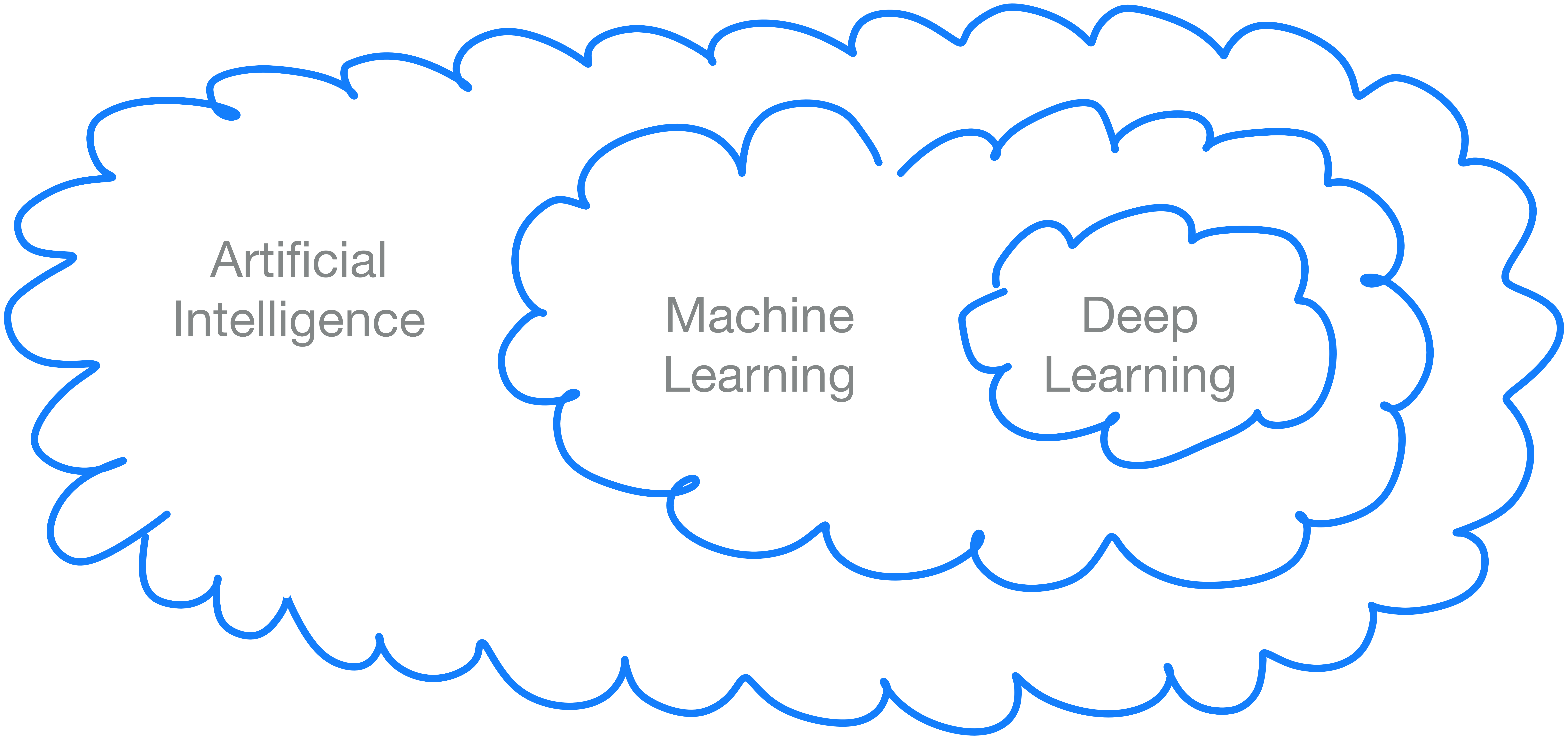
PAUL SCHERRER INSTITUT

Rasmus Ischebeck

MACHINE LEARNING IN PARTICLE ACCELERATORS

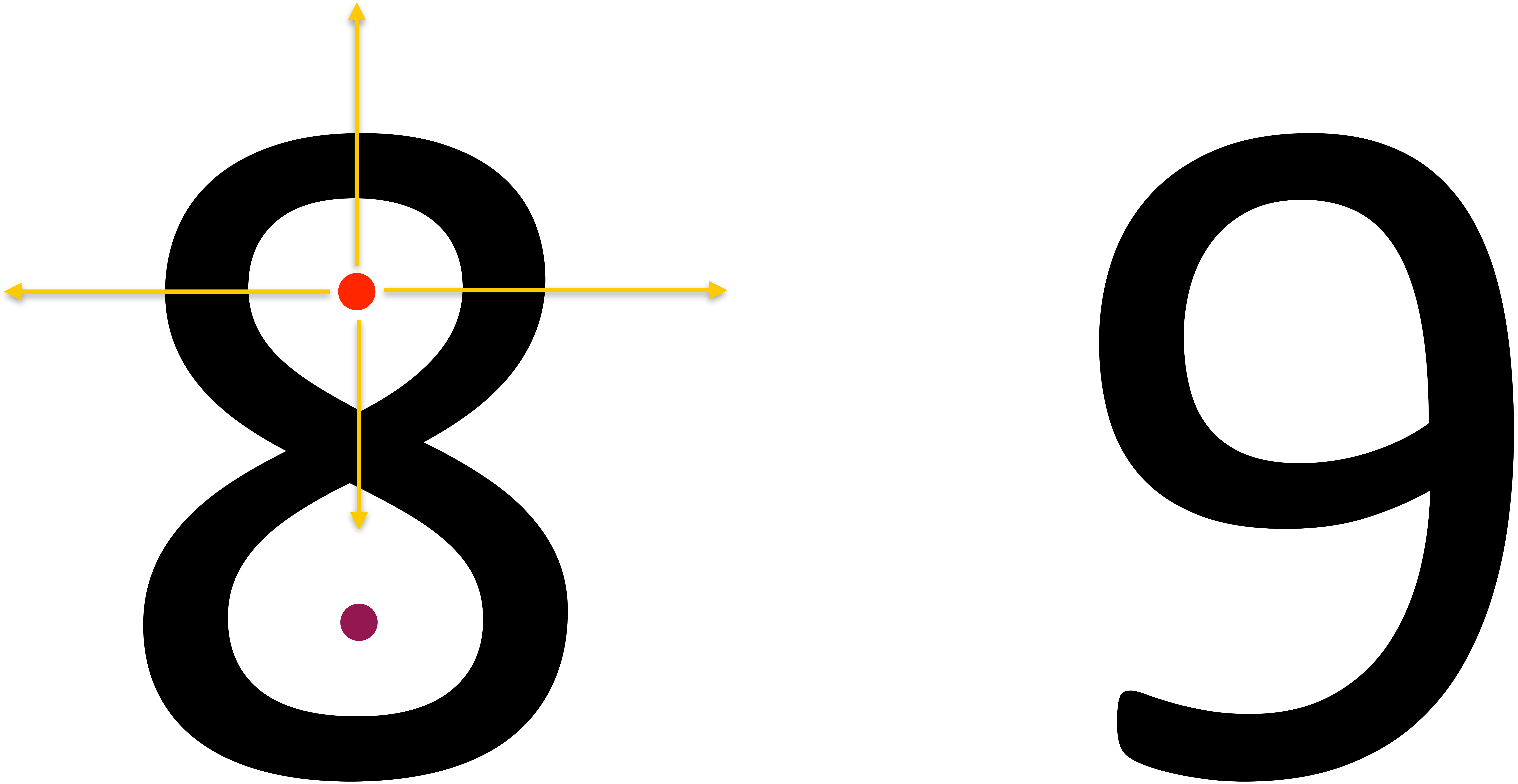
MACHINE LEARNING LANDSCAPE

- Before we start, a bit of context



MACHINE LEARNING LANDSCAPE

- What is an “eight”?



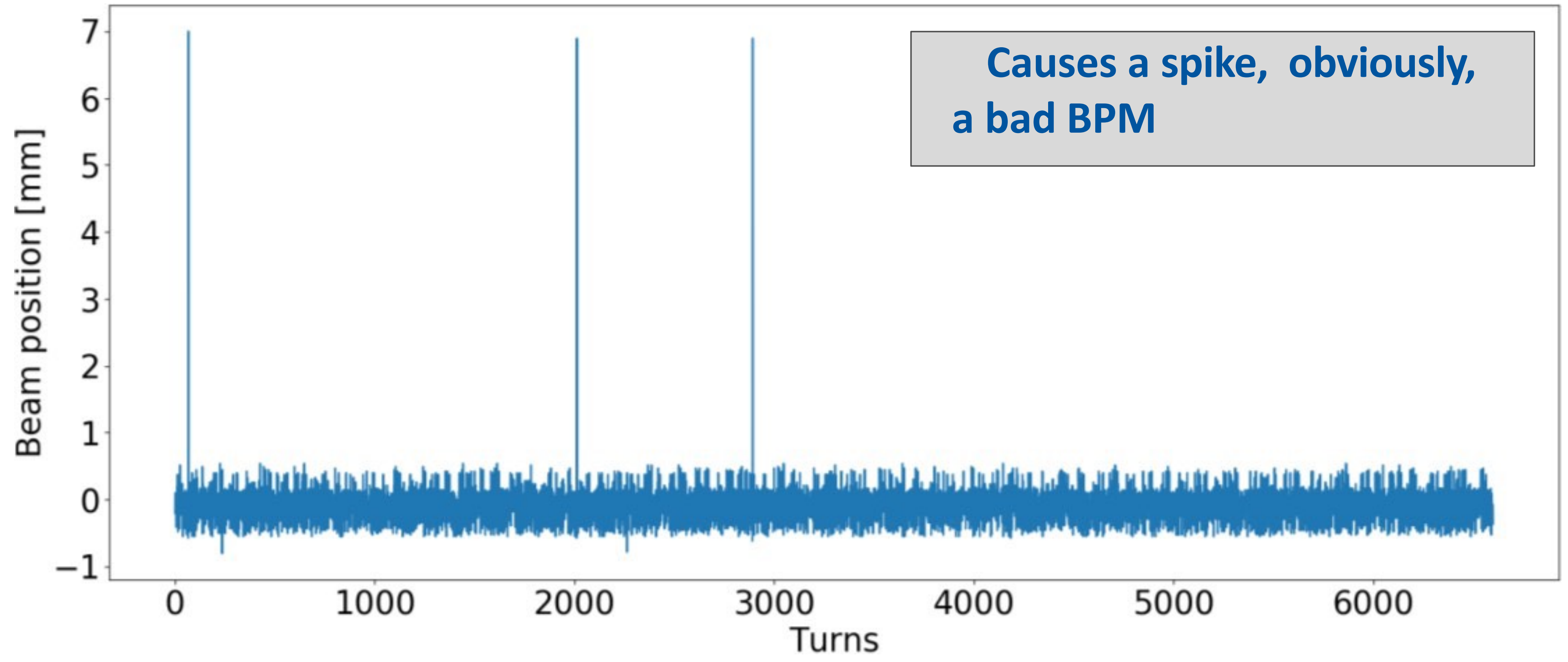
MACHINE LEARNING LANDSCAPE

1 2 3 4 5 6 7 8 9 0

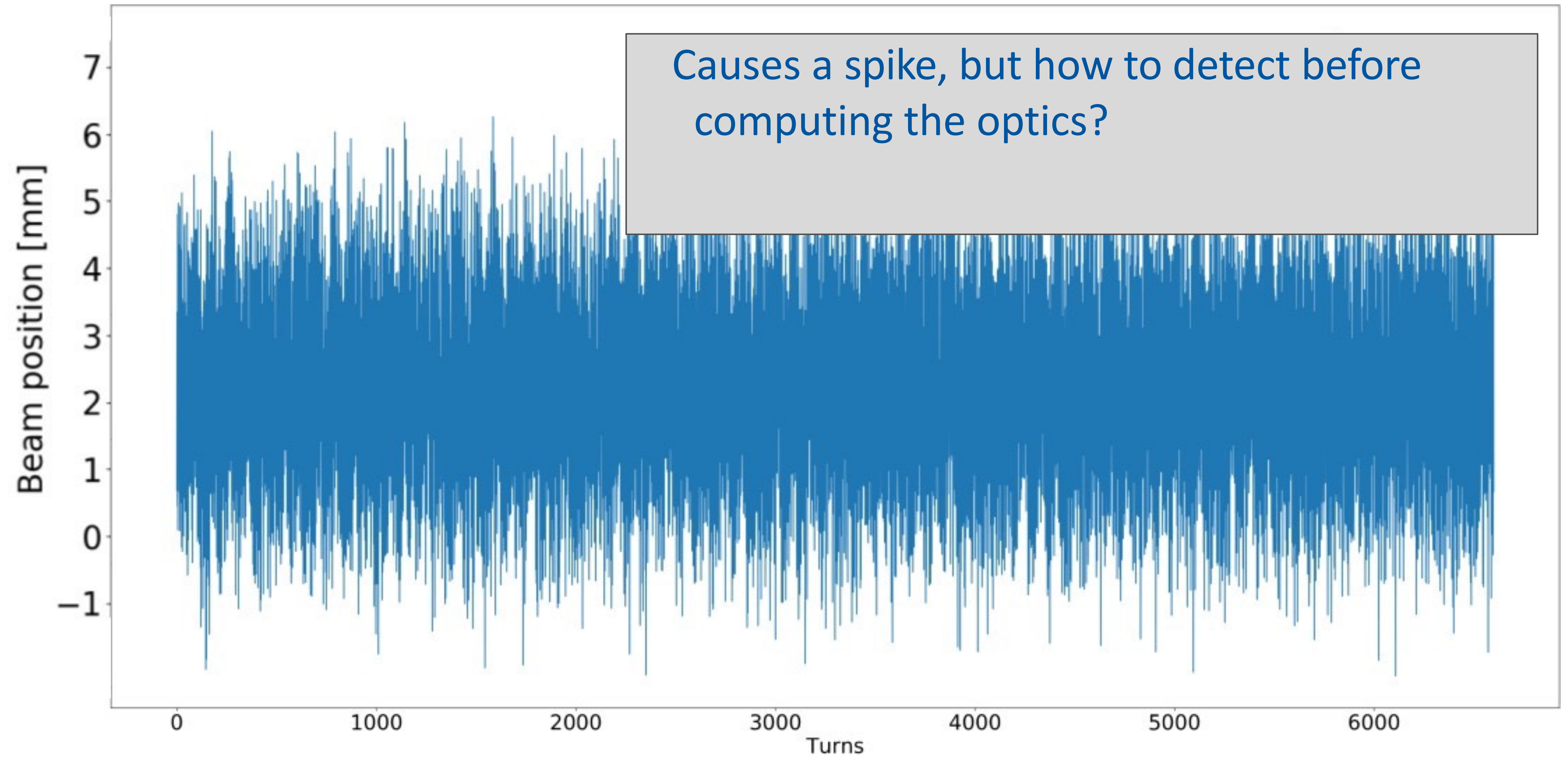
1 2 3 4 5 6 7 8 9 0

APPLICATIONS TO ACCELERATORS

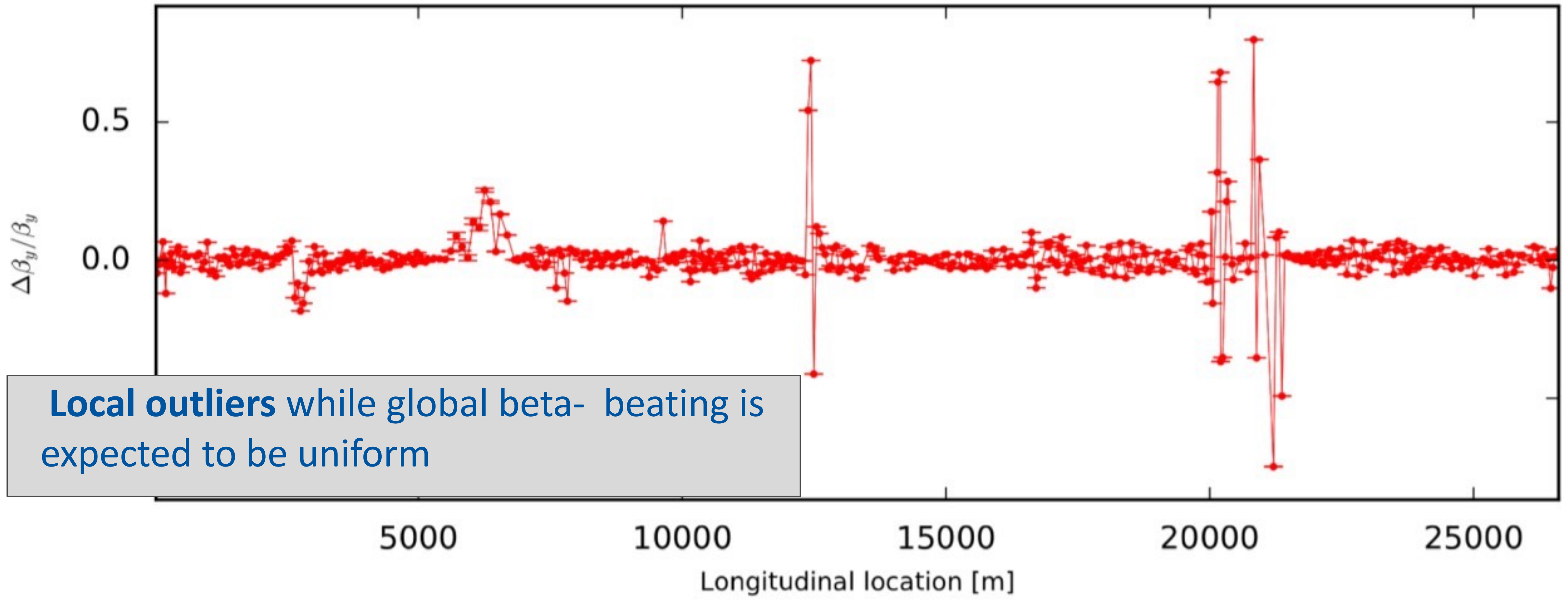
- What is a “faulty BPM”?



Causes a spike, obviously, a bad BPM



Causes a spike, but how to detect before computing the optics?



Local outliers while global beta- beating is expected to be uniform

“AN OBSERVATION WHICH DEVIATES SO MUCH FROM OTHER OBSERVATIONS AS TO AROUSE SUSPICIONS THAT IT WAS GENERATED BY A DIFFERENT MECHANISM."

