Machine Learning-based Modelling at the LHC and Muon Collider Studies

Elena Fol

CERN, Beams Department
(BE-ABP-LAF)

I.FAST 9th Low Emittance Rings Workshop 2024

13-16 February 2024

ML-based Modelling vs. Traditional methods

Traditional Modelling

Data →

$$\frac{\Delta\beta(s)}{\beta} \approx \sum_i \frac{\Delta k_i \beta_i}{2\sin(2\pi Q)} \cos(2\pi Q - 2|\phi(s) - \phi_i|)$$

Model

→ **Output**

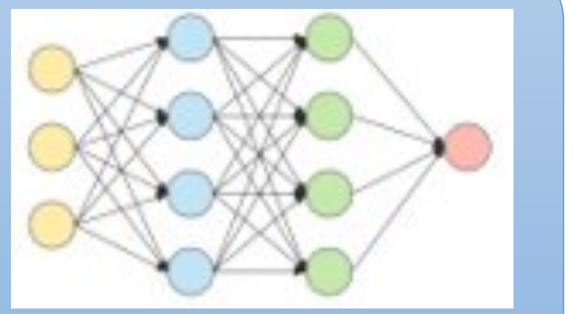
- Creating **manually** a set of commands / equations and rules
- Example: comparing simulations and measurements

Machine Learning approach

Data ***x*** →

Output ***y*** →

$$y = f\left(\sum x_i w_i + b\right)$$



→ ***Model***

- **Learn from data automatically**
- Model is developed by adjusting model's parameters **to explain the relation between given data and output**

ML-based Modelling in Particle Accelerators

Which limitations can be solved by ML?

- ➢ Direct measurements are not possible
- ➢ Analytical solution does not exist
- ➢ Computationally expensive simulations
- ➢ Non-linear, correlated sub-systems
- ➢ Rapidly changing environment

Machine Learning:

✓ ***Learn arbitrary models***

✓ ***Directly from provided data***

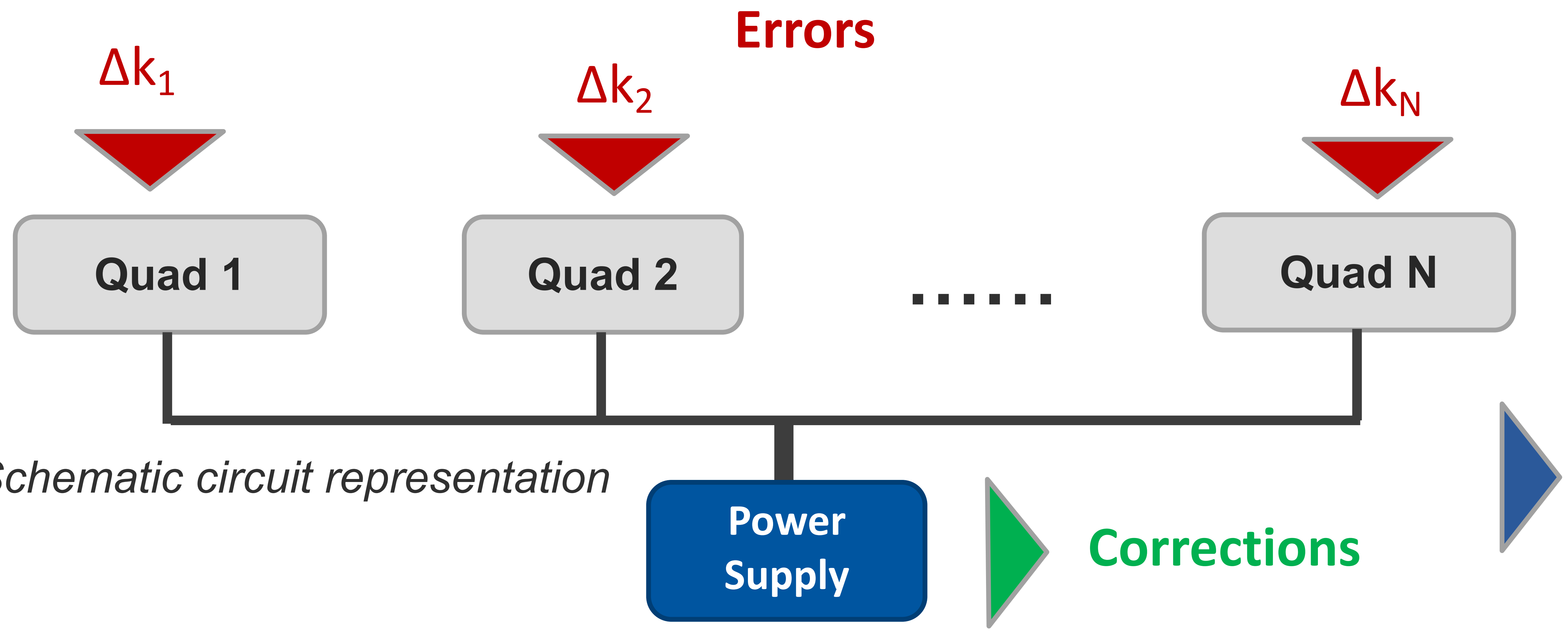
ML in accelerators modelling: Examples

- Speeding-up computationally costly **simulations**:
  Methods: **Clustering techniques**, **Gaussian Processes**, Supervised Learning (inverse) models
  Applications: Sample-efficient dynamic aperture estimation [1], electron beam size optimisation[2]
- **Operation** automation and online tuning:
  Methods: Bayesian optimization (using **Gaussian Processes**), Reinforcement Learning, **physics-informed NN** for modelling, **Clustering techniques**
  Applications: Tuning optics models in storage rings [3], beam trajectory steering [4], faulty BPMs detection [5]
- Virtual **Diagnostics**:
  Methods: Image-based analysis using **Convolutional NN** trained on simulations
  Applications: 6D phase space reconstruction [6]

[1] F.F. Van der Veken et al., "Using Machine Learning to Improve Dynamic Aperture Estimates", IPAC'21
[2] A. Edelen et al., "Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems", Phys. Rev. Accel. Beams **23**, 044601 (2020)
[3] A. Ivanov, I. Agapov, "Physics-Based Deep Neural Networks for Beam Dynamics in Charged Particle Accelerators", Phys. Rev. Accel. Beams **23**, 074601 (2020)
[4] V. Kain et al., "Sample-efficient reinforcement learning for CERN accelerator control", Phys. Rev. Accel. Beams, 23.124801 (2020)
[5] E. Fol et al., "Detection of faulty beam position monitors using unsupervised learning", Phys. Rev. Accel. Beams **23**, 102805 (2020)
[6] R. Roussel et al., "Phase Space Reconstruction from Accelerator Beam Measurements Using Neural Networks and Differentiable Simulations", Phys. Rev. Lett. **130**, 145001 (2023)

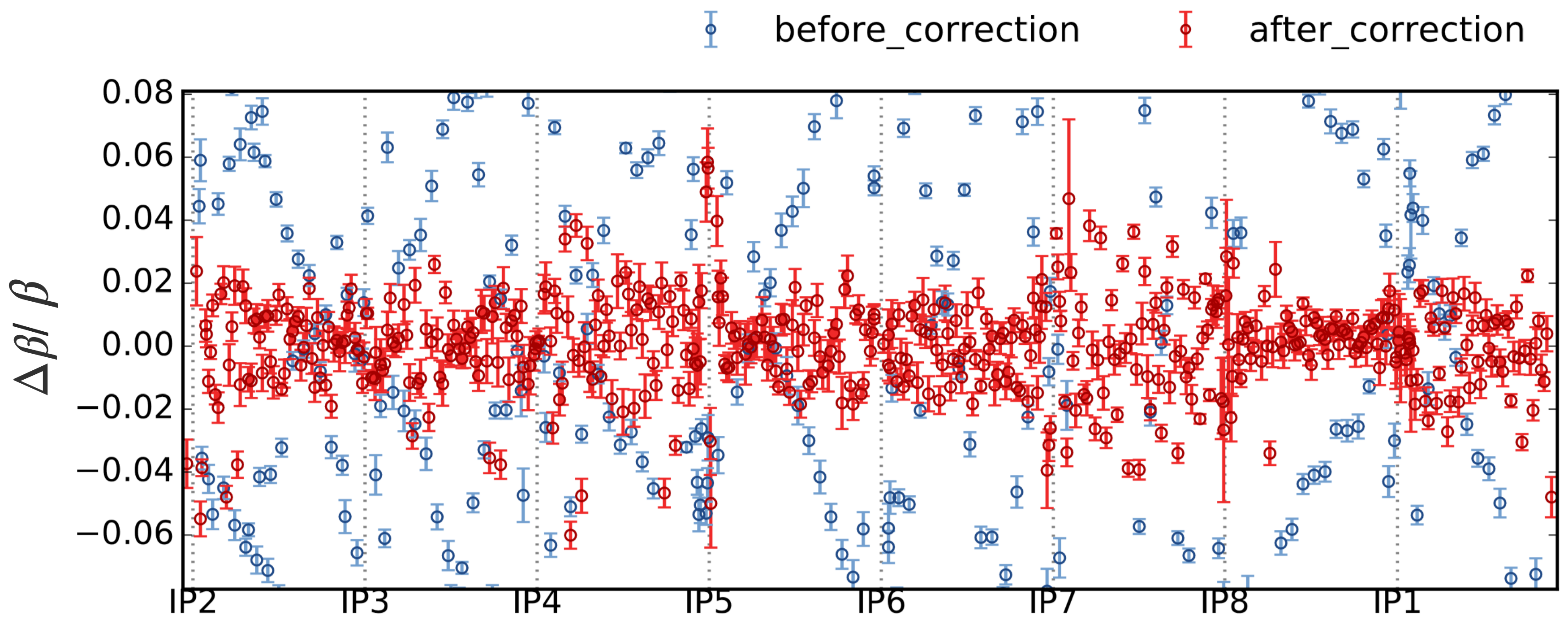
ML-based Modelling at the LHC

Optics corrections in the LHC using Supervised Learning



Schematic circuit representation

β ➡ Determined by **quadrupole arrangement and powering:** $\frac{\Delta\beta}{\beta} = \frac{\beta_{meas} - \beta_{model}}{\beta_{model}}$

