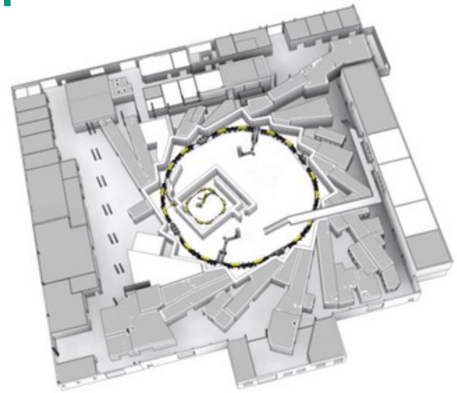
Karlsruhe Institute of Technology

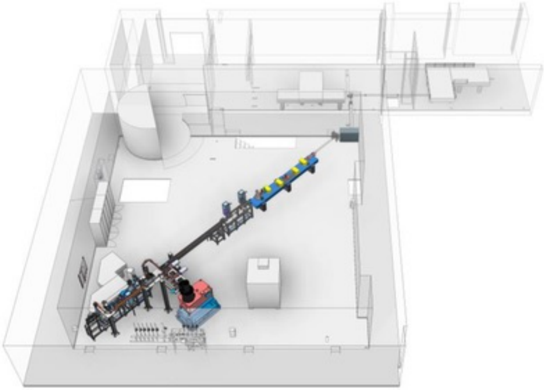
Machine learning to control and predict beam properties

Andrea Santamaría García[1], Erik Bründermann[2], Michele Caselle[3], Luca Scomparin[3],
Johannes Steinmann[2], Chenran Xu[2], Anke-Susanne Müller[1,2]

[1]Laboratory for Applications of Synchrotron Radiation (LAS)
[2]Institute for Beam Physics and Technology (IBPT)
[3]Institute for Data Processing and Electronics (IPE)

I.FAST Workshop 2022





Ultra-low emittance rings
Desiderata

Insights into beam dynamics

Non-destructive

Single-shot
Bunch-by-bunch

High temporal resolution

» Constant delivery of high quality, intense and stable photon beams to a variety of beamlines

Characterization of the beam in real time → Challenges of fast/online big data processing and analysis

What can machine learning do for us?

Karlsruhe Institute of Technology

[Very fast inference]

=

Making a prediction by evaluating an already trained model

Data driven

Prediction task
static

Predict the beam properties based on current machine parameters

- Surrogate models
- Virtual diagnostics

Optimization task
static

Achieve desired beam properties or states by tuning machine parameters

- Bayesian algorithms
- Optimizers

Control task
dynamic

Control the state of the beam in real time in a dynamically changing environment

- Reinforcement learning

Machine Learning

Learning style	Task	Popular algorithms
Supervised Learning — We know the input & output (labeled data)	Classification (discrete variables); Regression (continuous variables)	▪ Neural networks (e.g. stochastic gradient descent, backpropagation) ▪ Support Vector Machine ▪ K-nearest neighbor ▪ Decision Tree algorithms (e.g. Classification and Regression Tree) ▪ Random Forest (ensemble) ▪ Uni or multivariate, linear or logistic
Unsupervised Learning — We only know the input (unlabeled data)	Clustering	▪ K-means ▪ K-medians ▪ Expectation Maximization (EM) ▪ Hierarchical clustering
	Association	▪ Apriori algorithm ▪ Eclat algorithm
Reinforcement Learning	Control; Classification	▪ Model-free ▪ Value based (Q-learning) ▪ Policy based ▪ Deterministic or stochastic ▪ On or off policy ▪ Policy gradient or actor-critic ▪ Model based ▪ Learn the model ▪ Model given

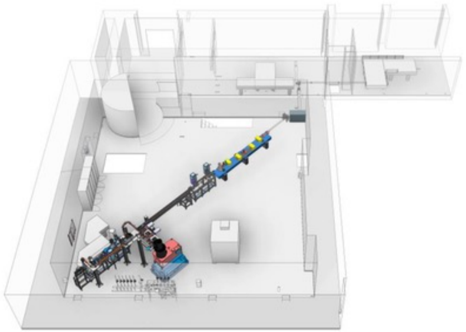
Deep Learning Networks
- Convolutional Neural Networks
- Recurrent Neural Networks
- Long Short-Term Memory Networks
- Autoencoders
- Deep Boltzmann Machine
- Deep Belief Networks

Bayesian Algorithms
- Naive Bayes
- Gaussian Naive Bayes
- Bayesian Network
- Bayesian Belief Network
- Bayesian optimization

Regularization, dimensionality reduction, ensemble, evolutionary algorithms, computer vision, recommender systems,
…

Accelerator facilities at KIT

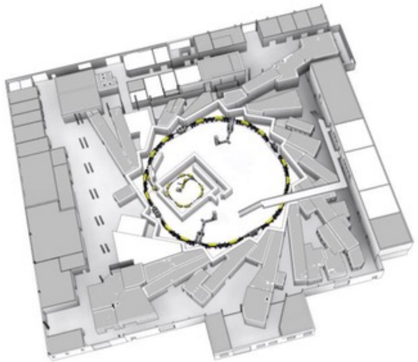
Karlsruhe Institute of Technology



FLUTE

Linac-based THz source

41 MeV top energy



KARA

Synchrotron light source and storage ring

2.5 GeV top energy

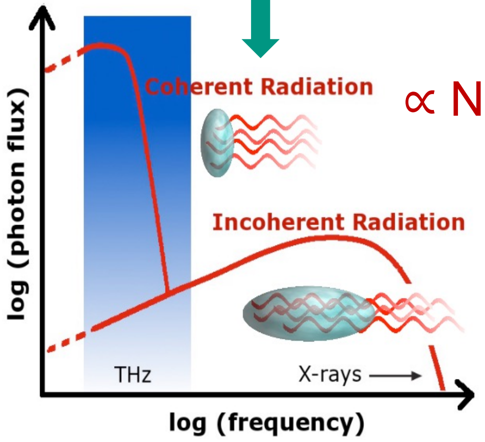
Tailoring THz radiation with machine learning

For more information on facilities: [“Introduction to KIT, KARA, and FLUTE”](), B. Härer (this workshop)

Terahertz coherent synchrotron radiation (CSR)

