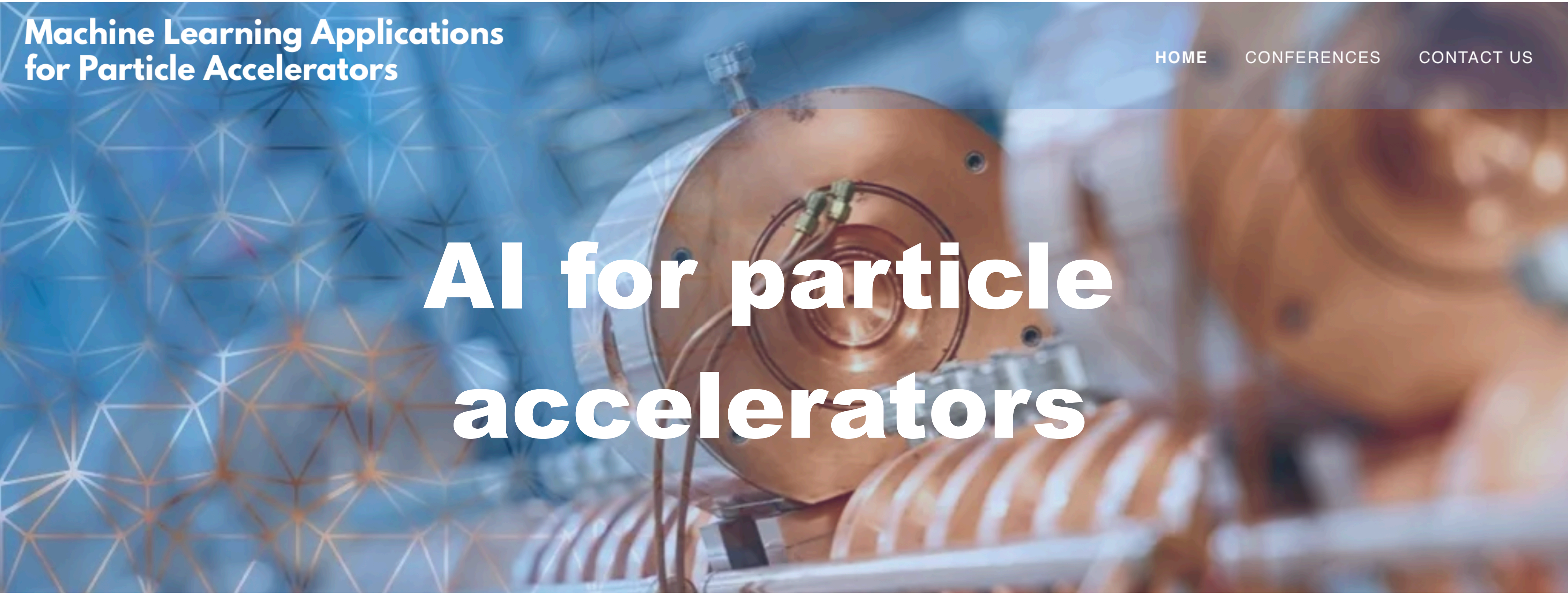
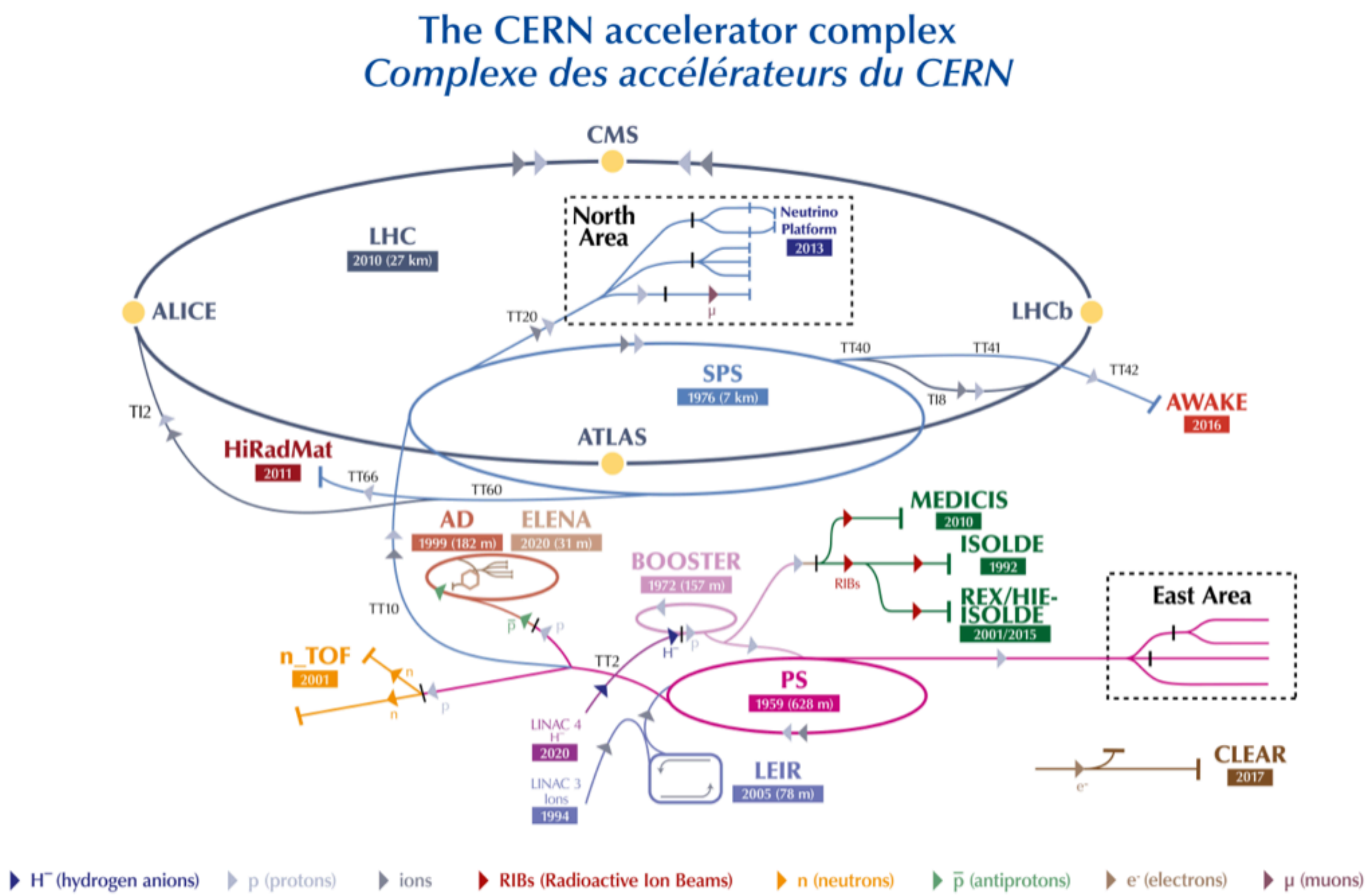

Machine Learning Applications
for Particle Accelerators
HOME
CONFERENCES
CONTACT US
AI for particle
accelerators

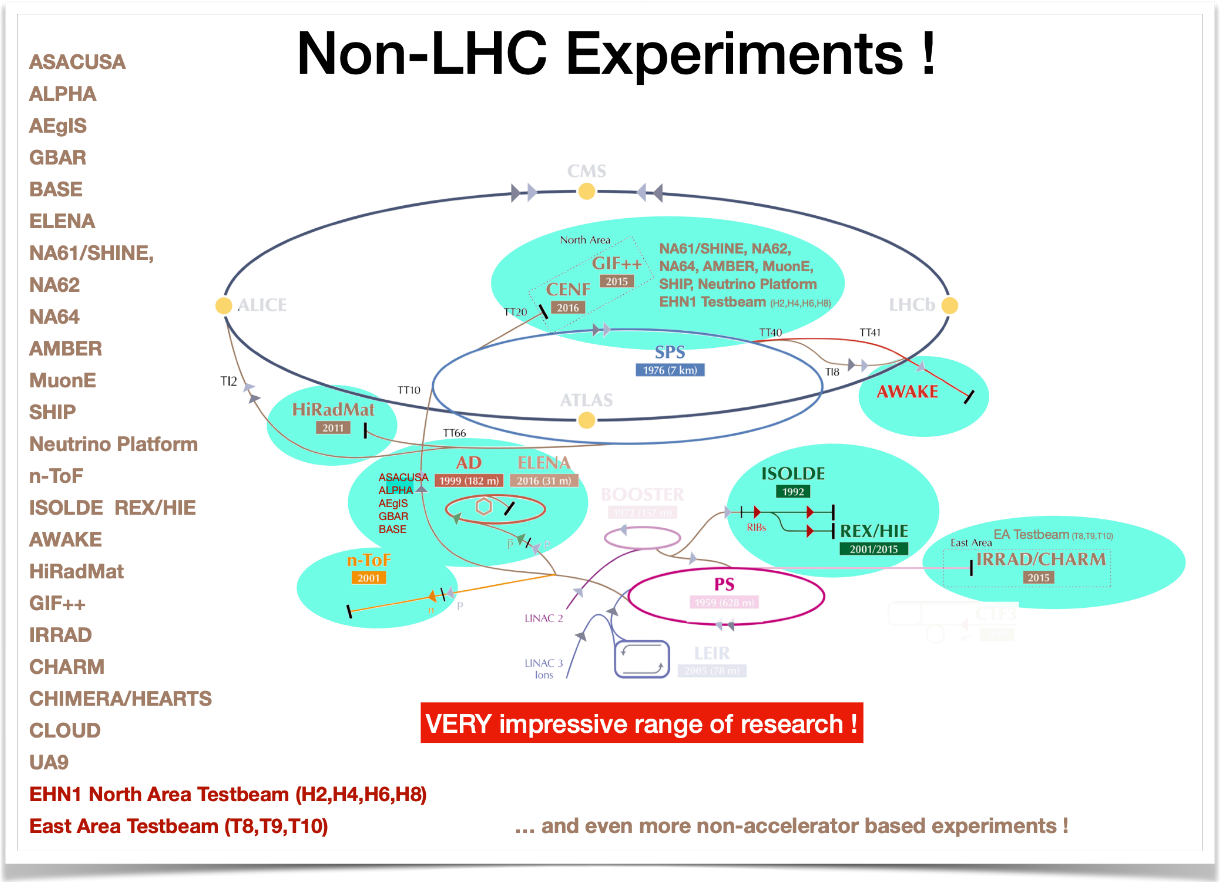
Motivation

The CERN accelerator complex



LHC and non-LHC physics

Not only LHC physics! → Many different beam types, production schemes,…



Screenshot from recent Chamonix CERN Accelerator Performance workshop Jan '24

Flexibility comes at a prize. Examples…

Summary talk *CERN Injector and Experimental Facility Workshop (IEF) '21*

2. Address reproducibility and availability

- Availability OK, under control of Groups. **Reproducibility** is critical concern with increasing flexibility and multi-destination operation
- Transmission problems and instability in beam delivery in many locations. "Need more time in 2022"→have to ensure this is there (add in schedule?) #A
- Addressing reproducibility relies on many factors including equipment, accelerator modelling and high-level controls approach