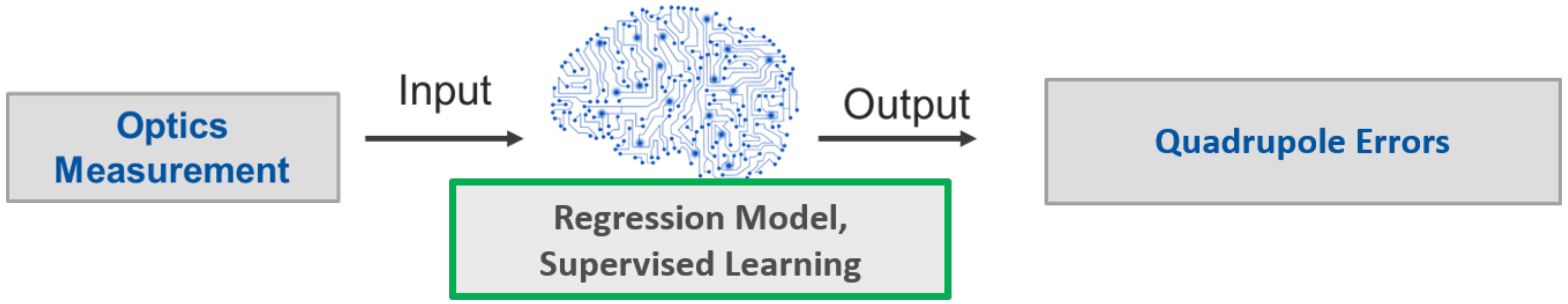


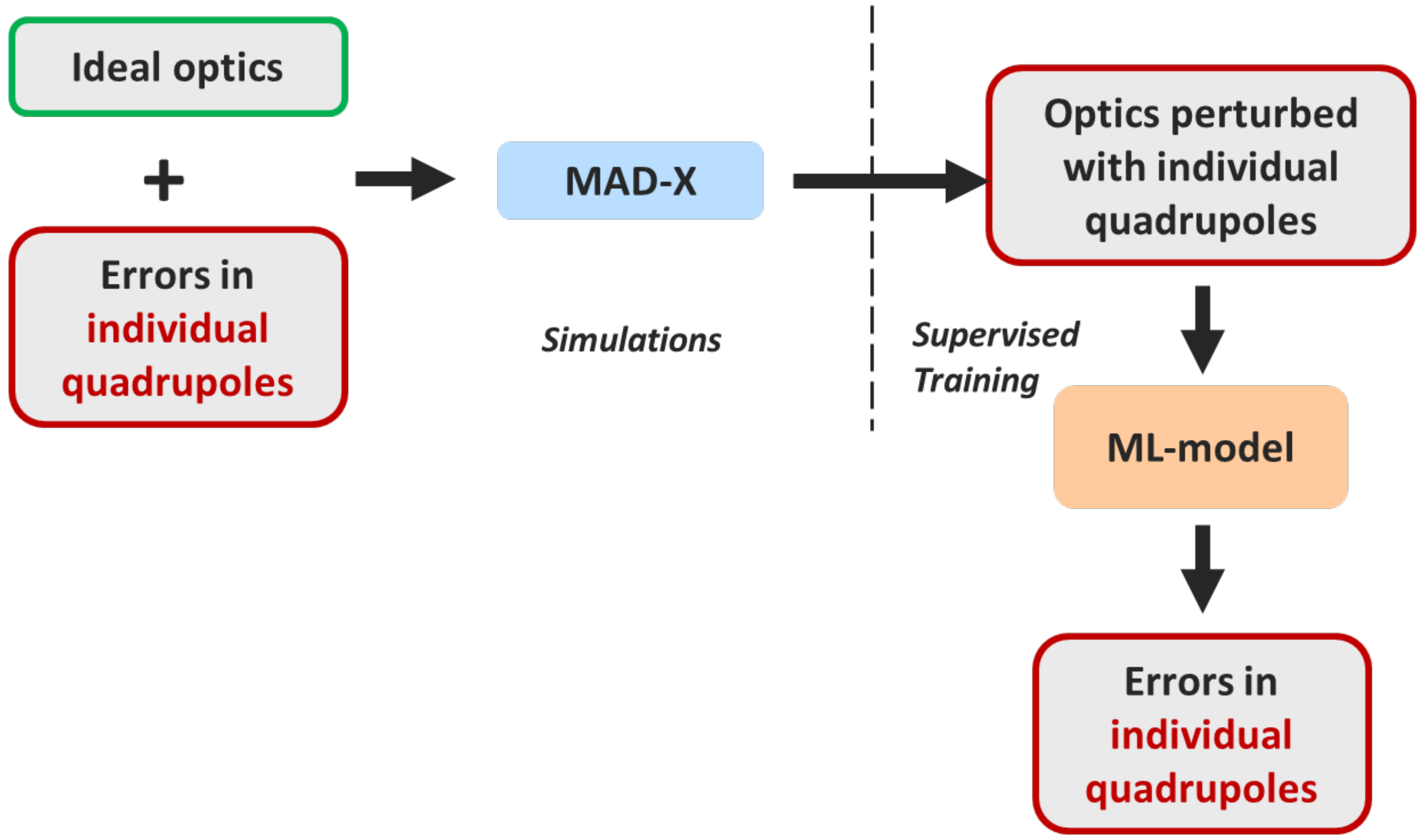
- Access to the magnets for direct measurements is not possible during operation.
  ➡ Beam-based measurements and corrections of lattice imperfections.
- Computed corrections provide **circuit settings to compensate measured beta-beating**

➡ What are the **actual individual magnet errors?**

➡ **Modelling of inverse relation between measured optics and magnet errors**

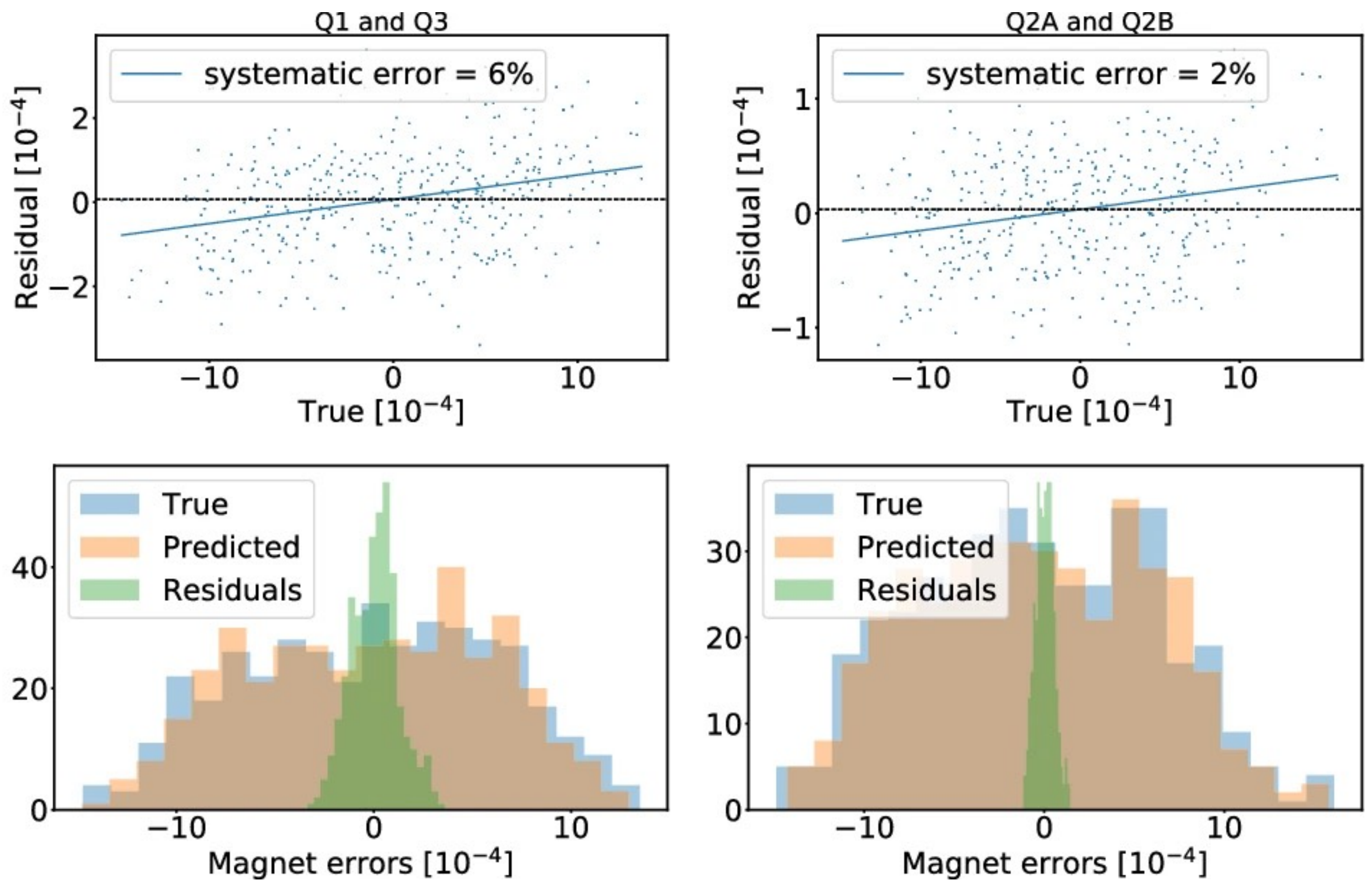
Optics corrections: prediction of magnets errors





Random Forest Regressor:

- Ensemble of **decision trees**: lower complexity vs. NN
- **1256 target** variables, **2048 input** variables
- Tested on simulations, historical data and **LHC commissioning**

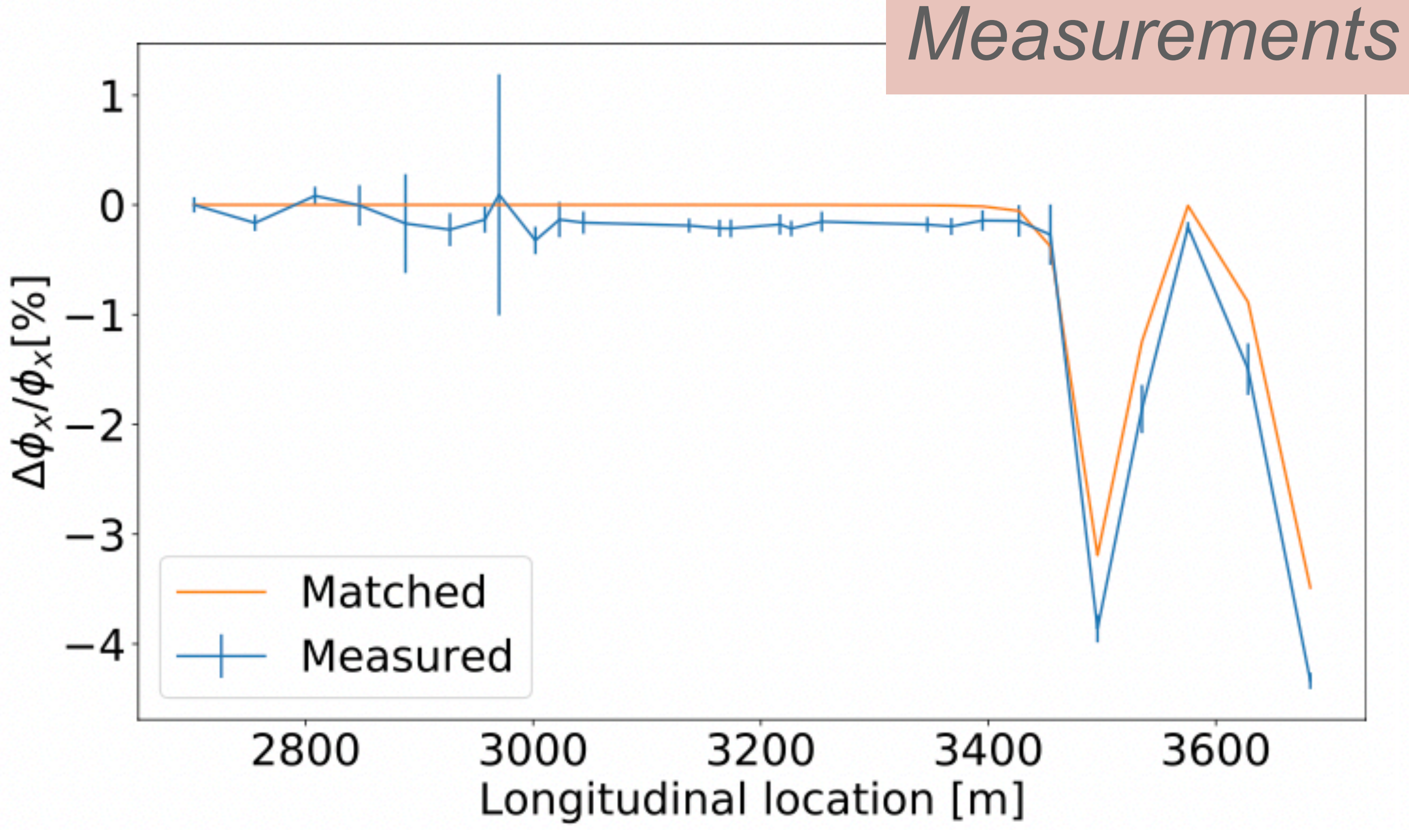


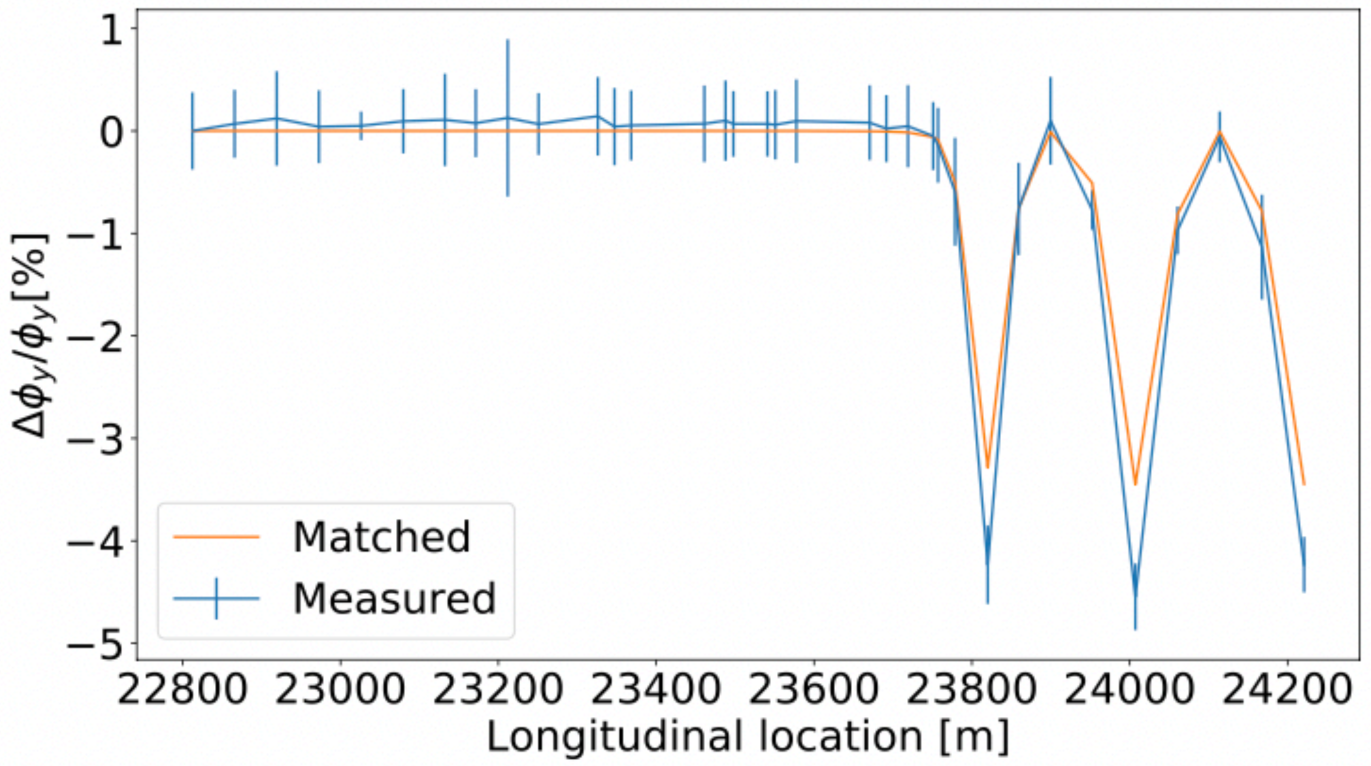
Published in: The European Physical Journal Plus volume 136, Article number: 365 (2021), "Supervised learning-based reconstruction of magnet errors in circular accelerators".

