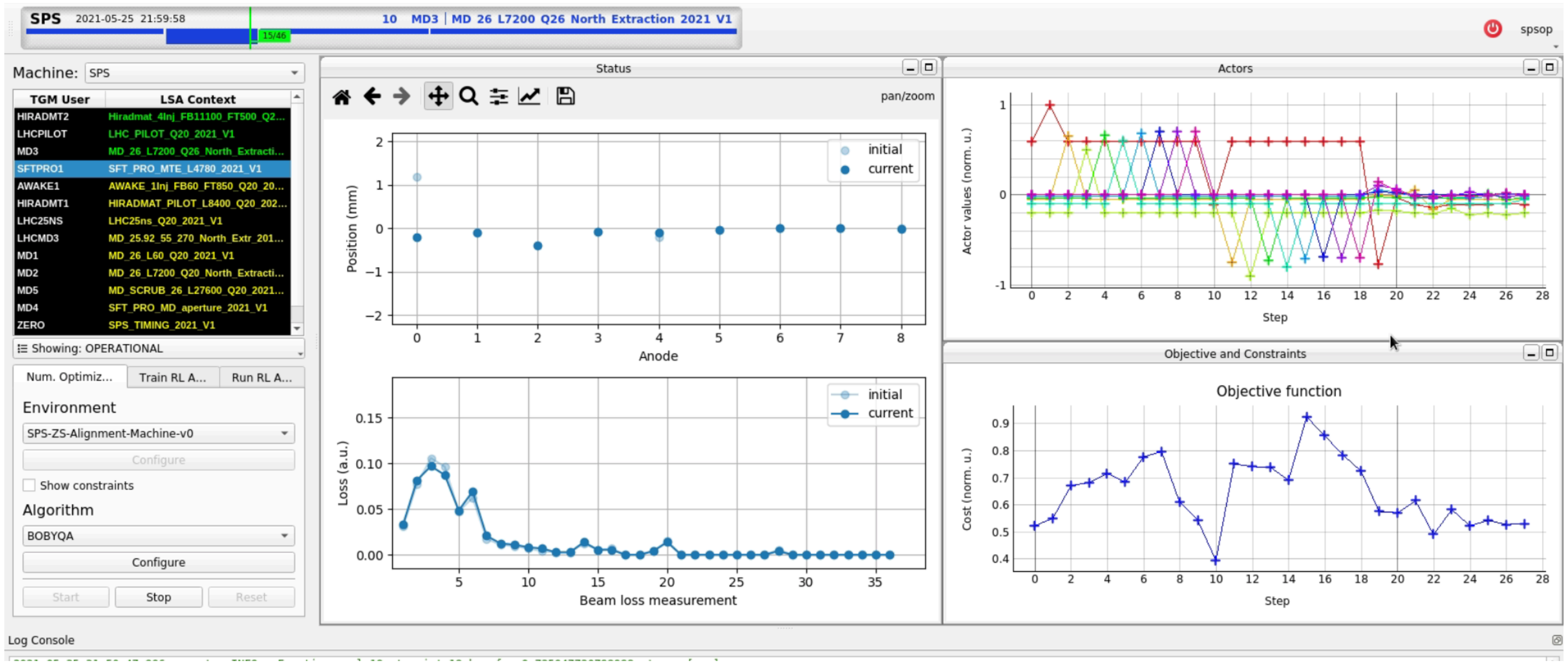
Recap:

★ before numerical optimisation: ~ 8 h

★ with Powell algorithm in 2018: 130 iterations, ~ 45 minutes

2021

9 DOF: ~ 30 iterations, ~ 10 minutes



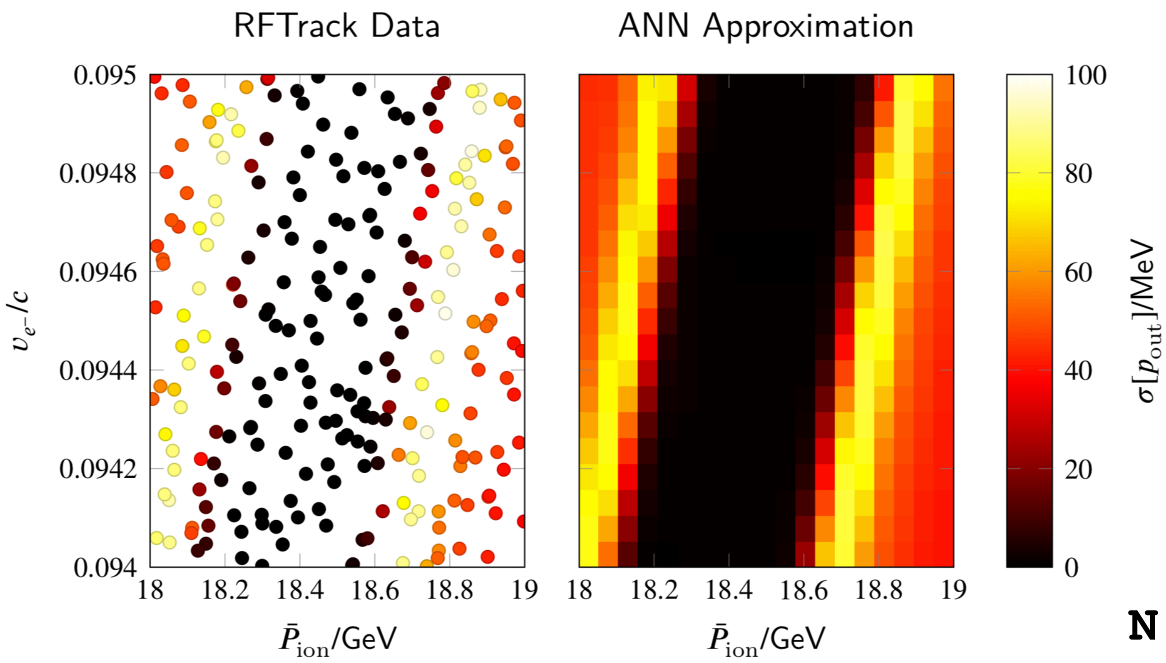
Example: dynamics model in ANN from simulation

E-cooling at LEIR. Simulation with RF track.

Final goal: train a Reinforcement Learning agent to optimally control e-cooling.

But: need to to interact with environment (i.e. RF track simulation). Becomes undoable. Too slow.

Solved with supervised learning.



N. Madysa, A. Latina

Model-predictive control to further exploit classical models or learned models

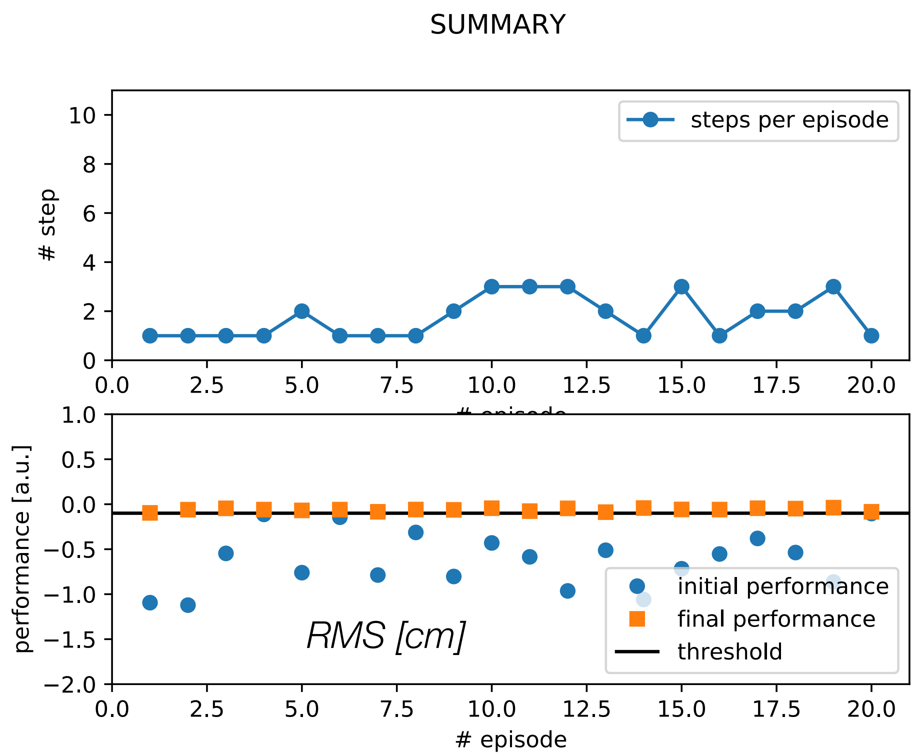
If model available, but cannot (easily) invert to find optimum correction for given state observation, can use e.g. Model Predictive Control (MPC)

Example: iLQR for AWAKE trajectory correction.