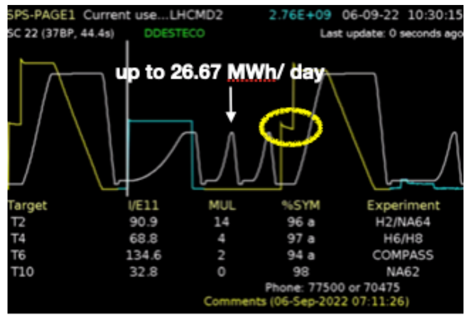
→ Current **beam scheduling** has severe impact on resources needed to run accelerators and on efficiency

- ✳ Statistics: 20-100 clicks to change supercycle = 2-25 min; 40-60 times/24 h

Input from CERN Joint Accelerator Performance Workshop'22

→ **Hysteresis** is severe limitation for efficiency and flexibility in most machines, current mitigation methods wasting energy

- ✳ ~ 15 % of yearly cost of SPS fixed target cycle for "waste" cycles and quasi-degauss Cycle MD1



Future accelerators? Like FCC...

The ***business-as-usual*** solution: FCC just larger LHC

- Brute force scale-up → using helicopters to reduce intervention times, more people, more sites,...
- (Financially excluded, luckily)

Future accelerators? Like FCC...

The ***business-as-usual*** solution: FCC just larger LHC

- Brute force scale-up → using helicopters to reduce intervention times, more people, more sites,...
- (Financially excluded, luckily)

The ***elegant*** solution: FCC to be run like a space telescope.

- Reinvent exploitation paradigm: hierarchical autonomous systems
- AI is key technology
- Management's preferred option

Future accelerators? Like FCC...

The ***business-as-usual*** solution: FCC just larger LHC

- Brute force scale-up → using helicopters to reduce intervention times, more people, more sites,...
- (Financially excluded, luckily)

The ***elegant*** solution: FCC to be run like a space telescope.

- Reinvent exploitation paradigm: hierarchical autonomous systems
- AI is key technology
- Management's preferred option

Need to use the next 10 years as test bed to be ready with adequate design choices

What can AI do for accelerators?

The vision…

Autonomous accelerators

Key words: optimal control and optimisation, anomaly detection and preventive maintenance, differentiable simulations, virtual diagnostics,…

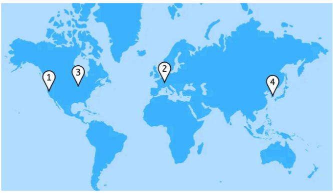
Optimised accelerator design

Key words: fast-executing simulations for optimisation algorithms, differentiable simulations,…

Generic AI for efficient research and development

Key words: AI assistants for code development, knowledge retrievable,…

Workshop series for ML for particle accelerators since 2018



4TH MACHINE LEARNING APPLICATIONS FOR PARTICLE ACCELERATORS (2024), GYEONGJU, SOUTH KOREA. HOSTED BY PAL

3RD ICFA BEAM DYNAMICS MINI-WORKSHOP ON MACHINE LEARNING APPLICATIONS FOR PARTICLE ACCELERATORS (2022), CHICAGO, USA. HOSTED BY BNL.

2ND ICFA MINI-WORKSHOP ON MACHINE LEARNING FOR CHARGED PARTICLE ACCELERATORS (2019), VILLIGEN, SWITZERLAND. HOSTED BY PSI.

1ST MACHINE LEARNING APPLICATIONS FOR PARTICLE ACCELERATORS (2018), MENLO PARK, USA. HOSTED BY SLAC.

Workshop 2025 will be at CERN!

Topics of ML workshop '24

We are pleased to announce the **4th ICFA Beam Dynamics Mini-Workshop on Machine Learning for Particle Accelerators** will be held in ***Gyeongju, South Korea***. The goal of this workshop is to help build a world-wide community of researchers interested in applying machine learning techniques to particle accelerators.

The workshop will consist of six topics:

1. Analysis & Diagnostics
2. Anomaly Detection / Failure Prediction
3. Infrastructure / Deployment Workflows
4. Optimization & Control
5. Modeling Approaches
6. Lessons Learned

Tutorials:

1. Reinforcement Learning
2. Model Adaptation / Up-keep
3. Transformers for Timeseries Prediction

Talks will include both accelerator physicists and computer scientists. This workshop has the following goals:

- Collect and unify the community's understanding of the relevant state-of-the-art ML techniques.
- Provide a simple tutorial of machine learning for accelerator physicists and engineers.
- Seed collaborations between laboratories, academia, and industry.

Please contact the organizers if you are interested in attending.

Accelerator ML community - Trends...

Focus shifting slowly from R&D only to AI at scale with full life cycle management.

→ Infrastructure/Deployment Workflows (MLOps) one of the longest sessions at last workshop

→ Discussion about standards: e.g. optimisation problem definition standards

→ Non-trivial life cycle management questions becoming important: "continual learning"

- Full tutorial about it.

The big new theme: LLMs...PACuna, Logbook search, AI assistants in the control room

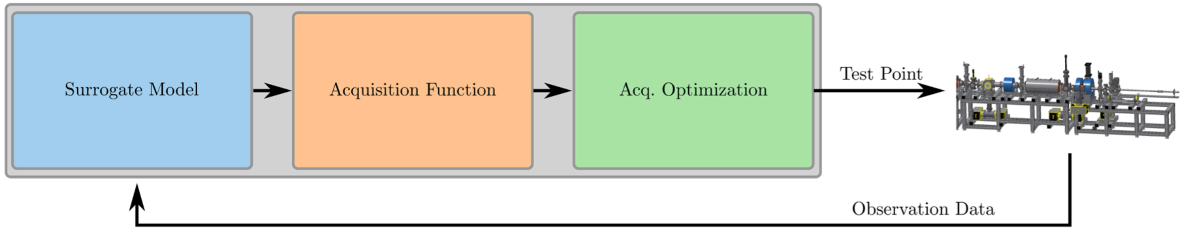
Progress on all fronts, will focus on **optimisation and control** in the next slides.

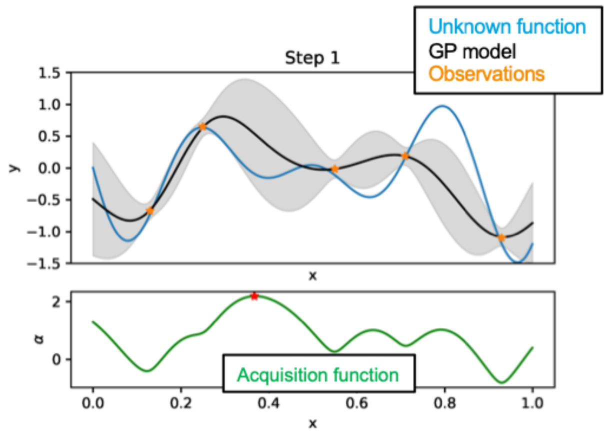
Bayesian Optimisation (BO)

Majority of talks in "optimisation and control" session about BO

Bayesian Optimisation (BO)

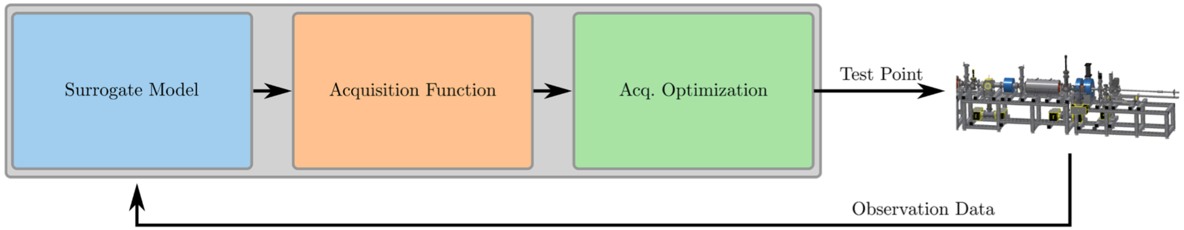
Majority of talks in "optimisation and control" session about BO

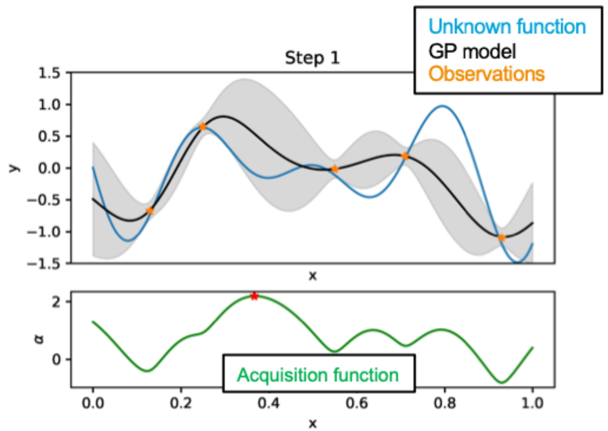




Bayesian Optimisation (BO)

Majority of talks in "optimisation and control" session about BO





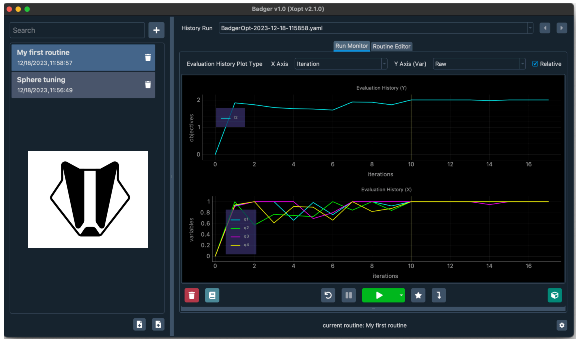
Impressive expertise in the community!