LHC commissioning 2022: beam optics corrections

Example: Corrections in Interaction Region 1, squeeze to β^* = 30 cm (challenging low beta optics)

Measurements





	ΔK_1 $[10^{-5}m^{-2}]$		
Magnet	**APJ**	**SbS**	**ML**
MQXA1.L1	-	1.23	1.23
MQXA1.R1	-	-1.23	-1.24
MQXB2.L1	1.15	1.22	-0.11
MQXB2.R1	-0.87	-1.22	0.18
MQXA3.L1	1.94	0.41	0.31
MQXA3.R1	-2.88	-0.7	-0.1

- ✓ Sufficiently accurate prediction **of magnet errors** directly from standard optics analysis data
- ✓ **Phase errors can be corrected** applying the errors with opposite sign as correction settings

- ✓ **Simultaneous local correction in all IRs within seconds.**
- ➢ **Potential to save operation time!**

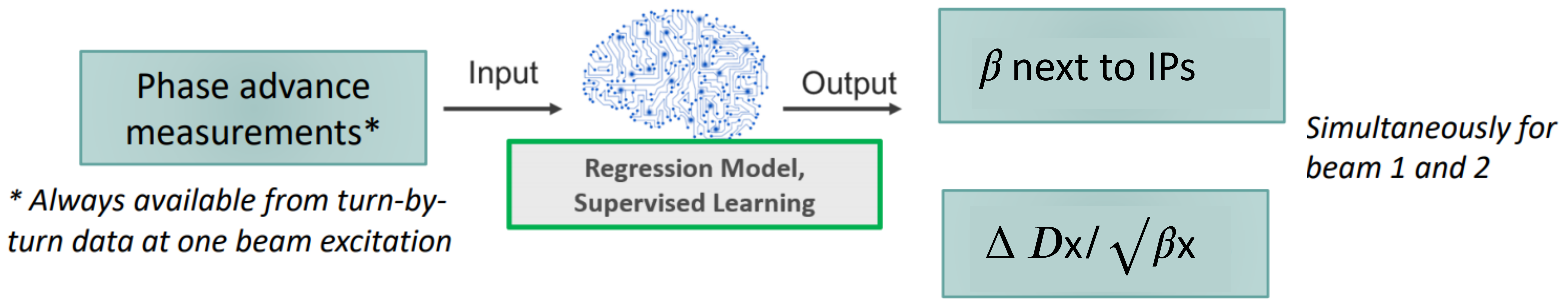
E.Fol et al.,"Experimental Demonstration of Machine Learning Application in LHC optics commissioning",IPAC'22 MOPOPT047

Virtual Optics Measurements

- How to reconstruct optics observables **without direct measurements?**

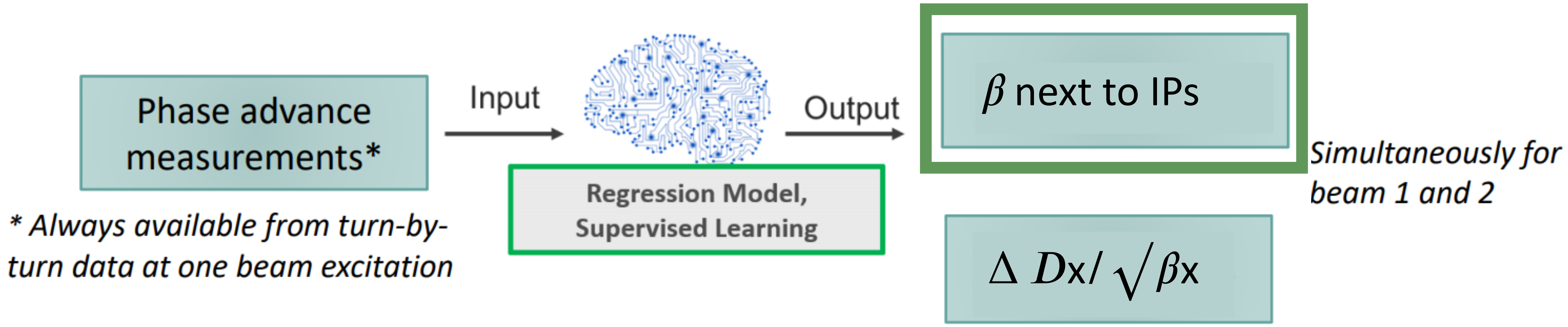
Virtual Optics Measurements

- ➢ How to reconstruct optics observables **without direct measurements?**



Virtual Optics Measurements

➢ How to reconstruct optics observables **without direct measurements?**



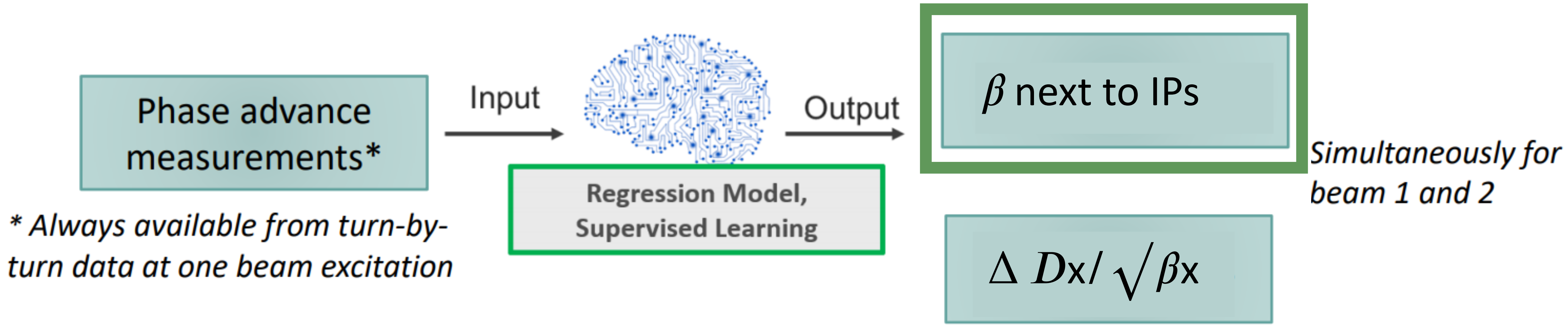
Measuring beta-function in Interaction Regions:

Traditional technique: **k-modulation:**

- Based on modulation of quadrupole current
- Time consuming
- Accuracy varies depending on tune measurement uncertainty, magnet errors and β^* settings.

Virtual Optics Measurements

➢ How to reconstruct optics observables **without direct measurements?**



Measuring beta-function in Interaction Regions:

Traditional technique: **k-modulation:**

- Based on modulation of quadrupole current
- Time consuming
- Accuracy varies depending on tune measurement uncertainty, magnet errors and β^* settings.

✓ β-functions left and right from IPs **within a few seconds vs. several minutes for k-modulation**

✓ Average accuracy: **5 % for β^* = 30 cm**.

➢ Tests during LHC commissioning 2022

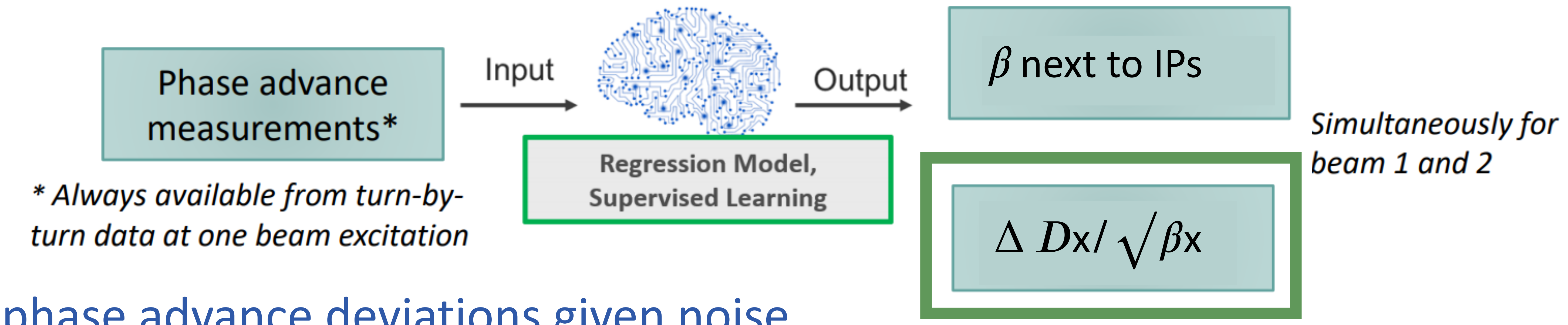
β^* = 30 cm

Location	K-mod β_x, β_y[m]	ML β_x, β_y [m]	$\Delta\beta/\beta_{kmod}$ x, y [%]
B1, IP1L	1262, 1074	1296, 1223	2.6, 13.8
B1, IP1R	1340, 1051	1268, 1197	5.3, 13.9
B1, IP5L	1388, 1552	1377, 1659	0.8, 6.9
B1, IP5R	1302, 1624	1369, 1642	5.2, 1.1
B2, IP1L	1406, 1773	1435, 1851	2.1, 4.4
B2, IP1R	1366, 1947	1412, 1893	3.4, 2.7
B2, IP5L	1511, 1364	1639, 1315	8.4, 3.6
B2, IP5R	1637, 1377	1632, 1303	0.3, 5.4

Virtual Optics Measurements

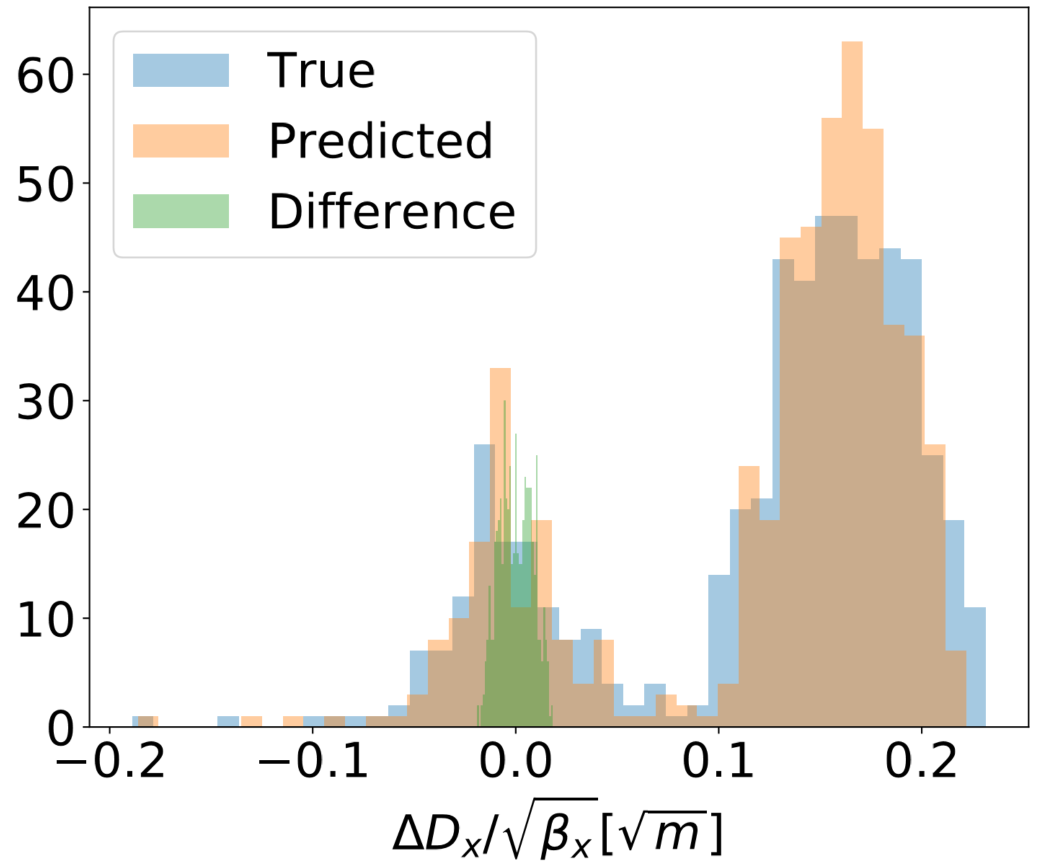
Horizontal Dispersion reconstruction:

- Computed by acquiring turn-by- turn data from **several beam excitations, shifting the momentum.**



- **Input**: simulated phase advance deviations given noise
- **Output**: normalized dispersion $\Delta\, Dx/\sqrt{\beta x}$
- Using **linear regression model**: 10 000 samples

***Simulation** example: Beam 1*



Virtual Optics Measurements

Horizontal Dispersion reconstruction:

- Computed by acquiring turn-by- turn data from **several beam excitations, shifting the momentum.**