Stephen Hawking

Relevant ML concepts and definitions

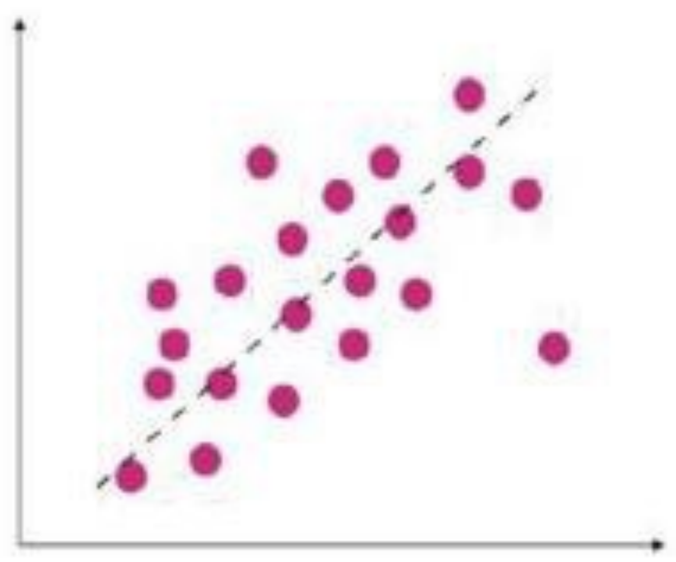
Supervised Learning

- **Input/output pairs** available
- Learn a mapping function, **generalizing for all provided data**
- Predict from **unseen data**

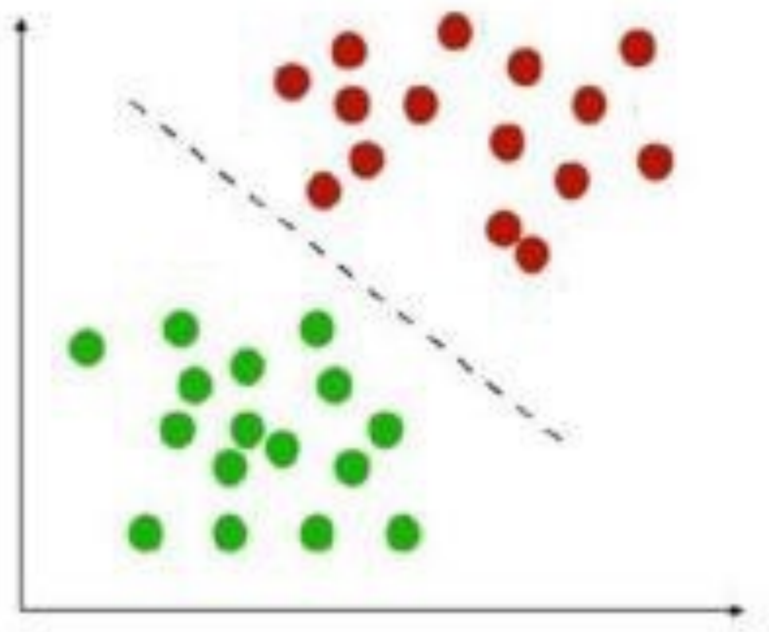
Unsupervised Learning

- **Only input** data is given
- Discover structures and patterns

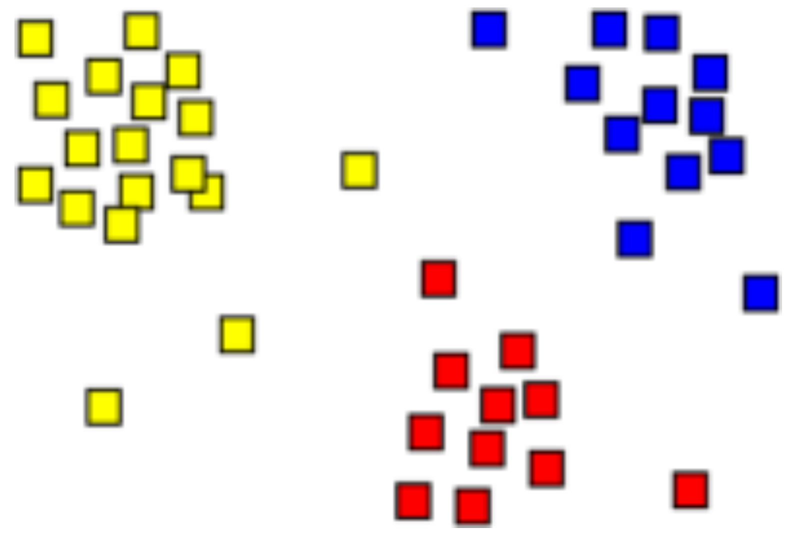
Regression



Classification



Clustering



What is “Learning”?

Elena Fol

MACHINE LEARNING OUTSIDE THE ACCELERATOR WORLD

- Image recognition

MACHINE LEARNING OUTSIDE THE ACCELERATOR WORLD

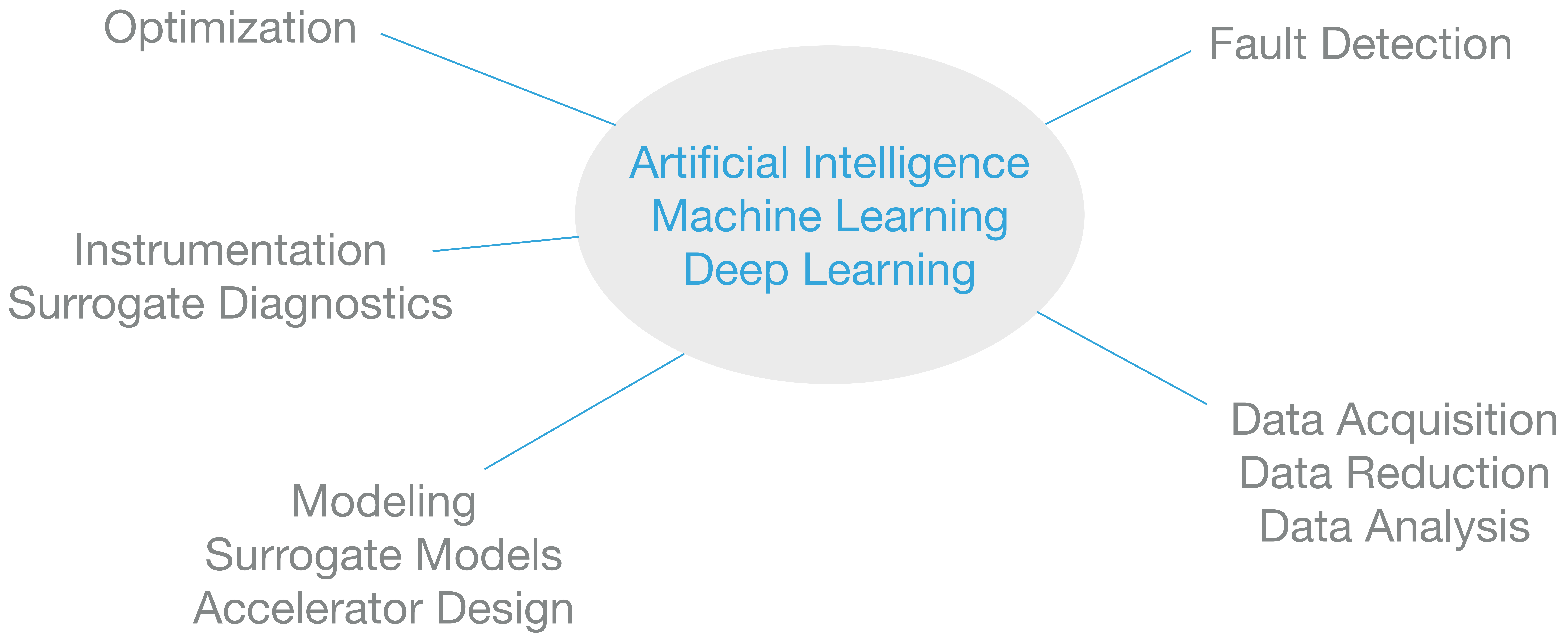
- Social media

MACHINE LEARNING OUTSIDE THE ACCELERATOR WORLD

- Marketing

"The machine learning algorithm wants to know if we'd like a dozen wireless mice to feed the Python book we just bought."

USE OF MACHINE LEARNING FOR ACCELERATORS



EXAMPLES OF APPLICATIONS OF AI/ML/DL

- Cases of problems that could be solved only with AI/ML/DL
  - … ?
- Applications enabled by AI/ML/DL that took less effort than manual coding
  - Safe optimization of accelerators
  - Virtual diagnostics
  - Detection of faulty diagnostics
  - Use of ML for accelerator design
  - Use for data analysis

APPLICATIONS OF ML TO ACCELERATOR OPERATION

SAFE OPTIMIZATION OF ACCELERATOR PERFORMANCE

- Two examples from PSI:
  - SwissFEL: optimization of the pulse energy while keeping losses in the undulators low
  - HIPA: optimization of an accelerator that is limited by beam loss
- One example from Elettra
  - Optimization of the FEL by straightening the trajectory

How does Safe Bayesian Optimization work?

ETH zürich

Learning & Adaptive Systems

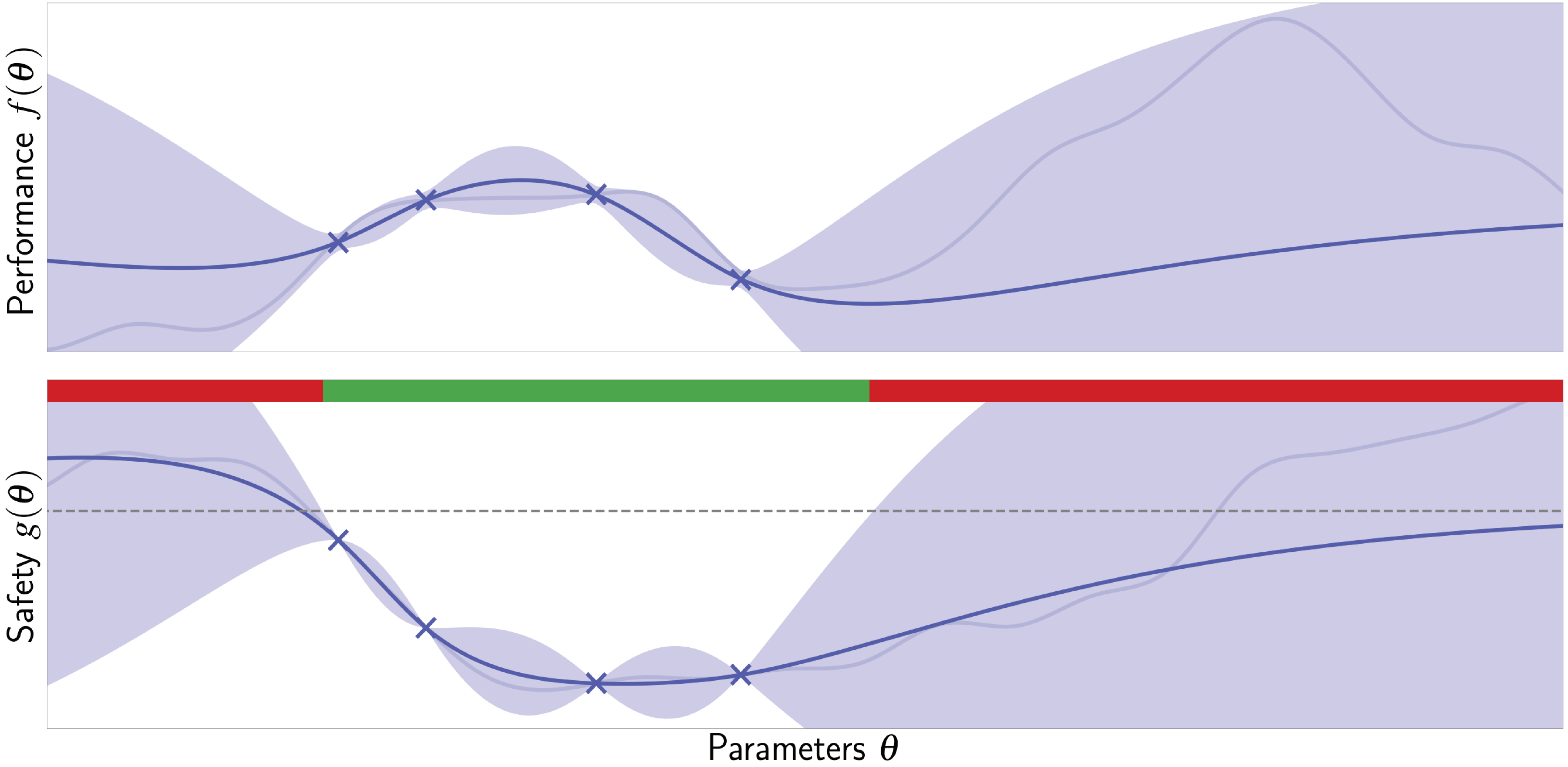


Figure courtesy of J. Kirschner

PAUL SCHERRER INSTITUT PSI

First Results

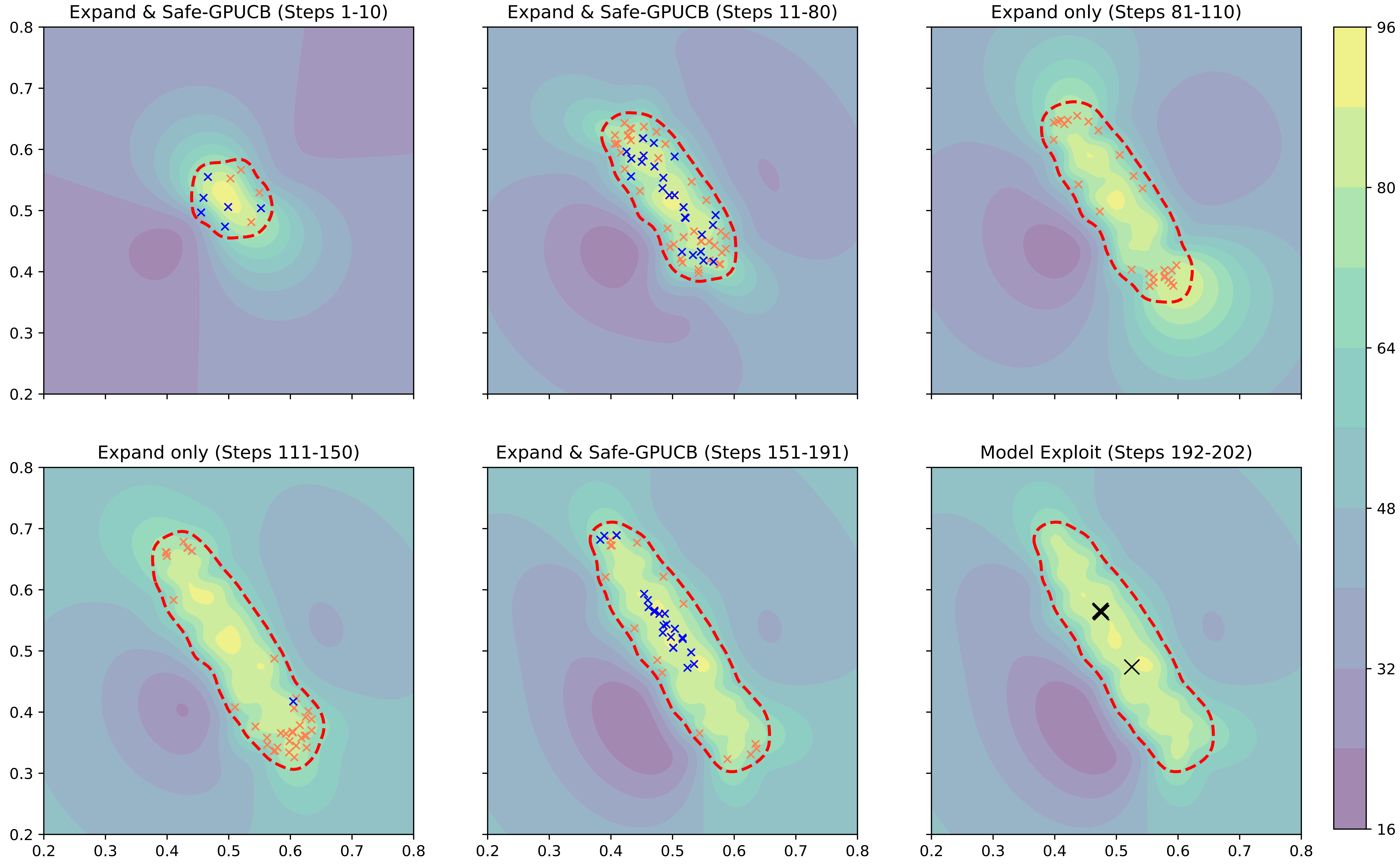


Figure courtesy of J. Kirschner

From Linear Least Squares to Gaussian Processes

Least squares regression in a Hilbert space $\mathcal{H}$:

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \sum_{t=1}^{T} \left(f(x_t) - y_t\right)^2 + \|f\|_{\mathcal{H}}^2$$

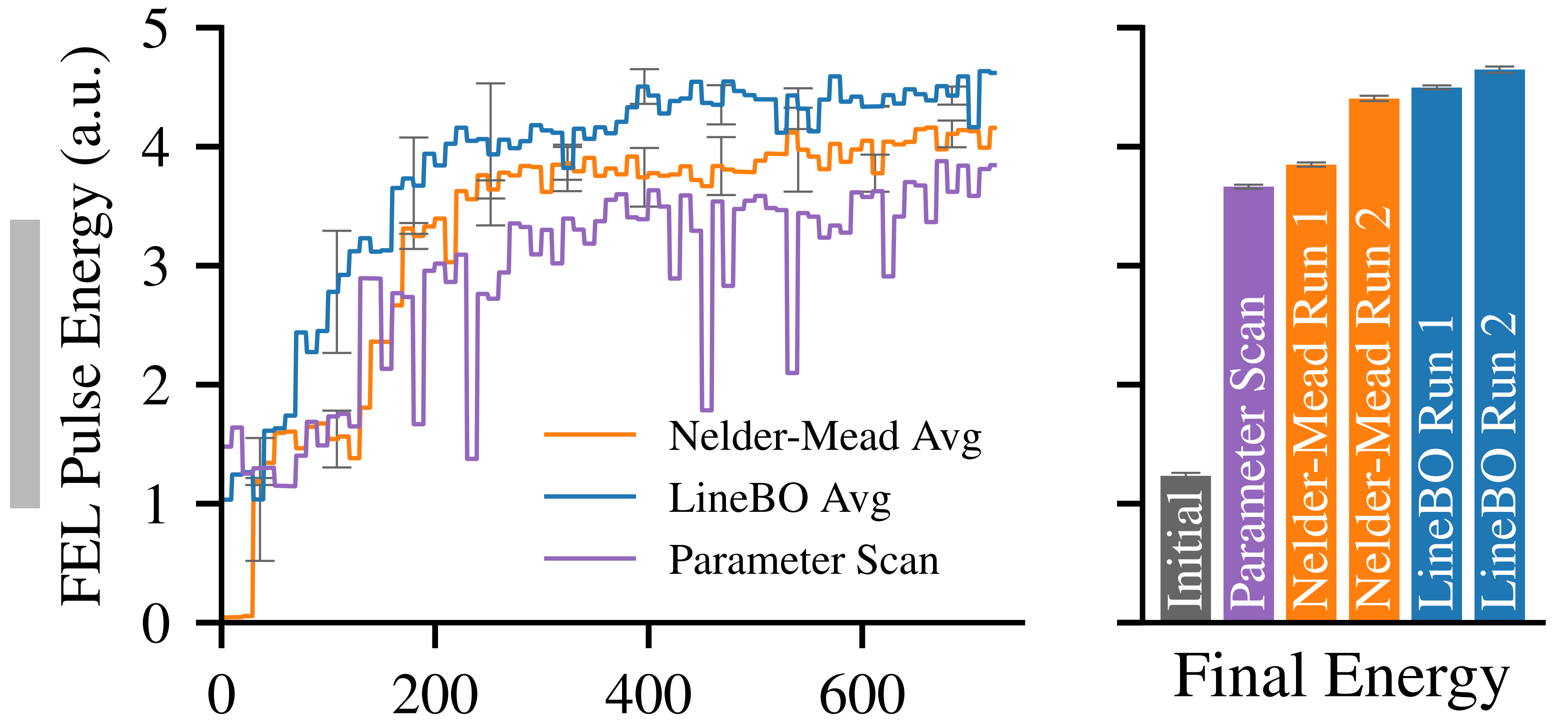
Closed form solution if $\mathcal{H}$ is a *Reproducing Kernel Hilbert Space*!