Short bunch

Increased radiation power



$\propto N_e^2$

$\propto N_e$

Image courtesy of A.-S. Müller

Synchrotron radiation spectral intensity

$$\frac{dI}{d\omega} = [N_e + N_e(N_e - 1)F(\omega)]\frac{dI_0}{d\omega}$$

Incoherent radiation

Coherent radiation

Single particle spectrum

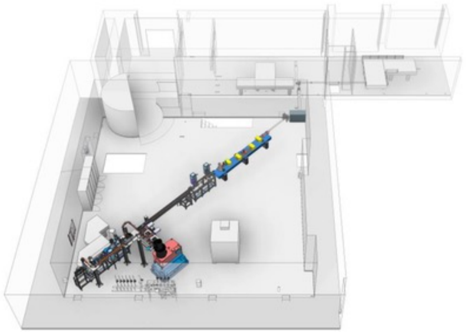
Form factor

$$F(\omega,\vec{n}) = \left| \int \rho(\vec{r}) e^{i\omega\, \vec{n}.\vec{r}/c}\, \mathrm{d}^3\vec{r} \right|^2$$

Highly dependent on the **shape** of the generating **charge distributions**

"Accelerator-Based THz Radiation Sources", A.-S. Müller & M. Schwarz, doi:10.1007/978-3-030-23201-6_6

Structure of the talk



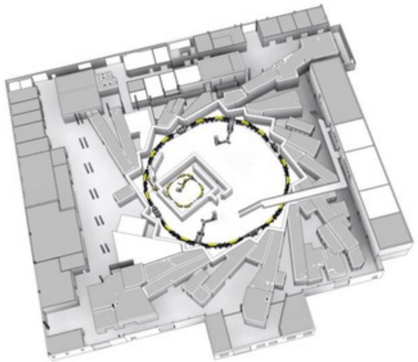
FLUTE

Linac-based THz source

41 MeV top energy

→ CSR enhancement with Bayesian optimization and surrogate models

Big thanks to the FLUTE team for their help!



KARA

Synchrotron light source and storage ring

2.5 GeV top energy

→ Control of the micro-bunching instability with reinforcement learning for tailored CSR

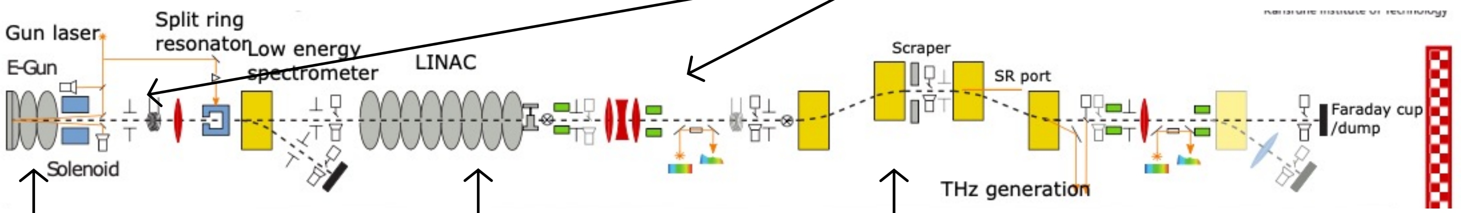
Bunch control at FLUTE - knobs

Will be varied in other studies

Photoinjector
- Laser pulse length = 700 fs (1 pC) or 2 ps (100 pC)
- Laser pulse shape = Circular
- Laser spot size = 250 µm radius
- Laser spot position = Centered in cathode

WEPAB289, IPAC21

Magnets
- Quadrupoles
- Solenoid



RF gun
- RF amplitude
- RF phase

Traveling wave structure
- Amplitude
- Phase

Chicane
- Bending radius

Bunch control at FLUTE - concept

Karlsruhe Institute of Technology

Surrogate model

input

1. RF gun phase
2. RF gun amplitude
3. Solenoid current
4. Bunch charge

output

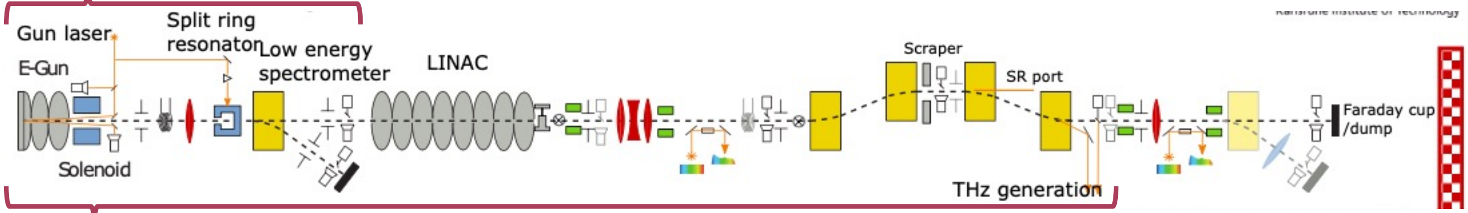
1. Mean energy
2. Energy spread
3. RMS bunch length
4. Beam size
5. Emittance
6. % remaining part.

10 steps/input
→10^4 sims
3 min/sim
40 sims in parallel
= 12.5 h ✅

Why not a surrogate model of the whole machine?

10 steps/input
→10^6 sims
10 min/sim
40 sims in parallel
= 6 months! ❌

Can inform/guide the optimization with smart initial guesses



Parallel Bayesian optimization

input

1. RF gun phase
2. RF gun amplitude
3. Solenoid current
4. Linac phase
5. Linac amplitude
6. Chicane bending radius

objective

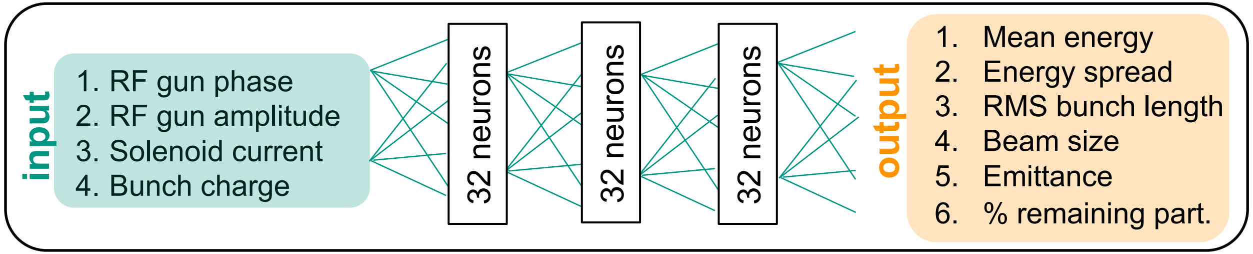
- Min. RMS bunch length after chicane
- Max. peak E-field of CSR pulse

observation

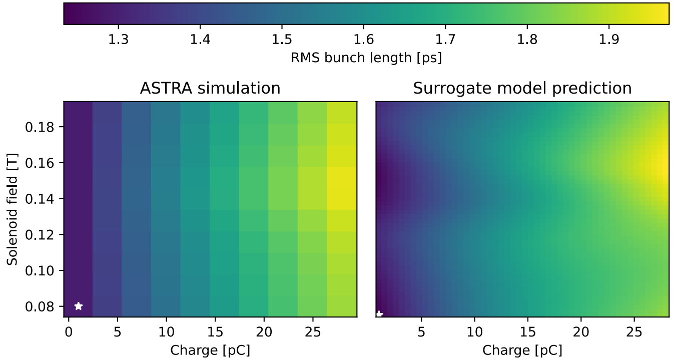
- Long. phase space
- Spectral intensity
- Form factor
- Bunch current profile
- THz pulse E-field