Phase advance measurements*

Input

Output

Regression Model, Supervised Learning

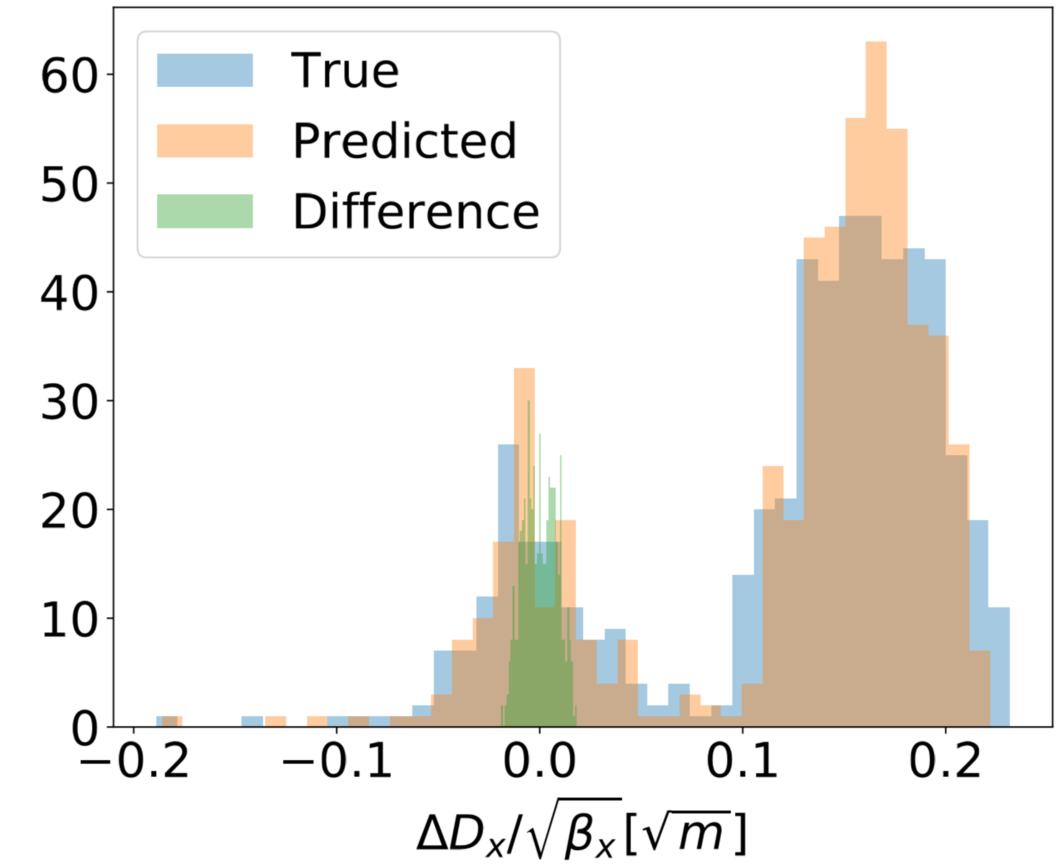
β next to IPs

$\Delta\, Dx/\sqrt{\beta x}$

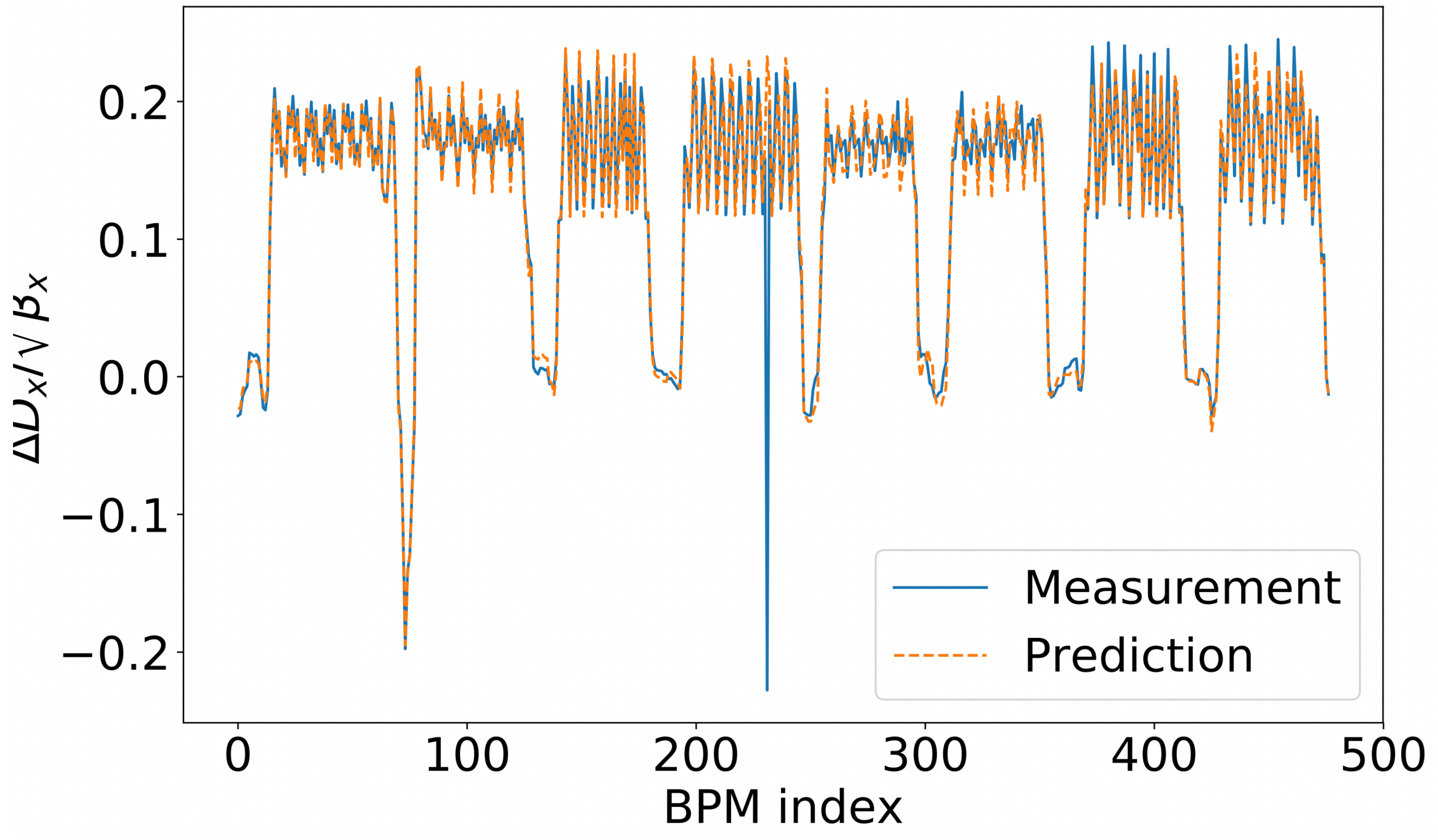
Simultaneously for beam 1 and 2

** Always available from turn-by-turn data at one beam excitation*

***Simulation** example: Beam 1*



***Measurement taken during LHC commissioning**, β^* = 30 cm*

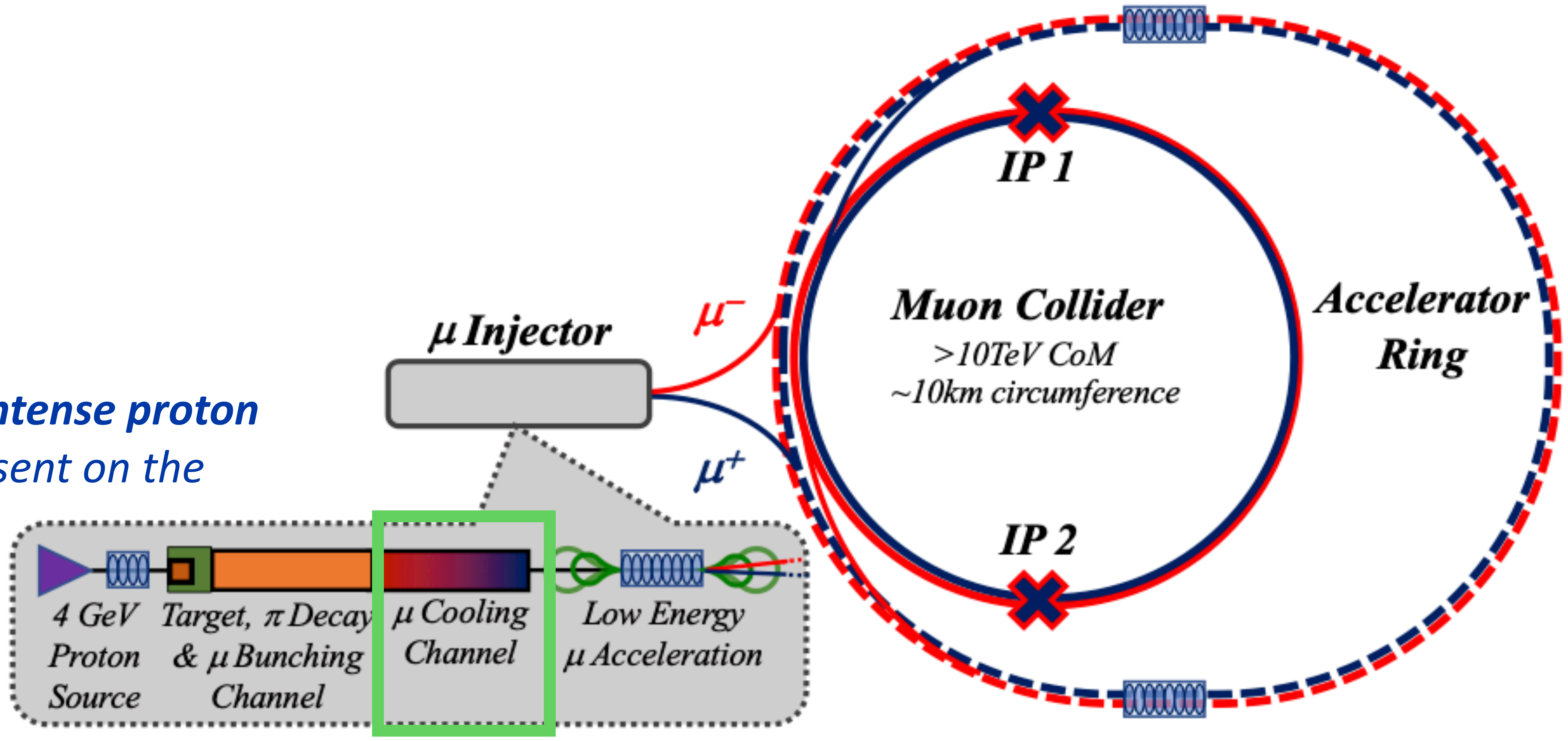


✓The relative **error of prediction is 5%** (beam 1) and **7%** (beam 2)

✓Potential speedup of machine commissioning for the same performance.

ML-based Modelling (and optimization) in Muon Collider Design

Muon Collider: overview



*Short **intense proton bunch** sent on the target*

Interaction with the target produces pions
*➡ **decay into muons***

*Muons are captured and **cooled to lower emittance***

$$\mathcal{L} \propto \gamma \langle B \rangle \sigma_\delta \frac{N_0}{\epsilon \epsilon_L} f_r N_0 \gamma$$

High energy

High field in collider ring

Large energy acceptance

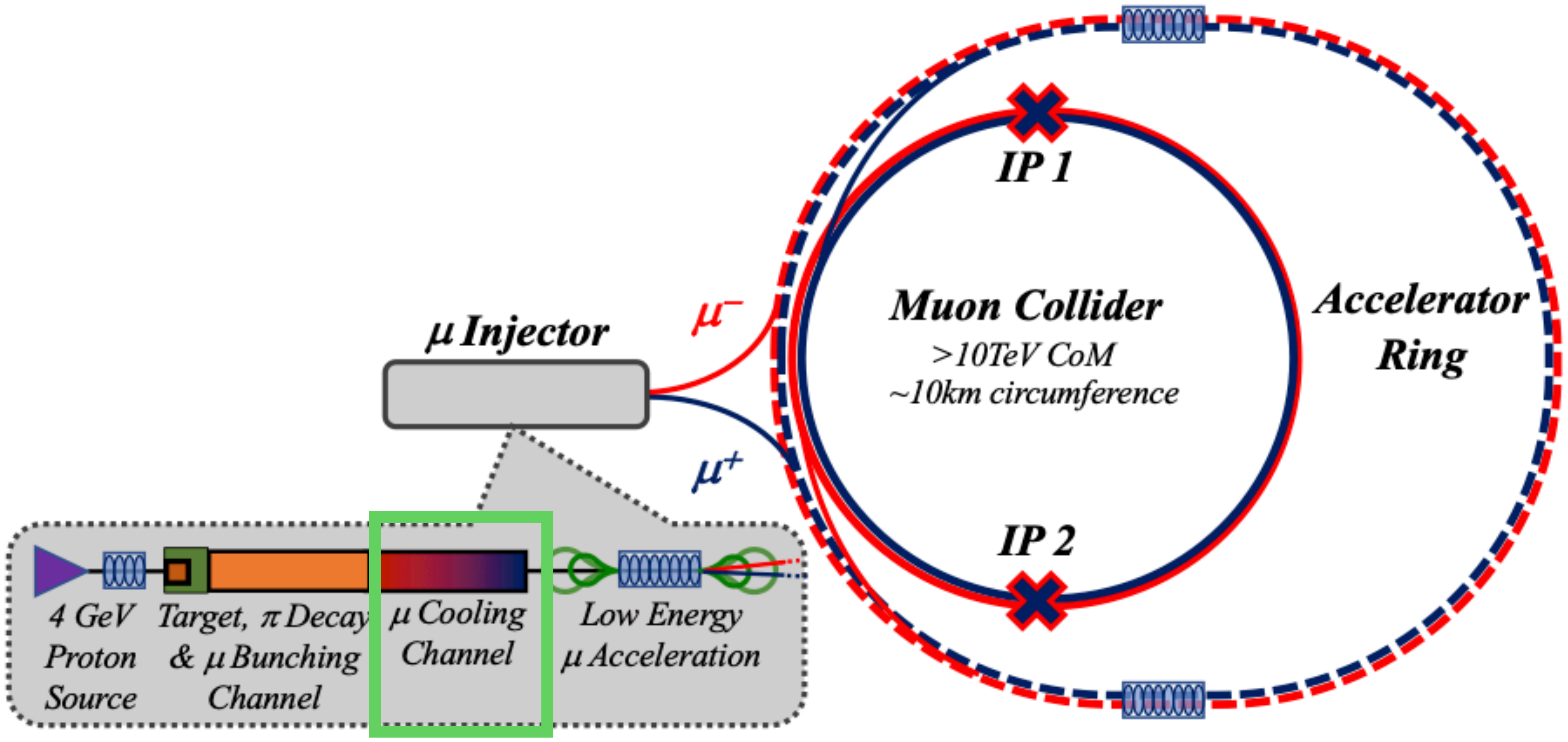
Dense beam

High beam power

https://muoncollider.web.cern.ch

Muon Collider: overview

International MUON Collider Collaboration



$$\mathcal{L} \propto \gamma \langle B \rangle \sigma_\delta \frac{N_0}{\epsilon \epsilon_L} f_r N_0 \gamma$$