Defined by a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

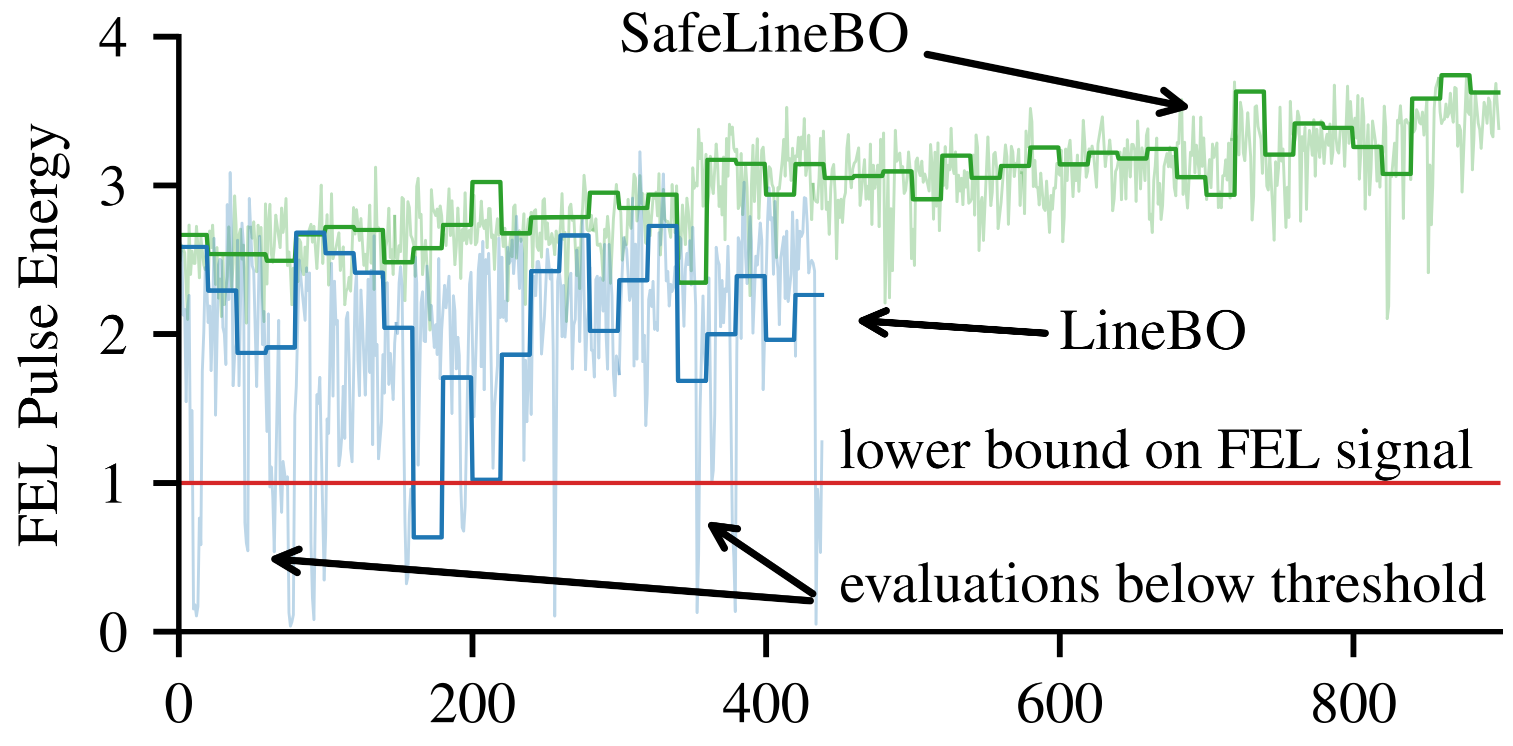
Example: **RBF Kernel** $k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$

Kernel characterizes smoothness of functions in $\mathcal{H}$.

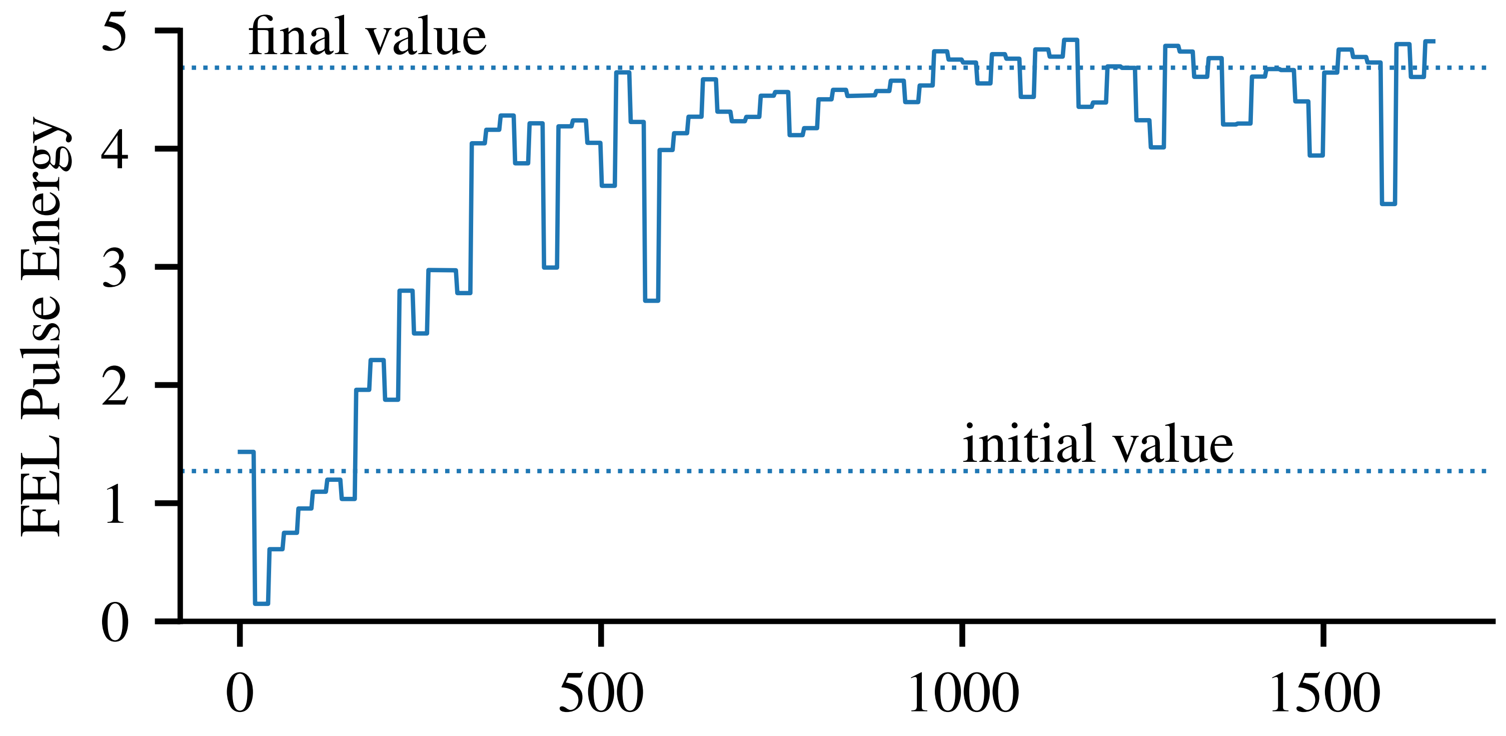
COMPARISON OF OPTIMIZATION ALGORITHMS



Benchmarking against other algorithms!
(24 Parameters)



Adding Safety Constraints
(24 Parameters)



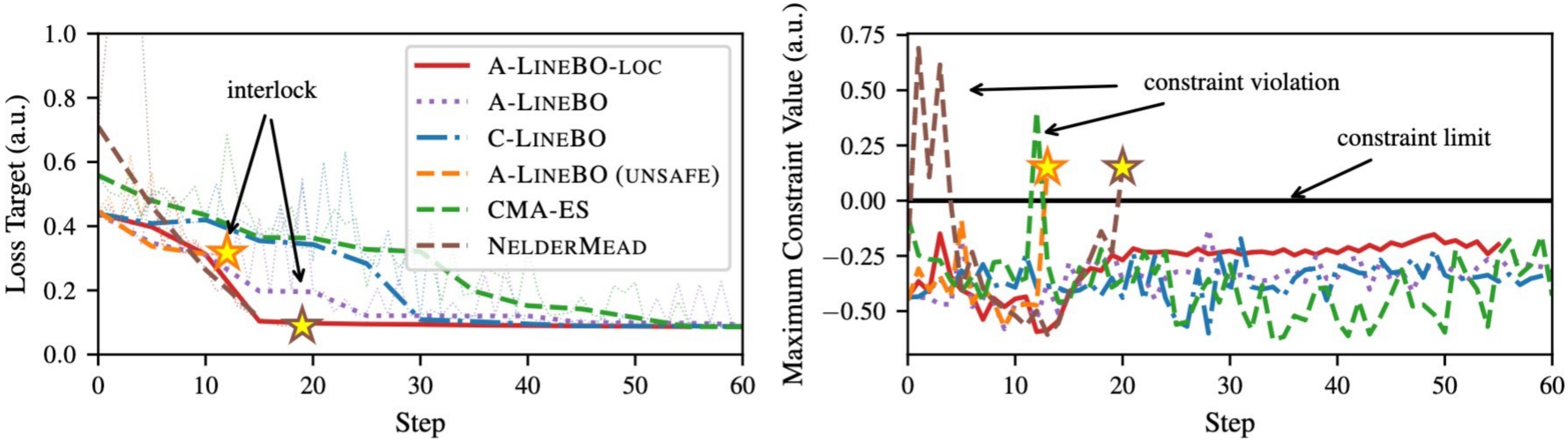
Tuning at about 1 Hz: 500 steps = 8 min

Still works fast with 40 Parameters!

8

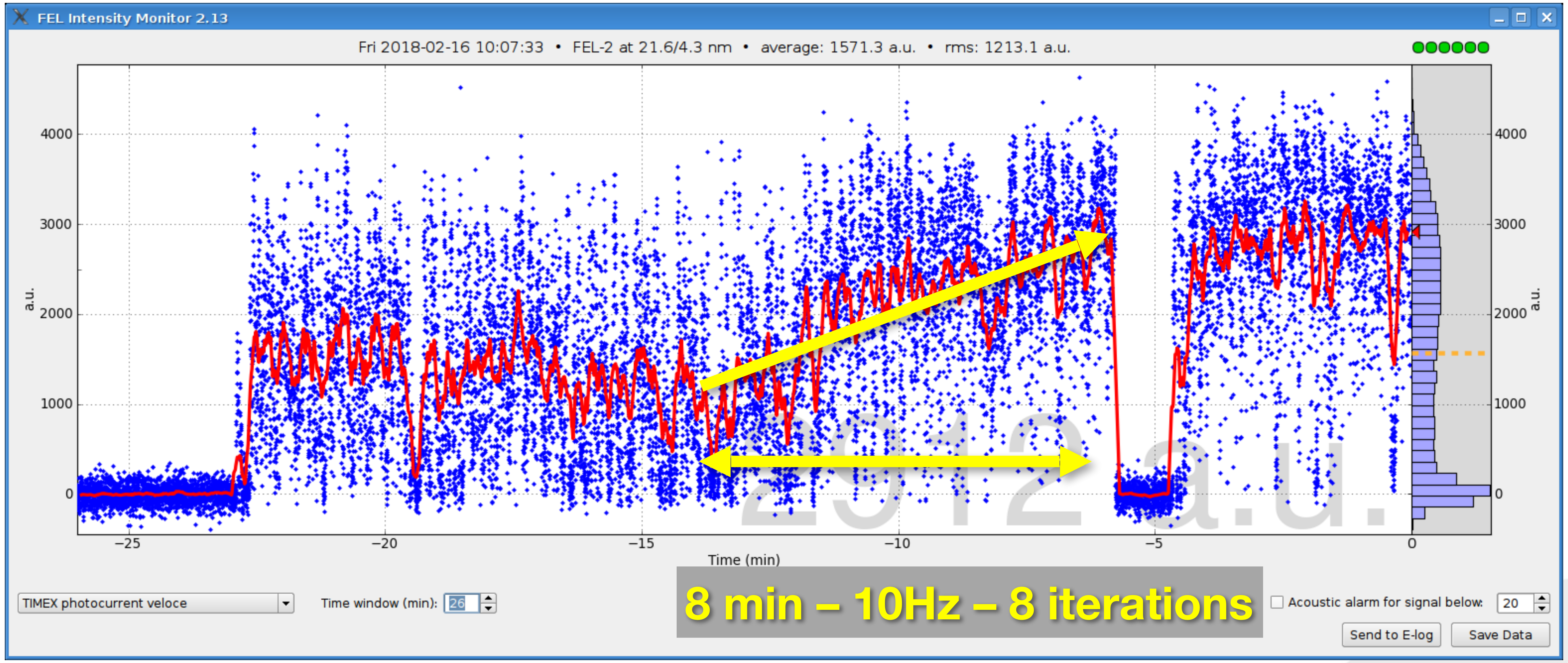
HIPA RESULTS

- safe variants competitive
- non-safe methods create interlocks (violate constraints)
  - proves constraints are working



Statistical Optimization of the electron beam trajectory in the undulators

- 13 Beam Position Monitors (26 variables)
- **Objective function:** FEL-2 intensity estimated by a photocurrent of a mirror installed in front of the experimental chamber
- **Passive mode** (no excitation)
- **3X** average FEL energy

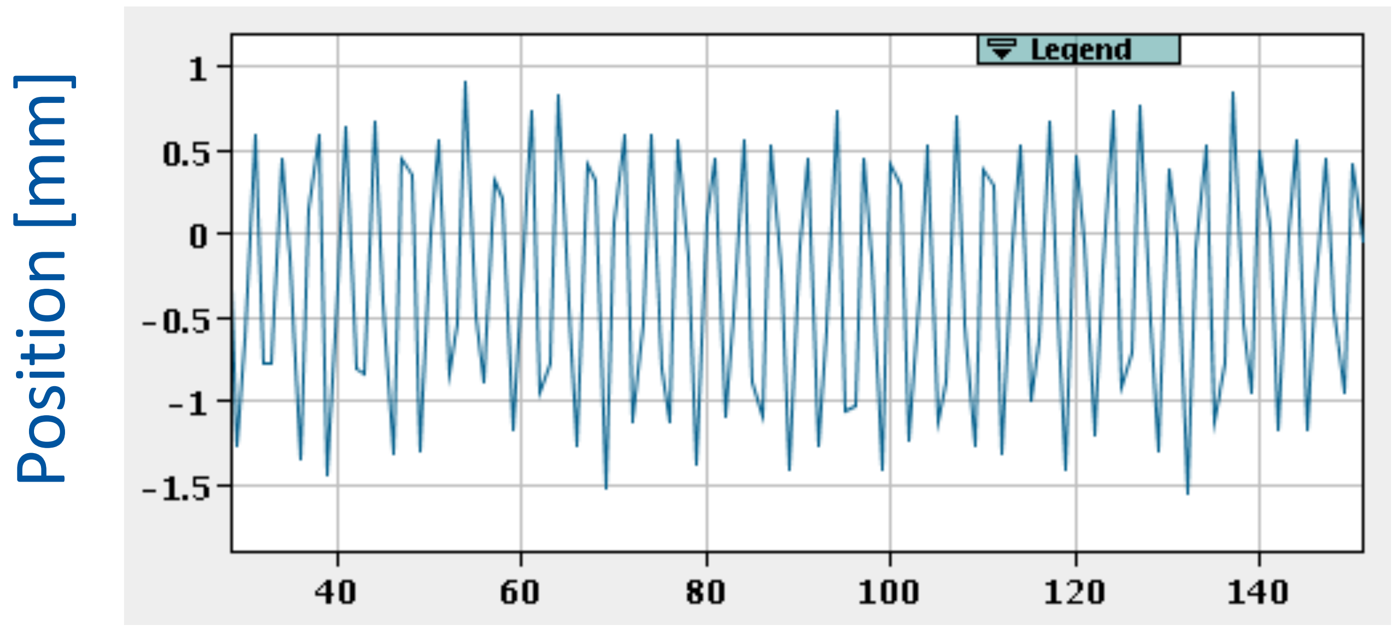


DETECTION OF FAULTY DIAGNOSTICS

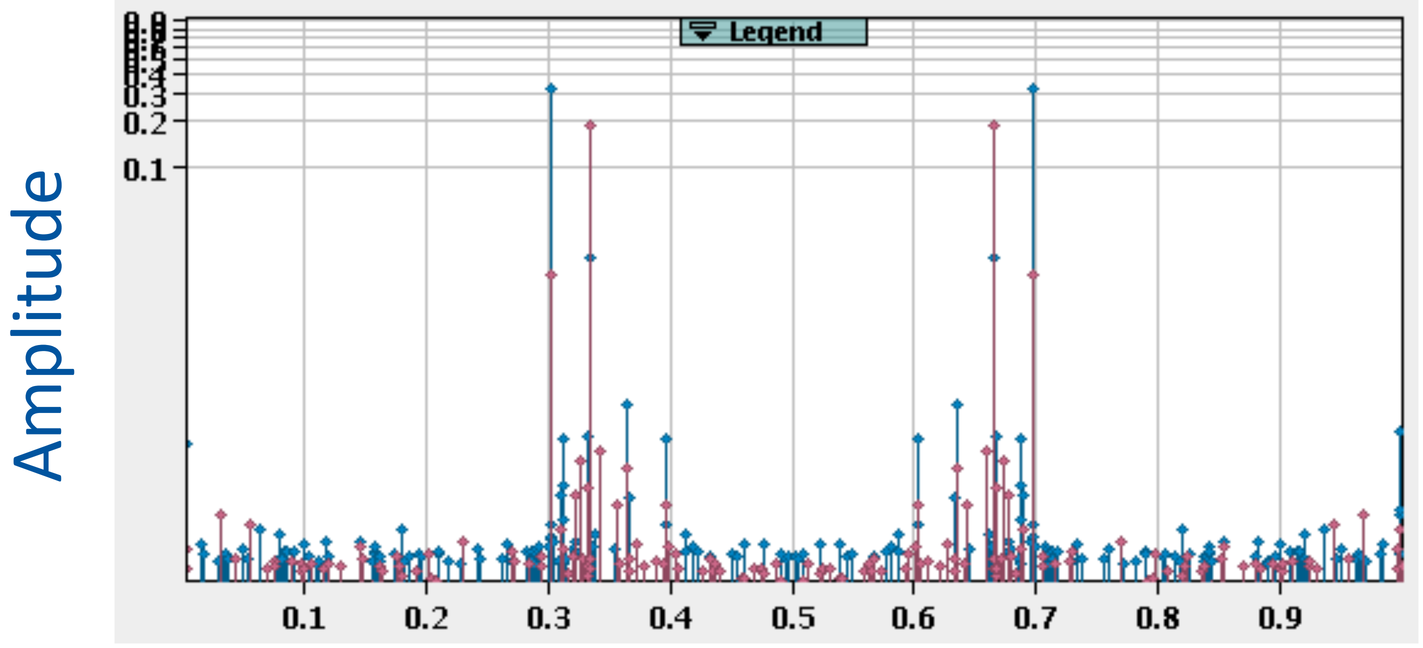
- One example from CERN:
  - Detection of faulty beam position monitors
- One example from DESY:
  - Detection of faults in the superconducting RF system