10k macroparticles
20 parallel evaluations
Max. 50 steps
~ 4-6 h

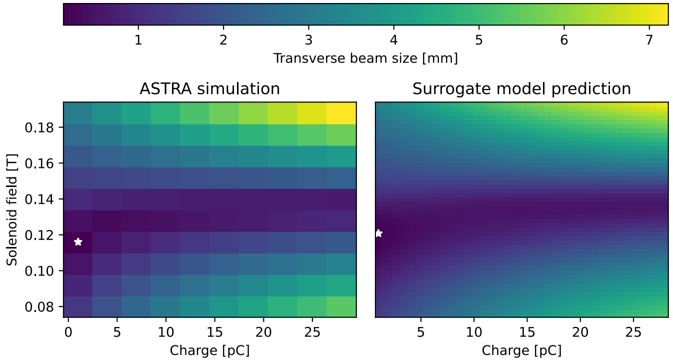
Surrogate model – design & visualizations



- Fully connected
- Activation function = tanh
- Optimizer = Adam
- Early stopping
- 150 epochs
- 10k training dataset & 1k validation dataset
- Uniform random sampling of 4D space

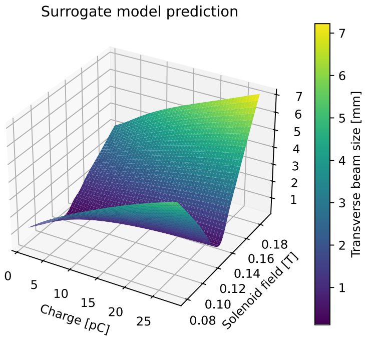


1.29 ps 1.23 ps



0.12 mm 0.13 mm

Higher solenoid field needed to compensate for radial space charge force

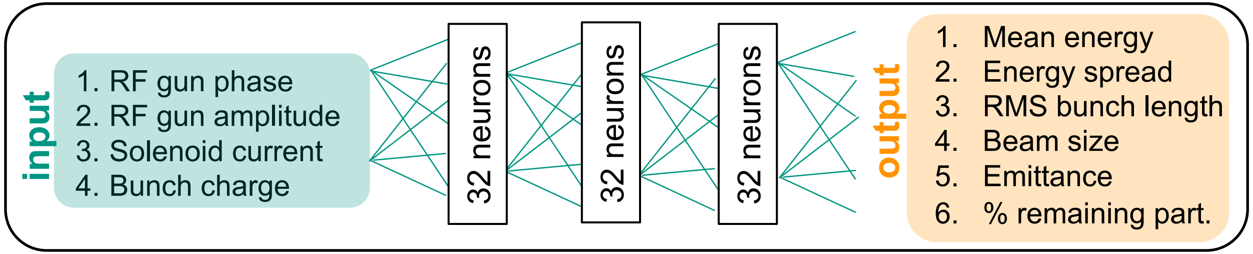


It is actually a 4D space, so 10^2 more values per 2D projection

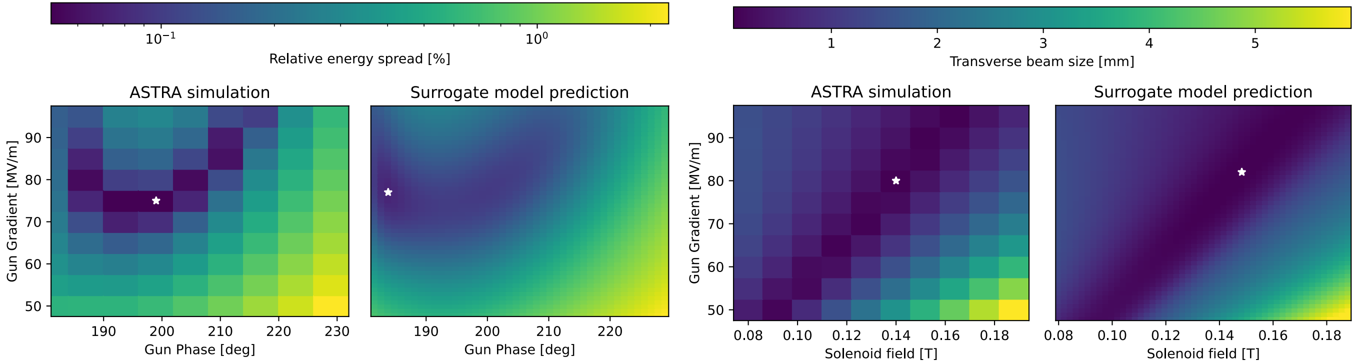
Developed by Chenran Xu

Surrogate model – design & visualizations

Karlsruhe Institute of Technology



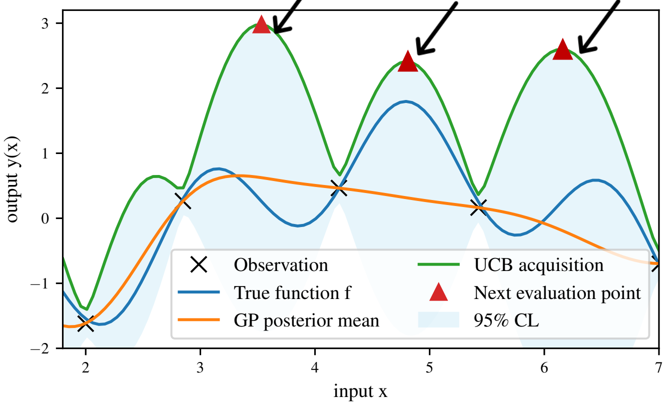
- Fully connected
- Activation function = tanh
- Optimizer = Adam
- Early stopping
- 150 epochs
- 10k training dataset & 1k validation dataset
- Uniform random sampling of 4D space