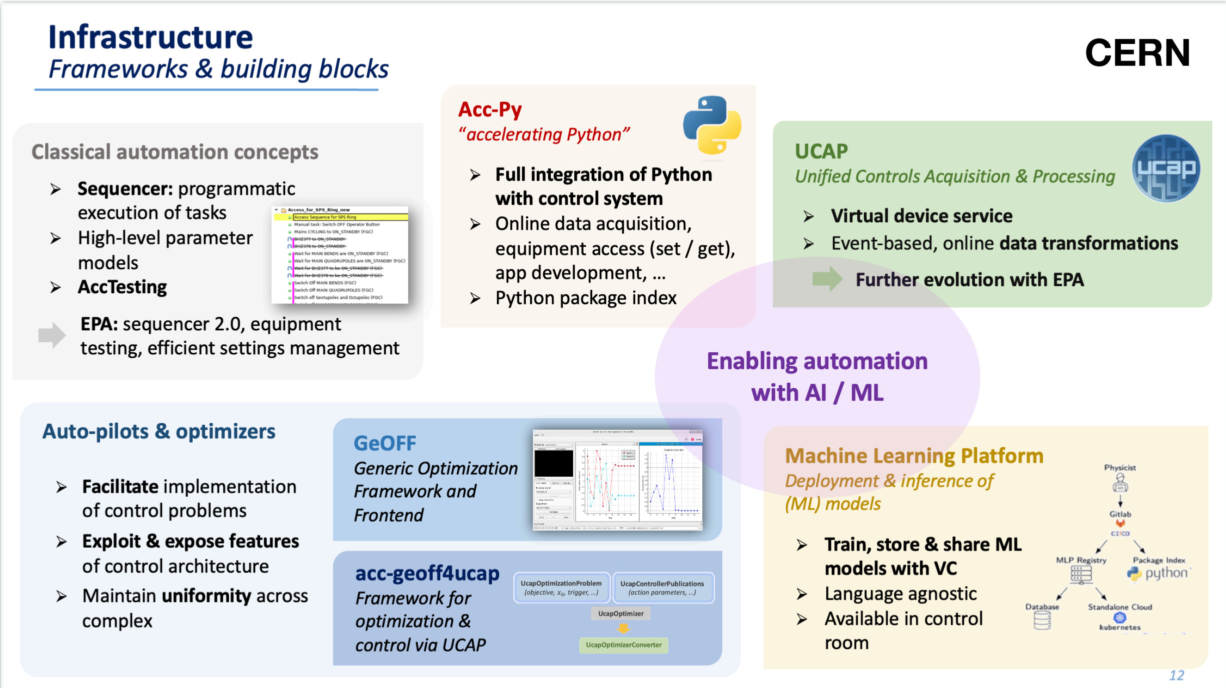
State-of-the-art BO:

- Model-based priors for various applications
- Multi-fidelity BO for laser plasma accelerators
- SafeOpt to include safety constraints - faster convergence with ModSafeOpt
- Information-based Bayesian Optimisation with virtual objectives
- …

Common tools and frameworks were key

SLAC





AI for particle accelerators, EuCAIF, V. Kain, 01-May-2024

State-of-the-art BO with **BoTorch**

- GPU accelerated
- versatile
- fully integrated with PyTorch, GPyTorch

BoTorch
Bayesian Optimization in PyTorch

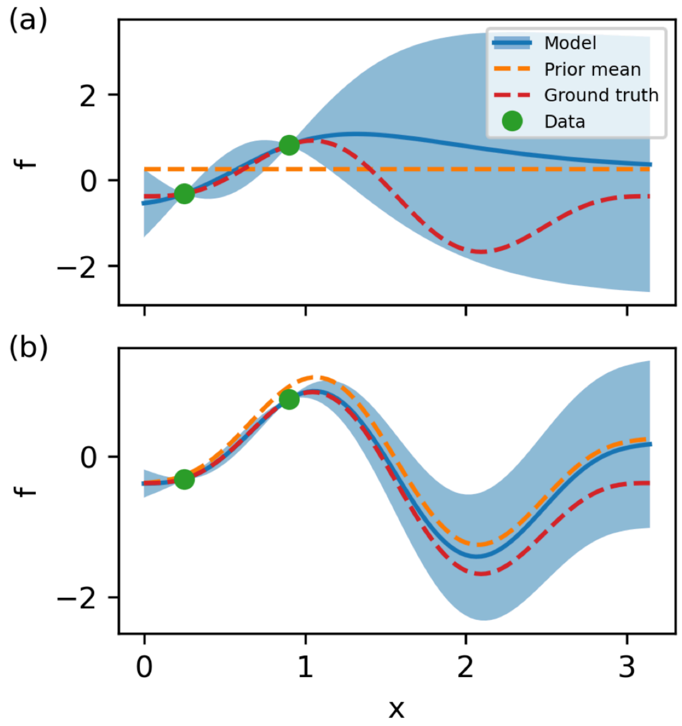
Example: BO with neural network mean priors

Including prior information through historical data not efficient with GPs.

→ Modify $p(A)$ in $p(A\,|\,B) \propto p(B\,|\,A)p(A)$

Instead of constant prior in GP → GP becomes model of model → much more sample-efficient

Model of non-constant mean $\mu(\mathbf{x})$ from simulation, ANN of historical data, other GP,…

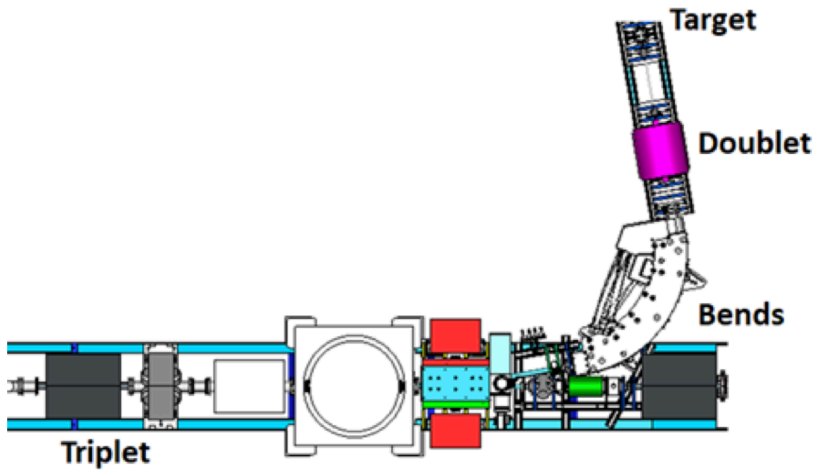


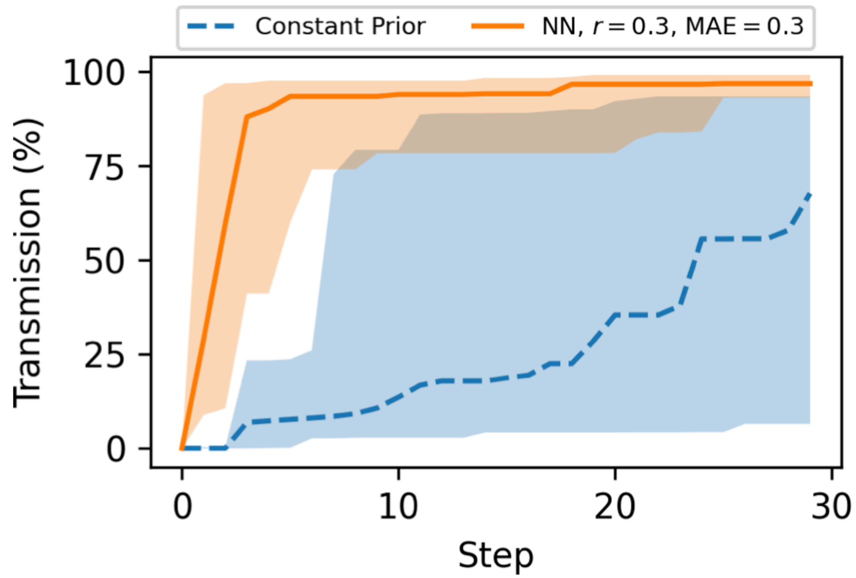
Example: BO with neural network mean priors

Example from ATLAS

- **Argonne Tandem Linear Accelerator System** for study of low-energy nuclear physics with heavy ions
- Optimise transmission to target: 5 DOF
  - Trained ANN from previous 3k dataset of ^{14}N run and used it for optimising ^{16}O run

Argonne Tandem Linear Accelerator System is a US Department of Energy User Facility dedicated to the study of low-energy nuclear physics with heavy ions