Use learned response = ANN.

Use iLQR for “inversion" of ANN

Was done in simulation.
Next week test for real

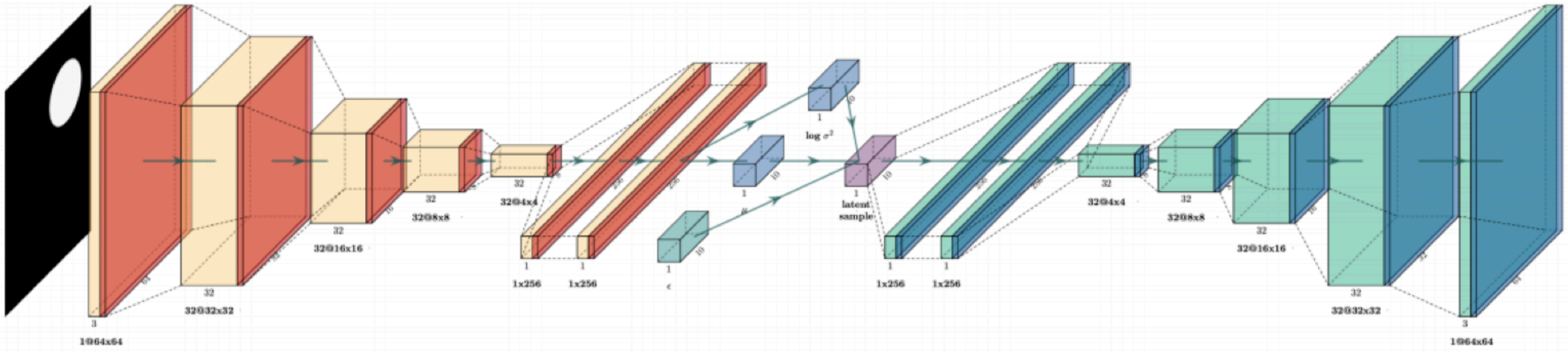


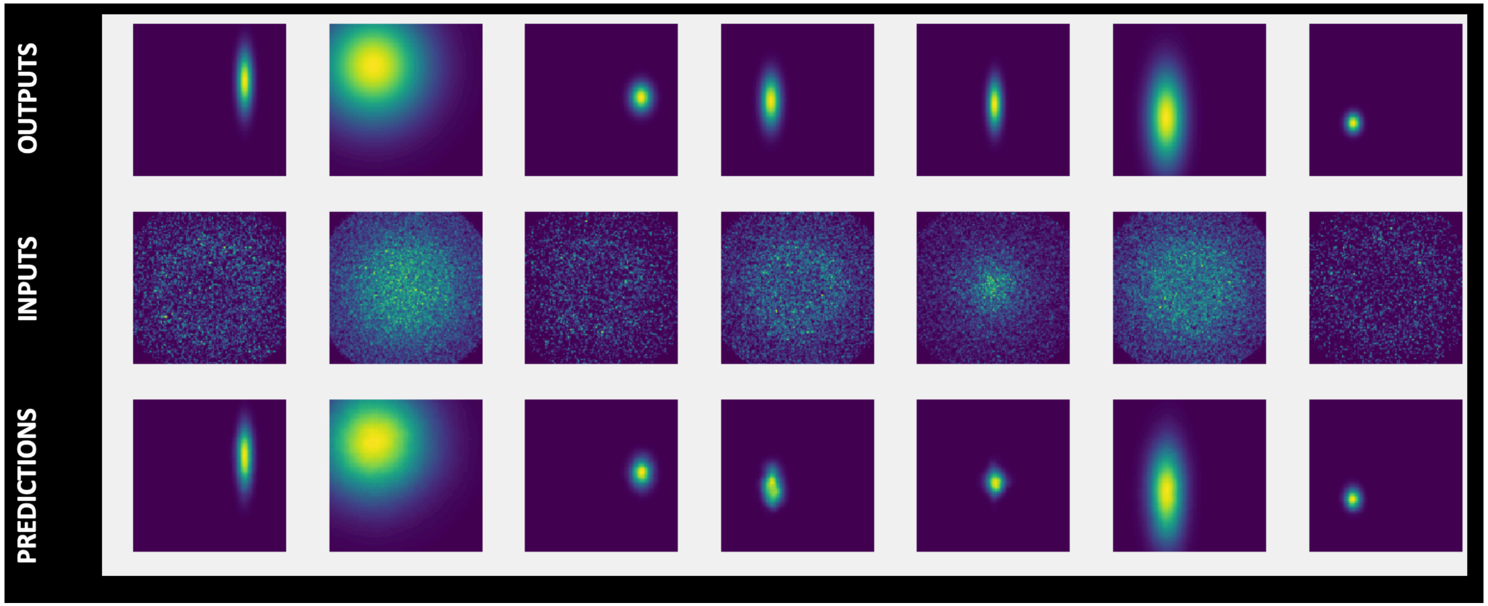
N. Bruchon et al

Computer Vision

Example: Next generation profile measurement?