Measuring the optics

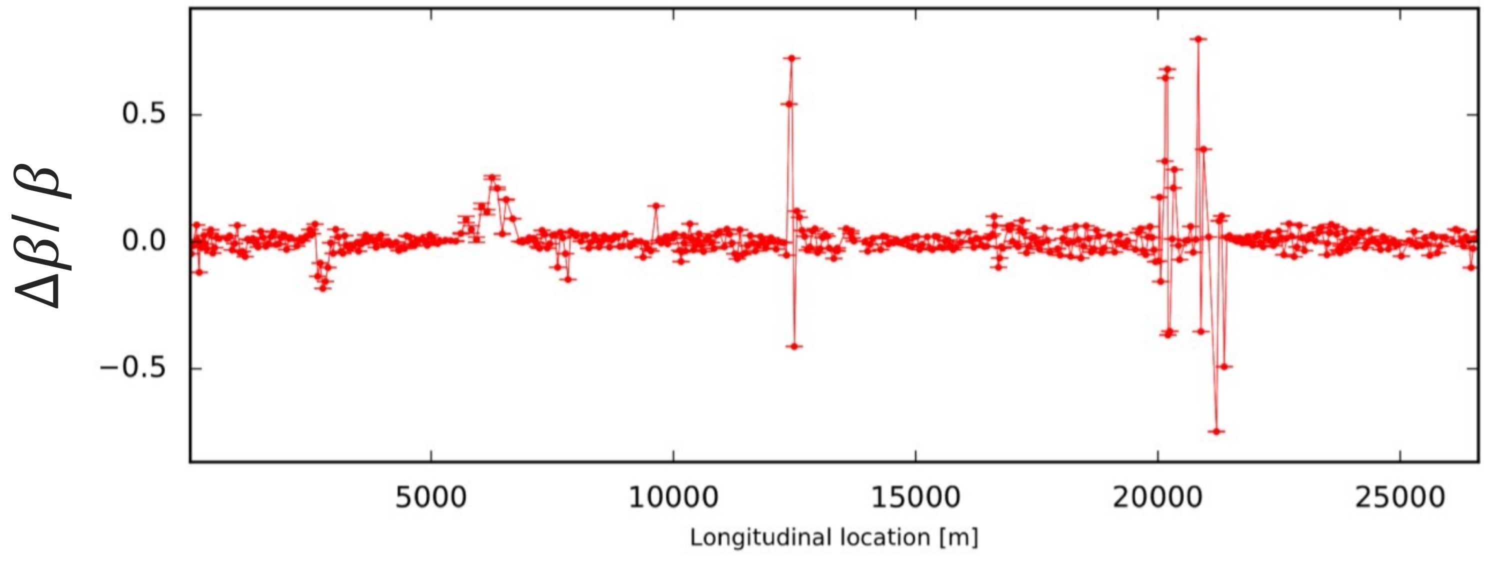
Turn-by-turn beam position



Spectrum



Optics



- Excite the beam to perform transverse oscillations.

➔ **Beam Position Monitors (BPMs) to measure the beam centroid turn-by-turn**

Denoising (SVD)
Signal cuts

- Harmonic analysis using Fast Fourier Transform (FFT)

Semi-automatic and manual cleaning of outliers

- Compute beta-beating and other optics functions

Unphysical values still can be observed

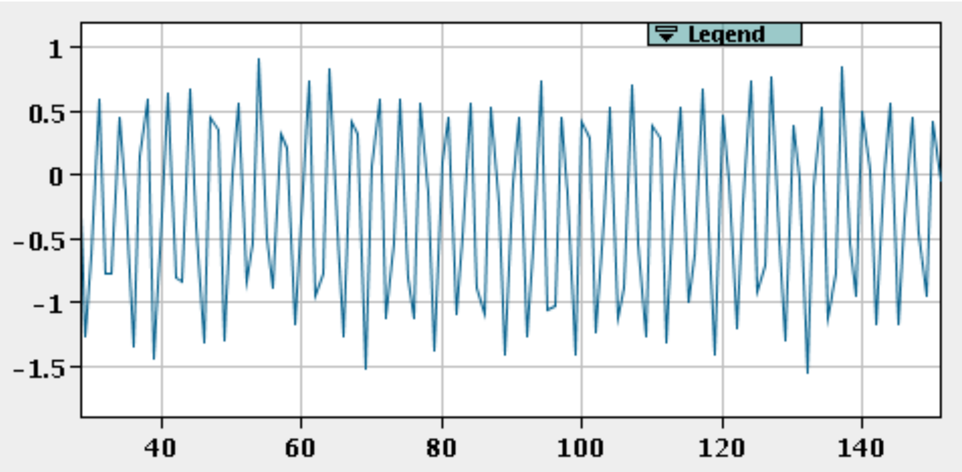
What are the limitations of traditional techniques?

Detection of faulty Beam Position Monitors

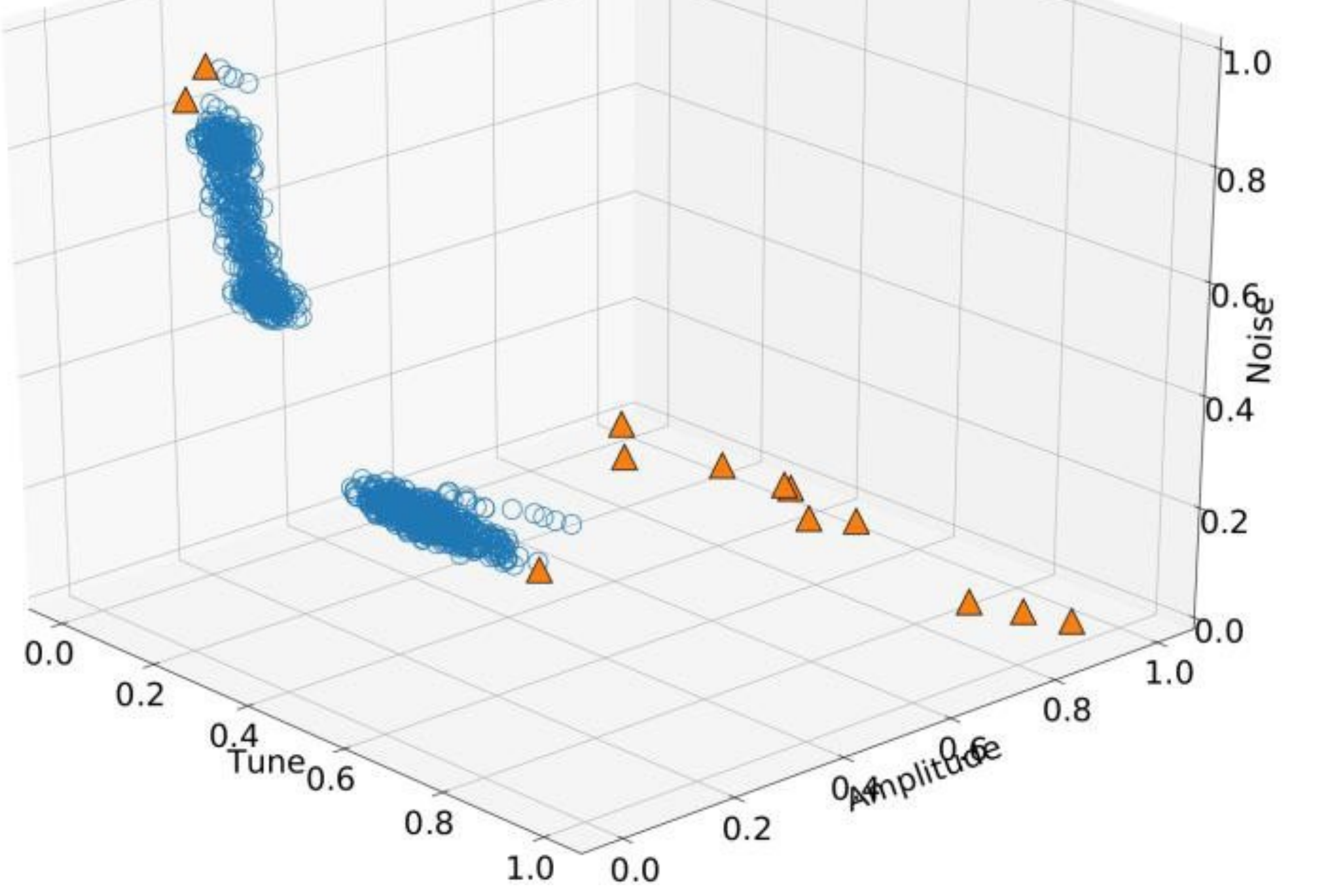
- Faulty BPMs **are a-priori unknown**: no ground truth ➔ **Unsupervised Learning**

- Applied clustering algorithms: DBSCAN[1], Local Outlier Factor[2], anomaly detection using **Isolation Forest**[3] implemented with *Scikit-Learn.*

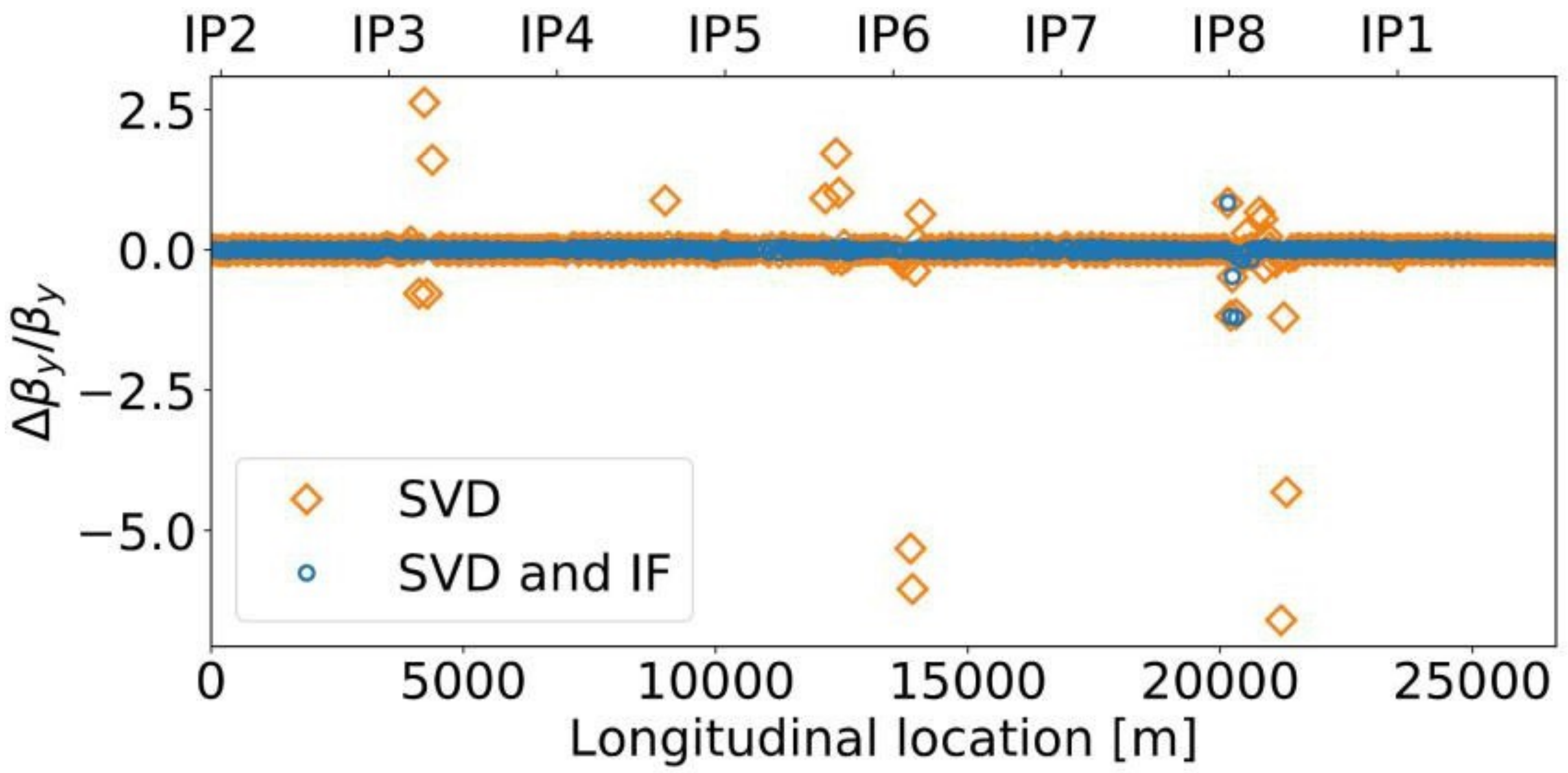
Harmonic analysis of all BPMs



Detection of faulty signal
prior to optics computation



Avoid the appearance of erroneous optics computation



- Outlier detection based on combination of several signal properties
- Immediate results

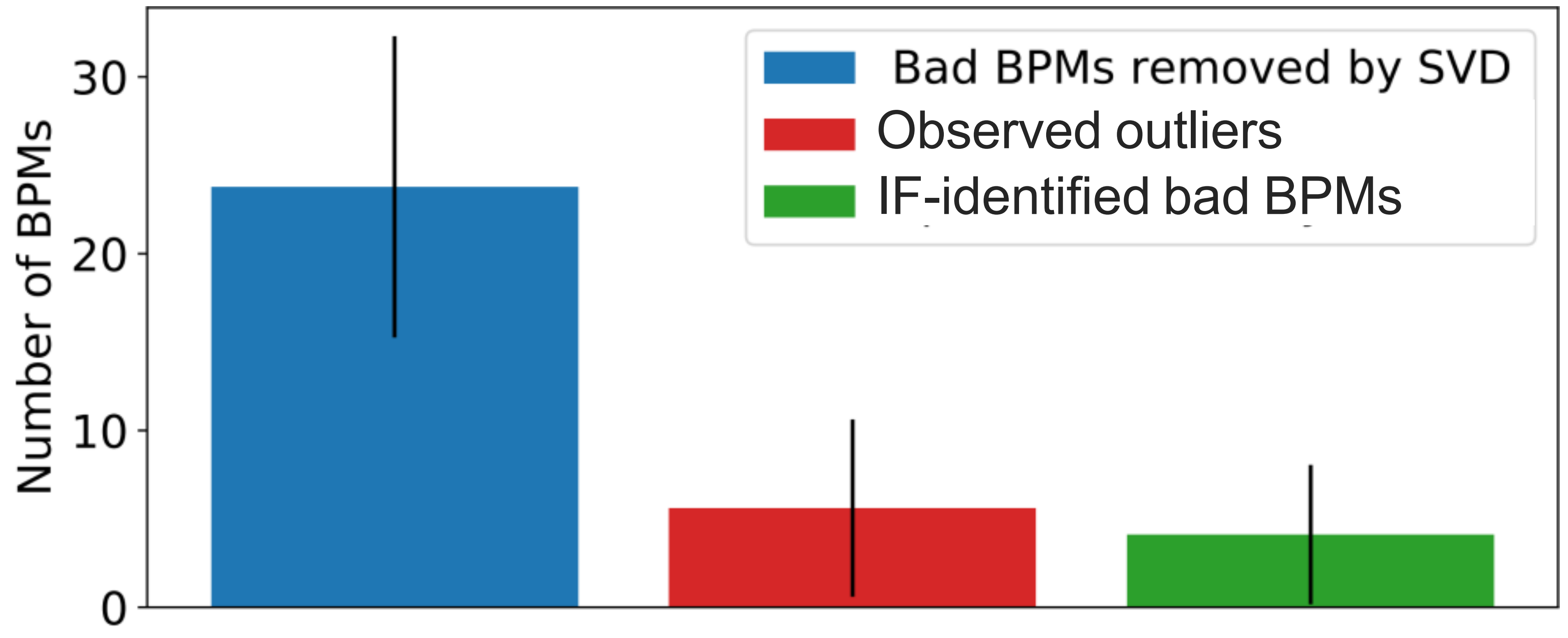
1. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise” Ester, M., H. P. Kriegel, J. Sander

2.Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000, May)., LOF: identifying density-based local outliers

3. Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. “Isolation forest.” Data Mining, 2008. ICDM‘08.

Detection of faulty Beam Position Monitors

Reduction of non-physical outliers in beta-beating:
Averaged cleaning results, optics measurements in 2018.



- Instant faults detection instead of offline diagnostics.
- Full optics analysis is possible directly during dedicated measurements session instead of iterative procedure of cleaning and analysis.

✓ **Fully integrated** into optics measurements at LHC
✓ **Successfully used in operation** under different optics settings.