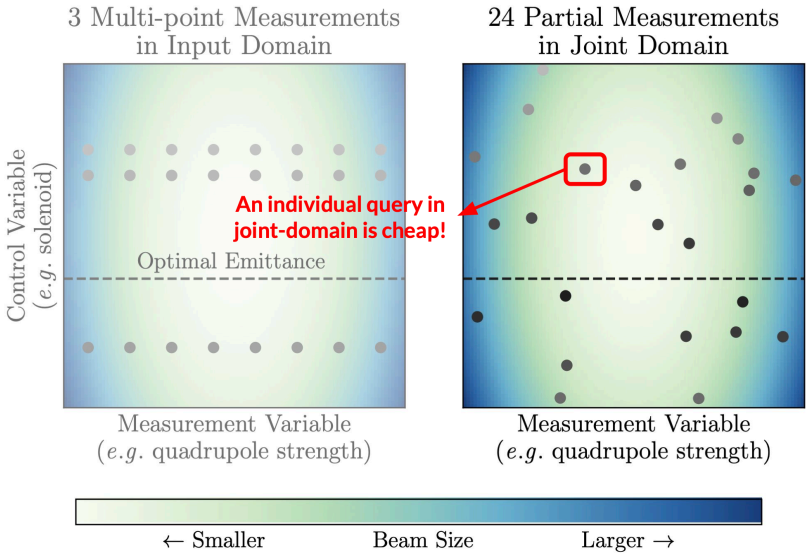
Courtesy T. Boltz et al, arXiv:2403.03225

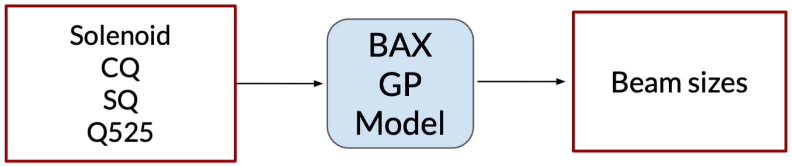
Example: info-based BAX - virtual objectives

Example for emittance tuning

- BAX (**B**ayesian **A**lgorithm e**X**ecution): https://willieneis.github.io/bax-website/
- Minimum emittance important for many applications: e.g. determines brightness of X-rays in FELs
- Classical methods slow due to multi-point queries: quadrupole scans for emittance evaluation

Courtesy S. Miskovich et al



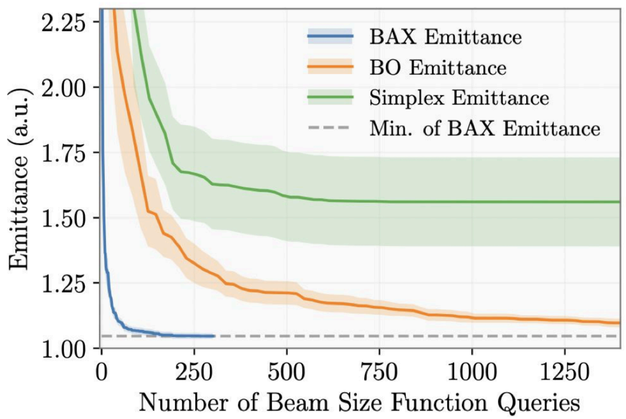


→ Virtual emittance scans on posterior samples of GP as input to acquisition function optimisation

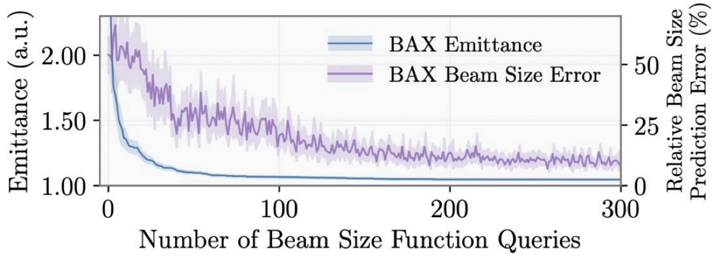
→ choose queries (settings) to maximise information and minimise emittance

Example: info-based BAX - virtual objectives

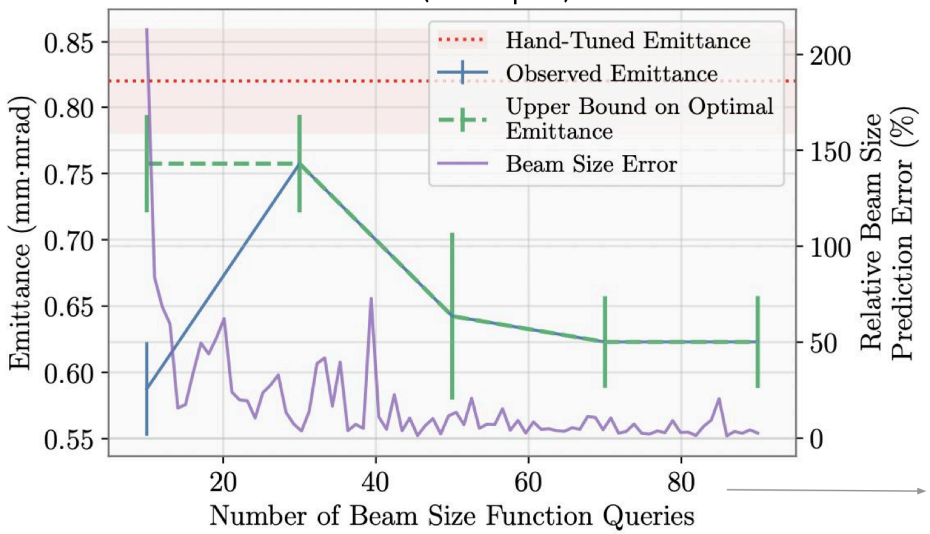
Example from noisy LCLS simulation @ SCLAC



Proxy for convergence



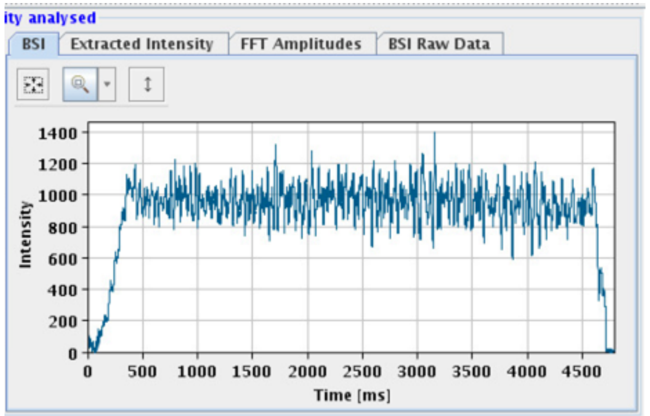
Live Optimization at LCLS
(250 pC)



Courtesy S. Miskovich et al

Example: Continuous control with BO → ABO

Controlling the $n \times 50$ Hz noise in the slow extracted spill to the North Area Experimental Hall at CERN SPS.



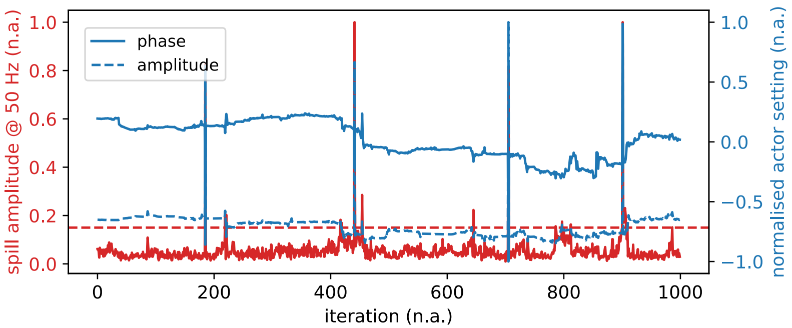
Modulate voltage of main quadrupoles at $n \times 50$ Hz to compensate.

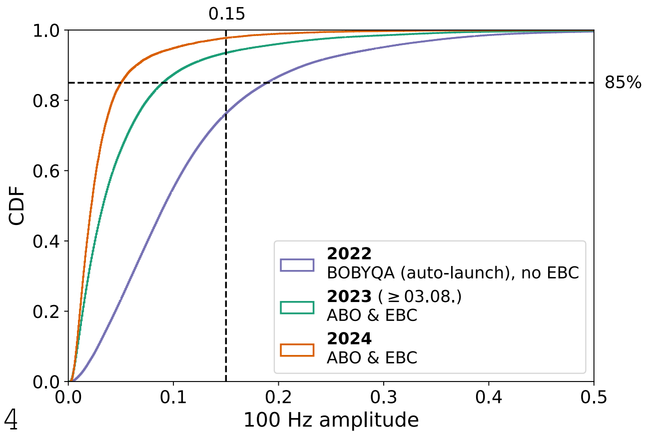
Spill noise changes over time following the European grid.

→ adaptive continuous control, **A**daptive **B**ayesian **O**ptimisation

Model objective function f as $f(\mathbf{x}, t)$. Spectral mixture kernel S for t and Matern for control parameters

Kernel: $k([t_1, \mathbf{x_1}], [t_2, \mathbf{x_2}]) = \theta_k \times S(t_1, t_2) \times M(\mathbf{x_1}, \mathbf{x_2})$



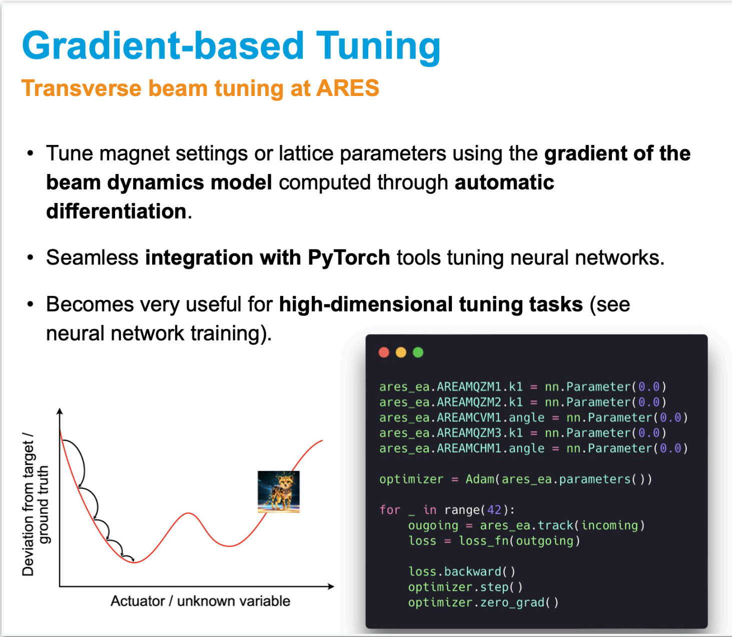


Will be key for:

SHiP
Search for Hidden Particles

Differentiable simulation codes

Optimisation algorithms work best and are most sample-efficient with gradient information of the objective function.