Variational Auto-encoders for radiation hard **Optical Fibre Imaging**





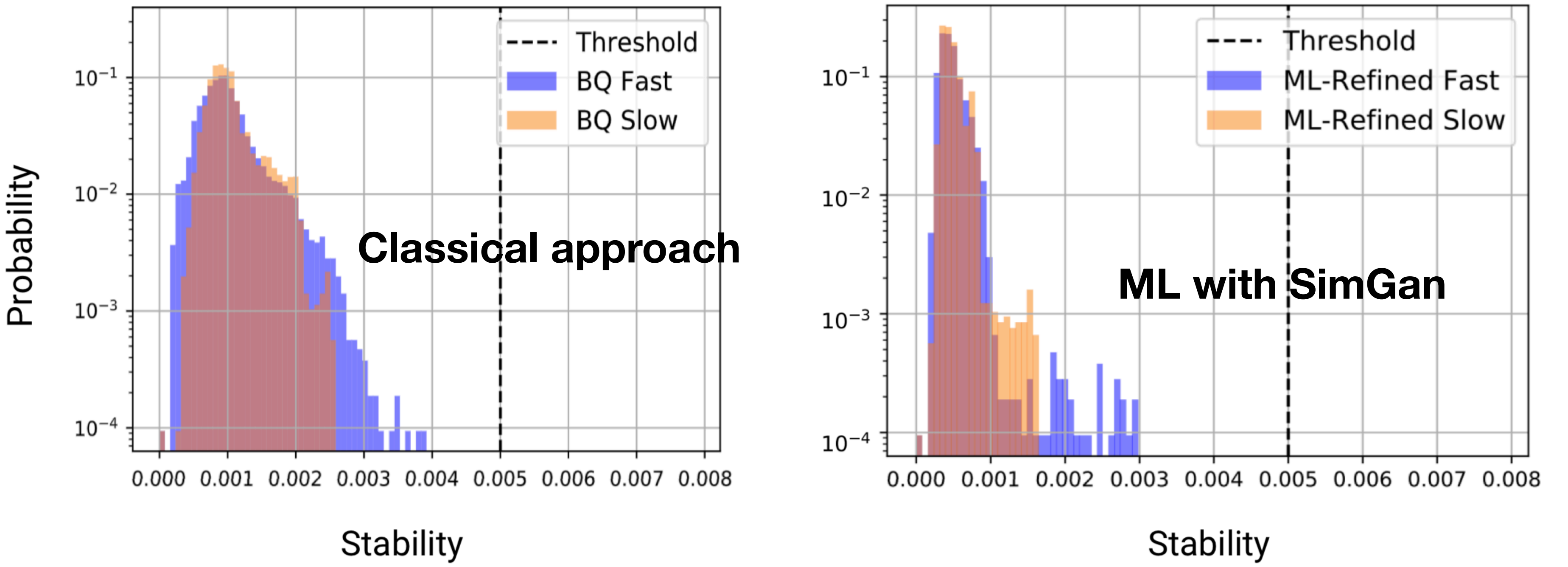
G. Trad

Interpreting the LHC tune signal

Tune estimation algorithm from BBQ spectra that does not get fooled by 50 Hz noise.

Refined Approach II

- What about the QFB stability metric?



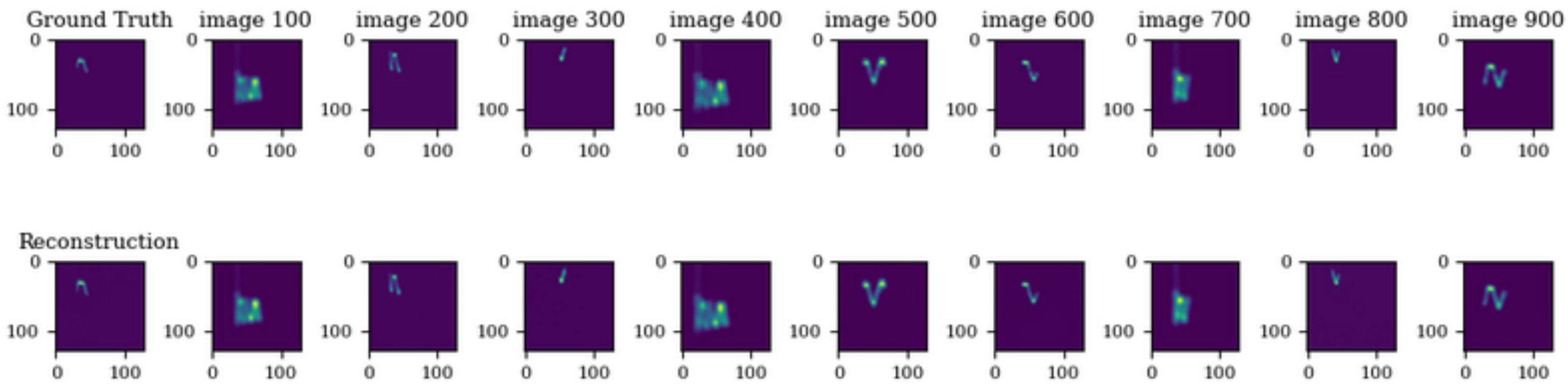
- ML-Refined gave the most stable estimates from all tune estimation systems attempted

L. Grecht

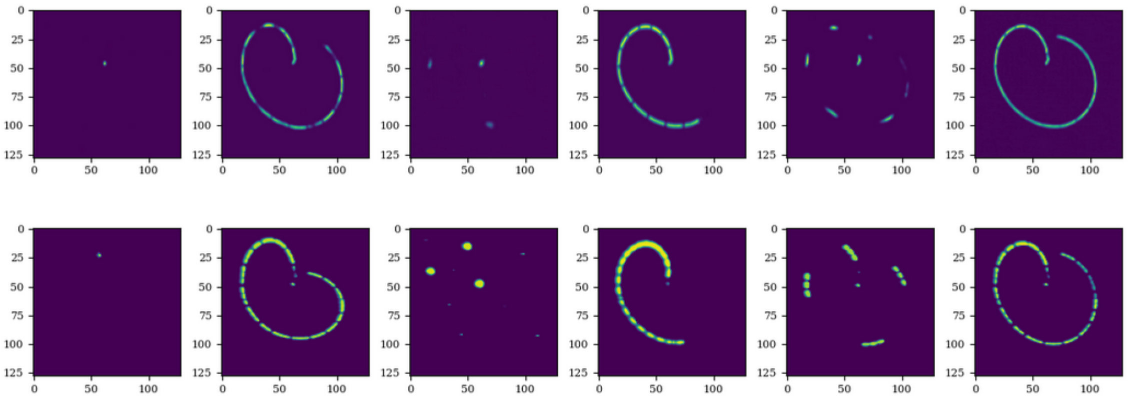
Example: interpreting beam dump pattern

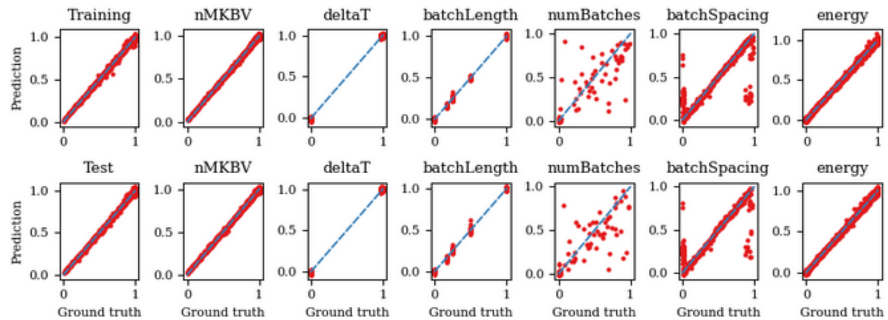
Classify dump kicker failures from the beam dump pattern images. SPS and LHC

Model results in simulated data



Model trained on simulations and applied on real data…and extract physical information about the system from images