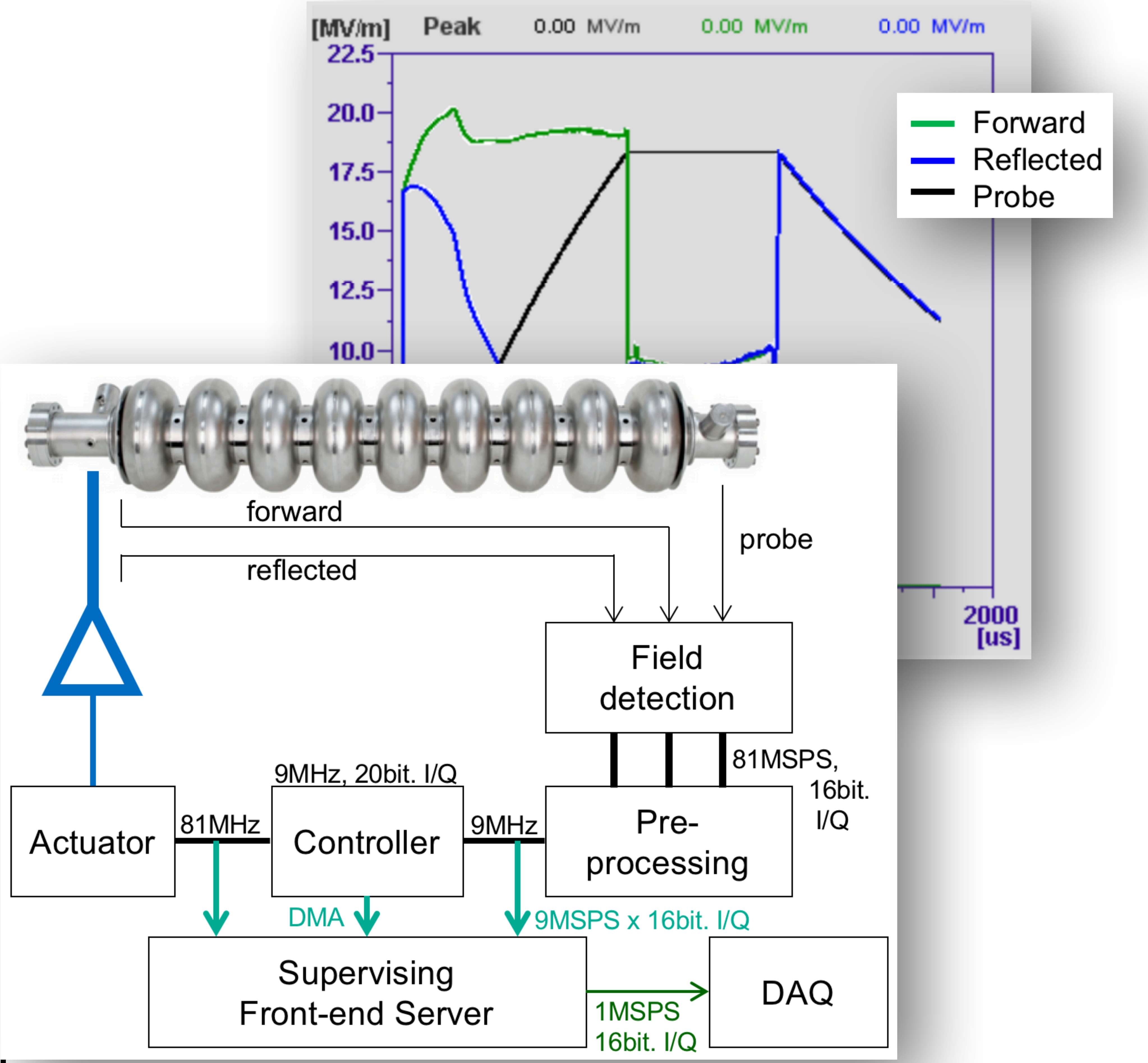
Published in: *Physical Review Accelerators and Beams:*
"Detection of faulty beam position monitors using unsupervised learning", Phys. Rev. Accel. Beams 23, 102805.

Areas for potential ML applications – Anomaly detection for the cavity signals

- Cavity fault detection requires: U_{for}, U_{ref}, U_{probe}
- Data rates to DAQ per cavity per pulse:
  - 2048 x 2 x 3 x 16bit = 24.6kB
  - Pulses per Day = 864000
  - 700 cavities → **604 Mio** events/day
  - Total data/day = **14.8 TB**
- **Good statistics** (ensemble & events)

Questions we like to address:

- How many cav./pulses behave normally
- Cav/Pulses out of nominal operation range
- Reliably quench detection and reaction
- Anomalies: due to parameter changes
- due to digital / communication/ readout

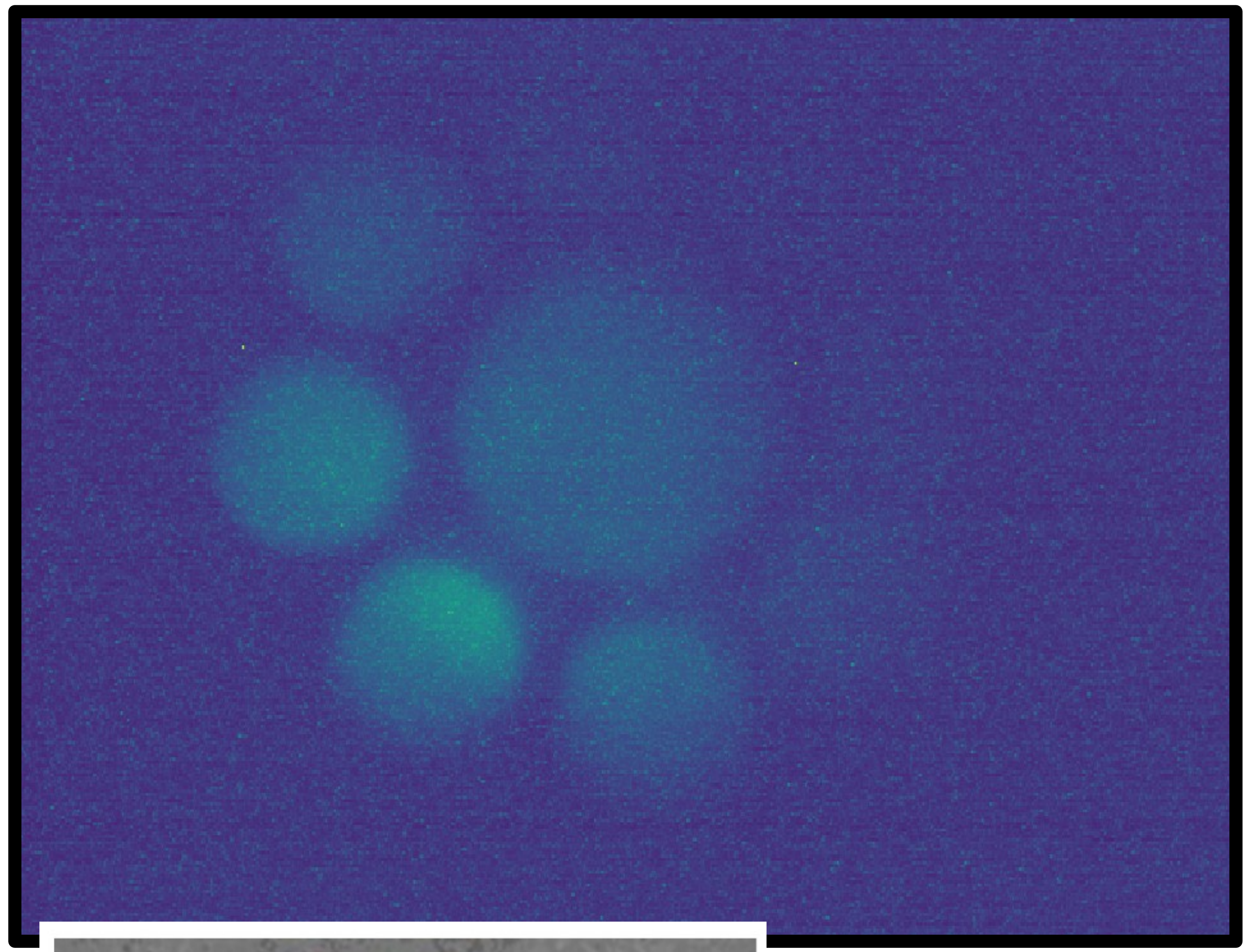


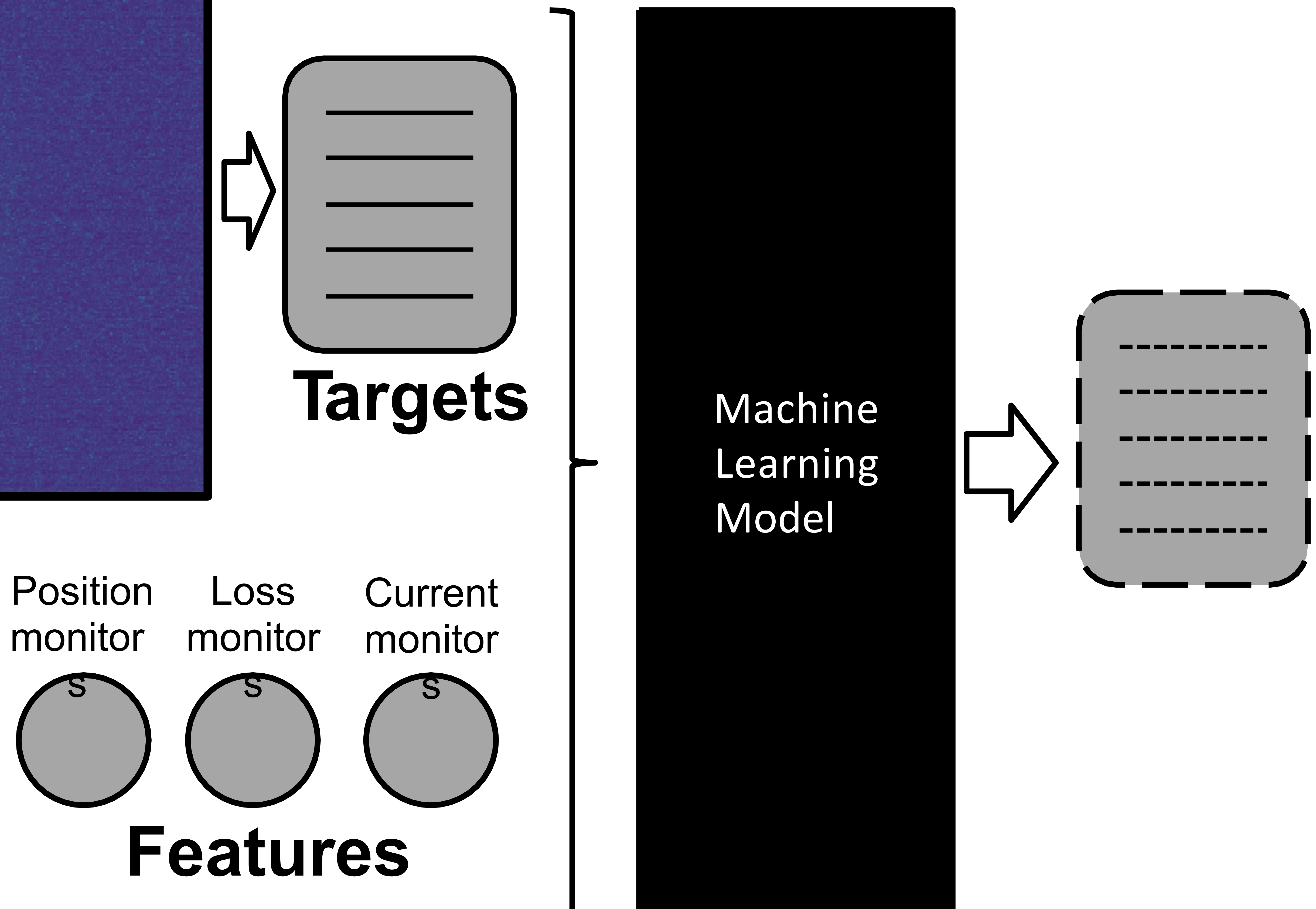
VIRTUAL DIAGNOSTICS

- Use ML to predict the response of an instrument
  - Invasive instruments (e.g. spectrometers, screens…)
  - Fragile instruments
  - Broken instruments

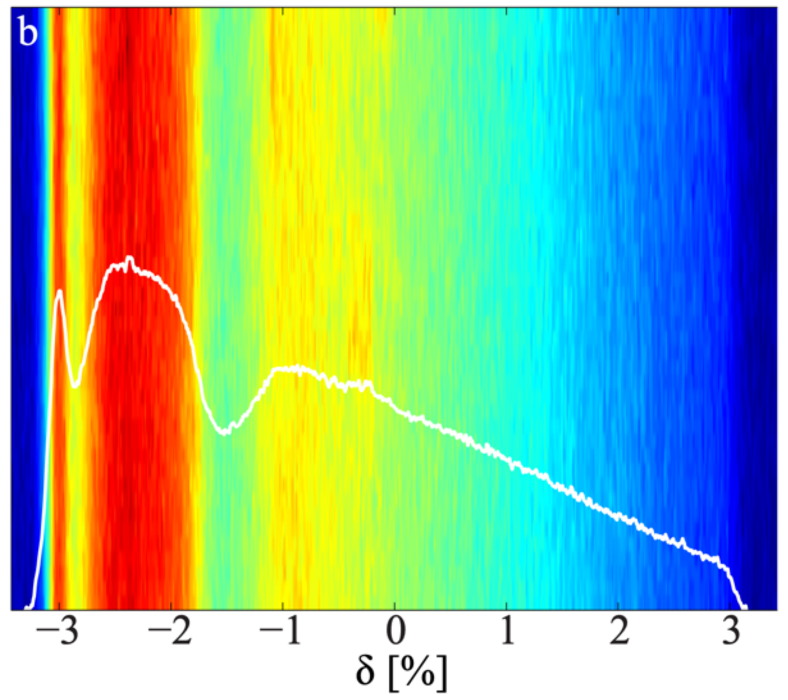
VIRTUAL DIAGNOSTICS AT SINQ

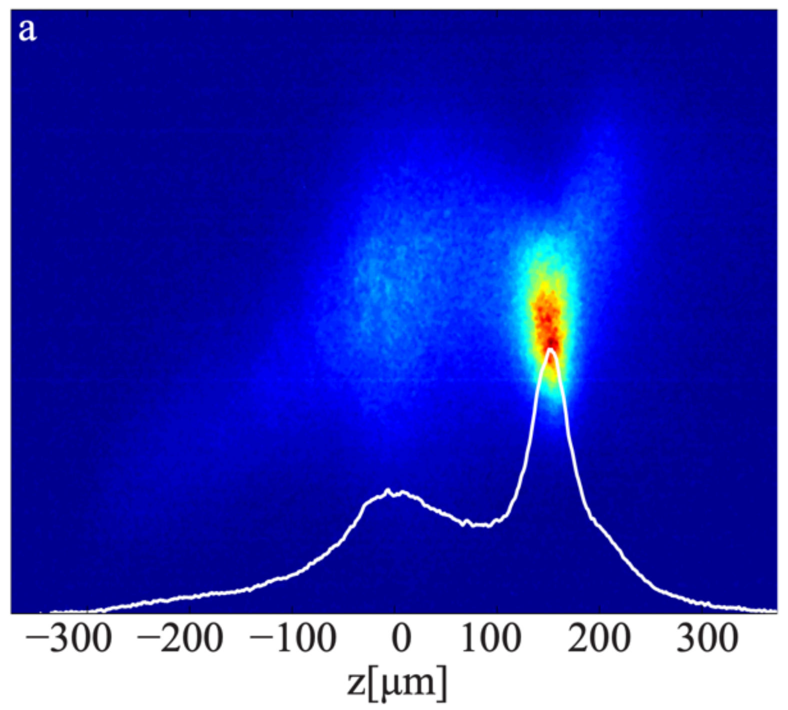
- The VIMOS system monitors the SINQ target beam spot with a metal grid.
- If the beam is focussed too much or changes too fast interlocks are triggered.
- This grid is degrading over time and cannot be replaced
- **Can we use other sensors to predict the images?**

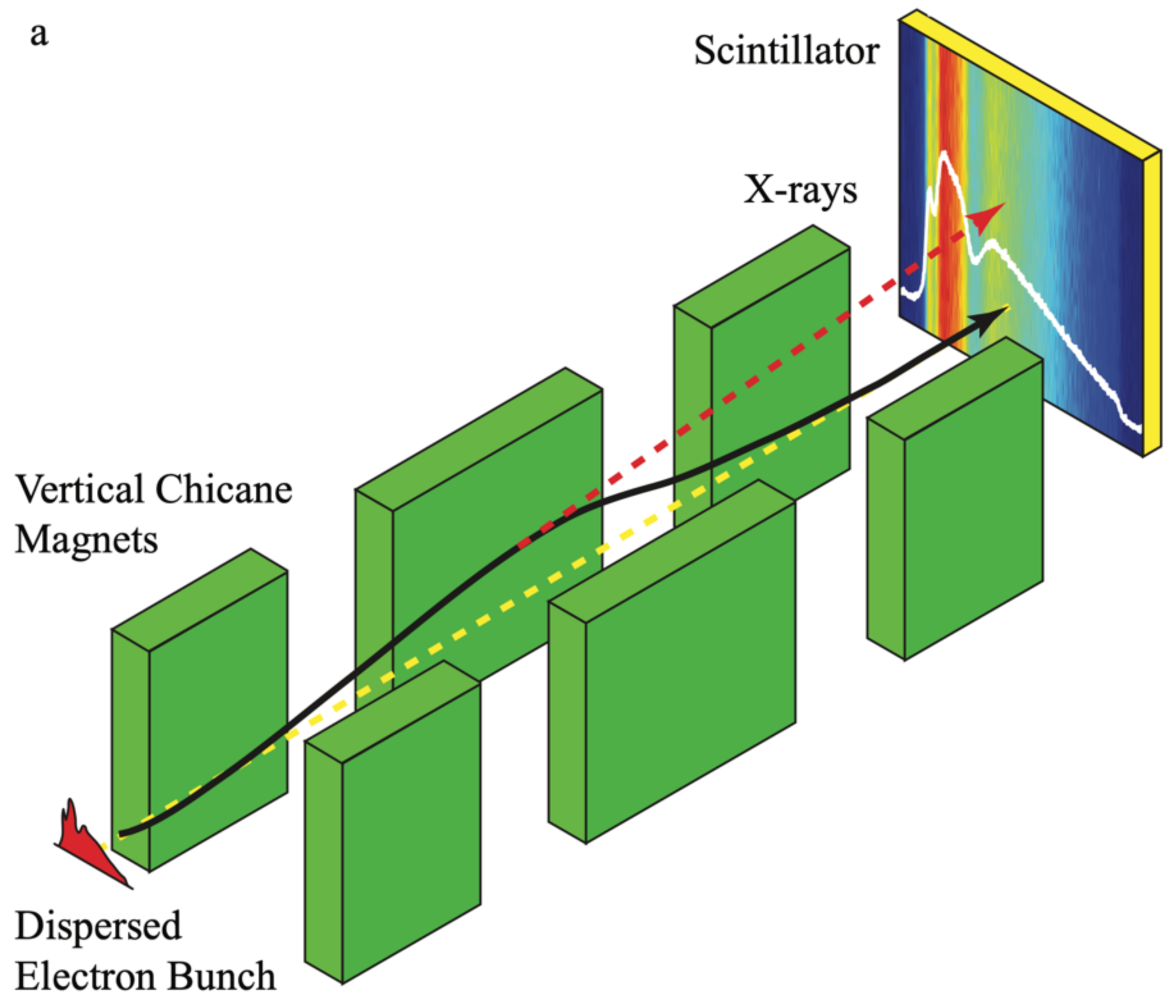


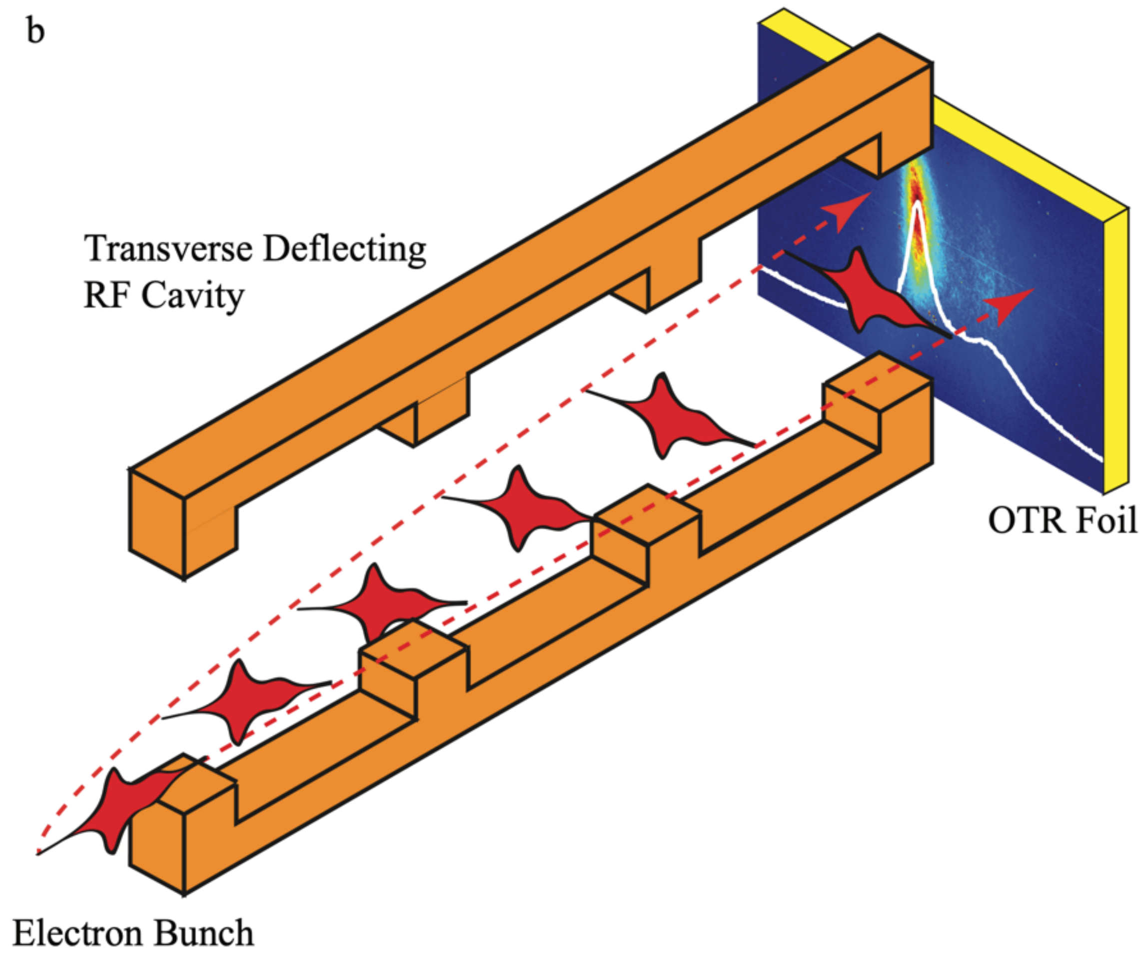


FACET: E-bunch profile prediction based on non-destructive measurements of energy spread spectrum. A. Scheinker and S. Gessner, “Adaptive method for electron bunch profile prediction,” *Physical Review Accelerators and Beams*, 18, 102801, 2015.