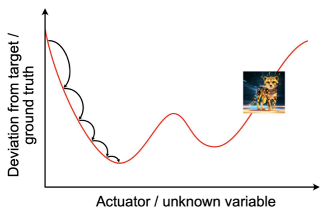
Differentiable simulation codes

Optimisation algorithms work best and are most sample-efficient with gradient information of the objective function.

Gradient-based Tuning

Transverse beam tuning at ARES

- Tune magnet settings or lattice parameters using the **gradient of the beam dynamics model** computed through **automatic differentiation**.
- Seamless **integration with PyTorch** tools tuning neural networks.
- Becomes very useful for **high-dimensional tuning tasks** (see neural network training).



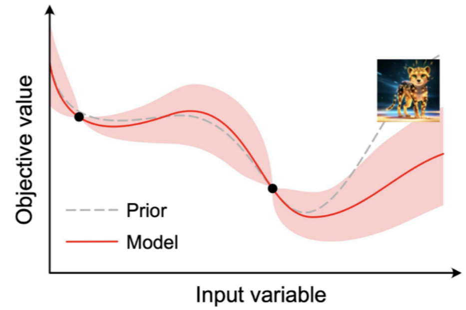

```
ares_ea.AREAMQZM1.k1 = nn.Parameter(0.0)
ares_ea.AREAMQZM2.k1 = nn.Parameter(0.0)
ares_ea.AREAMCVM1.angle = nn.Parameter(0.0)
ares_ea.AREAMQZM3.k1 = nn.Parameter(0.0)
ares_ea.AREAMCHM1.angle = nn.Parameter(0.0)

optimizer = Adam(ares_ea.parameters())

for _ in range(42):
    ougoing = ares_ea.track(incoming)
    loss = loss_fn(outgoing)

    loss.backward()
    optimizer.step()
    optimizer.zero_grad()
```

- A physics-informed prior can help **improve the performance of BO** by preventing over-exploitation.
- Cheetah's differentiability allows **efficient acquisition function optimisation** using gradient descent methods in modern BO packages like BoTorch.
- Has well-defined behaviour and **does not need data** to train like neural network priors.
- Can be used in **combination with gradient-based system identification** to overcome model inaccuracies.

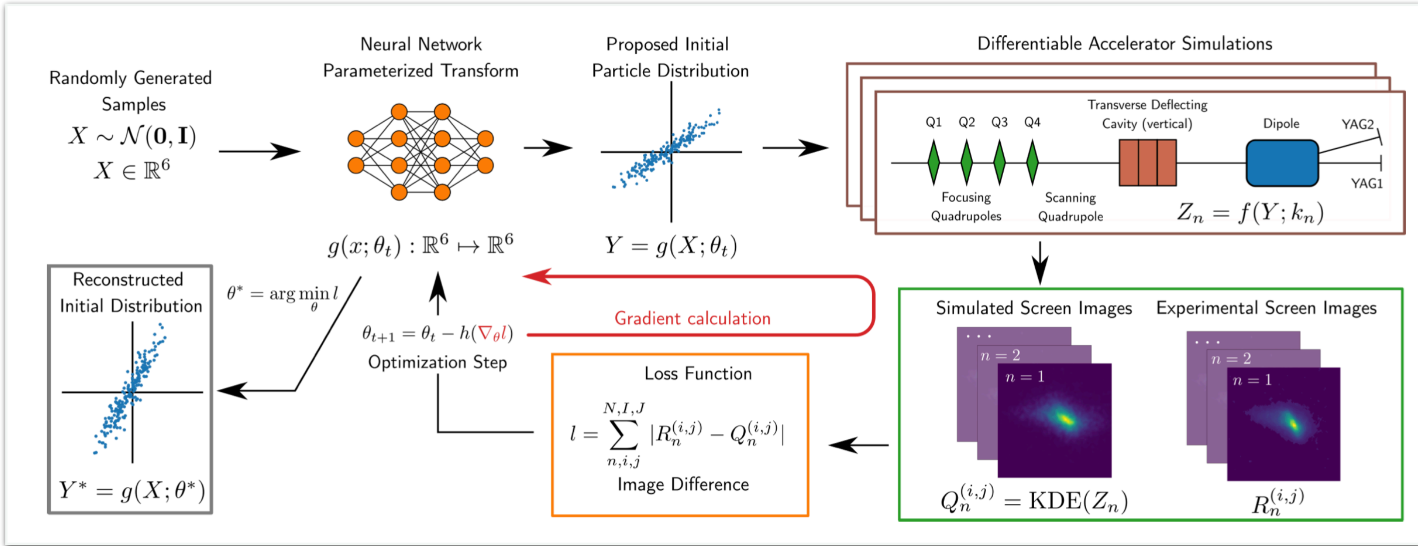


If you have differentiable codes….

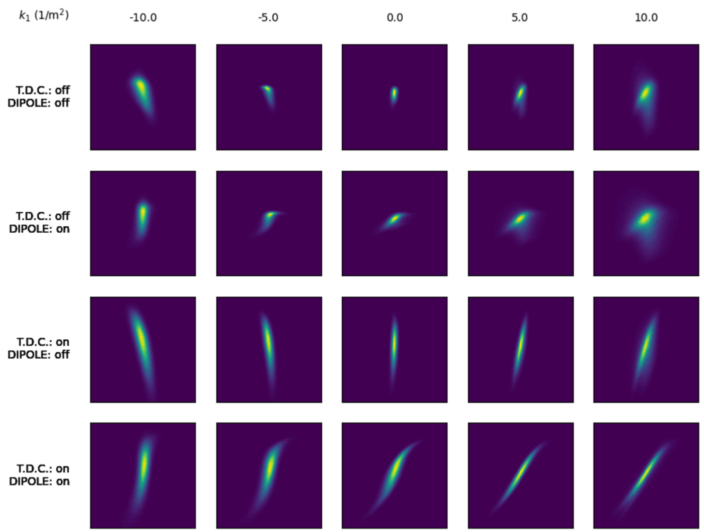
Example: generative phase-space reconstruction in 6D

These measurements are normally rarely done, too time consuming…

E.g. Spallation Neutron Source (SNS): 5×10^6 measurements over 36 h



Courtesy R. Roussel et al arXiv:2404.10853



Tested at Argonne Wakefield Accelerator (AWA):
only 20 measurements for full 6 D reconstruction.

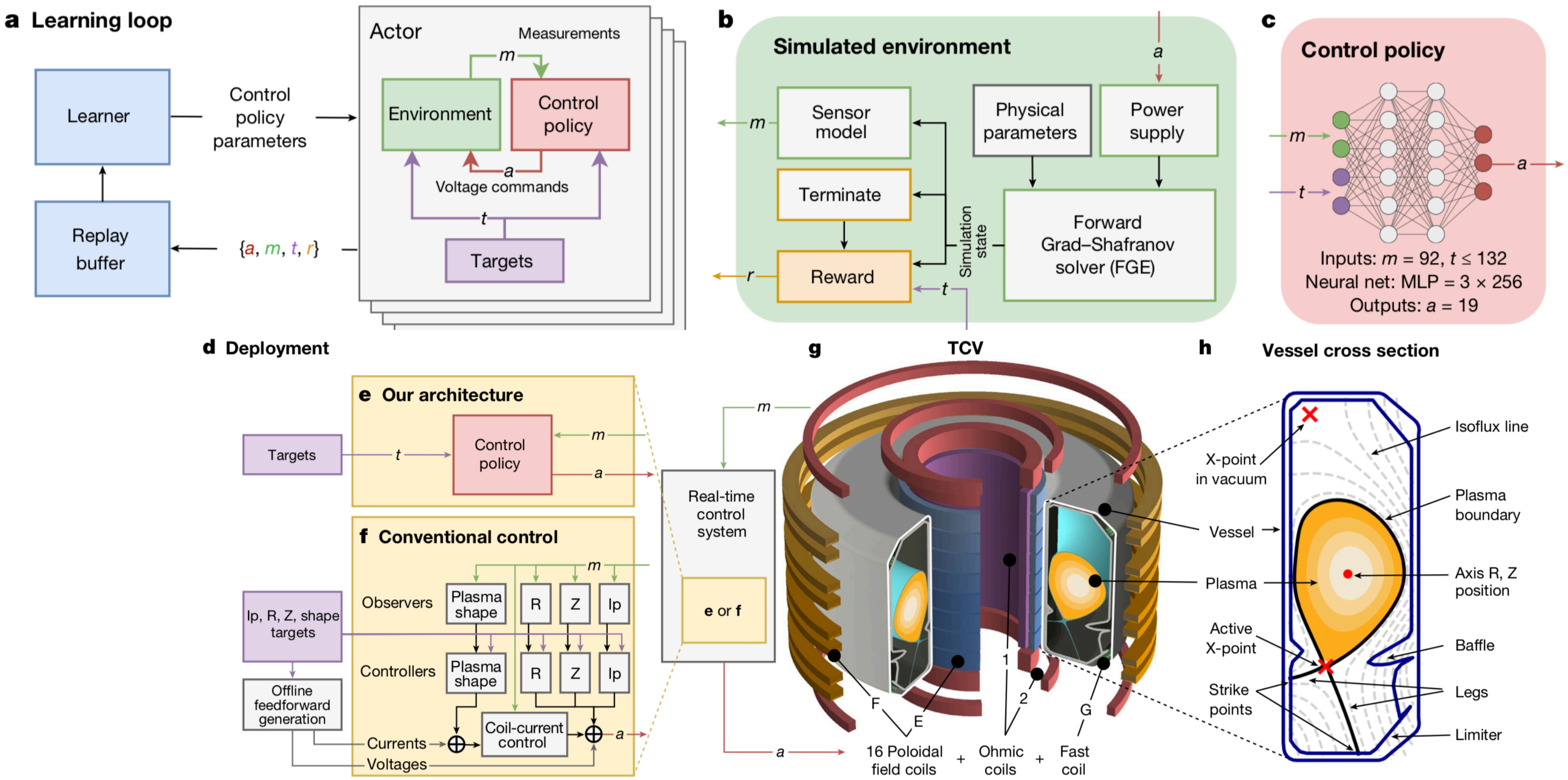
Reinforcement Learning (RL)