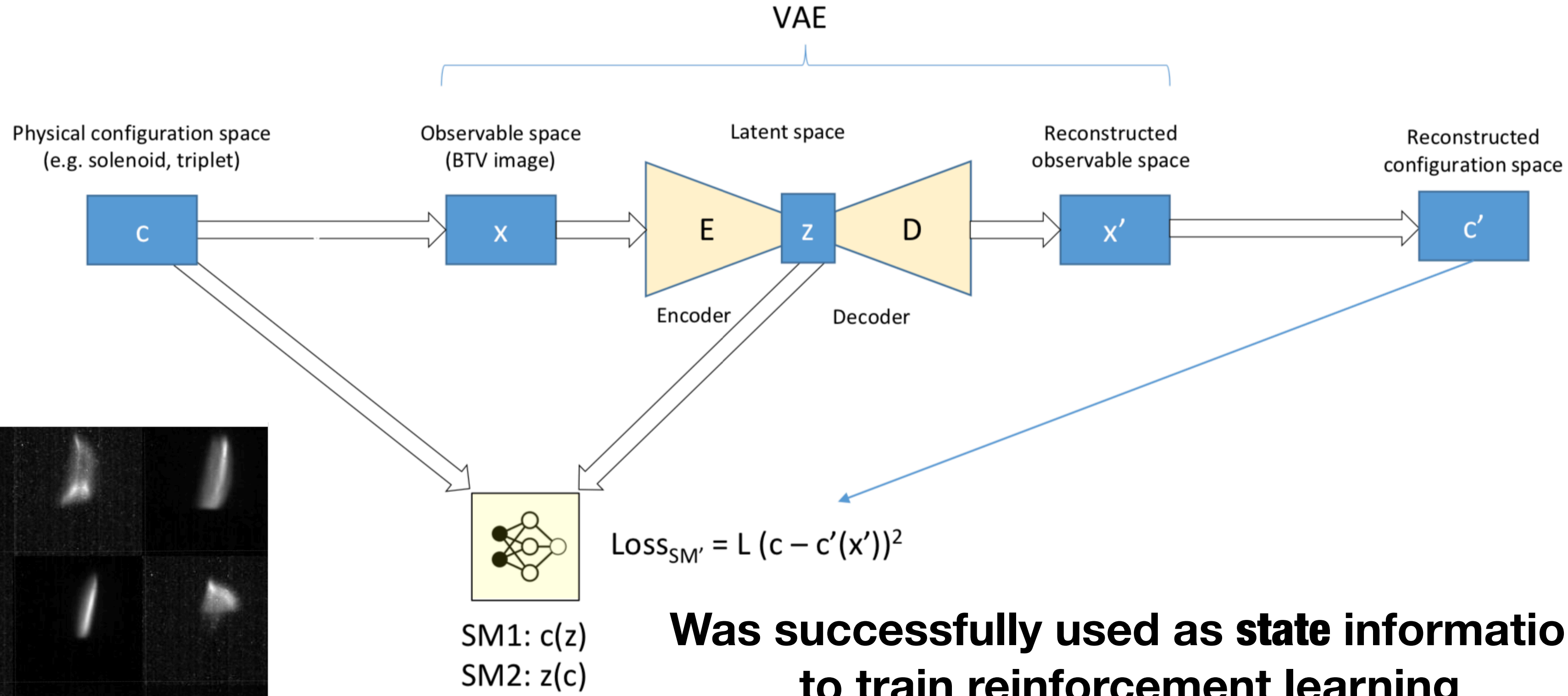




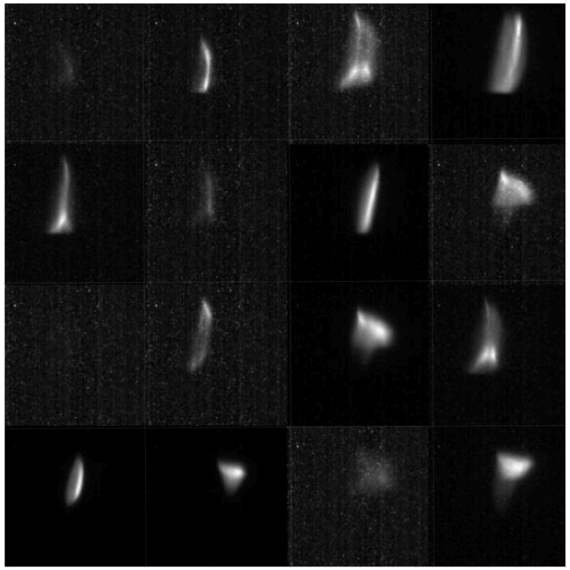
F. Velotti and B. Goddard

Example: interpreting BTV images

Building **surrogate models from BTV with images. Deep Fake AWAKE.**



The Data



Was successfully used as state information to train reinforcement learning agent to optimize spot size.

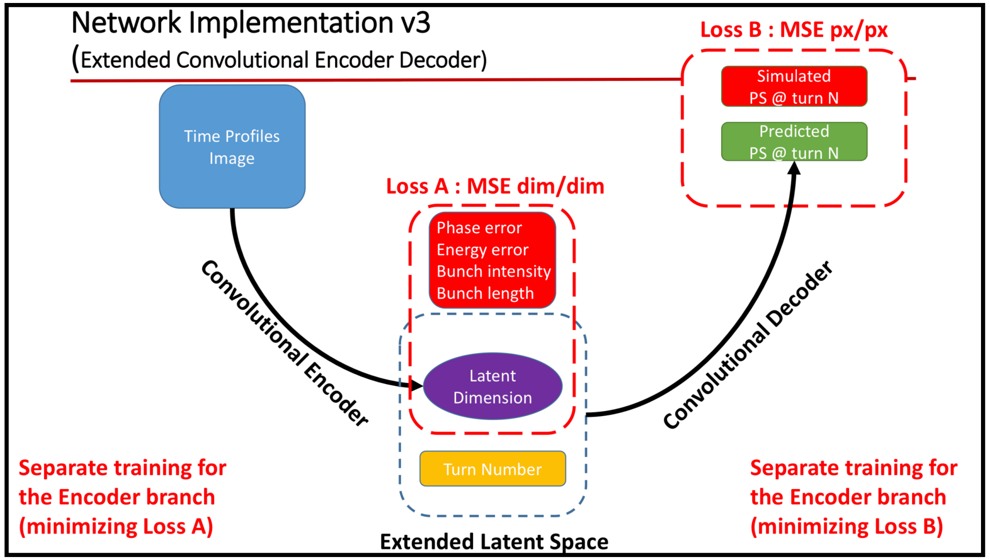
F. Velotti and B. Goddard

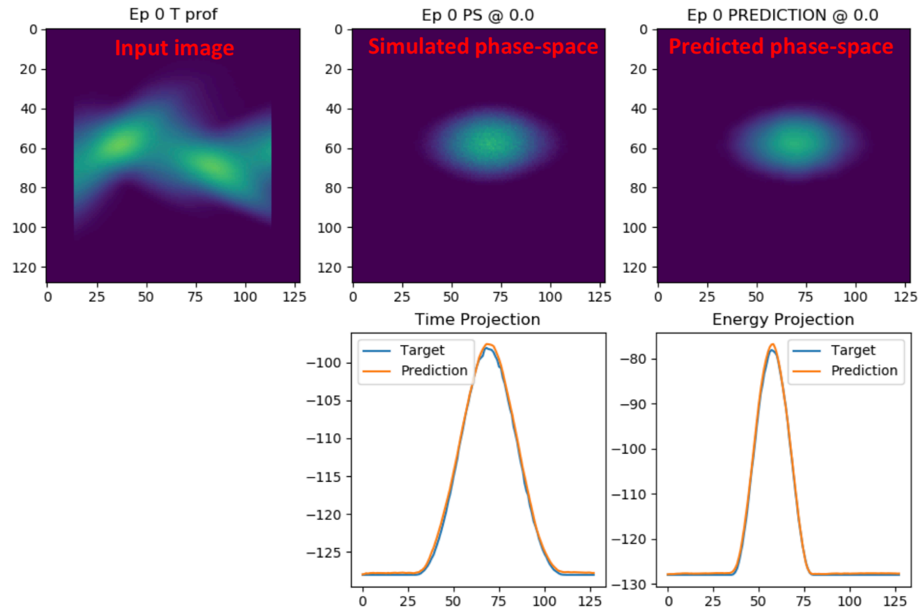
LHC bunch-by-bunch parameters from AI tomography

Longitudinal beam parameters at injection from fit of longitudinal profiles and tomography.

→ in the LHC online only possible for single bunch; too time consuming

… unless one uses ML.