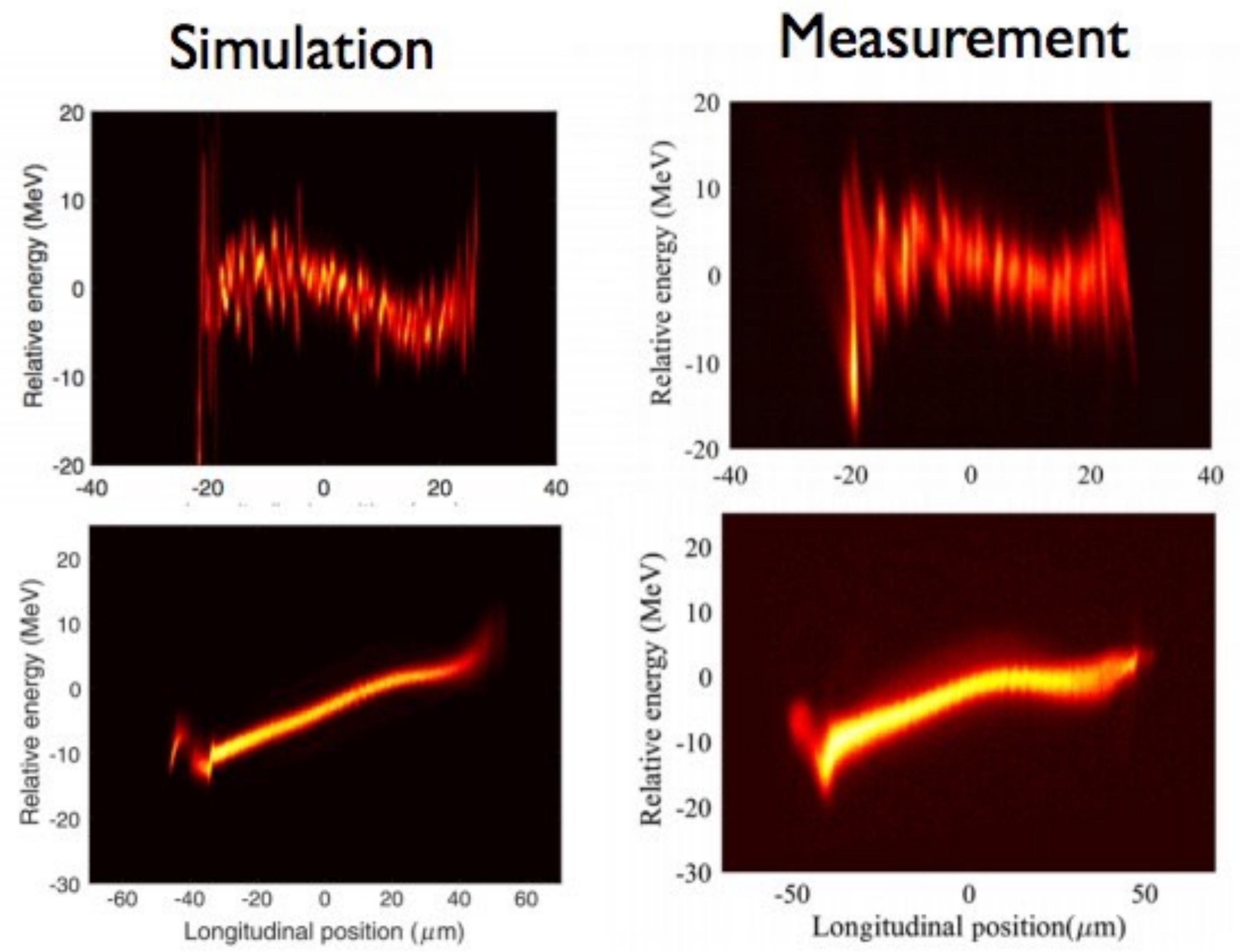




APPLICATIONS OF ML TO ACCELERATOR SIMULATION

Why add ML tools for simulations?

Accelerator simulations that include nonlinear + collective effects are powerful tools…



J. Qiang, et al., PRSTAB30, 054402, 2017

"10 hours on thousands of cores at the NERSC"

… but they are computationally expensive and don't always match the machine well

Impedes offline start-to-end optimization and control prototyping

Prohibits use as an online model (e.g. diagnostic / control applications)

Often takes much effort to replicate real machine behavior

but will still have this issue regardless

One approach: faster modeling codes

Simpler models (tradeoff with accuracy)

analytic calculations — *e. g. J. Galambos, et al., HPPA5, 2007*

Parallelization and GPU-acceleration of existing codes

HPSim/PARMILA — *X. Pang, PAC13, MOPMA13*

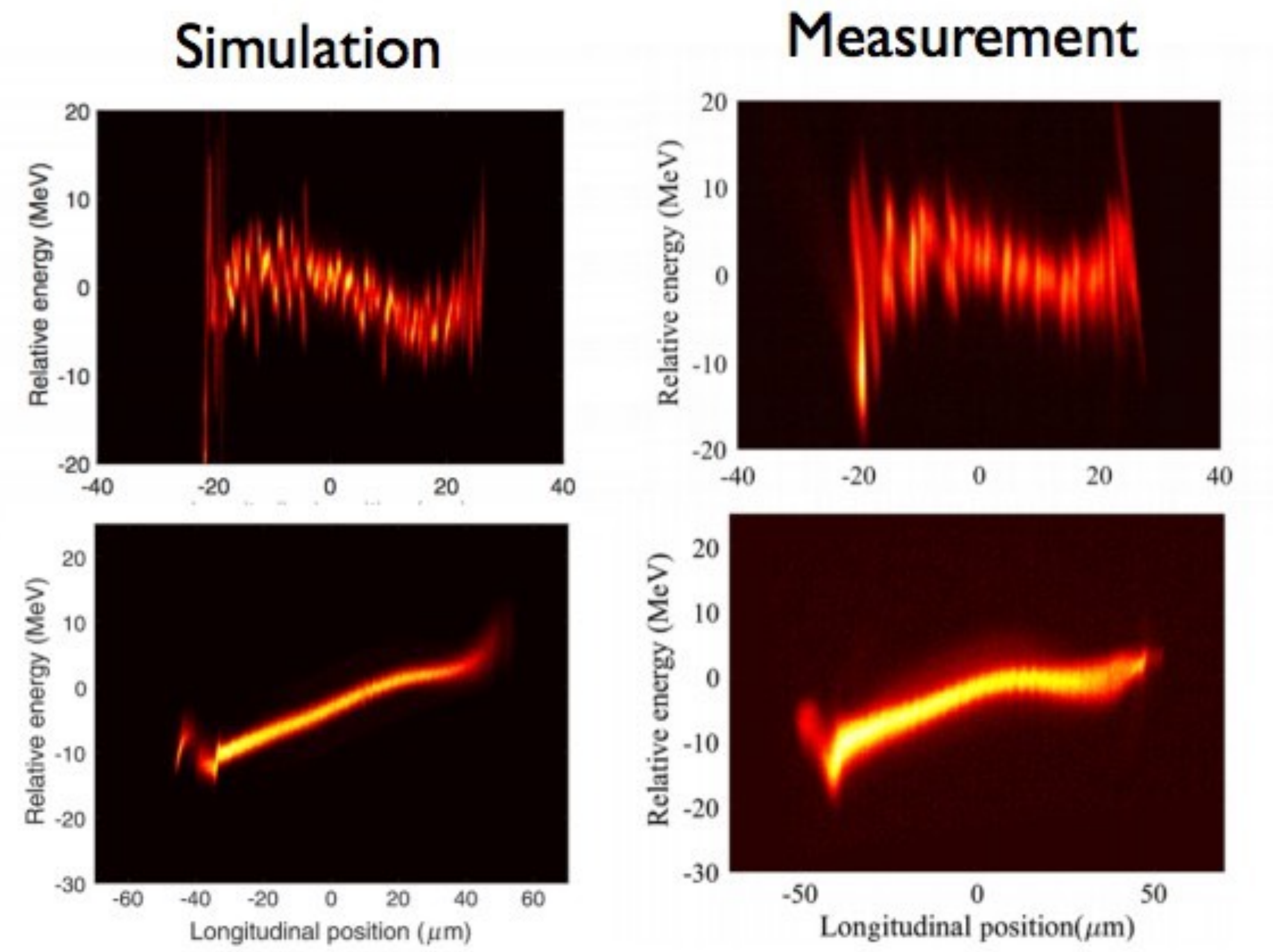
elegant — *I.V. Pogorelov, et al., IPAC15, MOPMA035*

Improvements to modeling algorithms

Lorentz-boosted frame — *J.-L. Vay, Phys. Rev. Lett. 98 (2007) 130405*

Why add ML tools for simulations?

Accelerator simulations that include nonlinear + collective effects are powerful tools…



J. Qiang, et al., PRSTAB30, 054402, 2017

"10 hours on thousands of cores at the NERSC"

… but they are computationally expensive and don't always match the machine well

Impedes offline start-to-end optimization and control prototyping

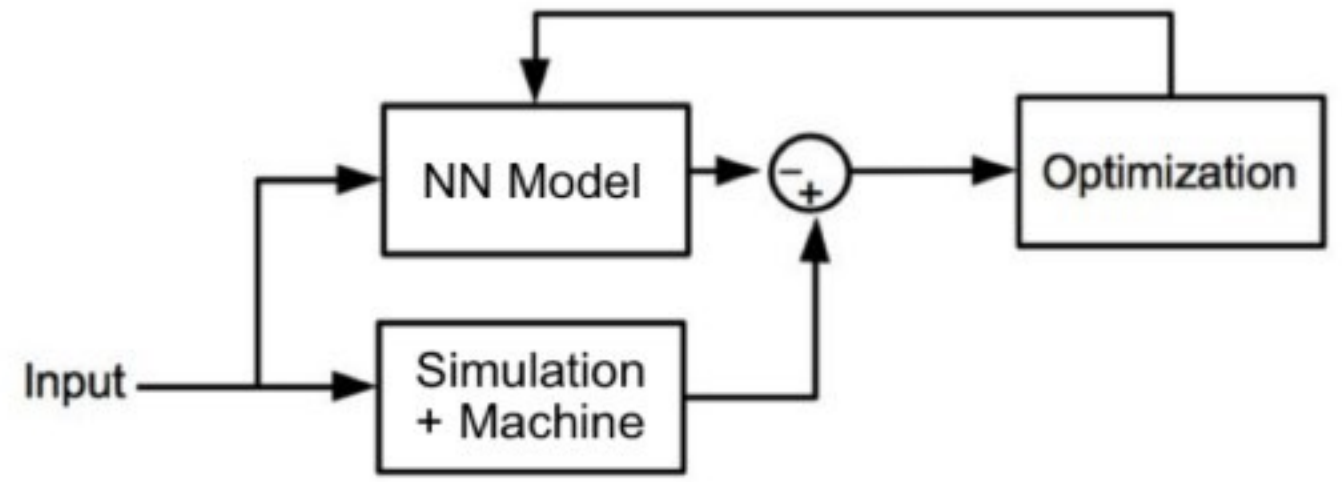
Prohibits use as an online model (e.g. diagnostic / control applications)

Often takes much effort to replicate real machine behavior

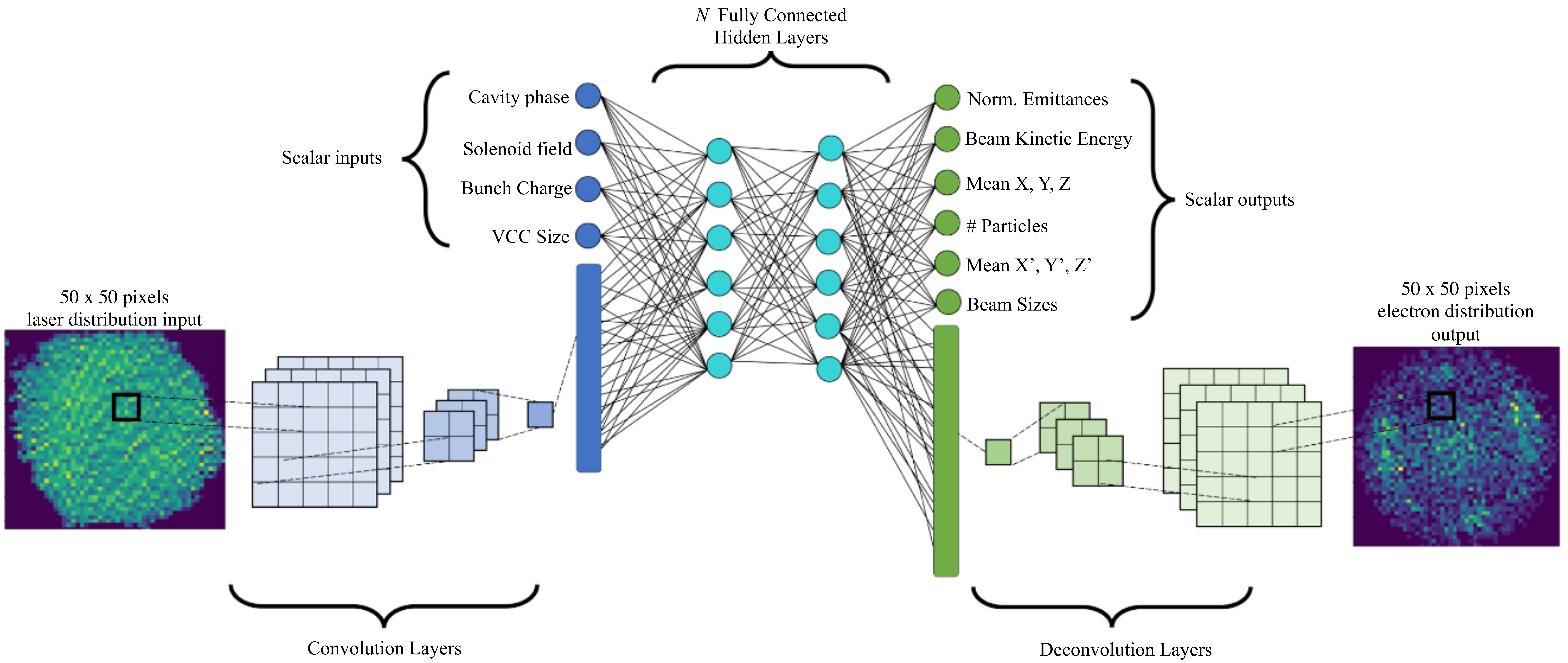
Complementary approach: ML model

Once trained, neural networks can execute quickly

Train on sparse sample from high-fidelity simulations
+
Train on measured data



OPTIMIZATION OF THE LCLS-II INJECTOR



Lipi Gupta et al 2021 Mach. Learn.: Sci. Technol. 2 045025

DA optimization for SLS 2.0

(Kranjčević, Riemann, Adelmann, Streun)

Multi-objective optimization with MOGA (`opt-pilot + tracy`):

GOOD: found tens of points with all objective function values better than the design solution (one point shown in Figure, right)

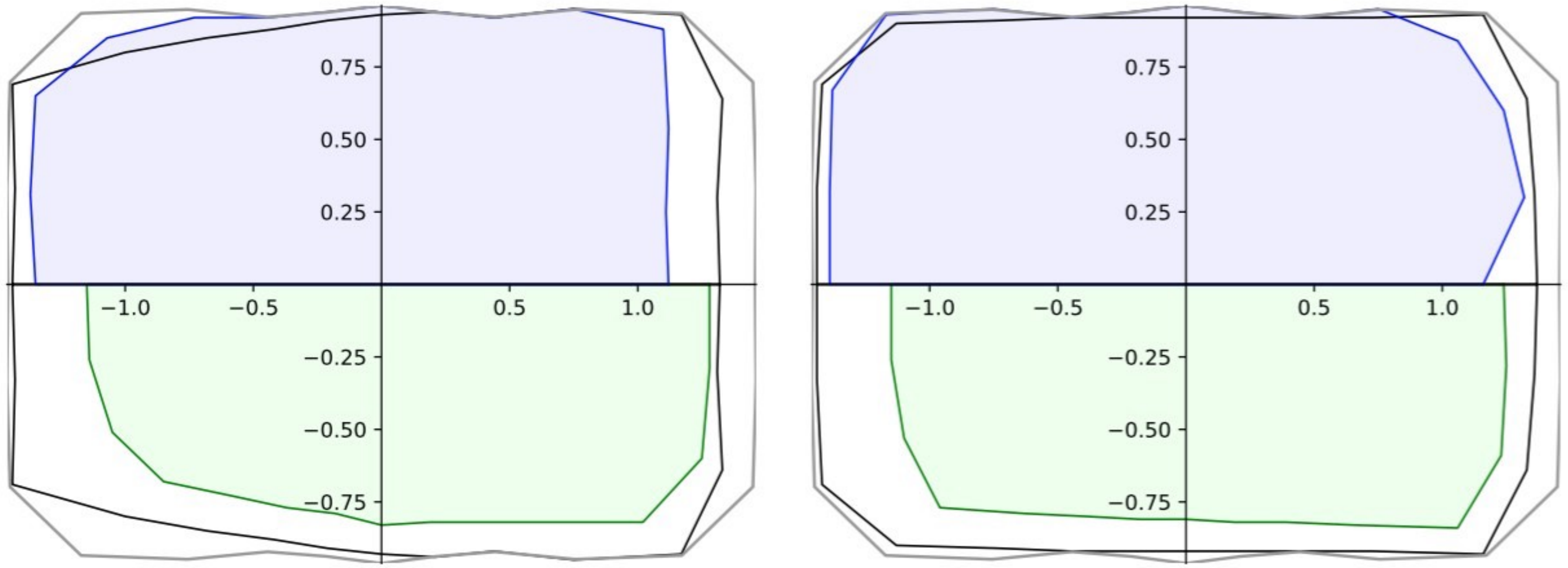


Figure: Left: design solution, right: newfound point. Transverse DAs at $\delta = -0.03$ (green), 0.03 (blue), and 0 (bold black line). For both points chromatic tune footprint and ADTS footprint constrained.