Reinforcement Learning (RL)

Magnetic control of tokamak plasmas through deep reinforcement learning

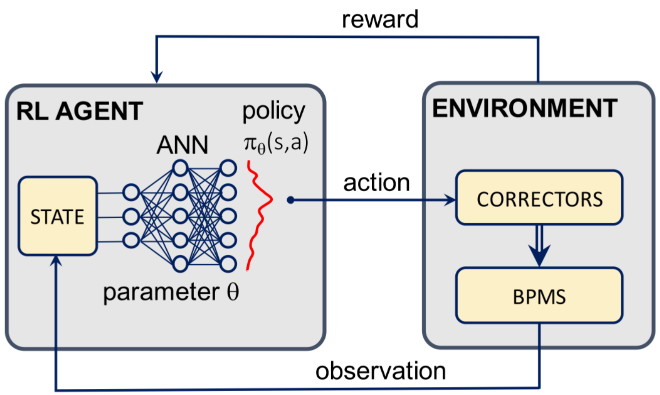
Time-varying, non-linear, multi-variate control problem

https://doi.org/10.1038/s41586-021-04301-9



Reinforcement Learning (RL)

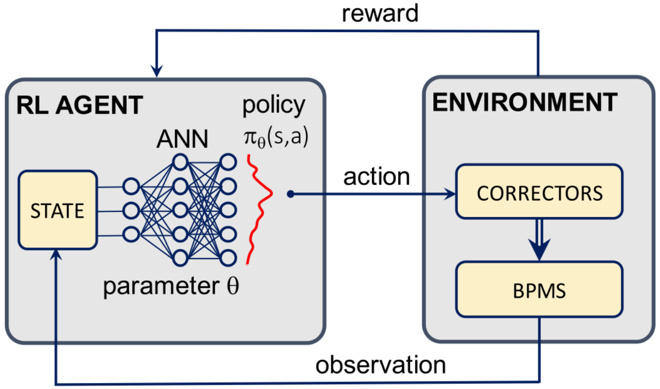
Learn dynamics (once and for all) through trial-and-error, no exploration after training!



RL setup for trajectory steering

Reinforcement Learning (RL)

Learn dynamics (once and for all) through trial-and-error, no exploration after training!



RL setup for trajectory steering

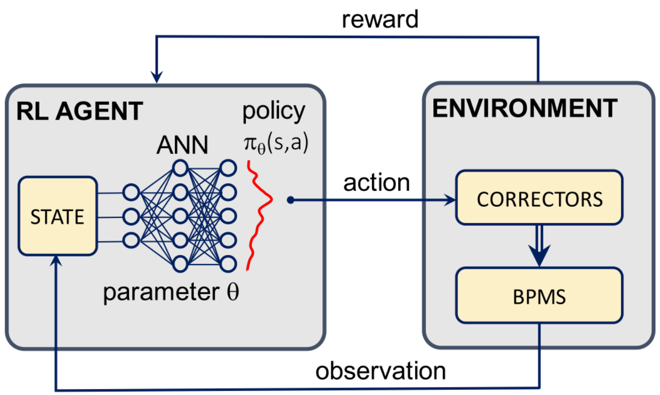
RL elegant (if not ideal) solution, but **online training** often not possible!

- Not sample-efficient enough
- Safety constraints

→ **RL (like MPC) needs to be built into accelerator design**.

Reinforcement Learning (RL)

Learn dynamics (once and for all) through trial-and-error, no exploration after training!



RL setup for trajectory steering

Next generation accelerators to be built <u>for</u> RL:

→ fast executing (accurate) simulation / digital twin
for training

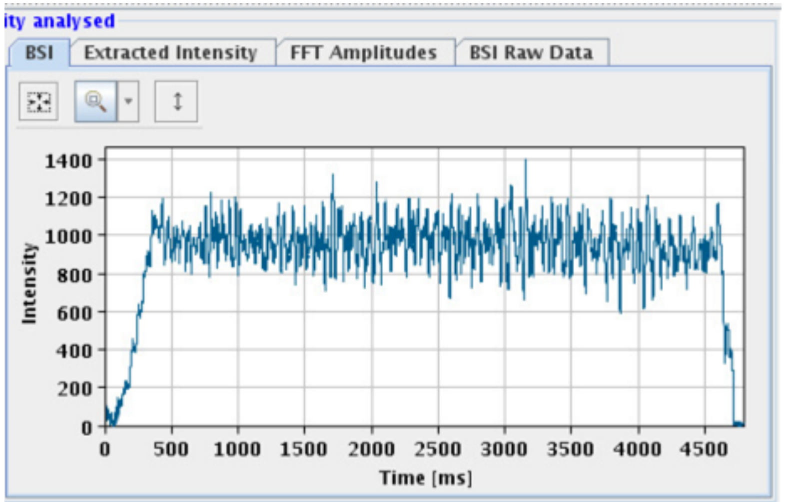
→ instrumentation designed with control algorithm

RL elegant (if not ideal) solution, but **online training** often not possible!

- Not sample-efficient enough
- Safety constraints

→ **RL (like MPC) needs to be built into accelerator design**.

Example: RL for $n \times 50$ Hz control at CERN SPS?



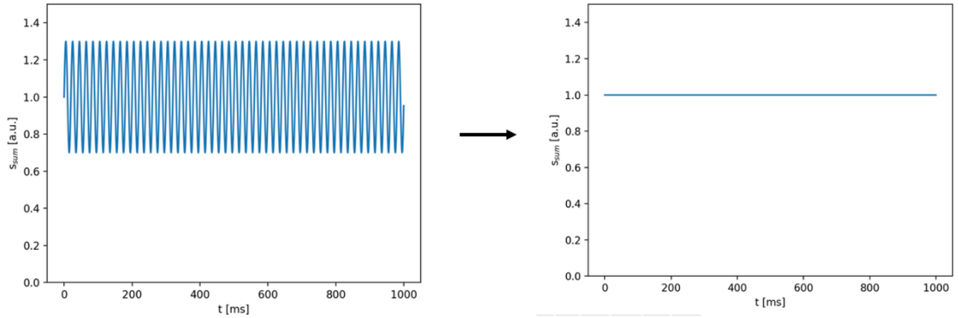
RL for $n \times 50$ Hz noise control?

Simulated environment:

$$\vec{s} = [A_{spill}, \phi_{spill}, A_{corr}, \phi_{corr}]$$

$$r = -\sqrt{A_{noise}^2 + A_{corr}^2 + 2A_{noise}A_{corr}\cos\Delta\phi}$$

Spill monitor signal (2kHz) for 1 s SHiP cycle



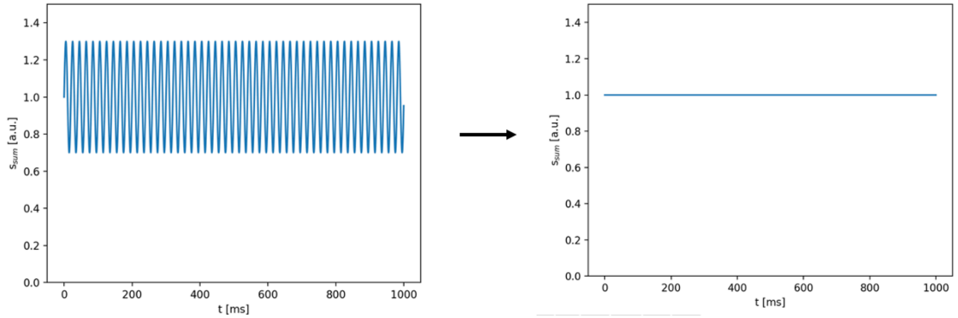
RL for $n \times 50$ Hz noise control?

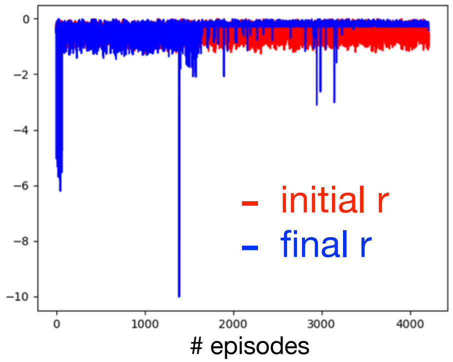
Simulated environment:

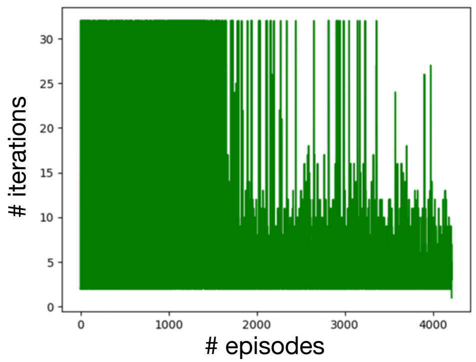
$$\vec{s} = [A_{spill}, \phi_{spill}, A_{corr}, \phi_{corr}]$$

$$r = -\sqrt{A_{noise}^2 + A_{corr}^2 + 2A_{noise}A_{corr} \cos \Delta\phi}$$

Spill monitor signal (2kHz) for 1 s SHiP cycle







→ TD3/SAC actually learns to optimise it.
Could work as controller…
But….

