Applications of Machine Learning in Accelerator Operation

S. Tomin and I. Agapov

TRIUMF-Helmholtz Workshop on Scientific Computing

HAMBURG, GERMANY
16 September 2019

Automation, optimization and simulation for accelerator facilities

S. Tomin and I. Agapov

TRIUMF-Helmholtz Workshop on Scientific Computing

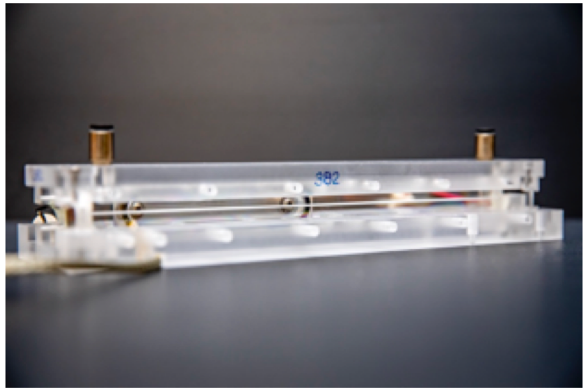
HAMBURG, GERMANY
16 September 2019

DESY accelerators highlights

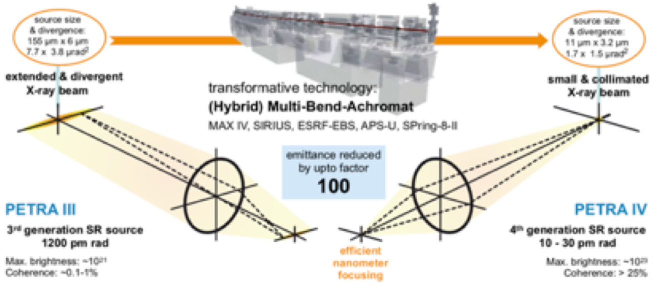
FLASH and European XFEL (operation since 2017)
World-leading Free Electron Lasers
providing high-intensity coherent x-ray radiation

https://accelerators.desy.de

Accelerator ARD such as plasma acceleration
Technology development for future accelerators

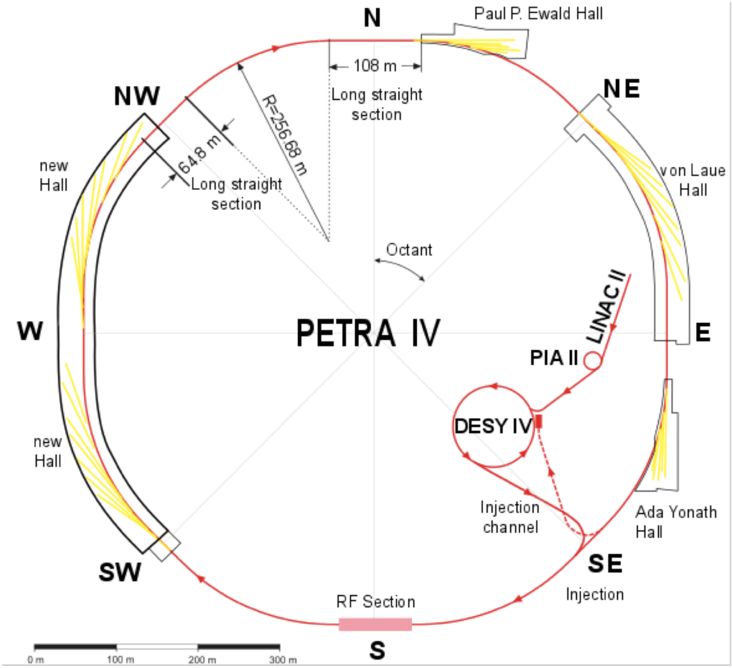


PETRA III is currently the most brilliant hard X-Ray synchrotron source worldwide
Providing the beams to the majority of DESY photon science users