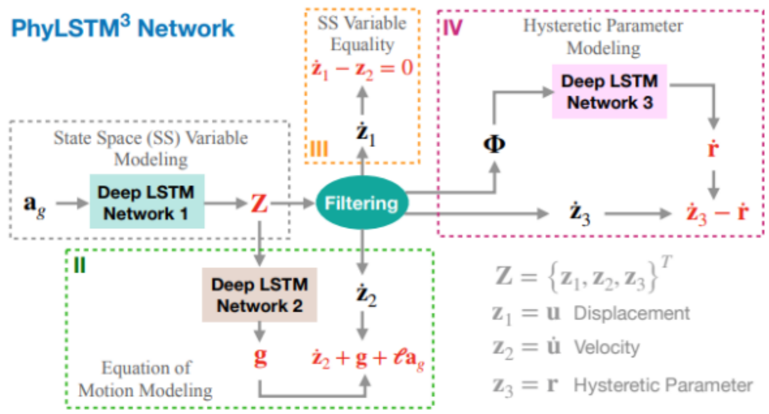
G. Trad and T. Argyropoulos

Trending and forecasting

Trending and forecasting

How to take the past into account to make sense of the present, near future or further future?

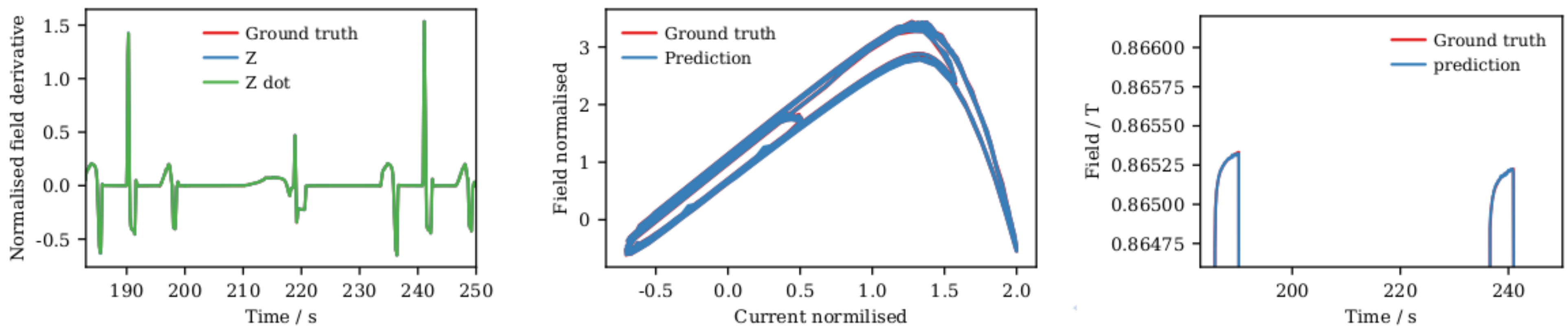
→ Recurrent networks, Physics guided NN, LSTMs, NARX,….



Goal to extend to all main types of magnets in the SPS.

Predicting the main SPS Quadrupole field depending on history

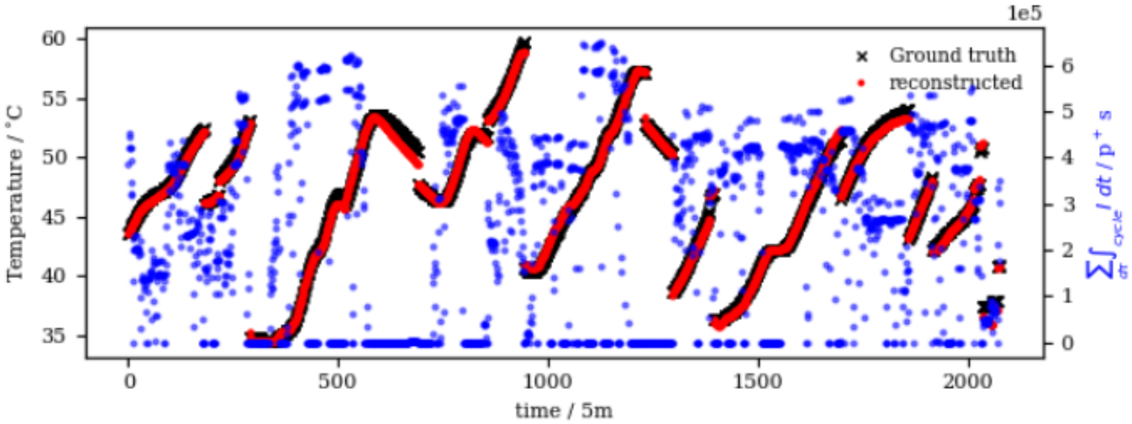
→ feed forward correction



F. Velotti, V. di Capua and M. Amodeo

Example: SPS injection kicker (MKP-L) temperature evolution prediction during scrubbing

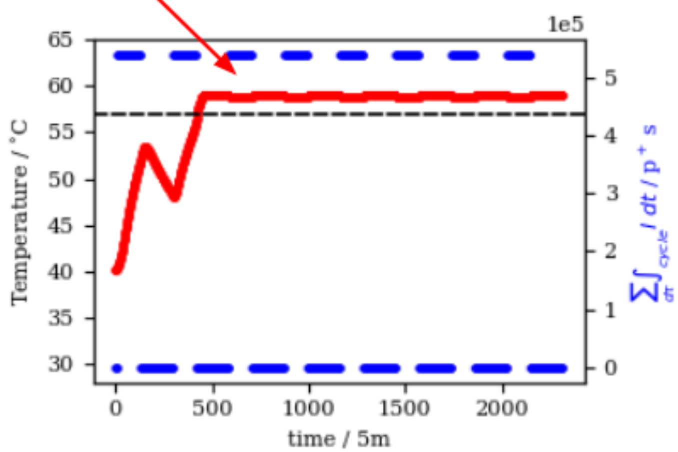
Optimise scrubbing versus cool-down time. Non-trivial…



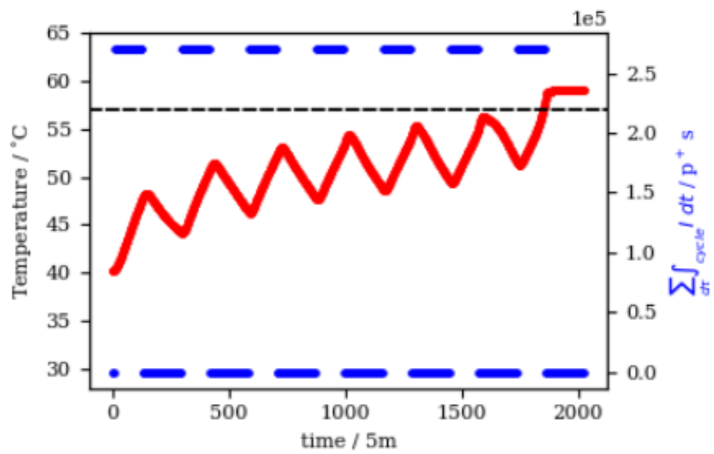
Model to predict temperature evolution of MKP-L as a function of integrated intensity in the machine…
help to guide SPS scrubbing planning during commissioning

Too much!

Case 1



Case 4



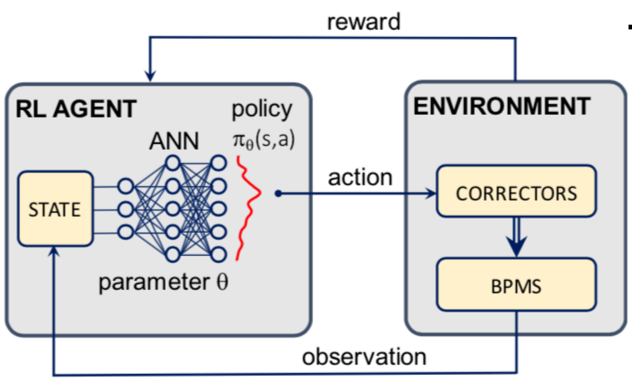
F. Velotti and B. Goddard

Reinforcement Learning

Automated tuning and optimisation.