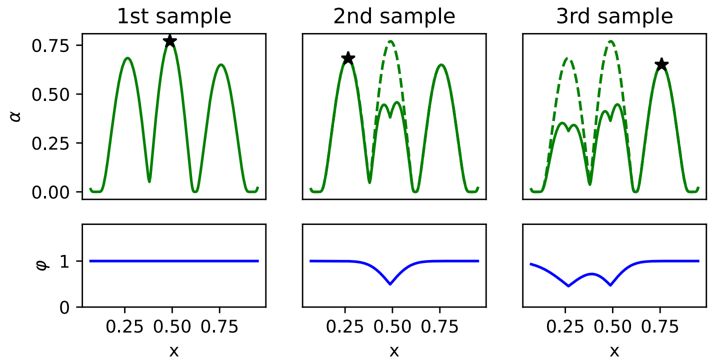
0.05 % 0.07% 0.076 mm 0.058 mm

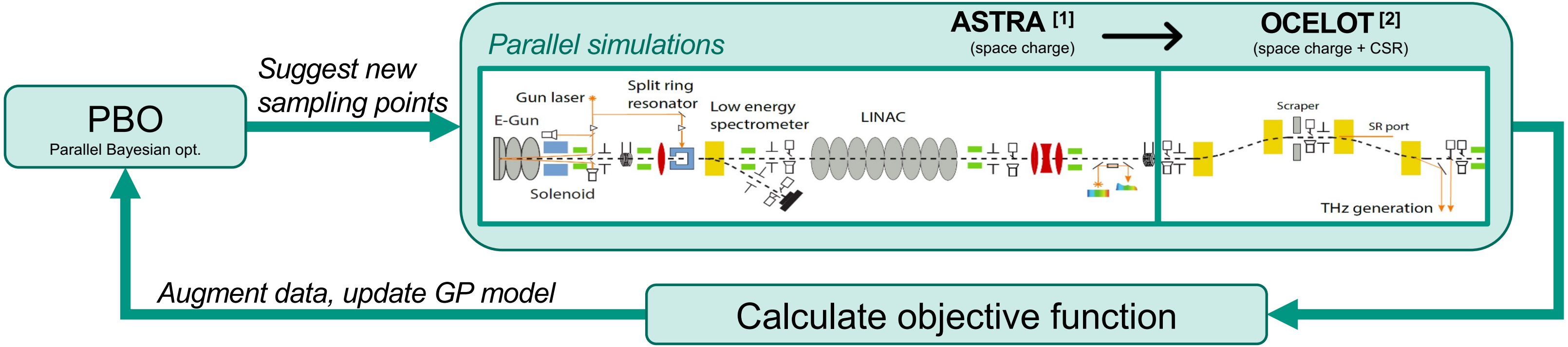
Developed by Chenran Xu

Parallel Bayesian optimization



A batch of points is selected to be processed in parallel

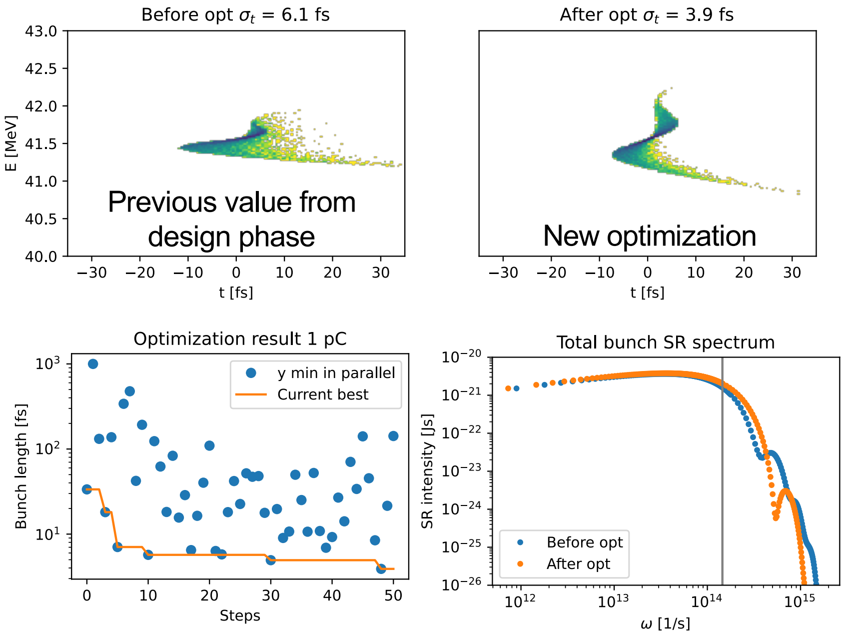




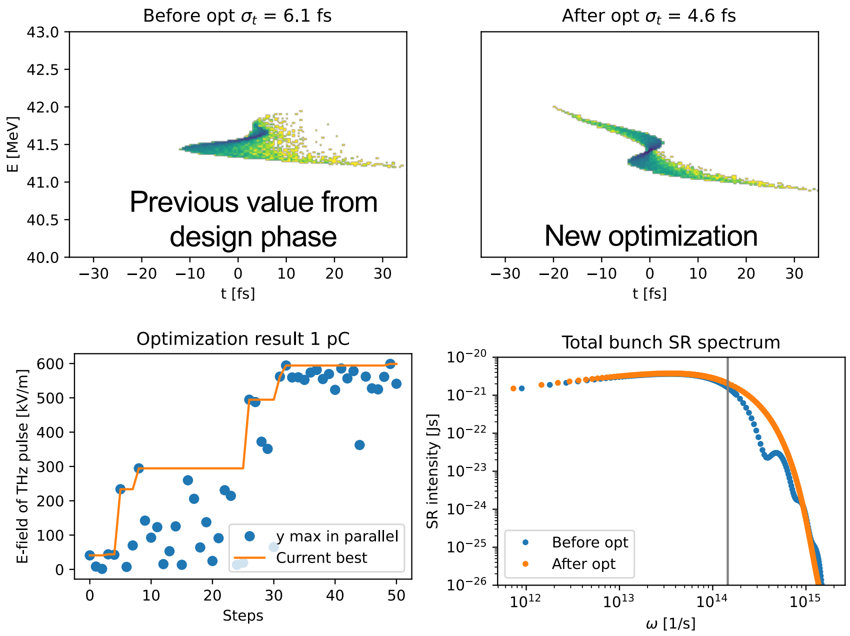
[1] Floettmann, K. ASTRA: A Space Charge Tracking Algorithm, http://www.desy.de/~mpyflo/ [2] Agapov, I.(2014) OCELOT: A software framework for synchrotron light source and FEL studies.