PETRA IV project to upgrade PETRA to
the diffraction-limited source

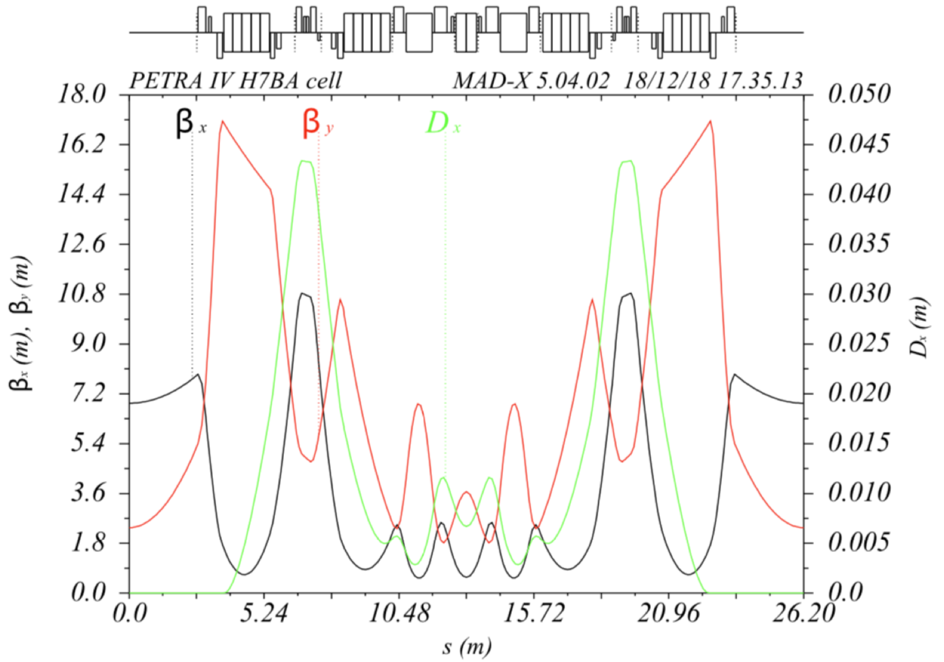
PETRA IV



Design parameter	PETRA III		PETRA IV	
Energy / GeV	6		6	
Circumference / m	2304		2304	
Emittance (horz. / vert.) / pm rad	1300 / 10		$<$ 20 / 4	$<$ 50 / 10
Total current / mA	100		200 [1]	80 [2]
Number of bunches	960	40	1600 *	80
Bunch population / 10^{10}	0.5	12	0.6	4.8
Bunch current / mA	0.1	2.5	0.125	1.0
Bunch separation / ns	8	192	4 / 20 (gaps)	96

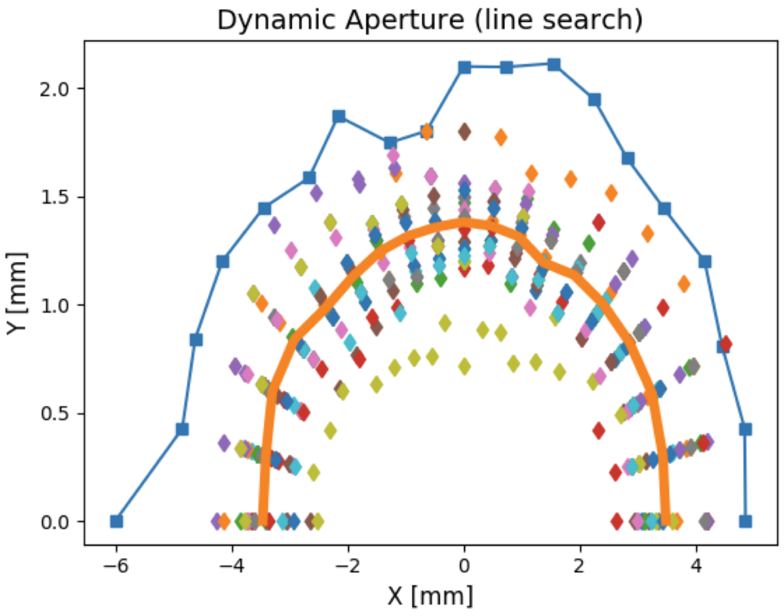
Challenges of 4G light source design

- Many tradeoffs to be made (vacuum chamber sizes, magnet strength, etc…) to stay technically feasible
- Nonlinear dynamics dominating (many options have close to 0 MA and DA)
- IBS and Touschek effects very prominent, need to operate with a lengthening cavity and round beams in high-intensity modes
- More demanding injection, extraction, and beam dumps