Numerical optimisers need to explore each time again. No memory.

Algorithms exist that implicitly or explicitly learn the underlying dynamics (model) and solve the control problem at the same time → Reinforcement Learning Agents



RL paradigm for trajectory/ orbit correction

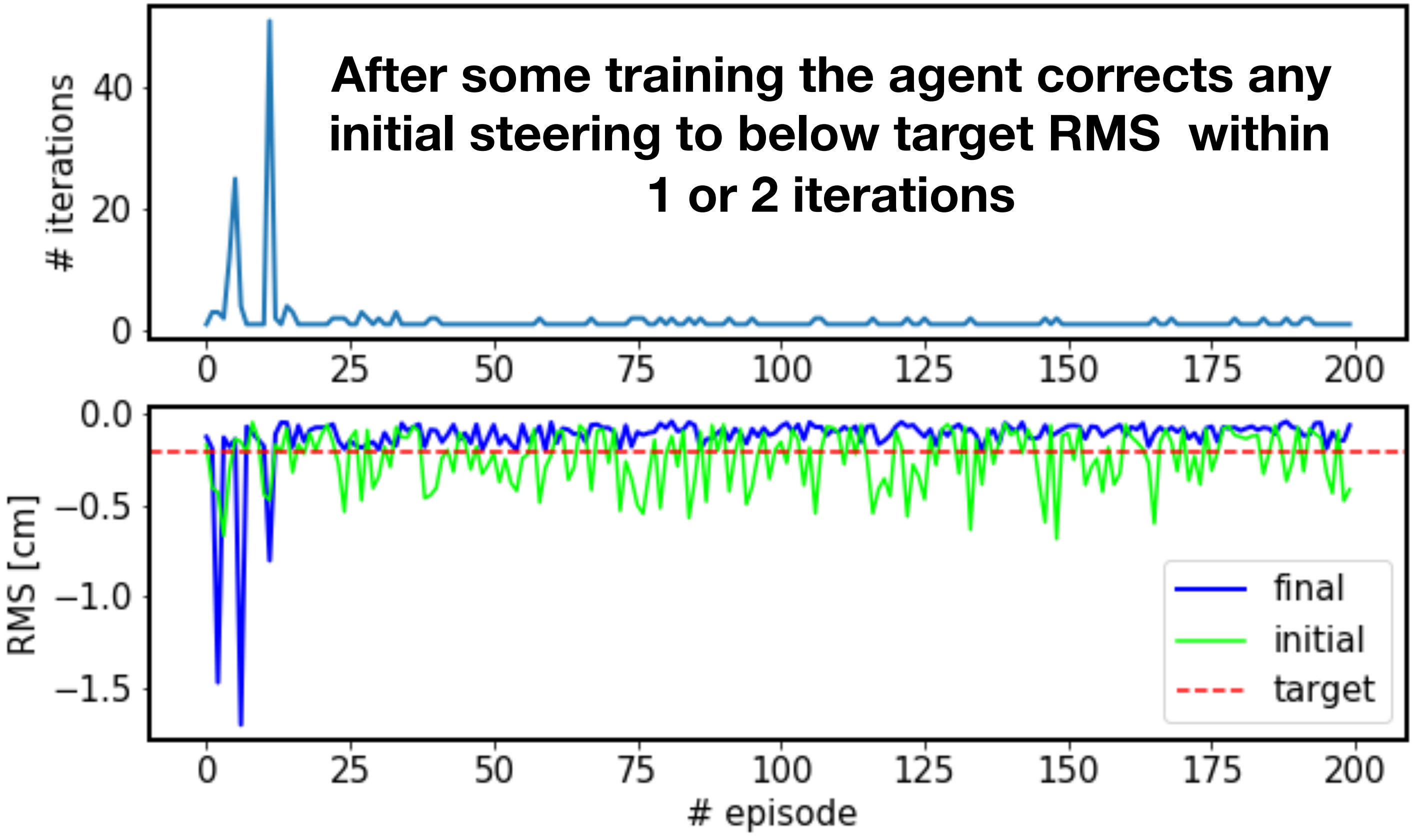
After a learning phase the agent can solve the problem within a few iterations.
It learns through **reward**.

Reinforcement Learning

Model-free online learning: sample-efficiency is key

Proof-of-principle: learn how to steer AWAKE e^{-} - line

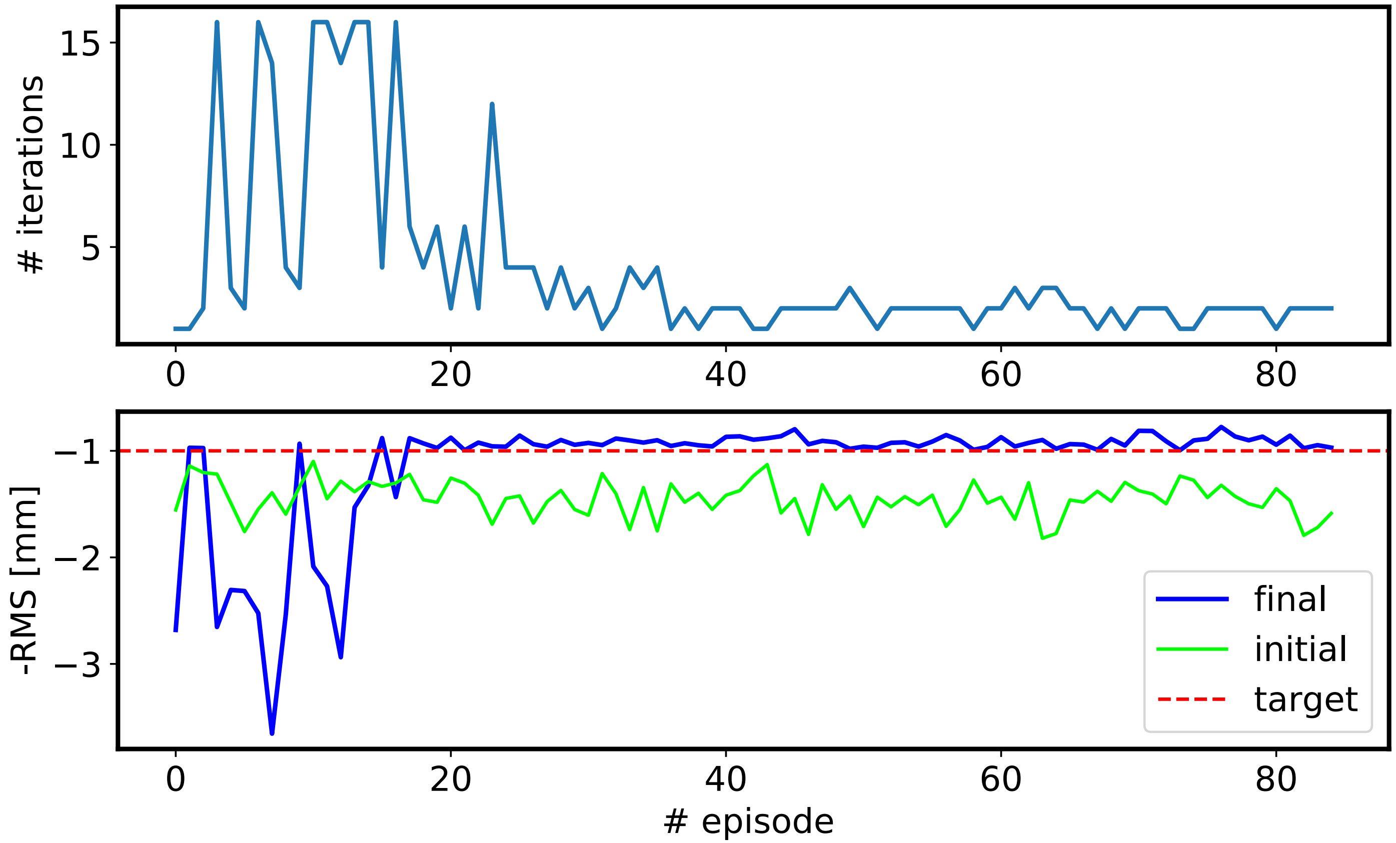
Q-learning with very sample-efficient NAF algorithm



Problem with 11 DOF

Example: agent for LINAC4 steering

Inexpensive way of learning any (also non-linear) response and solve control problem.