Emission optimization with Bayesian opt. – 1pC

Optimizing for minimum bunch length



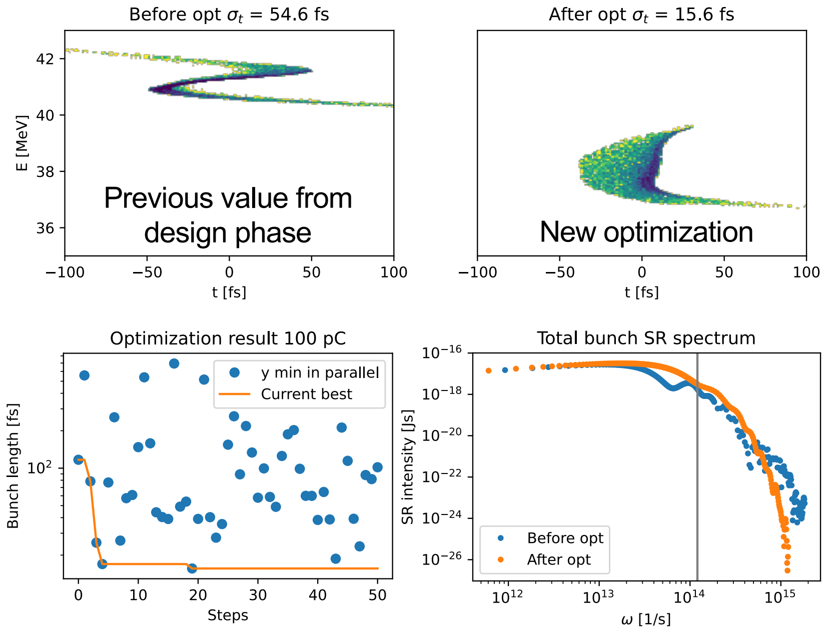
Optimizing for maximum E-field



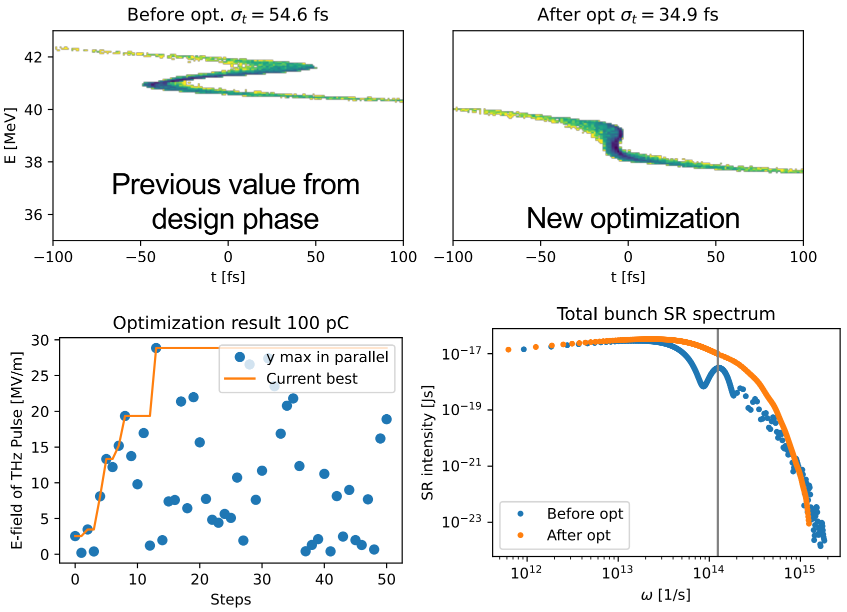
Courtesy of Chenran Xu

Emission optimization with Bayesian opt.- 100 pC

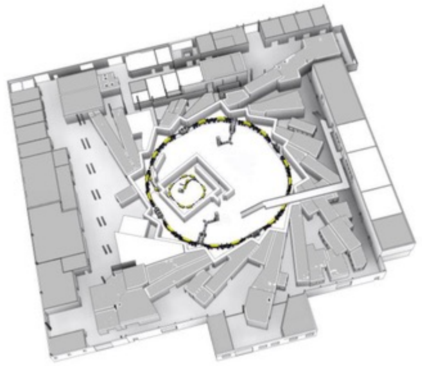
Optimizing for minimum bunch length



Optimizing for maximum E-field



Courtesy of Chenran Xu



KARA

Synchrotron light source and storage ring

2.5 GeV top energy

→ Control of the micro-bunching instability with reinforcement learning for stable/enhanced/damped CSR

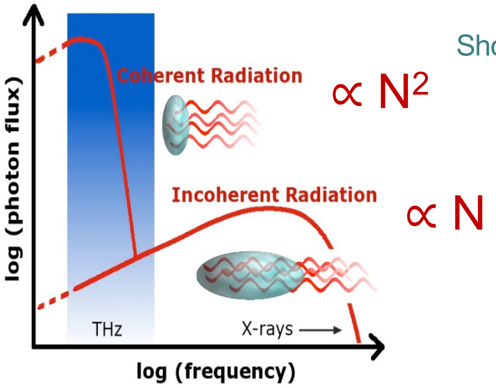
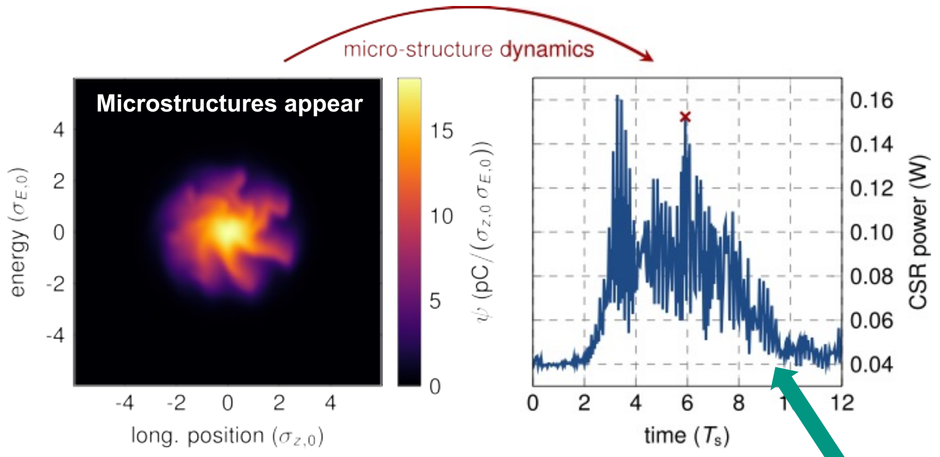


Image courtesy of A.-S. Müller

Short bunch mode (low-α)