Challenges of 4G light source commissioning

- Users demand 1 max 2 years dark time, tight schedule
- Machine extremely sensitive to errors, ~30 µm alignment required
- Involved (automated) startup procedures required to store beam and reach target emittance



Dynamic aperture after simulated startup procedure for PETRA IV

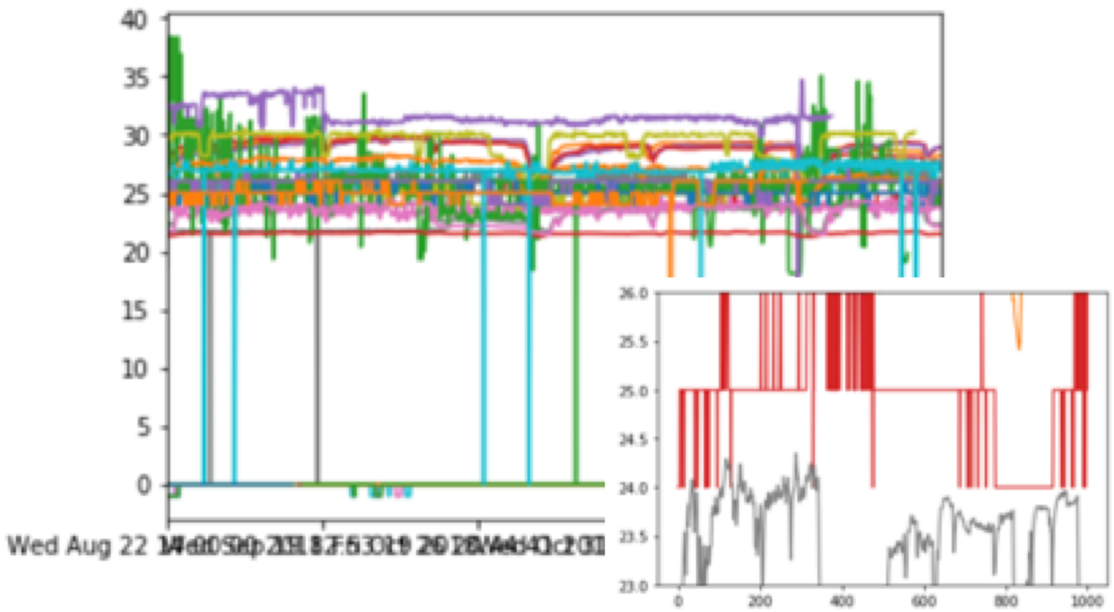
Uncorrected machine is typically unstable

Challenges of 4G light source operation

- Reliability demands grow (95% -> 99%)
- But machines are more sensitive with larger number of components
- We would like to meet availability goals and provide required beam-hours to users
- But at the same time we would like to keep doing accelerator physics and spend dedicated machine time on studies rather than machine setup
- This could only be successful if all standard procedures are highly automated
  - Startup of components (magnet cycling etc.)
  - Orbit correction
  - BBA
  - Optics measurement and correction
- Methods are well understood but (high-level) controls software not designed for autonomous operation

Big data and operational statistics

- A typical activity is to figure out what correlates to what through archive mining and try to improve things
- Problems:
  - Lengthy manual processing required to arrive at a meaningful dataset
  - Amount and quality of data after processing could be surprisingly low in comparison to the raw data



Typical pull from archive
All temperature sensors, Aug 22-Nov30 2018
Fraction of channel data missing (25%)
Unmatched timestamps
Data archived on thresholds
Bad readings/datapoints

ML and beam dynamics

- Although single-particle beam dynamics well developed, several directions still considered art, such as:
  - Multi-parameter matching in high dimensions (e.g. MBA. cell design)
  - Nonlinear aberrations (beyond simple ideas like low-order achromats, reducing sextuple strength, -I)
- A common line of reasoning — build fast “surrogate models” based on NN trained on simulated data