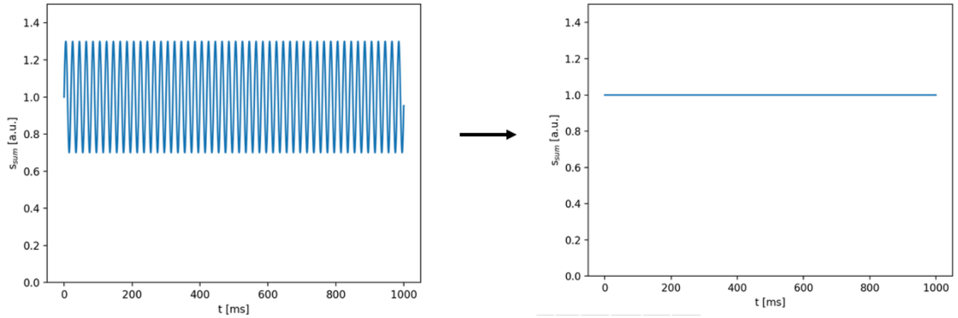
RL for $n \times 50$ Hz noise control?

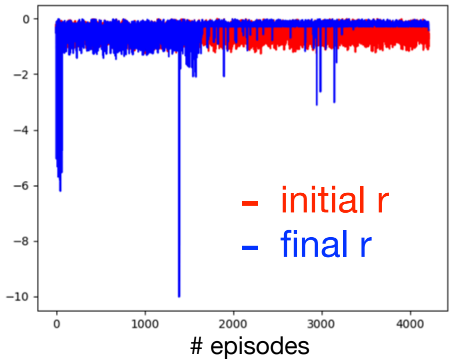
Simulated environment:

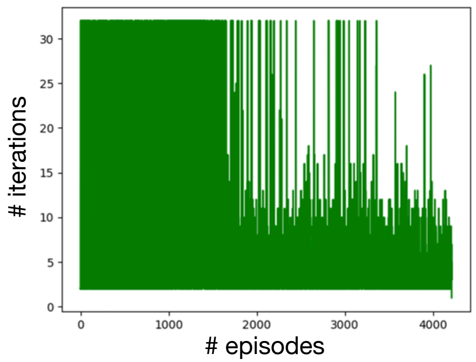
$$\vec{s} = [A_{spill}, \phi_{spill}, A_{corr}, \phi_{corr}]$$

$$r = -\sqrt{A_{noise}^2 + A_{corr}^2 + 2A_{noise}A_{corr}\cos\Delta\phi}$$

Spill monitor signal (2kHz) for 1 s SHiP cycle





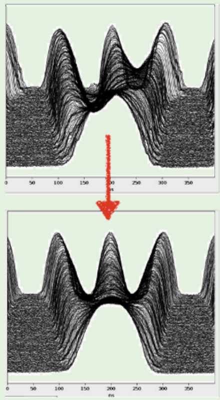


→ TD3/SAC actually learns to optimise it.
Could work as controller…
But….

→ Can we transfer? How does V_{QF} translate to A_{corr}?

→ Training on the machine takes too long. Also, how to change A_{noise}, ϕ_{noise}?

RL @ CERN - a selection



PS

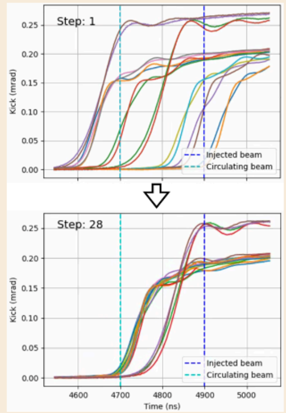
- Correct RF **phase & voltage** for **uniform bunch splitting** (LHC beams)
- Successful **sim2real** & fully **operational**
- **Multi-agent** (SAC) & **CNN** for initial guess
- **Next: continuous** controller (UCAP)

A. Lasheen, J. Wulff

PS to SPS

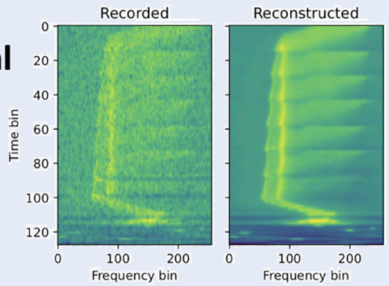
- Adjust **fine delays** of SPS **injection kicker**
- RL agent (PPO) trained on **data-driven dynamics model**
- Ready for **sim2real test**

M. Remta, F. Velotti

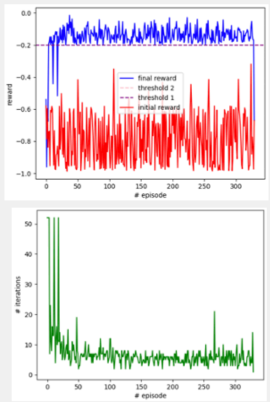


LINAC3 / LEIR

- **PhD project** *(B. Rodriguez)*: control LINAC3 cavities for **optimal injection efficiency** into LEIR
- RL state based on **VAE-encoded Schottky** spectra
- Agent trained on **data-driven dynamics model**



V. Kain, N. Madysa



SPS

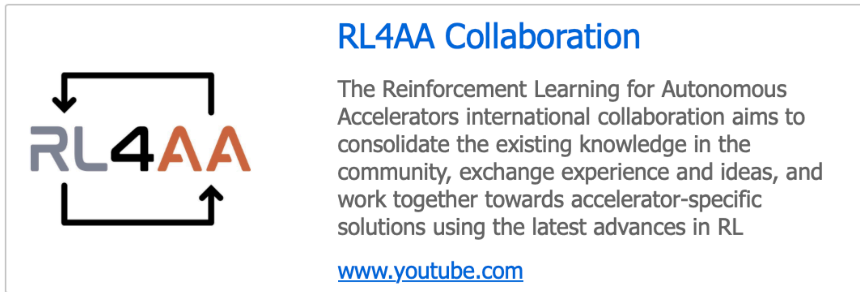
- **Steer DC beams** in TT20 TL using **split-foil** secondary emission **monitors**
- Works well in simulations, **with noise and varying emittances**
- Ready for **sim2real test**

N. Bruchon, V. Kain

Courtesy M. Schenk

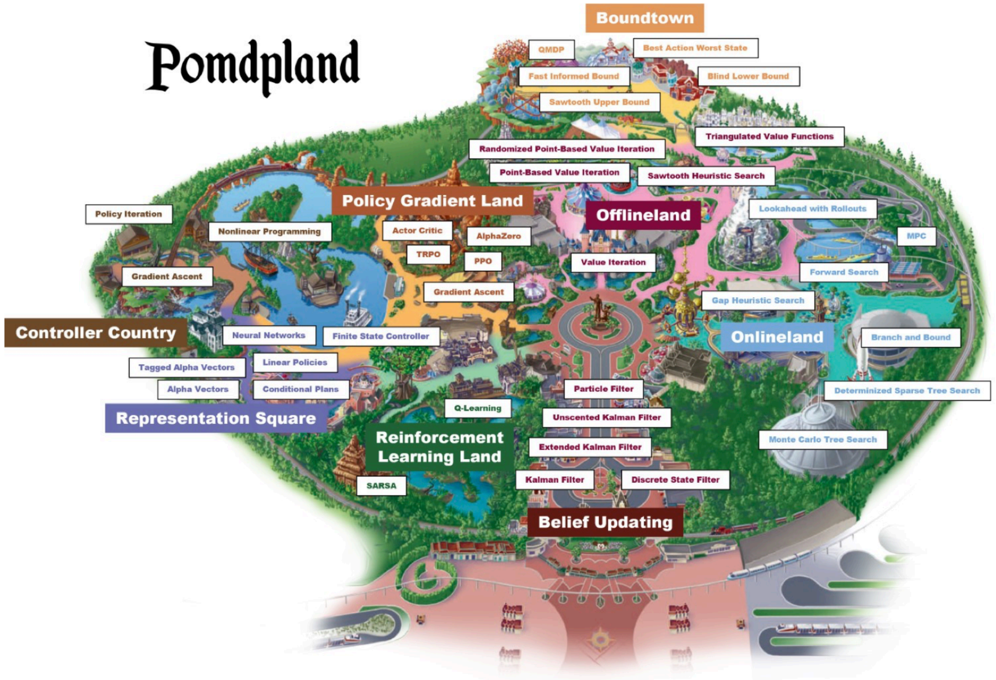
RL4AA - workshop

Pushing the frontiers of RL for accelerators → autonomous accelerators.



Advanced RL concepts

How to deal with time-varying systems, partially observable systems (POMDP).



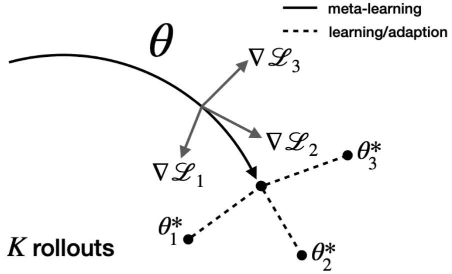
From Mykel Kochenderfer

META-RL i.e. MAML

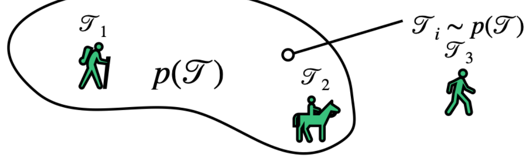
Require $p(\mathscr{T})$: distribution over tasks

Require α, β: step size hyper-parameters

1. randomly initialise θ
2. **while not done do**
3. sample batch of tasks $\mathscr{T}_i \sim p(\mathscr{T})$
4. **for each** $\mathscr{T}_i$ **do**
5. Sample with policy θ: $\mathscr{D}^{tr}_{\mathscr{T}_i} \sim \mathscr{D}_{\mathscr{T}_i}$
6. Evaluate $\nabla_\theta \mathscr{L}(\theta, \mathscr{D}^{tr}_{\mathscr{T}_i})$ with respect to K rollouts
7. Compute adapted parameters with gradient descent: $\phi_i = \theta - \alpha \nabla_\theta \mathscr{L}(\theta, \mathscr{D}^{tr}_{\mathscr{T}_i})$
8. Sample with new policy ϕ_i: $\mathscr{D}^{test}_{\mathscr{T}_i} \sim \mathscr{D}_{\mathscr{T}_i}$
9. Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_i \mathscr{L}(\phi_i, \mathscr{D}^{test}_{\mathscr{T}_i})$