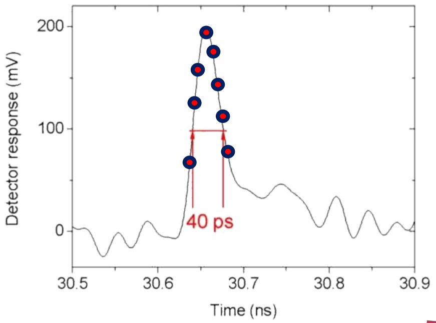
Plot courtesy of T. Boltz

CSR power fluctuation

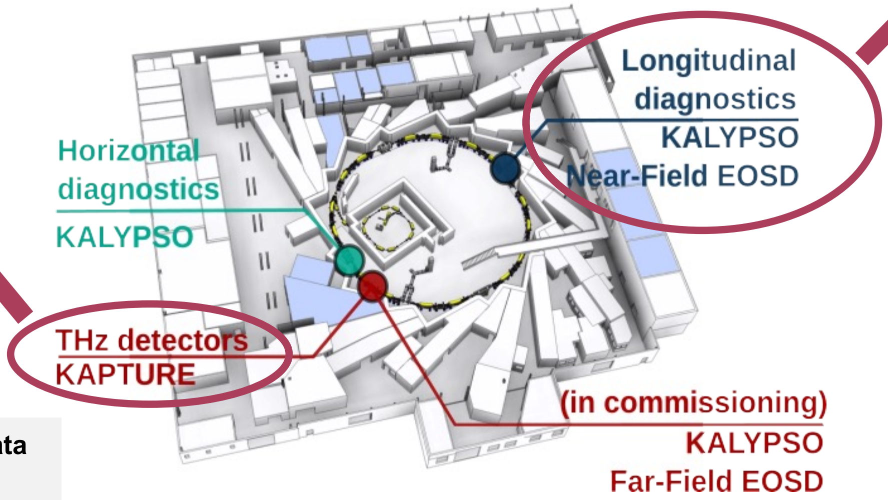
[“Micro-Bunching Control at Electron Storage Rings with Reinforcement Learning”, T.Boltz]

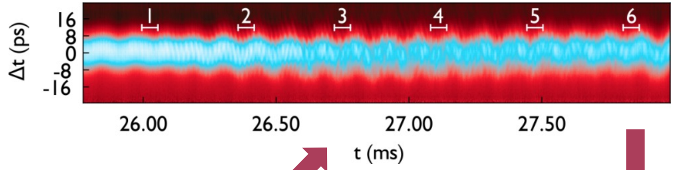
Real-time, high-repetition data acquisition

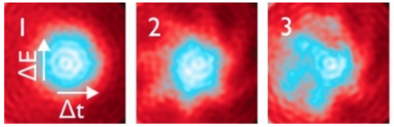
State-of-the-art diagnostics



KARA







Phase space density reconstruction

S. Funkner; DOI: [10.1103/PhysRevAccelBeams.22.022801](https://doi.org/10.1103/PhysRevAccelBeams.22.022801)

Collaboration with the **Institute of Data Processing and Electronics (IPE)**
Michele Caselle

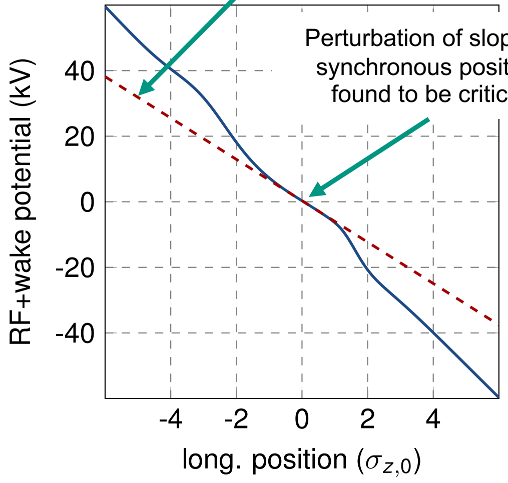
For more information on diagnostics: [“Turn-by-turn and bunch-by-bunch diagnostic developments at KIT”](), J. Steinmann (this workshop)

Influencing the micro-bunching instability

CSR self interaction

Continuously changing

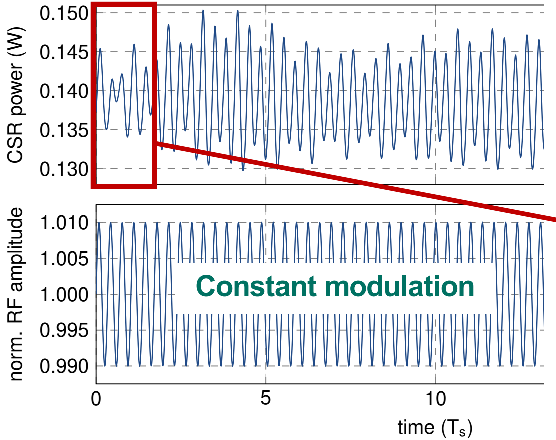
$$V_{eff}(q) = V_{RF}(q) + V_{CSR}(q)$$



Compensate the effect of the CSR perturbation by **modulating the RF voltage (amplitude)**

$$V_{RF} = \hat{V}(t)\sin(2\pi f_{RF} t)$$

$$\hat{V}(t) = \hat{V}_0 + A_{mod} \sin(2\pi f_{mod} + \varphi_{mod})$$



Initial damping, but quickly out of sync…we need **dynamic control**!

Images courtesy of T. Boltz

Applying reinforcement learning

Karlsruhe Institute of Technology

Action

$$\hat{V}(t) = \hat{V}_0 + A_{mod}\ \sin(2\pi f_{mod} + \varphi_{mod})$$

Reward

R = $\mu_{CSR} - w\ \sigma_{CSR}$ where w is a weight

Observable (state definition)

Charge distribution (simulation, KALYPSO)

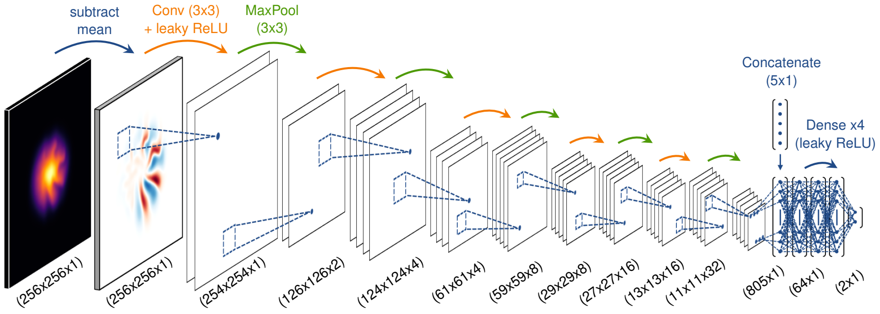
Input: (256x256) matrix + (5x1) feature vector