BAD: for detailed lattice models the optimization needs to be faster

DA optimization for SLS 2.0

(Kranjčević, Riemann, Adelmann, Streun)

Run time and solution quality comparison for different methods:

	opt-pilot + tracy	SM (30k)	SM + re-train (20k)	SM + re-train (5k)
nof pts better	31	0	148	87
run time	48 h	11 h 21 min	8 h 52 min	3 h 10 min
speedup	1.0	4.2	5.4	15.1

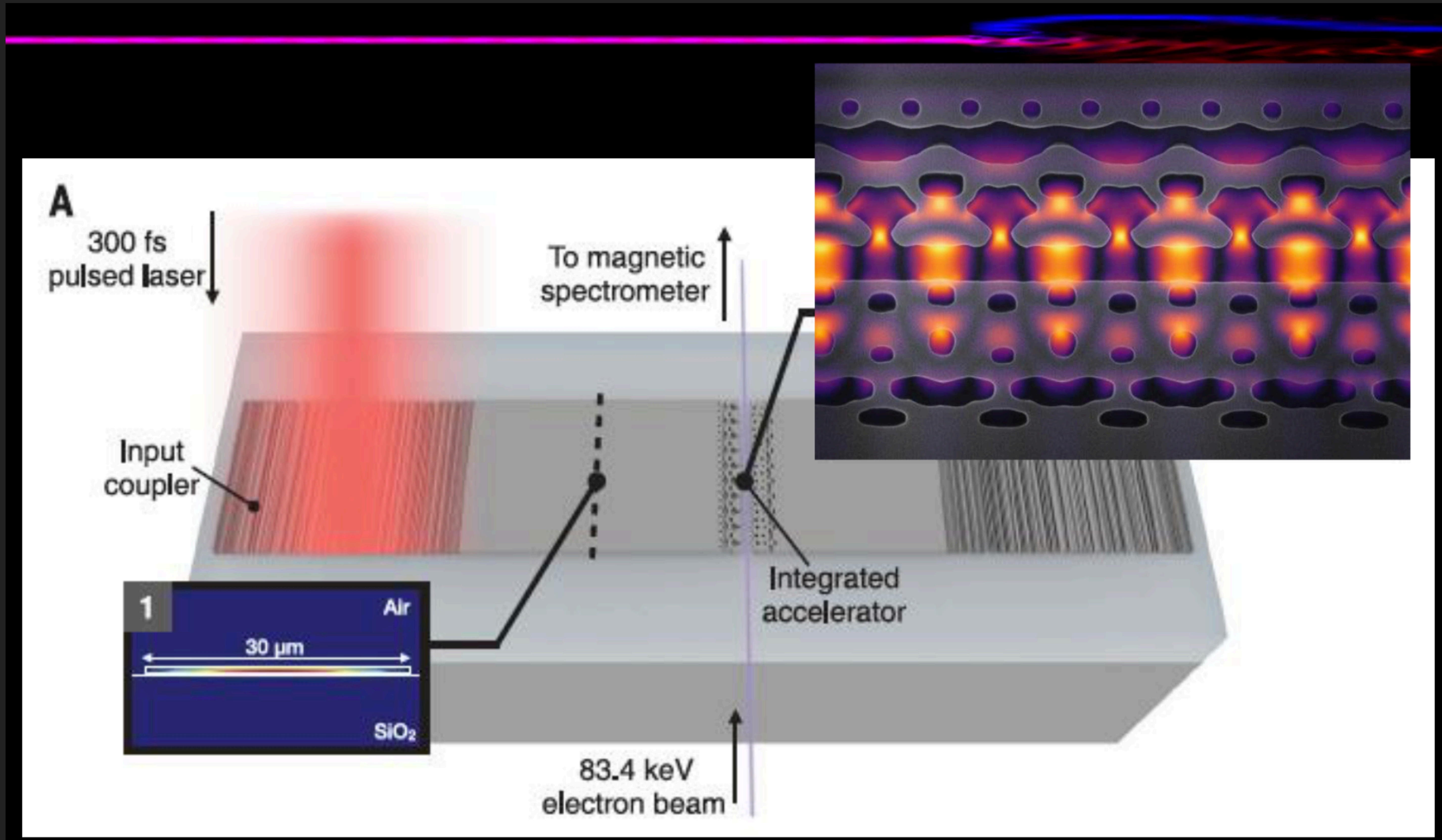
Columns: methods (the number in the parentheses is the combined size of the samples used for training)

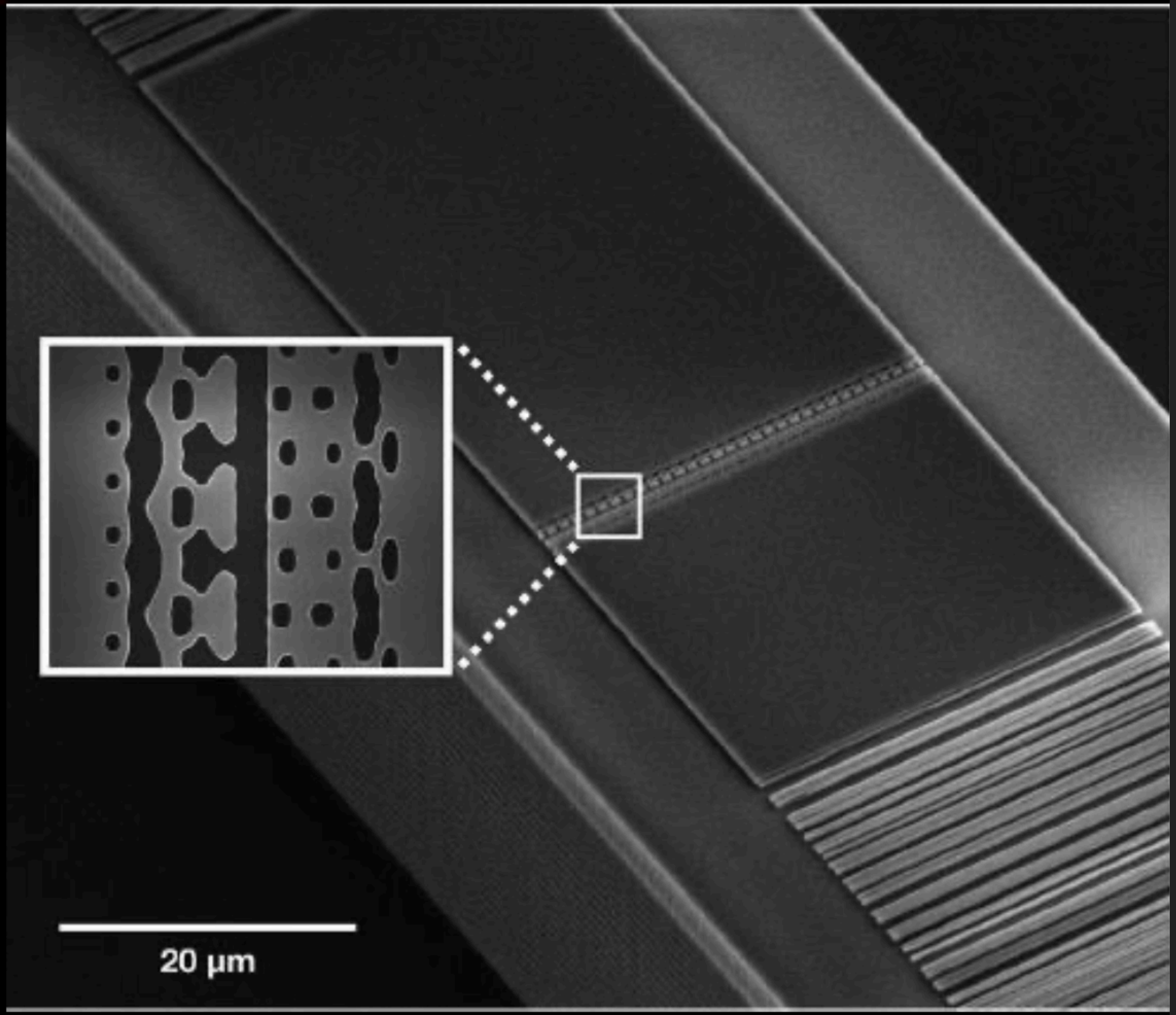
Rows:

- 'nof pts better' is the number of design points (magnet configurations) in the last gen. that satisfy the constraints and have all objectives better than the design solution
- run times include: evaluating points used for training, training, optimization and re-eval of 10 % of the points in the last gen.

https://arxiv.org/abs/2008.04151

INVERSE DESIGN OF ACCELERATING STRUCTURES



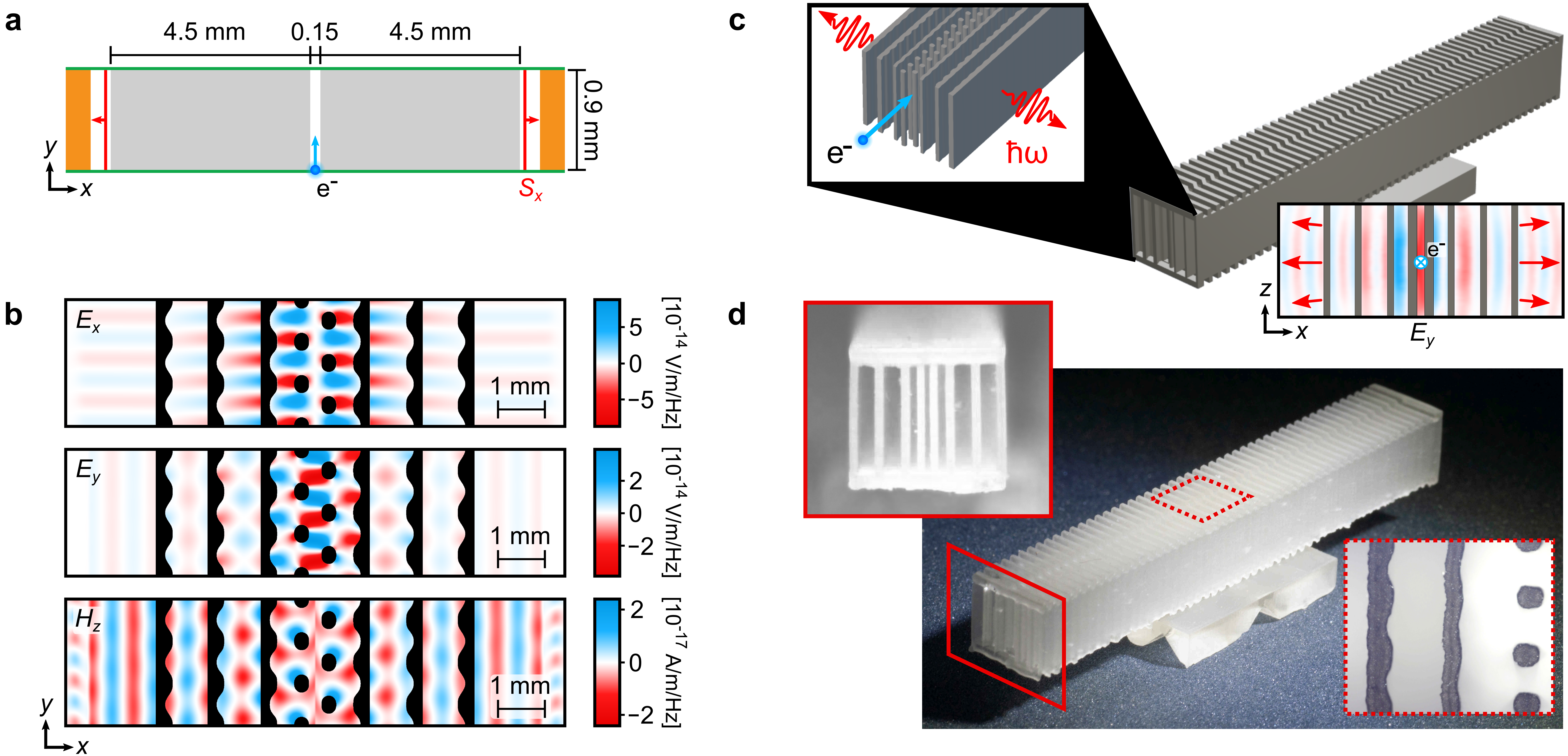


- > 30MeV/m acceleration gradient
- 10000-fold reduction in size of accelerator
- 30 µm long accelerator stage

ACHIP project (B. Byer, P. Hommelhoff)

Jelena Vučković, Neil Sapra: Science 367, 79–83 (2020)

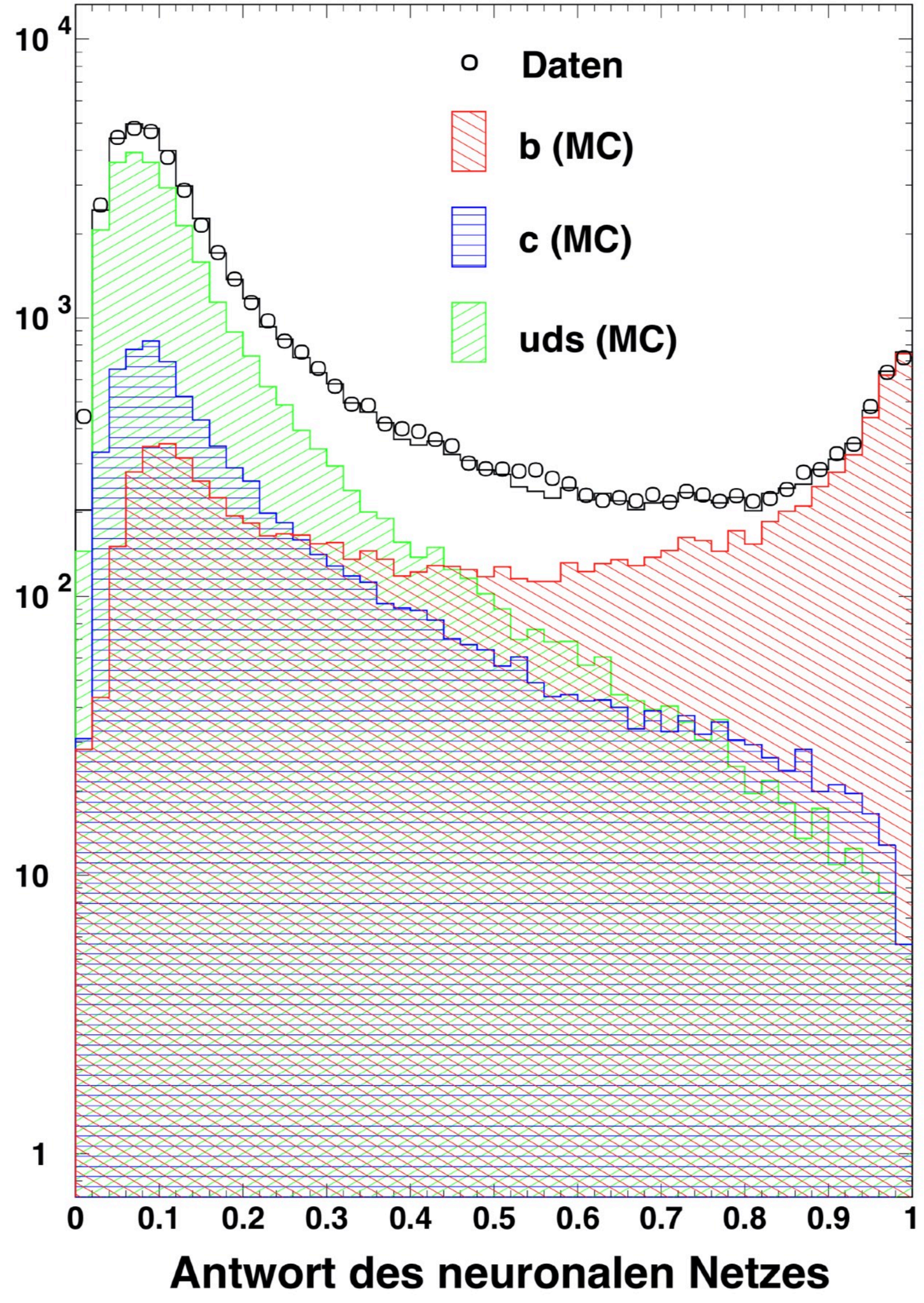
INVERSE DESIGN OF THz STRUCTURES



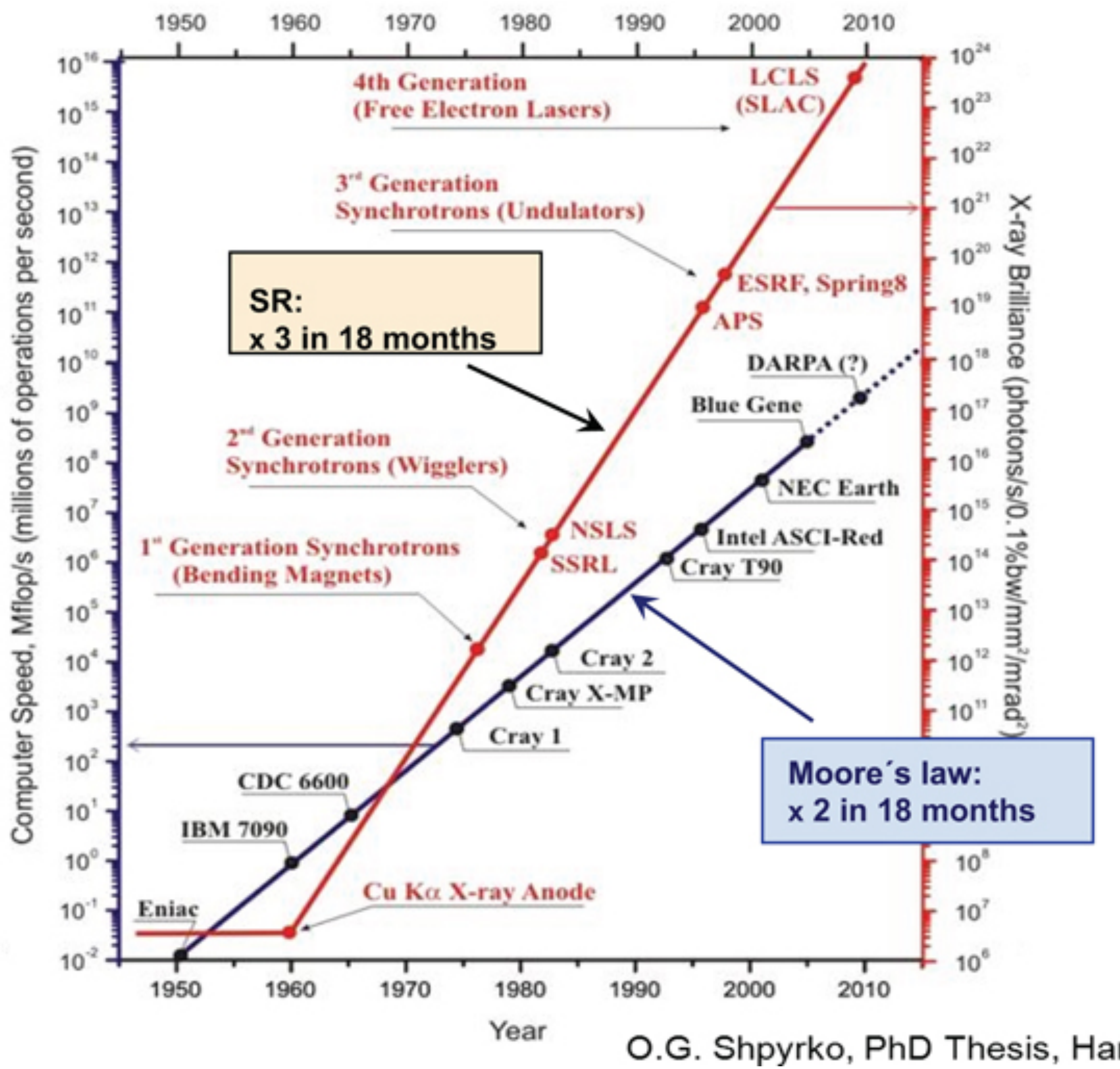
DATA ACQUISITION AND ANALYSIS

MACHINE LEARNING FOR HEP DATA ANALYSIS

- Improving interpretation of detector data (2002)