Sanity check — FODO

Script uses OCELOT python module

In [2]:
```python
# try to learn FODO stability
import os,sys
from ocelot import *
from pylab import *
QF = Quadrupole(l=0.1, k1=0.1)
QD = Quadrupole(l=0.1, k1=-0.14)
D = Drift(l=1.0)
fodo = (QF,D,QD,QD,D,QF)
lat = MagneticLattice(fodo)
```

```python
n_train = 10000

x_train = np.zeros([n_train,3])
y_train = np.zeros([n_train,1])

for i in range(n_train):
    QF.k1 = np.random.rand()
    QD.k1 = -np.random.rand()
    D.l = 2.0 * np.random.rand()

    x_train[i,0] = QF.k1
    x_train[i,1] = QD.k1
    x_train[i,2] = D.l
    lat.update_transfer_maps()
    tws = twiss(lat, Twiss())
    if tws is None:
        y_train[i,0] = 0
    else:
        y_train[i,0] = 1
```

```python
import keras
from keras.models import Sequential
from keras.layers import Input, Dense, Dropout
from keras.utils import to_categorical
from pylab import *

#fixing a duplicate openmp dylib on mac???
import os
os.environ['KMP_DUPLICATE_LIB_OK']='True'

model = Sequential()
model.add(Dense(64, input_dim=3, activation='relu'))
#model.add(Dropout(0.5))
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss='binary_crossentropy', optimizer='sgd',metrics=['accuracy'])
model.fit(x_train,y_train, epochs=30, batch_size=16)
```

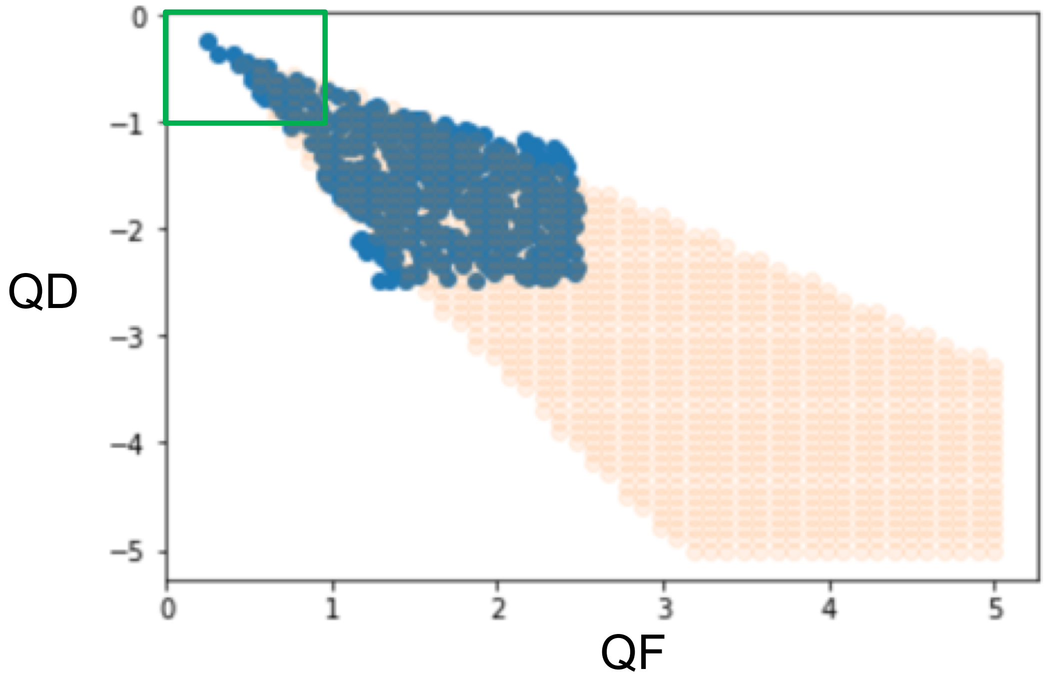
```
Using TensorFlow backend.

Epoch 1/30
10000/10000 [==============================] - 2s 211us/step - loss: 0.3254 - acc: 0.9241
Epoch 2/30
10000/10000 [==============================] - 2s 153us/step - loss: 0.2748 - acc: 0.9313
Epoch 3/30
10000/10000 [==============================] - 2s 152us/step - loss: 0.2614 - acc: 0.9313
```

```python
n_test = 5000

x_test = np.zeros([n_test,2])
y_test = np.zeros([n_test,1])

D.l = 3.9

for i in range(n_test):
    QF.k1 = 5*np.random.rand()
    QD.k1 = -5*np.random.rand()

    x_test[i,0] = QF.k1
    x_test[i,1] = QD.k1
    lat.update_transfer_maps()
    tws = twiss(lat, Twiss())
    if tws is None:
        y_test[i,0] = 0
    else:
        y_test[i,0] = 1
```