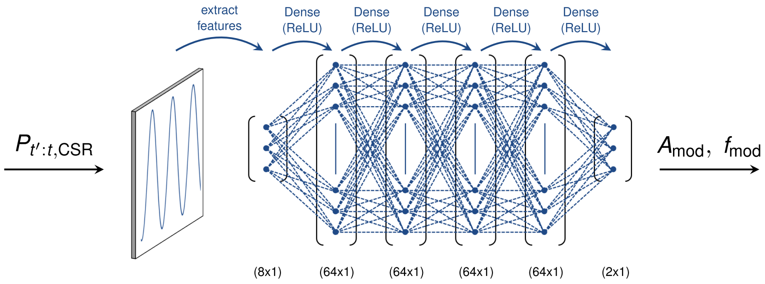
Observable (state definition)

CSR signal (simulation, KAPTURE)

Input: (8x1) feature vector

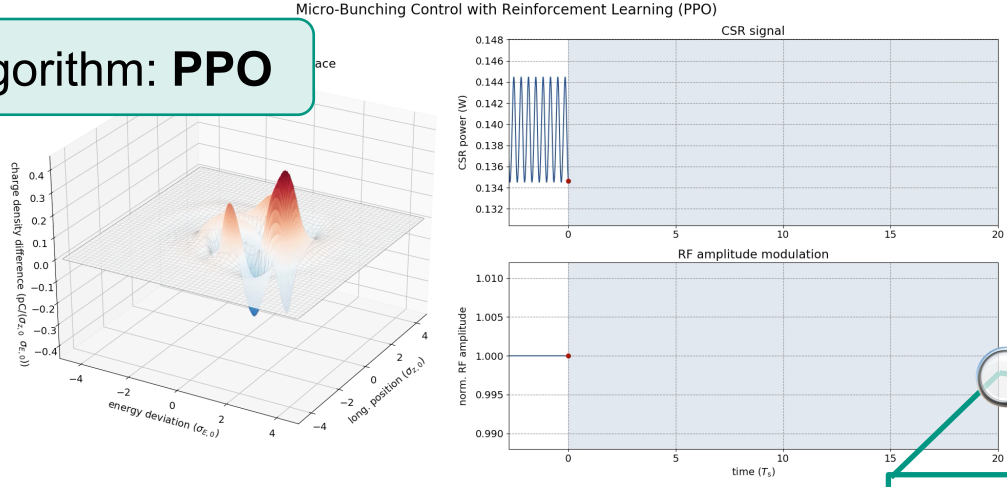
Easier to measure & process

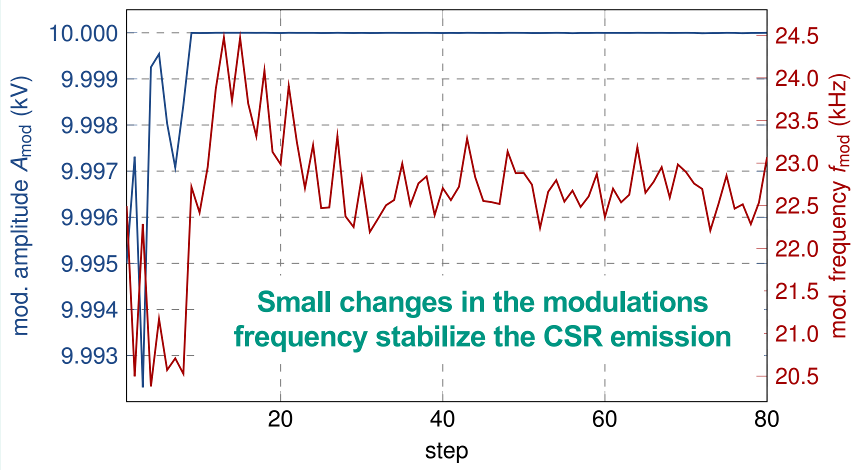




Images courtesy of T. Boltz

Algorithm: **PPO**



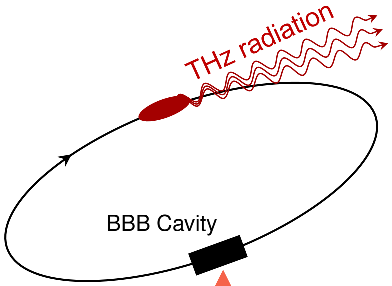


Courtesy of T. Boltz

In practice: we need hardware!

Fast feedback for real-time optimization

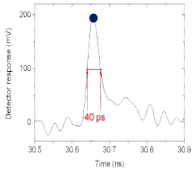
Collaboration with the **Institute of Data Processing and Electronics (IPE)**
Michele Caselle

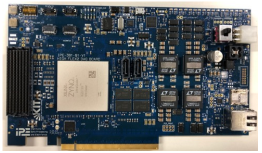


Detector

Schottky diode
50 GHz – 2 THz

KAPTURE II
Fast digitizer





HighFlex 2 DAQ card
Xilinx ZYNQ

Custom modular readout DAQ optimized for AI applications

BBB

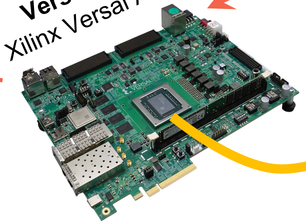
Bunch-by-bunch feedback system

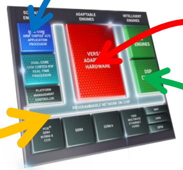
Δt < 1 ms

Ethernet 1GbE
Future: 380 Gb/s

RF modulation computed by Versal

Versal board
Xilinx Versal ACAP





Scalar processor (ARM-PS) for the implementation of the “critic” neural network

Large **FPGA**/DSP for data management

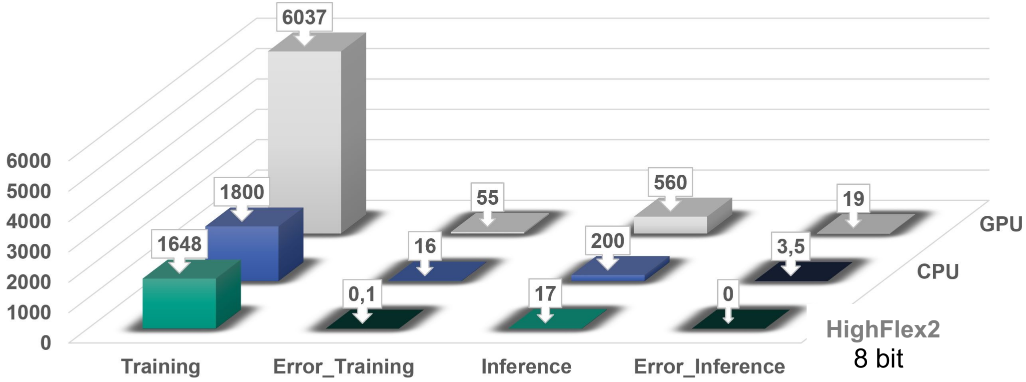
Dedicated programmable **AI-engines** for the deployment of fast inference “actor” neural network + feature extraction

400 tiles x 8 FP multiplications x 0.5 GHz
→ up to 1.6 Tera FP operations per second !!!