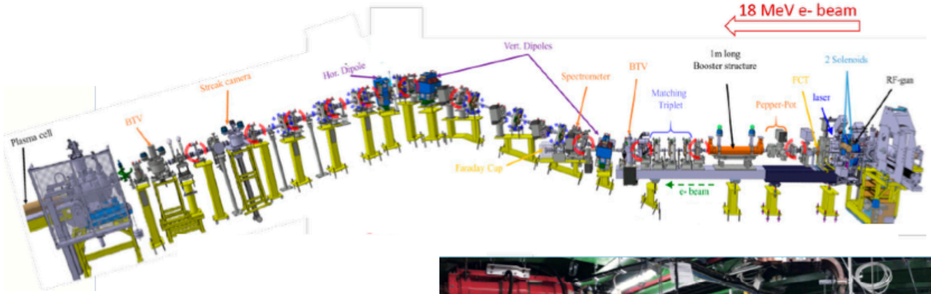
$$\mathscr{L}_{\mathscr{T}_i}(\theta) = -\mathbb{E}_{s_t, a_t \sim \pi_\theta, \mathscr{T}_i}\left[\sum_{t=1}^{H} R_i(s_t, a_t)\right]$$

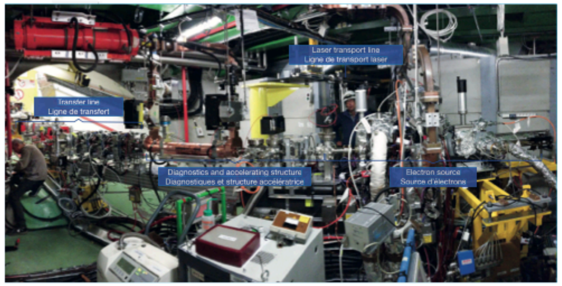


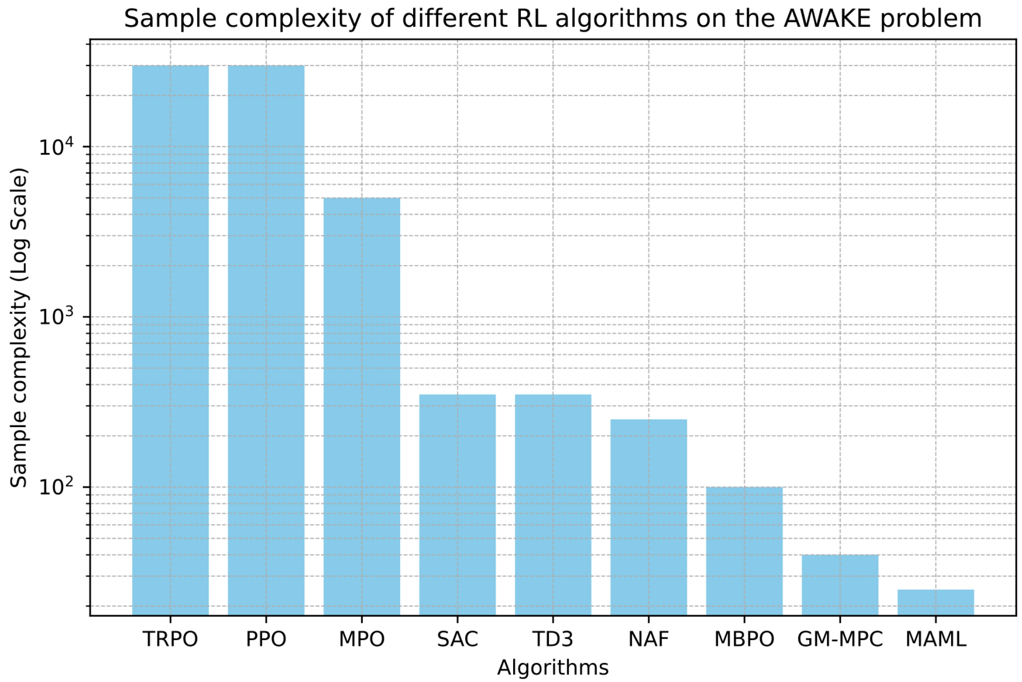
META-RL allows for few shot adaption

META-RL for accelerators

Tested algorithm on for AWAKE electron line steering at CERN.







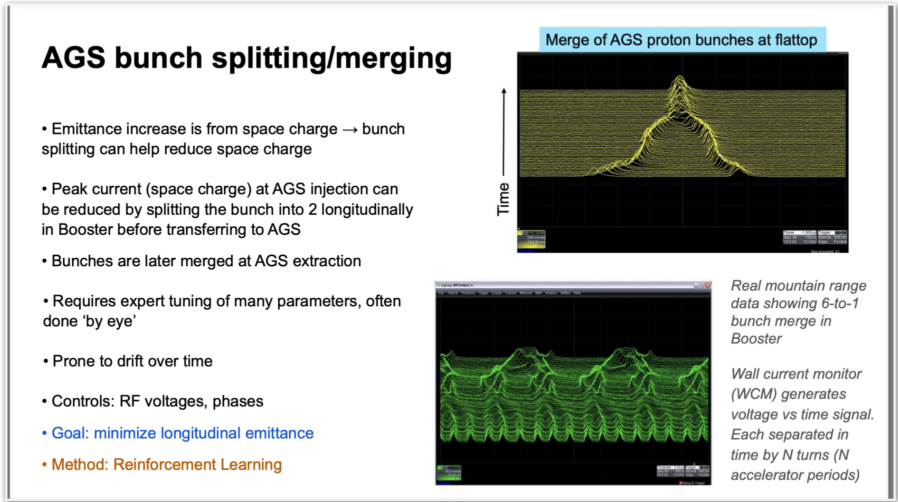
Maximise proton polarisation @ RHIC → towards EIC

Strategic and ambitious!

Numerous sub-projects: from improved modelling to fast and slow control and optimisation.

Main areas to optimise:

- Booster injection/capture
- AGS bunch splitting/merging scheme
- AGS spin resonance compensation



Efficient Particle Accelerators (EPA) project @ CERN

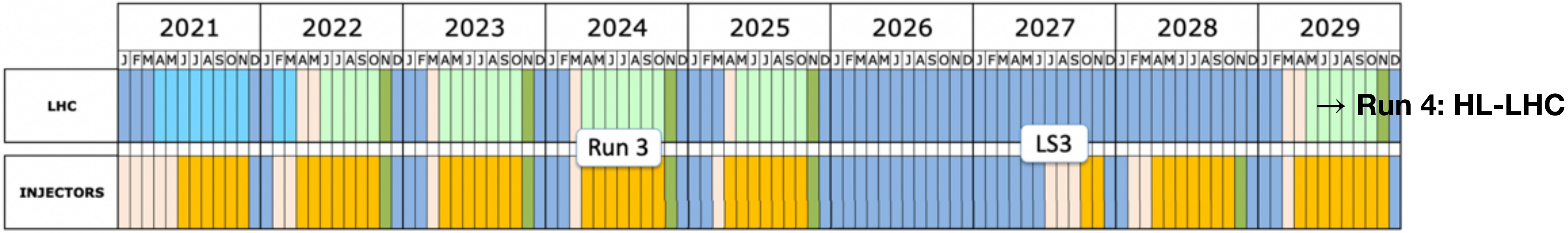
→ **automating accelerator exploitation**

Approved in autumn 2023 after pre-study in Efficiency Think Tank (ETT)

9 work packages: ETT recommendations and controls infrastructure evolution.

WP1 Dynamic Beam Scheduling	WP2 Automated LHC Filling	WP3 Automated Parameter Control & Optimisation
WP4 Hysteresis Compensation	WP5 Next Generation Sequencer	WP6 Efficient Settings Management
WP7 Automated Equipment Testing	WP8 Automate Equipment	WP9 Data Processing Framework

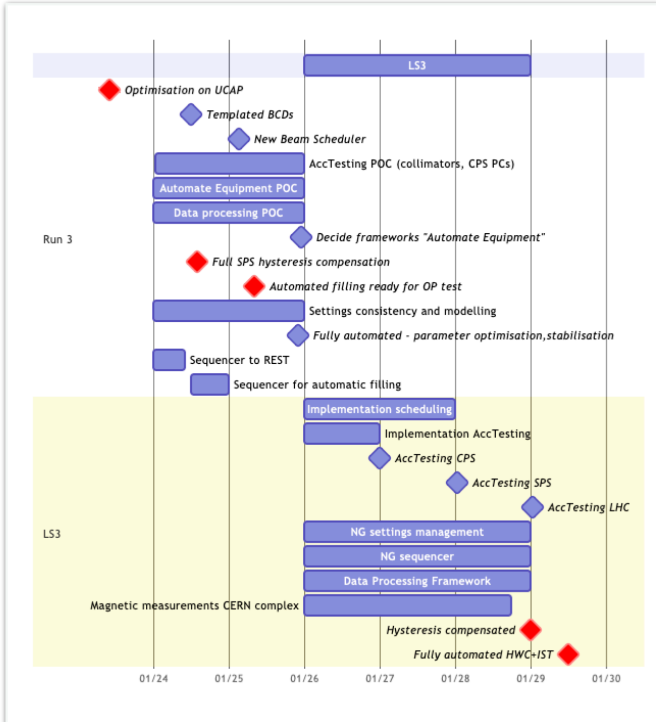
Efficient Particle Accelerators (EPA) project @ CERN



Time bounded project (5 years): improvements ready for **HL-LHC (2029)**

EPA is preparing a new CERN **accelerator exploitation paradigm**

→ **blazing the trail for FCC**



Main takeaways

AI is changing how we exploit particle accelerators and will enable new possibilities.

The ICFA ML workshop series created a global community for ML for particle accelerators

- → cross-fertilised and inspired labs.

Status: state-of-the-art AI in R&D and small scale operational installations

Next (hopefully): AI at scale - tackling infrastructure questions. From "works once" to "works always"

- Can only come with strategic ambitious projects! There are some now…