High energy

High field in collider ring

Large energy acceptance

Dense beam

High beam power

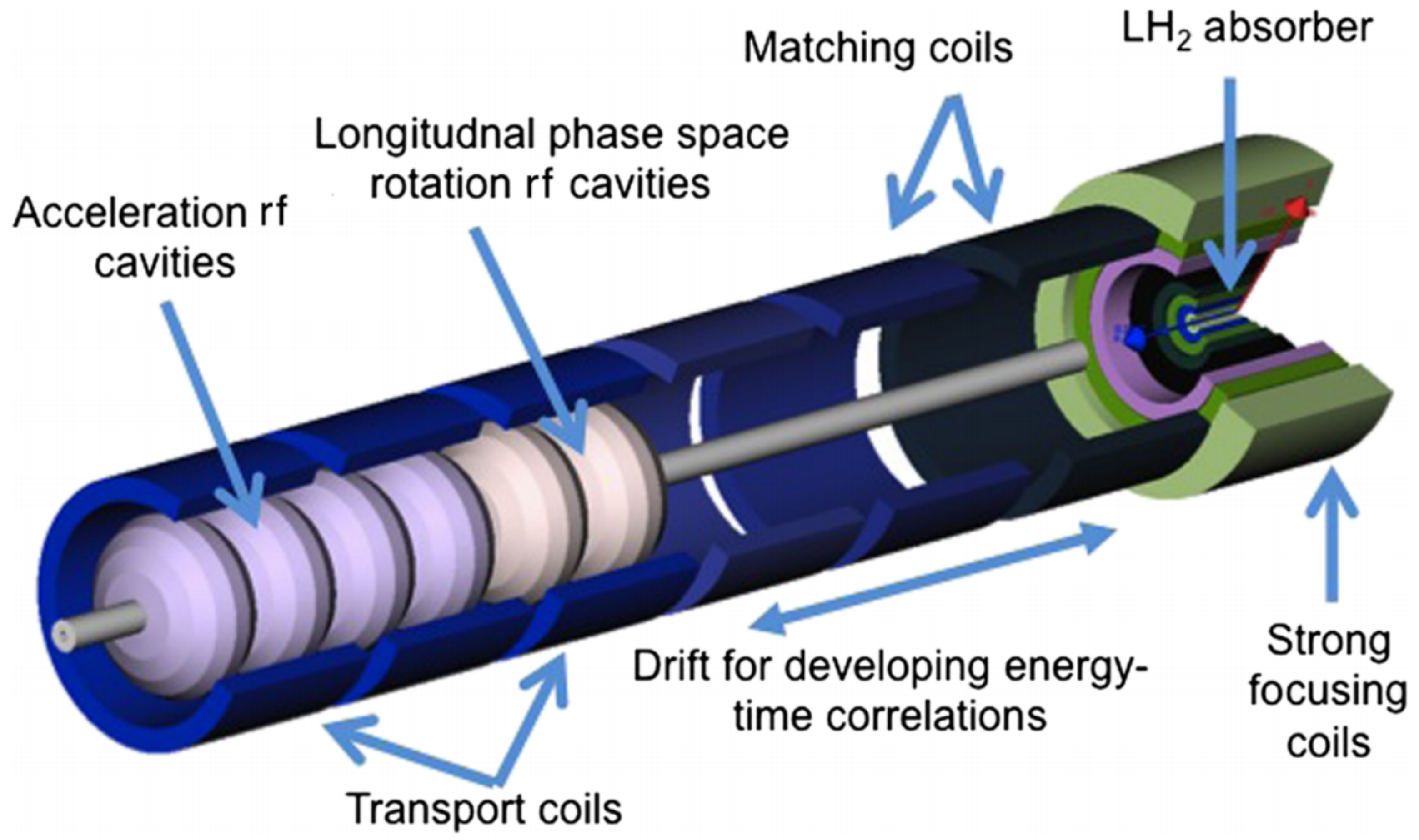
Parameter	Unit	3 TeV	10 TeV	14 TeV
L	10^{34} cm^{-2}s^{-1}	1.8	20	40
N	10^{12}	2.2	1.8	1.8
f_r	Hz	5	5	5
P_{beam}	MW	5.3	14.4	20
C	km	4.5	10	14
<B>	T	7	10.5	10.5
ε_L	MeV m	7.5	7.5	7.5
σ_E / E	%	0.1	0.1	0.1
σ_z	mm	5	1.5	1.07
β	mm	5	1.5	1.07
ε	μm	25	25	25
$\sigma_{x,y}$	μm	3.0	0.9	0.63

- **Ionisation cooling** (the reduction of occupied phase-space by muons)**: the only technique compatible with muon's lifetime**, demonstrated by [MICE collaboration]
- **Final Cooling Channel:** reduction of transverse emittance on the cost of longitudinal emittance growth

https://muoncollider.web.cern.ch

Challenges and objectives of Final Cooling

Re-accelerating, rotating the beam
=> Restore P_Z , reduce σE



energy loss due to the interaction with absorber material
=> Reduction of $\epsilon_\perp$

Challenges and objectives of Final Cooling

Lowering transverse emittance on the costs of :

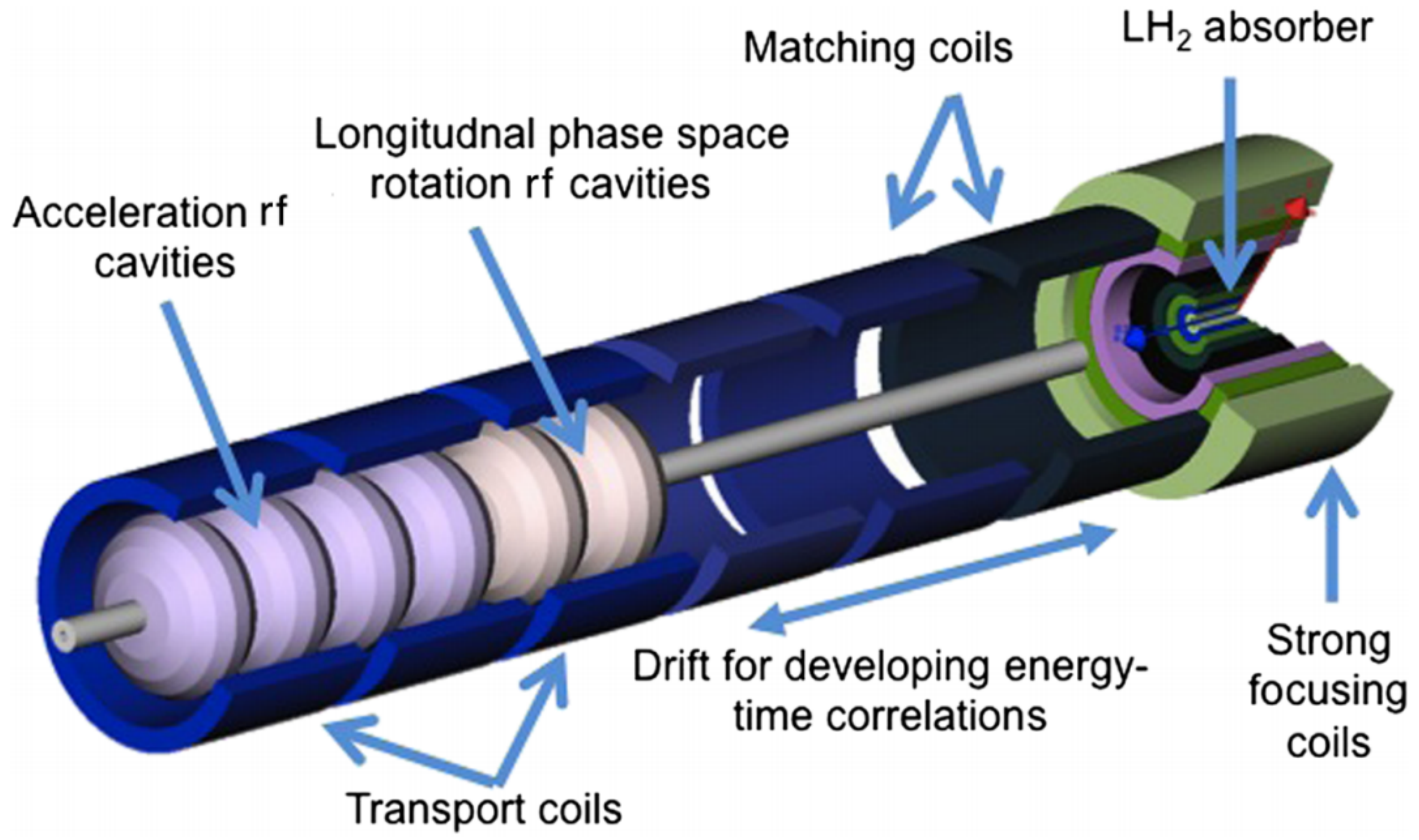
- Longitudinal emittance growth
- Bunch length increasing: challenging RF set-up
- Energy spread
- Particle losses due to decays and energy loss

$$\boxed{\frac{d\varepsilon_T}{ds} = -\frac{1}{\beta^2 E}\frac{dE}{ds}\varepsilon_T + \frac{\beta\gamma\beta_T}{2}\frac{d\theta_0^2}{ds}}$$

Energy loss term (Cooling)

Multiple scattering term (Heating)

Re-accelerating, rotating the beam => Restore P_Z, reduce σE



energy loss due to the interaction with absorber material

=> Reduction of $\epsilon_\perp$

- Achieved in previous studies*: $\varepsilon_\perp$ = 55 µm, with $\varepsilon_\parallel$ = 76 mm, transmission 50%
- Target is $\varepsilon_\perp$ = 25µm => higher solenoid field, **optimization**

Challenges and objectives of Final Cooling

Lowering transverse emittance on the costs of :

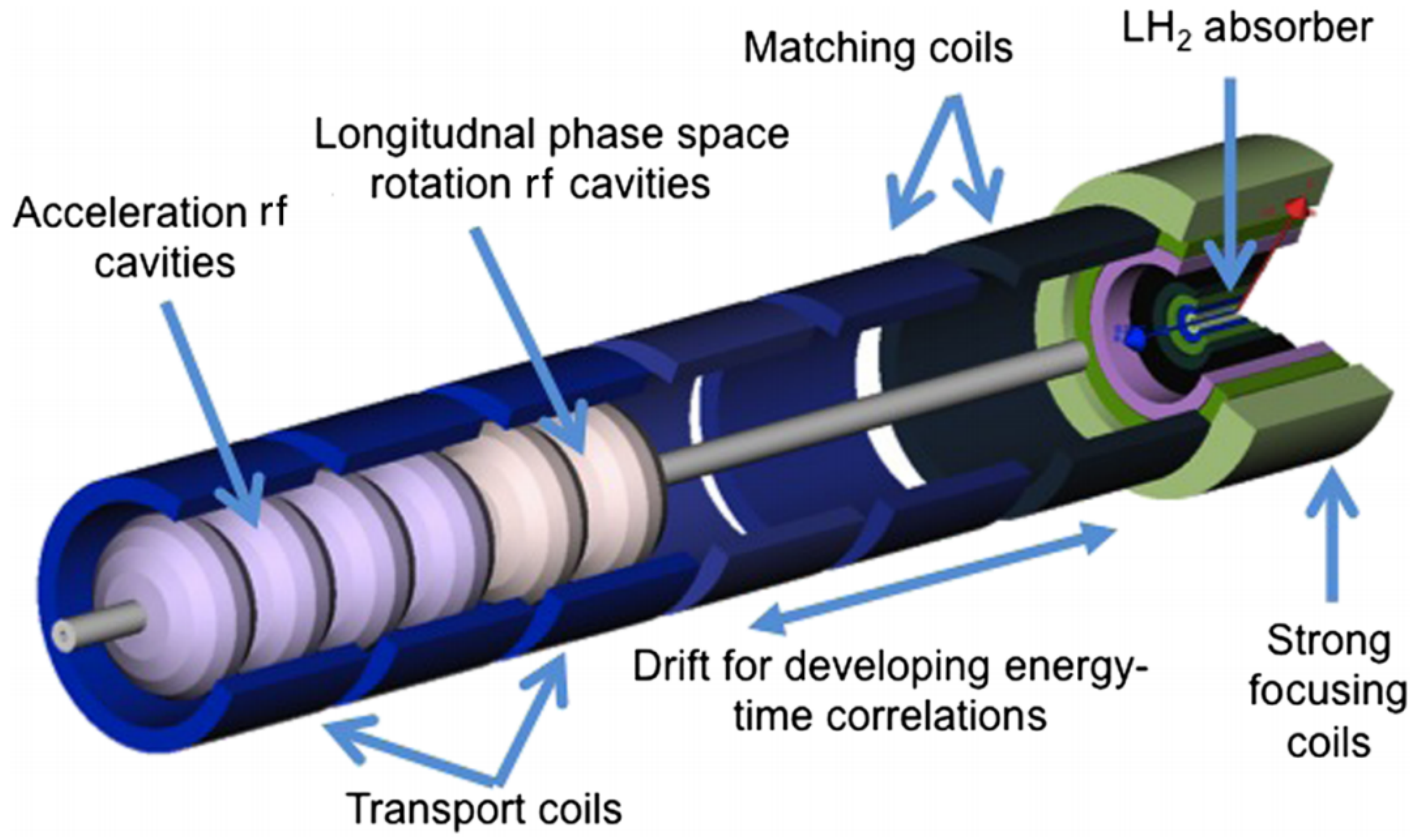
- Longitudinal emittance growth
- Bunch length increasing: challenging RF set-up
- Energy spread
- Particle losses due to decays and energy loss

$$\boxed{\frac{d\varepsilon_T}{ds} = -\frac{1}{\beta^2 E}\frac{dE}{ds}\varepsilon_T + \frac{\beta\gamma\beta_T}{2}\frac{d\theta_0^2}{ds}}$$