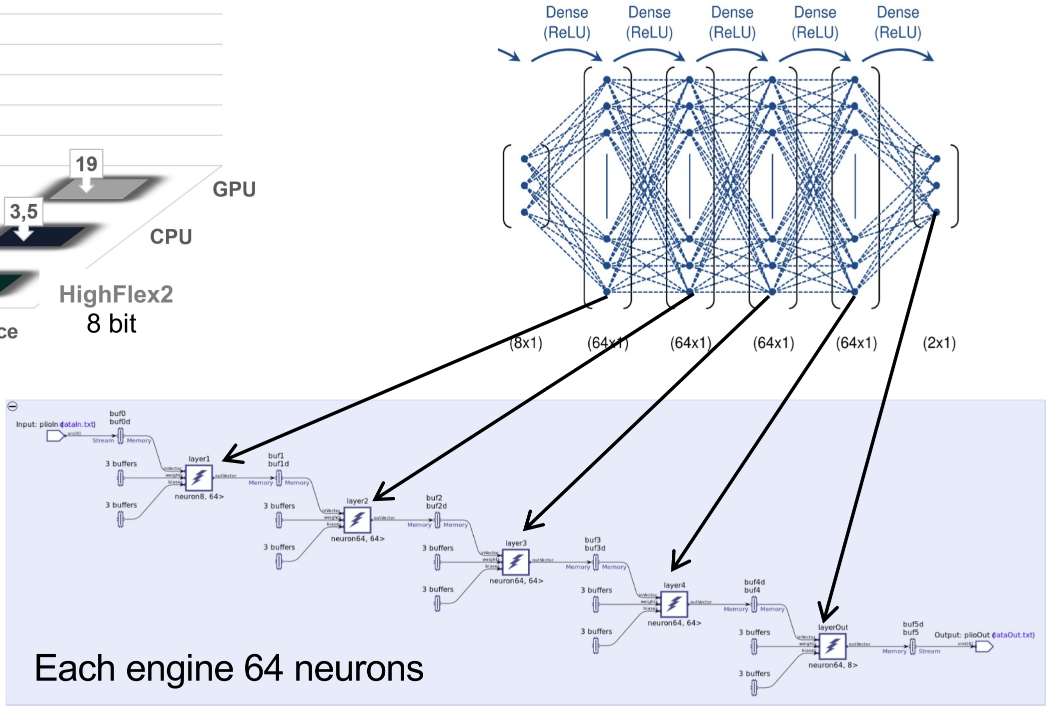
All linked by high bandwidth Network on a Chip (NoC)
full customization of dataflow

Images courtesy of M. Caselle

How fast can neural networks run?



Preliminary result: 4.5 µs
With Versal (32 bit floating point computation)



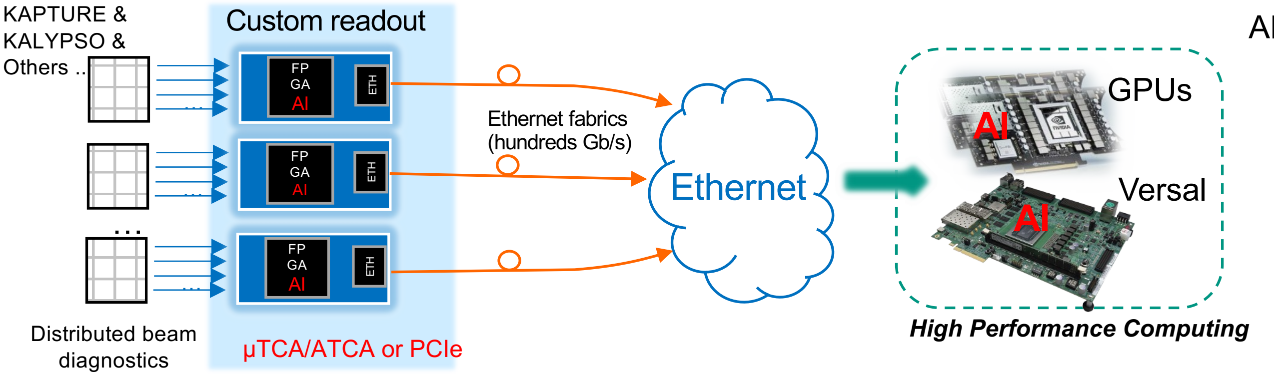
Images courtesy of E. Bründermann, M. Caselle, L. Scomparin

W. Wang, M. Caselle, et al IEEE TNS, https://doi.org/10.1109/TNS.2021.3084515 (2021)

The future: distributed DAQ-AI architecture

High-performance distributed /modular ML for physics experiments

- Heterogeneous FPGA/GPU architecture based on off-the-shelf devices(i.e. Ethernet, GPUs, etc)



AI processing tiers

- Hardware:
  - On ARM processors (Scalar Engines)
  - FPGA (Adaptable Engines)
  - AI Engines
  - DSP Engines
- Software
  - CPUs/GPUs in HPC

Slide courtesy of M. Caselle

Conclusions

Surrogate models

- Helpful in the design and commissioning phases (probing possible working points)
- Can give a smart starting point to optimizers to reduce optimization time
- Can be used as a virtual diagnostic with experimental input
- Can be partially re-trained with experimental data
- Curse of dimensionality: training only worth it for a limited number of parameters

Parallel Bayesian optimization

- Speeds up optimization considerably
- Gives you a stochastic model of your machine
- Helpful in the design and commissioning phases (probing possible working points)
- Can be extended to multiple objectives (Pareto front)

Reinforcement learning

- Extremely promising for online control of instabilities
- Requires hardware development for experimental implementation

This is only a small part of machine learning…

Many more promising applications to be studied!

andrea.santamaria@kit.edu