16 DOF

Many examples since: AWAKE auto-matching, LHC tune control, control of PS RF manipulations,…

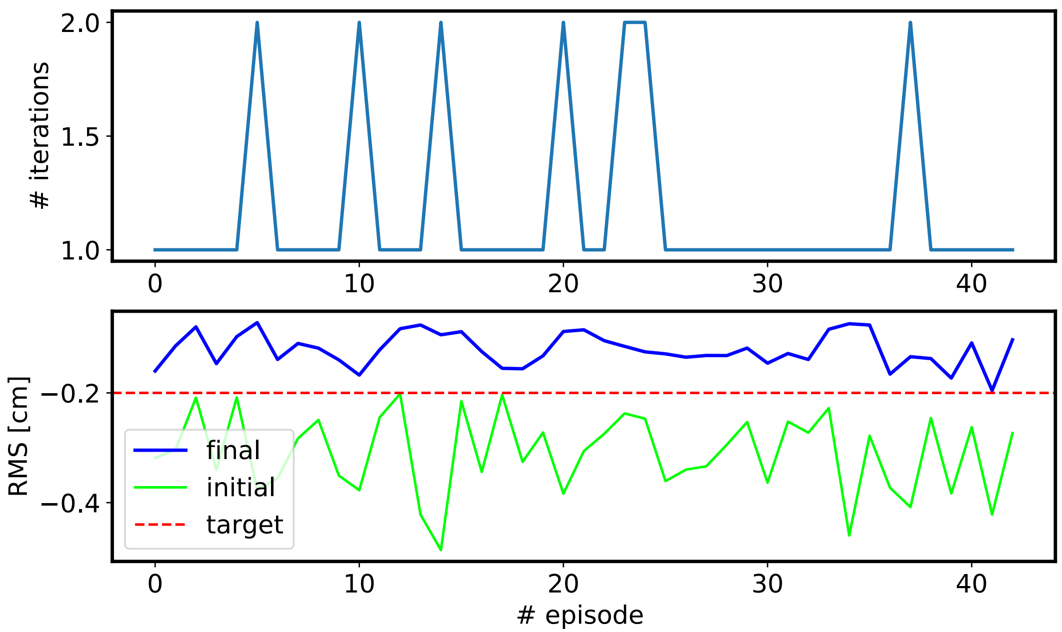
Train on simulation and apply on machine?

2 ways to circumvent the *sample-efficiency* issue.

→ Model-based RL: learn explicitly the model and train agent at the same time; tested in 2020 successfully at AWAKE

→ Train on simulation, apply on machine: typically relies on high level parameter control system

AWAKE training on simulation for trajectory steering; validation of trained agent on machine

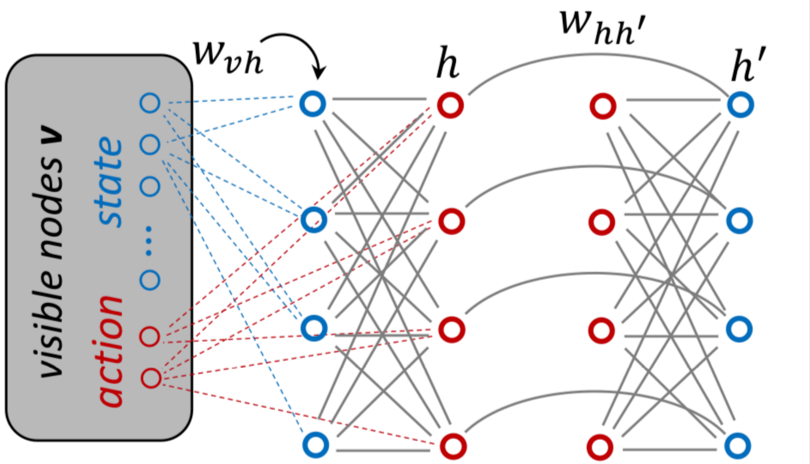


Quantum RL

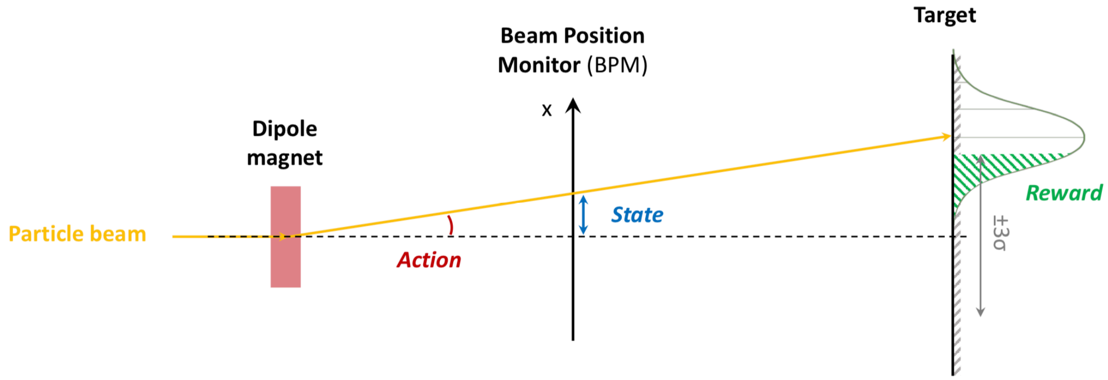
Collaboration with CERN OpenLab

M. Schenk and V. Kain

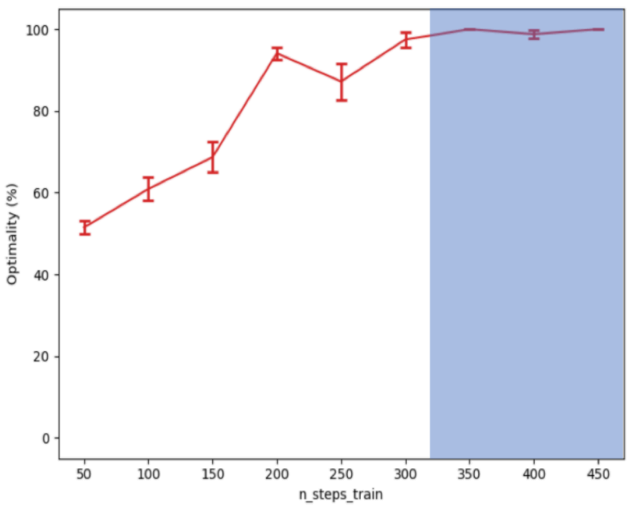
Clamped QBM



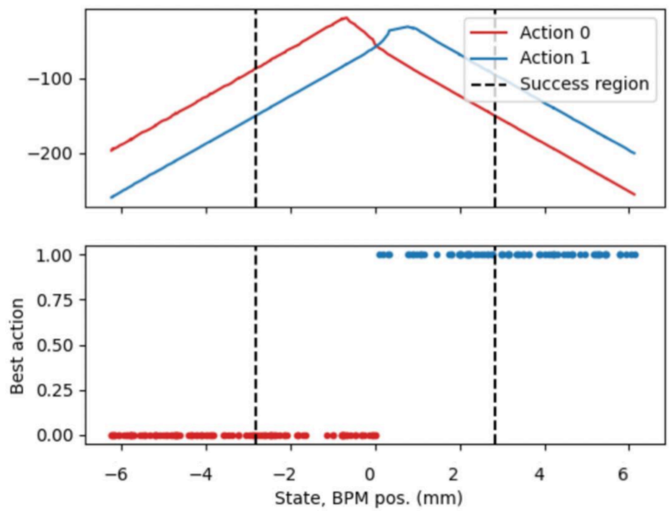
$$\hat{Q}(s,a) \approx -F(\boldsymbol{v}) = -\langle H_{\boldsymbol{v}}^{\mathrm{eff}} \rangle - \frac{1}{\beta} \sum_{c} \mathbb{P}(c|\boldsymbol{v}) \log \mathbb{P}(c|\boldsymbol{v})$$