Sanity check — FODO

Training set was QF in [0,1] QD in [-1,0] L in [0,2]

Stability diagram check on the test set
QF in [0,5] QD in [-5,0] L=3.9

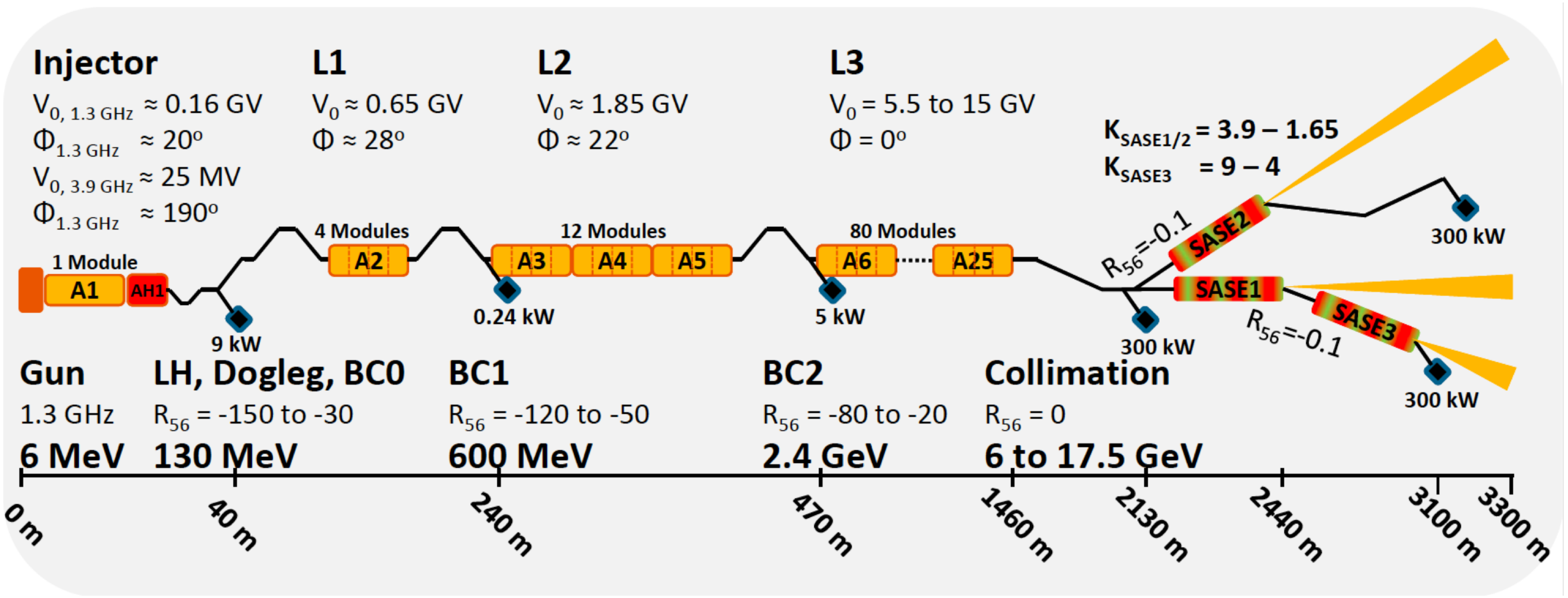


- The trained NN generalizes surprisingly well beyond training set parameters!
- However only within a certain range, and validity range needs checks
- Practical application not clear

Summary computing and ML for storage rings

- Design of next generation of storage rings is challenging and requires significant computing power to evaluate performance parameters
- GPU computing has potential to speed up storage ring beam dynamics calculations by several orders of magnitude, open ways to more advanced optimisation and speed-up the optics design process. Unfortunately, 4th generation light source projects (such as ESRF EBS, APS-U, PETRA IV) are all in late project phases to benefit from any breakthroughs here
- We don’t see much potential of ML for accelerator design or beam dynamics studies
- With increasing complexity and sensitivity (to temperature, ground motion etc.) of next-generation storage rings, we see potential of ML techniques such as reinforcement learning in the area of control and automation
- Proper facility monitoring and technical data curation remains a major challenge