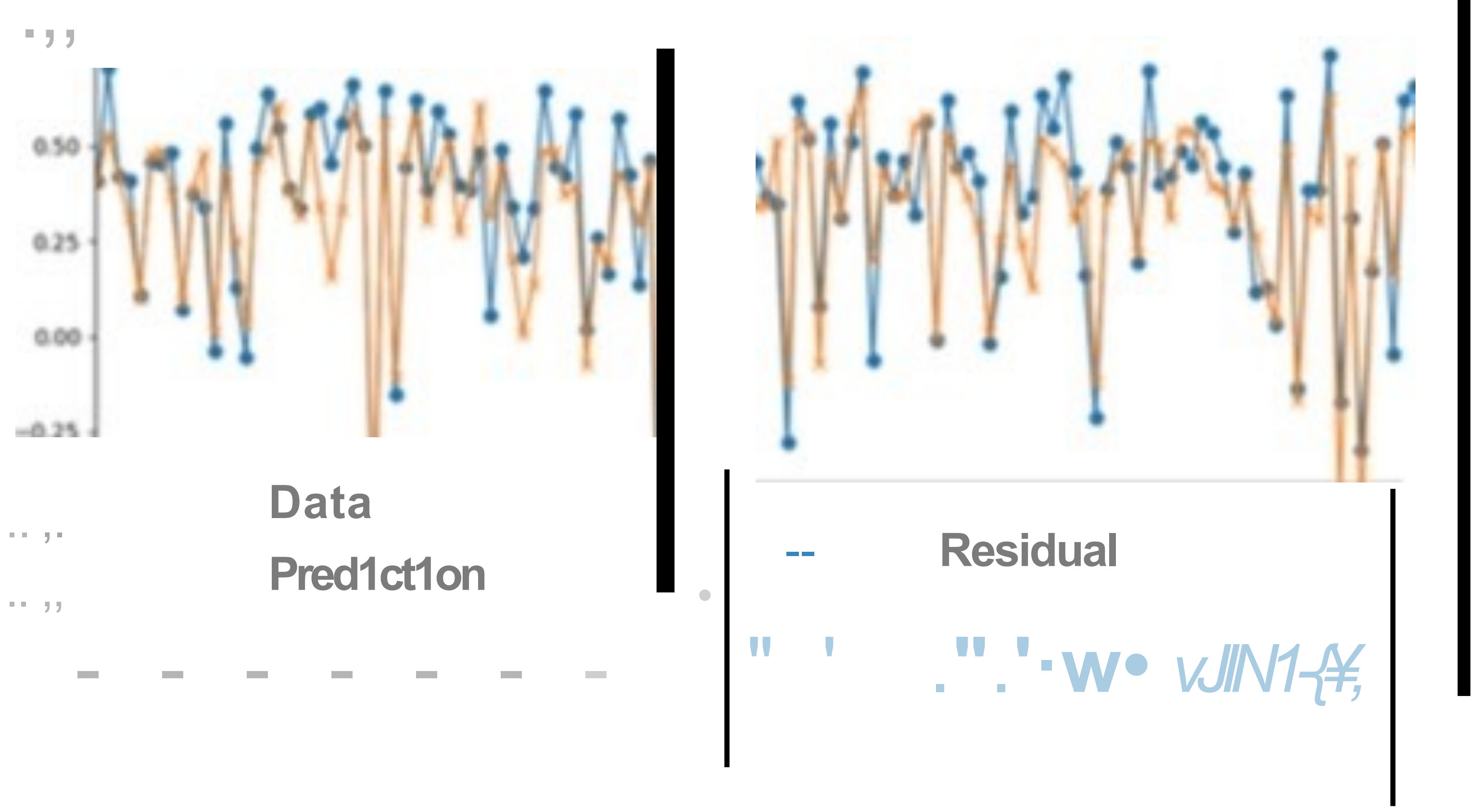
WE HAVE A PROBLEM…



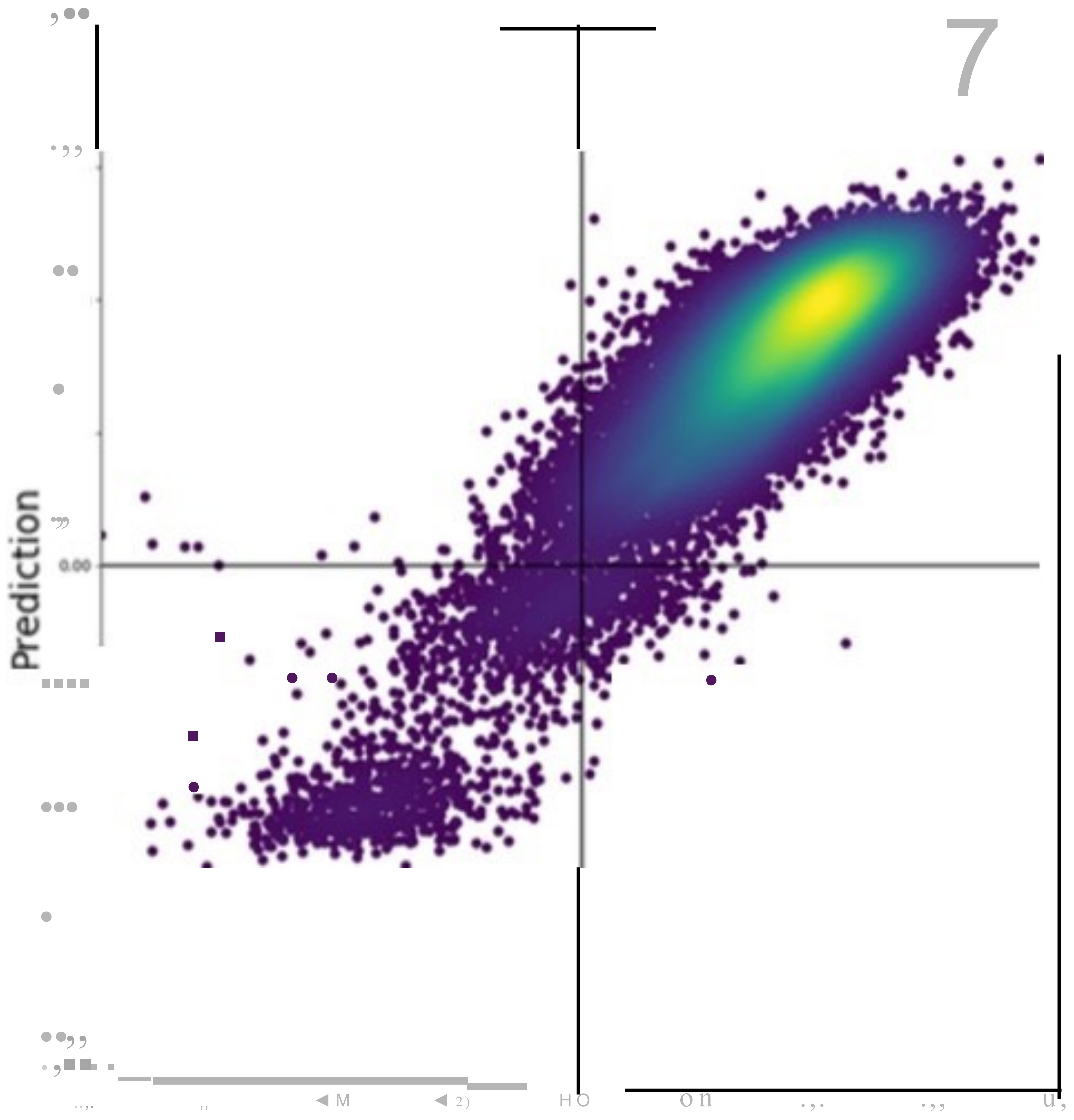
O.G. Shpyrko, PhD Thesis, Harvard 2005

SwissFEL Pump-Probe Delay Prediction using ML
Sven Augustin - PSI Fellow (LSF)

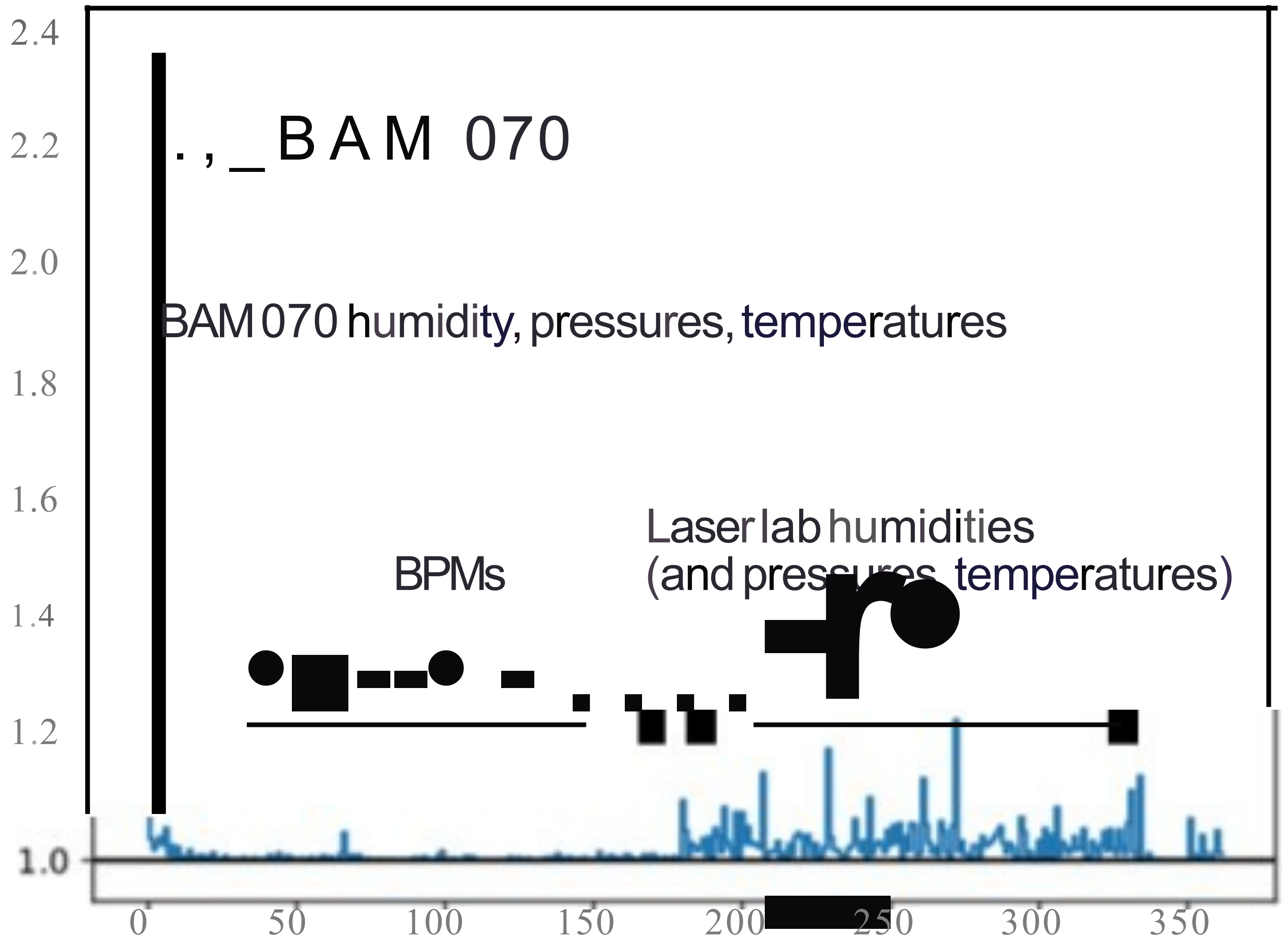
Trends are followed with tiny residual:



Input and output nicely correlated



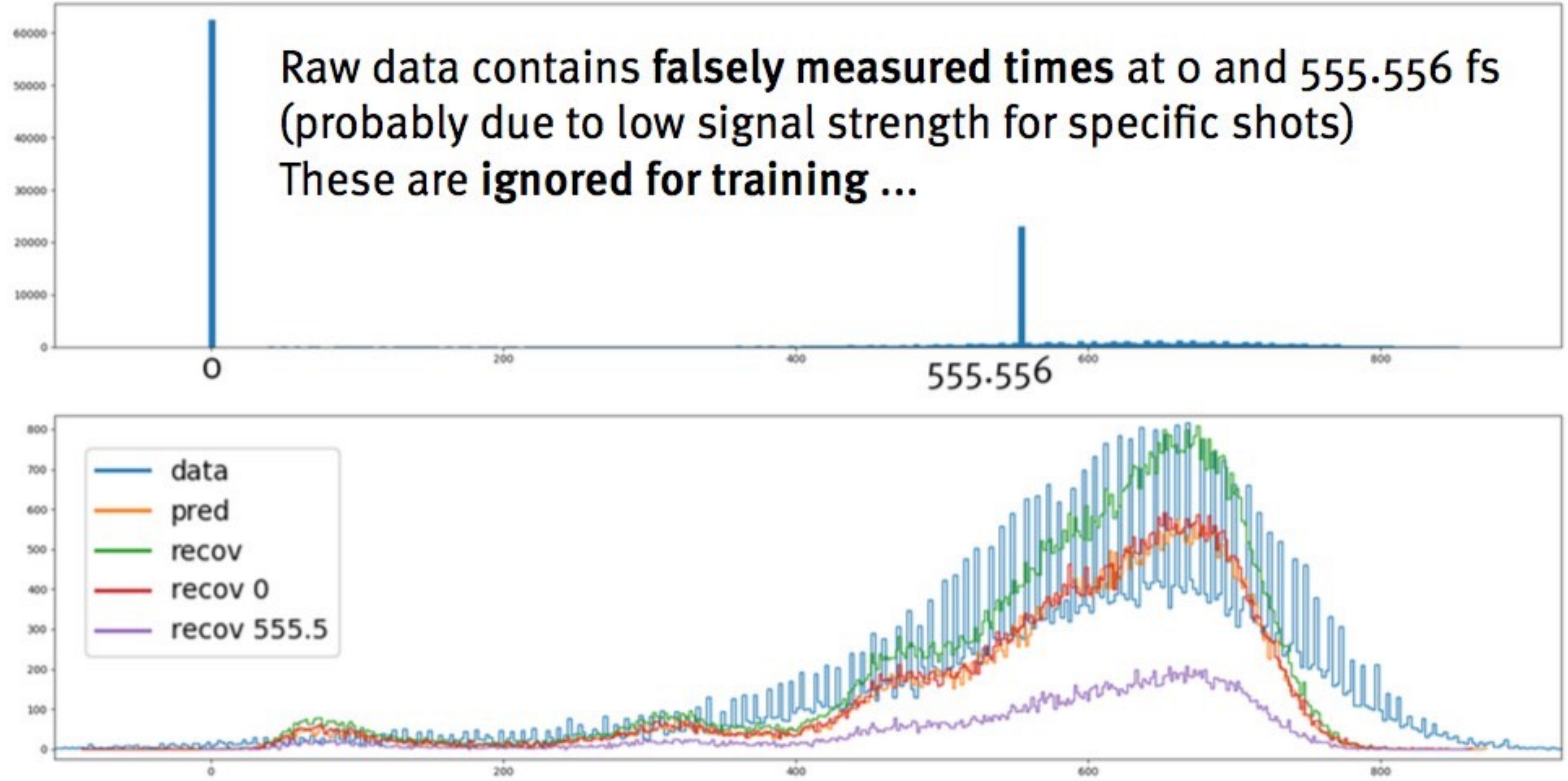
Feature Im ortance



In collaboration with A. Adelmann (LSM) and C. Bostedt (LSF)

SwissFEL Pump-Probe Delay Prediction using ML

Sven Augustin – PSI Fellow (LSF)



... and **predicting for these points** yields distributions similar to the input!

→ Model **recovers** false measurements for low intensity!

(which is about half of the data in this data set)

SUMMARY

- Use machine learning…
  - …when you have lots of data, and the algorithm to analyze it is not obvious
  - …when you have pre-classified cases (be it from simulations, from other detectors, or from manual classification)
  - …when speed matters
  - …when other methods are more effort, or more expensive
- ML is easy and cheap

THANK YOU

- Andreas Adelmann
- Andreas Krause
- Andreas Streun
- Auralee Edelen
- Benedikt Hermann
- Elena Fol
- Giulio Gaio
- Jelena Vučković
- Jochem Snuverink
- Johannes Kirschner
- Neil Sapra
- Nicole Hiller
- Peter Wienemann
- Raimund Kammering
- Sven Augustin