Energy loss term (Cooling)

Multiple scattering term (Heating)

Re-accelerating, rotating the beam => Restore P_Z, reduce σE



energy loss due to the interaction with absorber material

=> Reduction of $\epsilon_\perp$

- Achieved in previous studies*: $\varepsilon_\perp = 55$ µm, with $\varepsilon_\parallel = 76$ mm, transmission 50%
- Target is $\varepsilon_\perp = 25$µm => higher solenoid field, **optimization**

• How to **speed up** simulations-based design optimization? • How to **estimate initial optimization parameters**? • Robust **emittance estimation** during optimization?	▸ Surrogate models ▸ Feature Importance Analysis with Decision Trees ▸ Bayesian Optimization ▸ Clustering and anomaly detection

Final Cooling: Optimization Strategy

I. Estimate **optimal momenta and absorber lengths** in every cell, with objective $\epsilon_{\perp} = 25\mu m$.

- Nelder – Mead
- Using cooling equations* as objective function

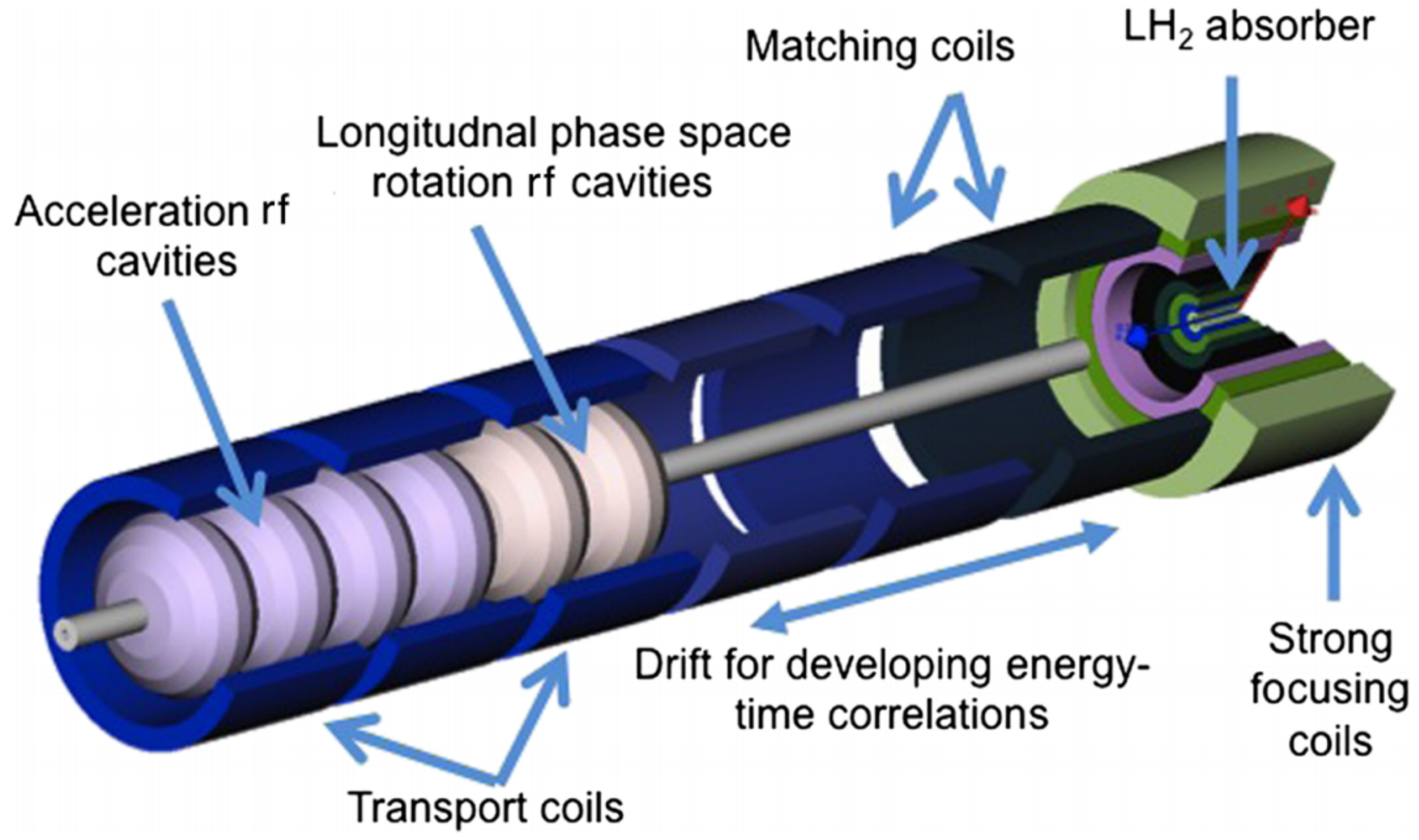
II. **Optics control**, ensure low beta-function in absorber by **optimizing solenoid field and matching coils**

- Numerical optimization, simulations
- Surrogate model (Random Forest)

III. **Optimize acceleration and rotation** of the bunch after absorber (simplified RF model)

IV. **Optimize a realistic RF system**: frequencies, phases, gradients to **control the longitudinal dynamics**

- Bayesian Optimization, BOBYQA
- Clustering to for robust emittance estimation



- Global optimization: would have **14 parameters** to optimize in **each cell**
- Expected to need **~17 cells in total**

▶ Step-by-step approach, testing different optimization algorithms

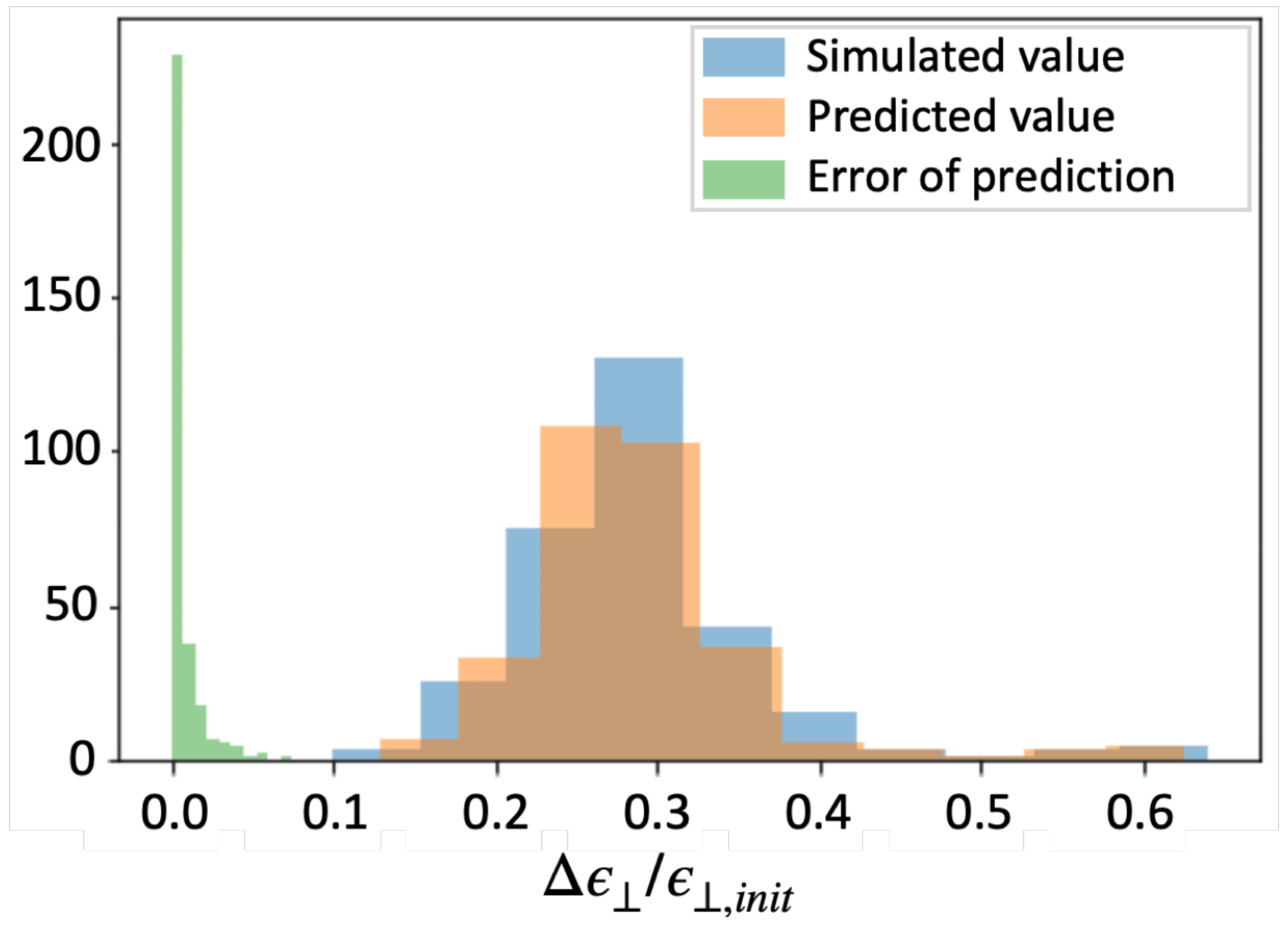
Optimizing solenoid fields: Surrogate Modeling

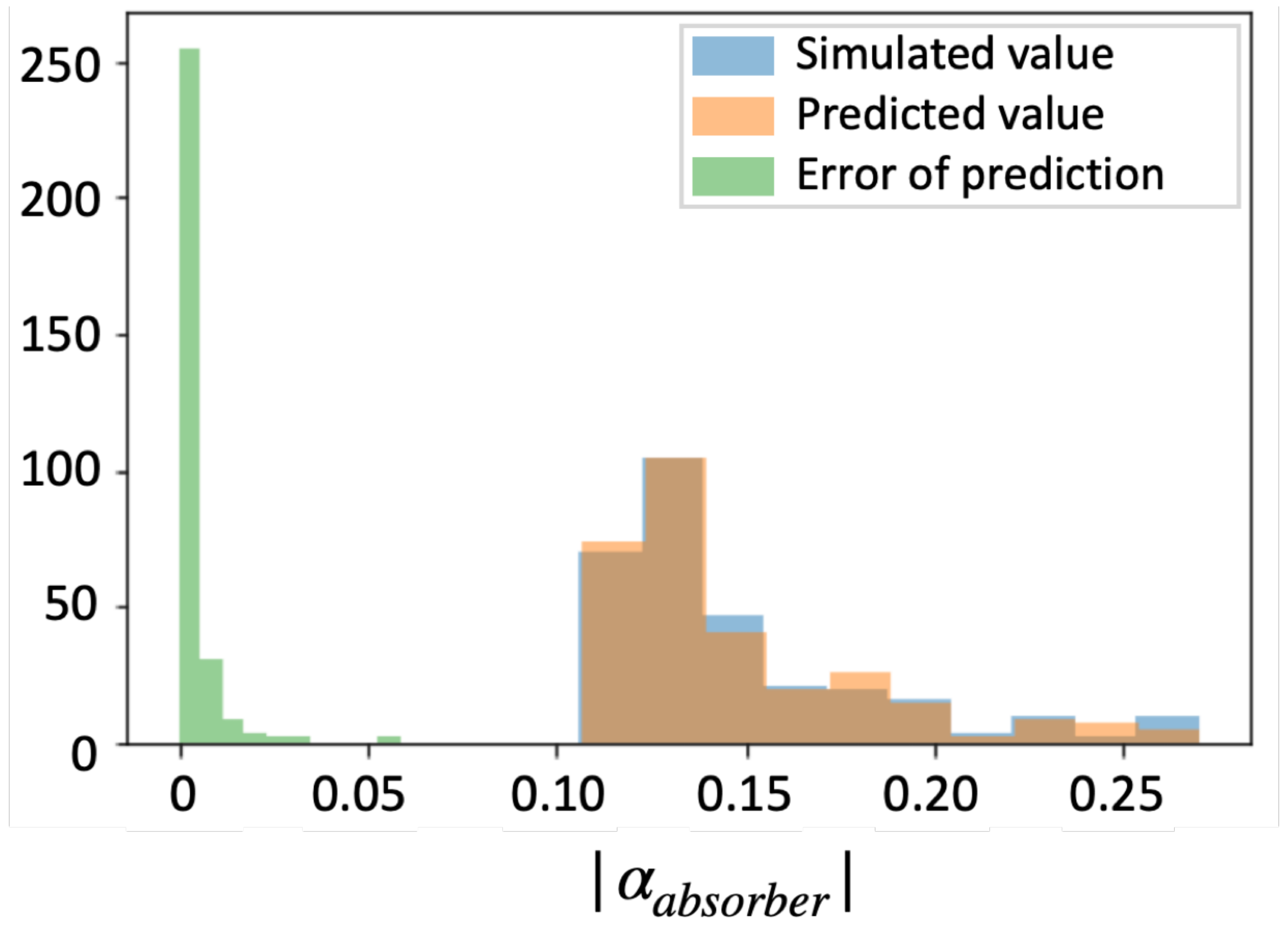
Proof of concept:

1. Run numerical optimisers, **systematically saving the data** (results of tracking simulations using ICOOL)

2. Train a **surrogate model** (Random Forest Regressor):
- input = parameters of the solenoid field in a cooling cell
- output = optics observables

3. **Replace time-costly simulations** with ML model, find optimal parameters





- ✓ Compute optimization function from ML-model prediction
- ✓ Optimization in a **few minutes instead of ~1.5 hours** for 200 steps using tracking simulations

Longitudinal phase-space optimization: Bayesian Optimization

- **Free parameters:**
  - Absorber (liquid hydrogen) thickness
  - Drift length
  - Number of accelerating RF cavities, rf phase
  - Number of rotating RF cavities, rf phase
  - B-field in RF region to match the field in the cooling cell and the change in momentum
  - **Objective function** : $\frac{\epsilon_{\perp}\epsilon_{||}}{\Delta N}$,

    obtained using RF-Track simulation code

    (developed by A. Latina https://gitlab.cern.ch/rf-track)

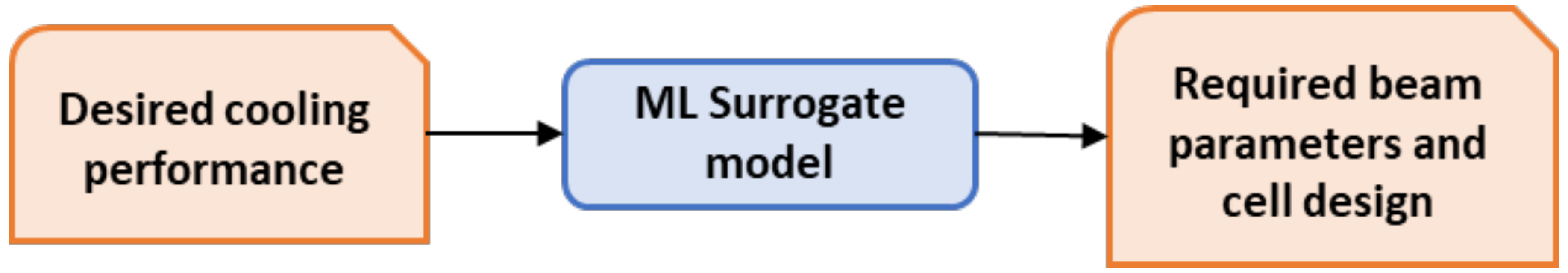
Longitudinal phase-space optimization: Bayesian Optimization

- **Free parameters:**
  - Absorber (liquid hydrogen) thickness
  - Drift length
  - Number of accelerating RF cavities, rf phase
  - Number of rotating RF cavities, rf phase
  - B-field in RF region to match the field in the cooling cell and the change in momentum
  - **Objective function** : $\frac{\epsilon_{\perp}\epsilon_{||}}{\Delta N}$,

    obtained using RF-Track simulation code

    (developed by A. Latina https://gitlab.cern.ch/rf-track)

- **Optimization procedure:**
  - Run optimization for a cell, a few iterations
  - Create a surrogate model to estimate the initial parameters
  - Bayesian Optimization*, BOBYQA



➡ Fast design estimate

➡ **Use as initial guess** for optimisation algorithms (optimal solution is found within fewer steps)

- * **Update probabilistic model** based on function evaluation
- Optimise an acquisition function (e.g. expected improvement) for sampling the new optimisation step
- Balance exploration and exploitation by controlling parameters of acquisition function
- Surrogate Model: Boosted Decision Trees
- Skopt implementation (https://scikit-optimize.github.io)

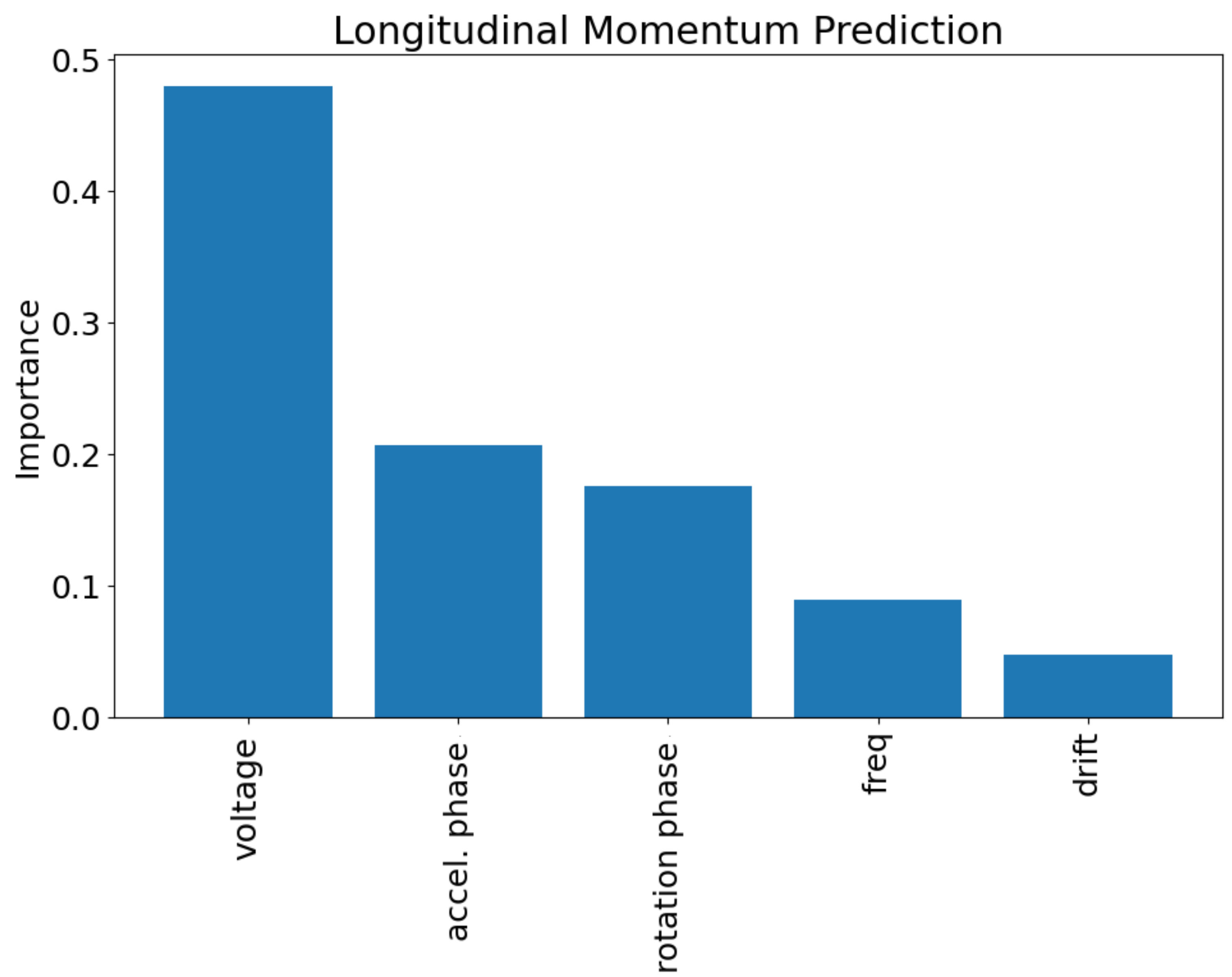
Model interpretability: permutation features importance

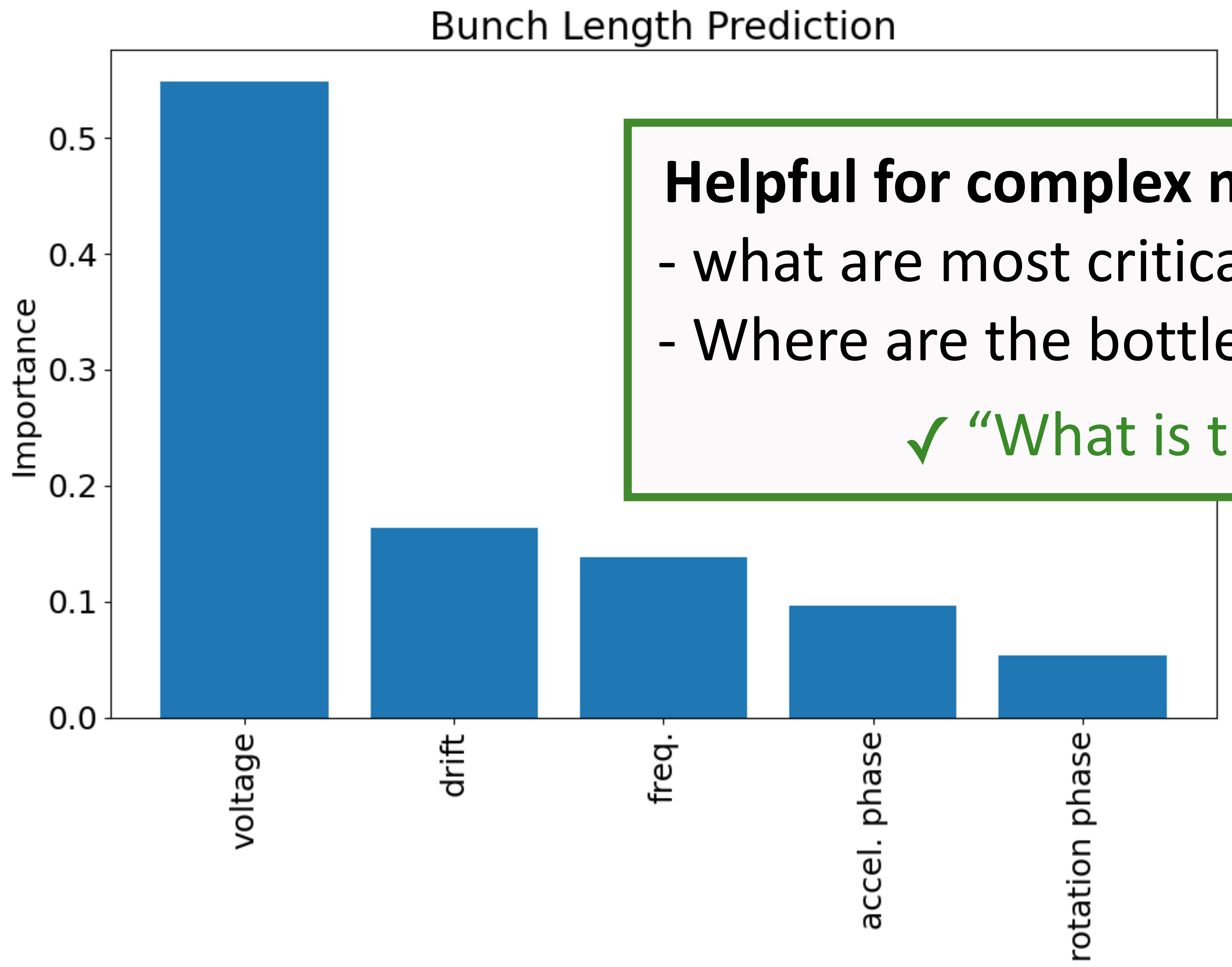
<u>Feature permutation</u>

- Measuring how much **model's performance decreases** when each **feature is randomly shuffled**
- Identify **which features have greatest impac**t on model's output

<u>Example: optimization of RF in cooling cells:</u>

- Model created from optimization data: <u>Input</u>: **cell set up**, <u>output</u>: **beam parameters** at the end of a cooling cell