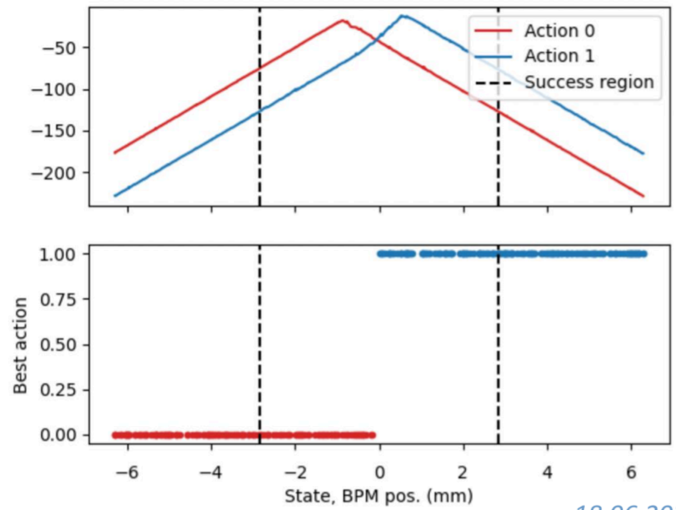
DQN (w/ experience replay, 50)
stable-baselines3, ~300 steps



FERL (w/ experience replay, 10)
SQA, ~120 steps



FERL (w/ experience replay, 7)
D-Wave train. -> SQA eval., ~120 steps



MLP

Storing and loading ANNs in the CERN control system

```python
"""Publishing new model parameters."""

from mlp_client import Client, Profile, AUTO
from my_model import MyModel

model = MyModel()
model.learn(...)

client = Client(Profile.PRO)
client.publish_model_parameters_version(
    model,
    name="TD3-on-SPS-ZS-Alignment",
    version=AUTO,  # Generate new version on server side!
)
```

```python
model = client.create_model(
        model_class=MyModel,
        params_name="sps_sftpro",
        params_version=AUTO,
)
```

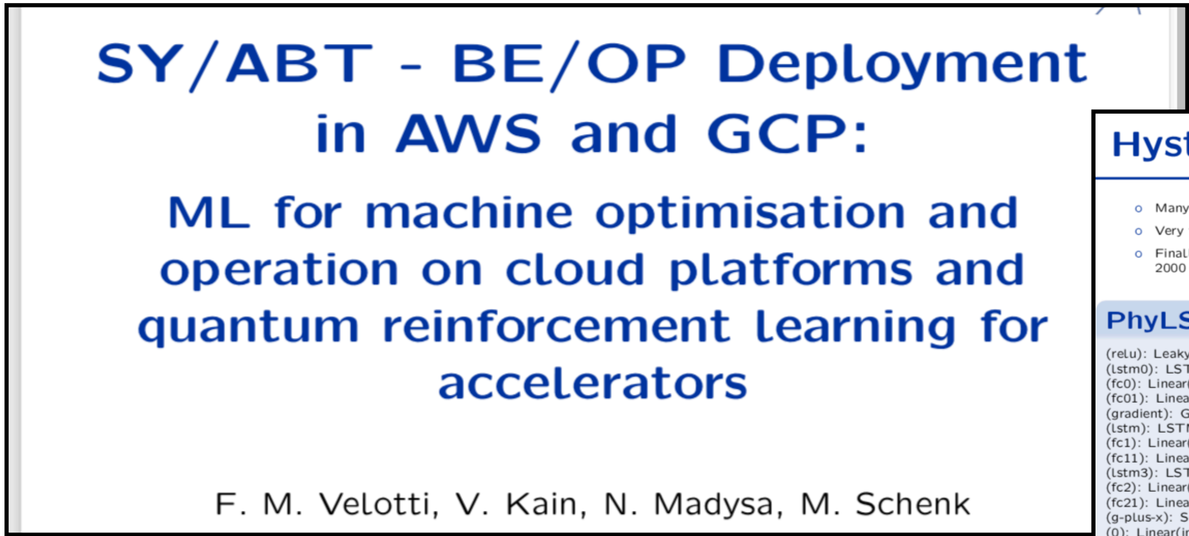
R.Gorbonosov, J.B. de Martel, N. Madysa

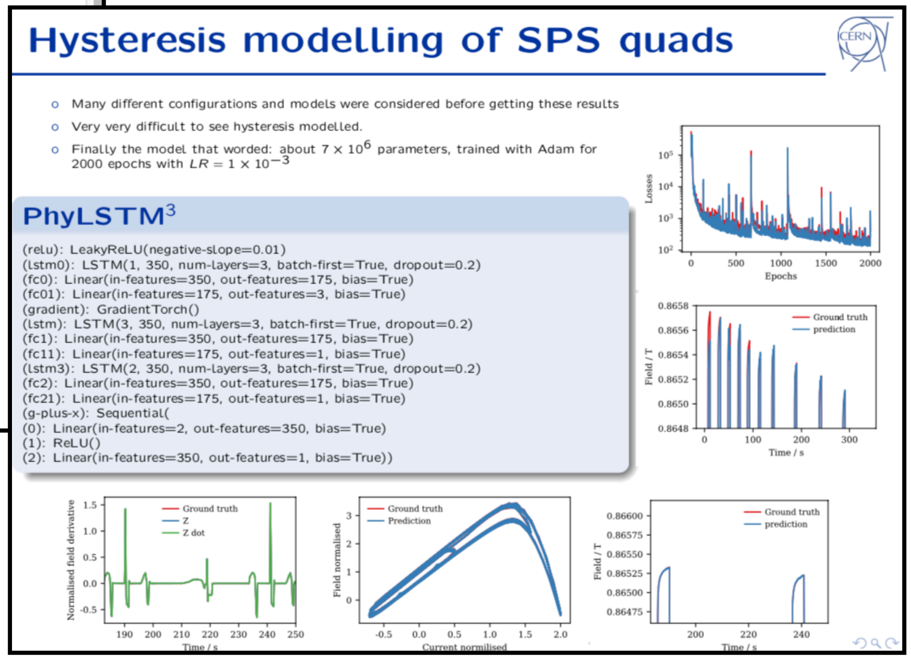
CloudBank

Members of ML community forum participated at cloud broker pilot project with CloudBank using GCP and AWS.

→ access to GPUs and quantum computers

CERN Pilot - Kain	20201016-kain	2020-10-16	2021-06-30	$0.00	$0.00 (100%)	Verena Kain





Status

Many of our accelerator problems, that were intractable/ needed lots of resources to solve are “straight forward” to solve with ML or other advanced algorithms.

→ successful demonstration in number of examples

Requires expertise and embracing new technologies.

→ ML community forum (former ML coffee) helped to establish sufficient expertise in many ML domains as well as frameworks and infrastructure.

Next step: self-driving accelerator project?