The European XFEL Accelerator Overview



Nominal parameters

- Electron beam energy up to 17.5 GeV
- Pulse rep. rate 10 Hz
- Bunches per pulse 2700
- Intratrain rep. rate 4.5 MHz
- Bunch charge 0.02 – 1 nC

- Electron bunches in a single pulse are distributed by a fast kicker system to three SASE undulators

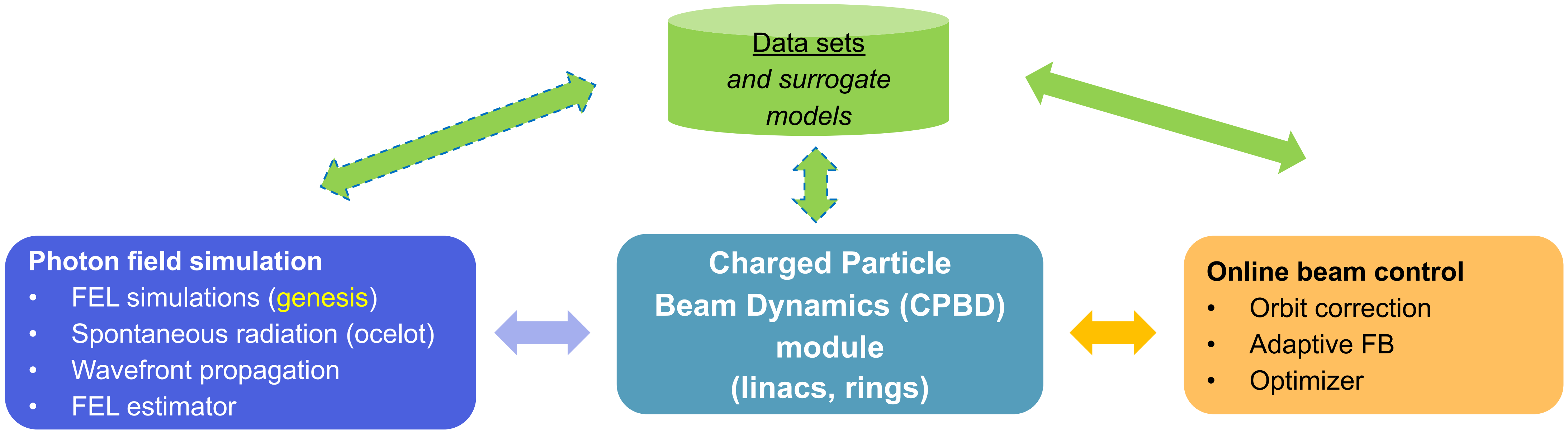
W. Decking et al, "Commissioning of the European XFEL", https://doi.org/10.18429/JACoW-IPAC2017-MOXAA1

The European XFEL challenges

- S2E simulation
  - Understanding the properties of the electron and photon beams.
    - ► Collective effects play an important role in the electron beam dynamics and in many cases define the electron beam properties, e.g. energy chirp, emittances, slice energy spread
  - Crucial for the operation, e.g. defining compression scenario
  - Studies of new FEL schemes (beam dynamics + FEL), e.g. two color with corrugated structure.
  - But S2E simulations are time consuming, typically needs complicated software setup
- Optimization of accelerator performance
  - FEL performance is highly sensitive to hundreds of free tuning parameters, e.g. the beam orbit and optics, RF settings, undulator gaps etc,