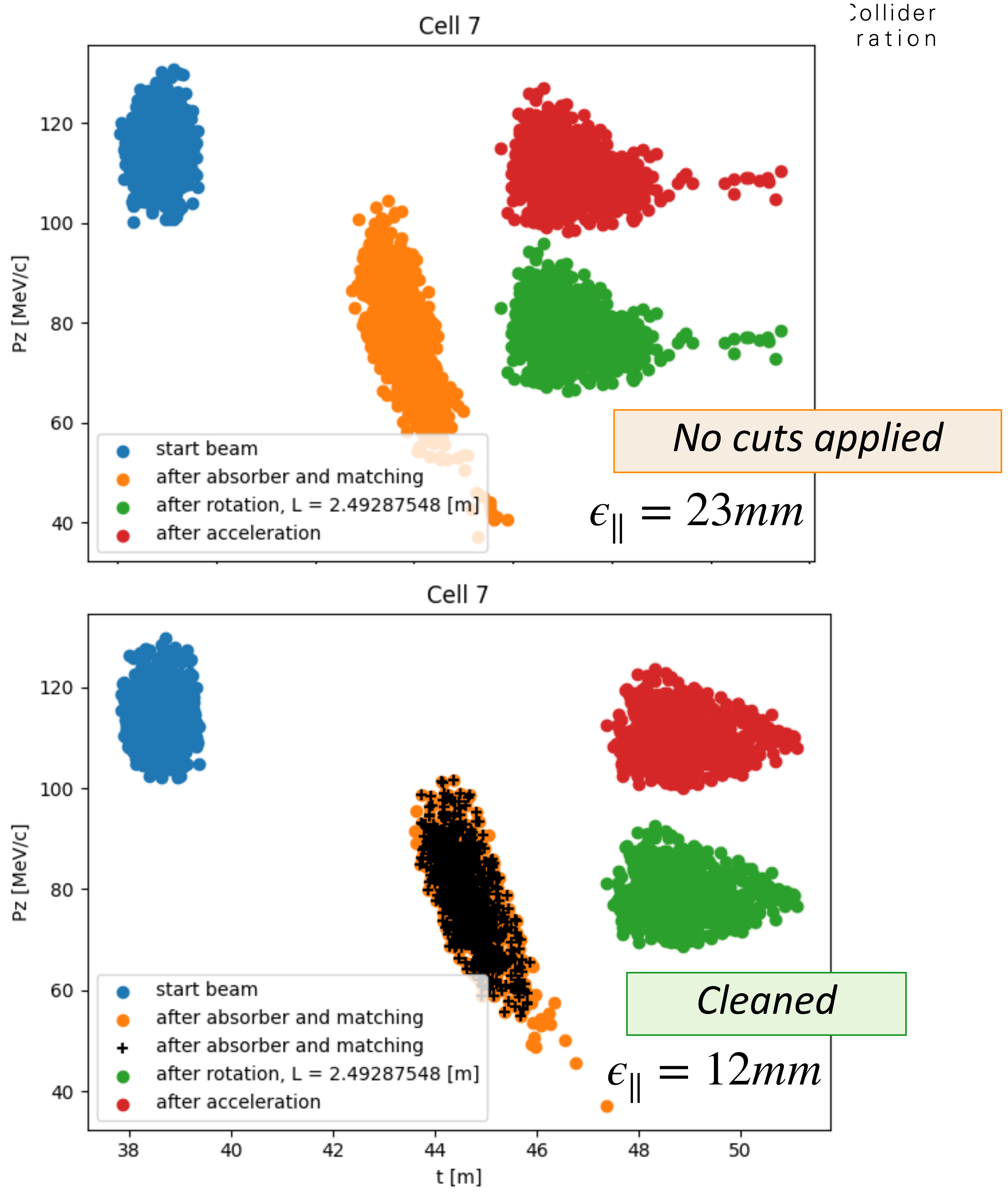
Helpful for complex models:
- what are most critical parameters to be optimised?
- Where are the bottle necks?

✓ "What is this model actually learning?"

Final cooling optimization: robust emittance estimation

Objective function : $\dfrac{\epsilon_{\perp}\epsilon_{||}}{\Delta N}$

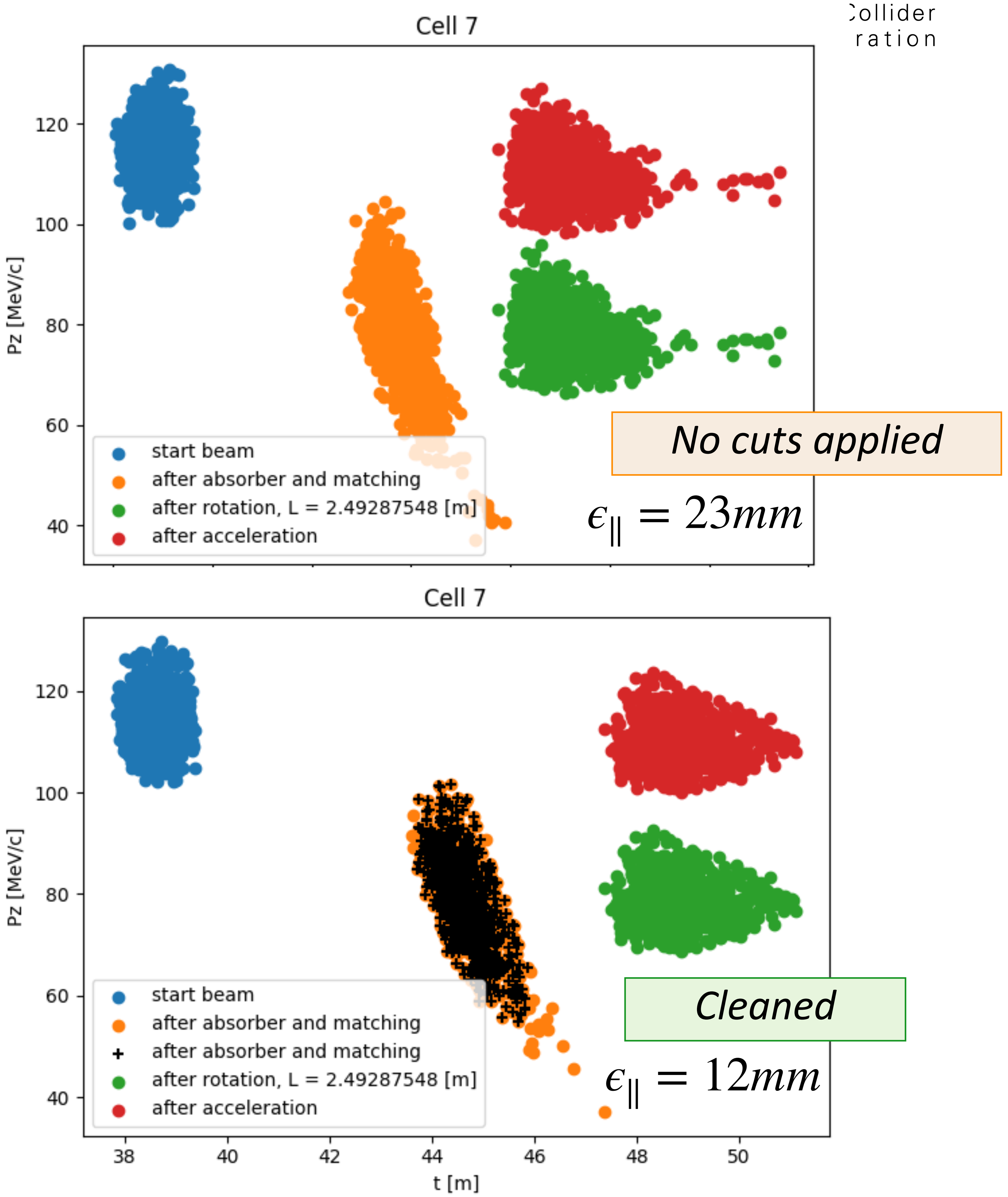
- Too high emittance can be caused by a few "outliers"
- Traditional "3 sigma-cut" not always reliable, especially towards the end of the channel
- Robust algorithm to exclude the outliers before evaluating the emittances?



Final cooling optimization: robust emittance estimation

Objective function : $\dfrac{\epsilon_{\perp}\epsilon_{||}}{\Delta N}$

- Too high emittance can be caused by a few "outliers"
- Traditional "3 sigma-cut" not always reliable, especially towards the end of the channel
- Robust algorithm to exclude the outliers before evaluating the emittances?
- Comparing anomaly detection techniques, density-based clustering
- **Unsupervised Learning (no data, no training needed), fast-executable**



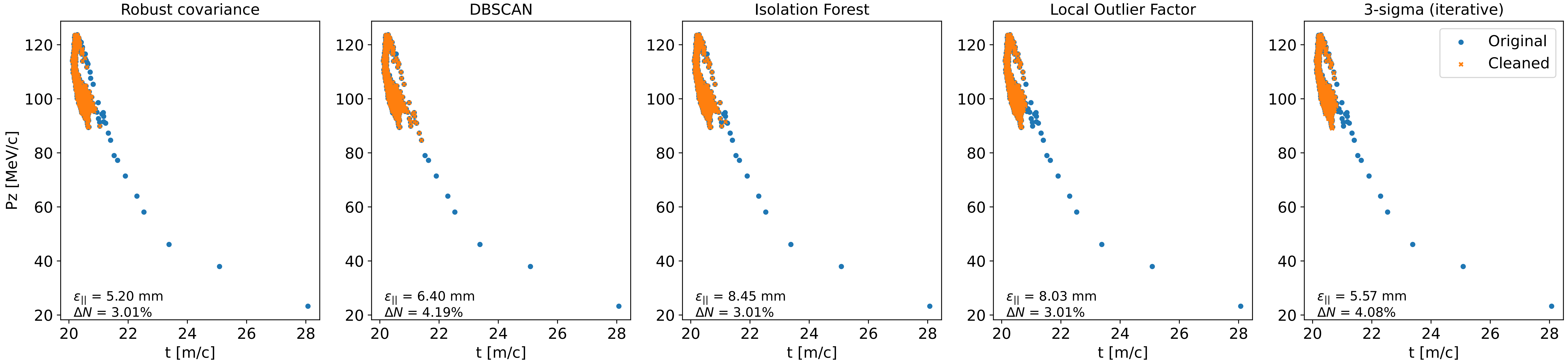
✓ Minimum Covariance Determinant (MCD): robust estimator of covariance

- Detecting "lost" particles based on the whole **6D phase space**
- Provides a "clean" covariance matrix

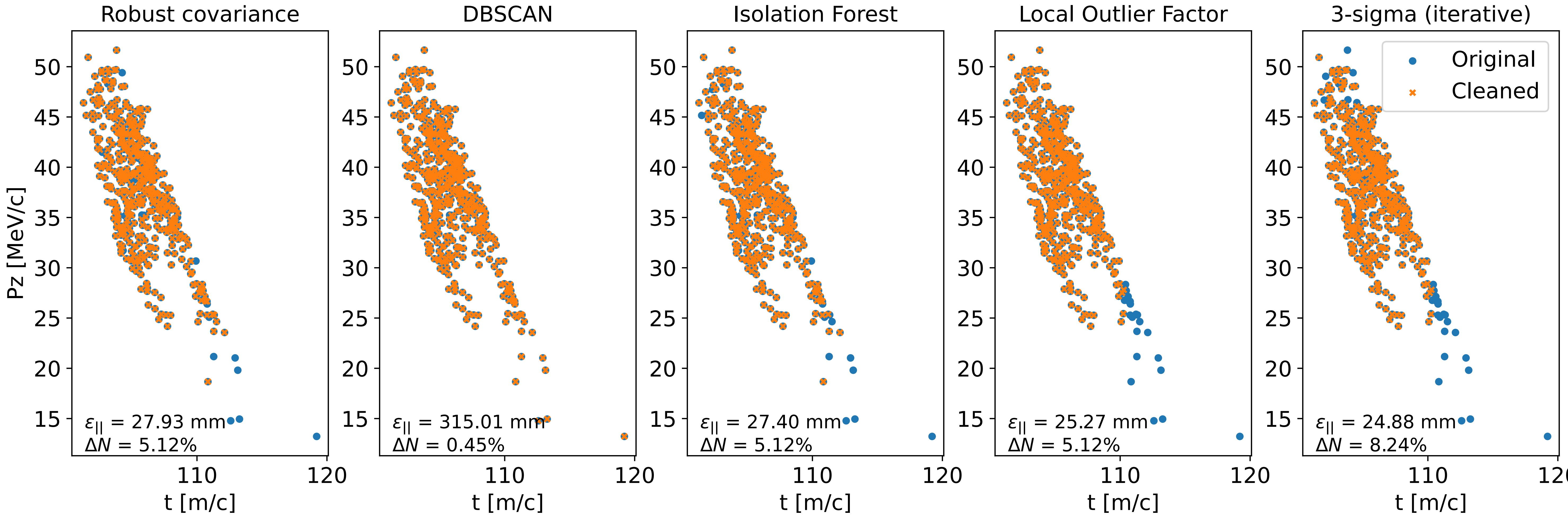
➡ Direct computation of emittances and optics observables possible

Robust Emittance Computation vs. Other techniques

Example: cell 3



Example: cell 6



Cooling performance

Preliminary

	$\epsilon_\perp [\mu m]$	$\epsilon_{||}$ [mm]	N [%]
3σ-cut	39	85	35
IF	33	82	33
MCD	35	80	38
No cuts	46	106	42

CERN

Summary

LHC Optics Commissioning

- A toolbox of ML methods shows a great potential to **save operational time**
- Providing closer look at the **individual magnet errors** and causes of optics perturbations
- **Fast virtual diagnostics** for time-consuming measurements

Muon Collider Design (Final Cooling Channel)

- **Surrogate models** for both, fast objective function evaluation and estimation of initial parameter
- **Bayesian Optimization** combining modelling and optimization
- **Anomaly detection techniques** for robust emittance analysis
- "Proof-of-concept": **Opening several opportunities for accelerator design studies:** identification of most critical parameters for collider performance (e.g. **feature importance analysis**, but also dimensionality reduction techniques)
- **Fast-executable methods** for changing requirements as design evolves

Practical Advice

- Start with **simpler models** - they are **easier to tune and interpret**. Neural Networks are not always the perfect solution!
- Numerical Optimisers are powerful tools and can be made even more efficient using **surrogate models** - **save and structure your data!**
- Not all ML algorithms need large amount of data - consider translating your problem as **Unsupervised Learning** task (e.g. anomaly detection)

Thanks a lot for your attention!