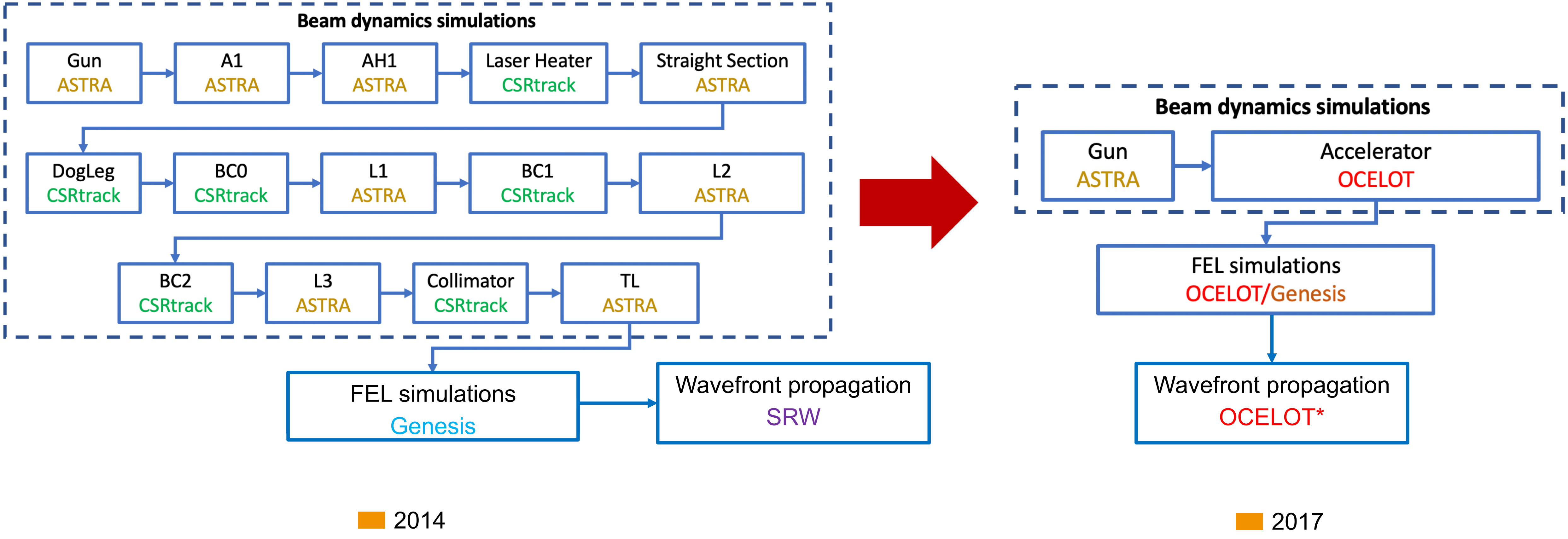
OCELOT

OCELOT toolkit overview

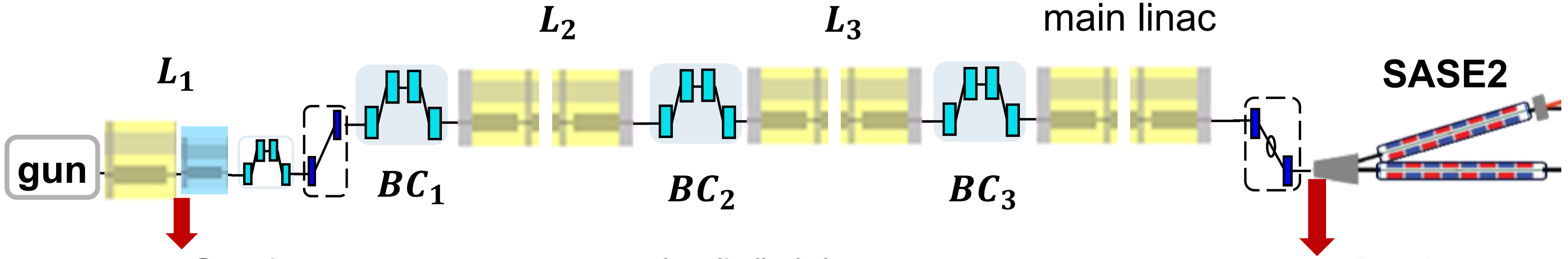


- Started as simulation project (spontaneous radiation, FEL) at European XFEL. *I. Agapov et al., NIM A. 768 2014*
- Beam dynamics module was developed (linear optics, collective effects, second order effects, optim. techniques).
  - *S.Tomin et al.* doi.org/10.18429/JACoW-IPAC2017-WEPAB031
- FEL applications. *S. Serkez et al, Journal of Optics, Volume 20, Number 2, 024005, 2018*
- Turned into more on-line control-oriented development
  - *Optimizer, orbit correction tool, adaptive feedback*
  - *arXiv:1704.02335,* S. Tomin et al., doi:10.18429/JACoW-ICALEPCS2017-WEAPL07
- Everything in **Python**. Focus on simplicity. Implement only physics
- Open source (On GitHub **https://github.com/ocelot-collab/ocelot**)

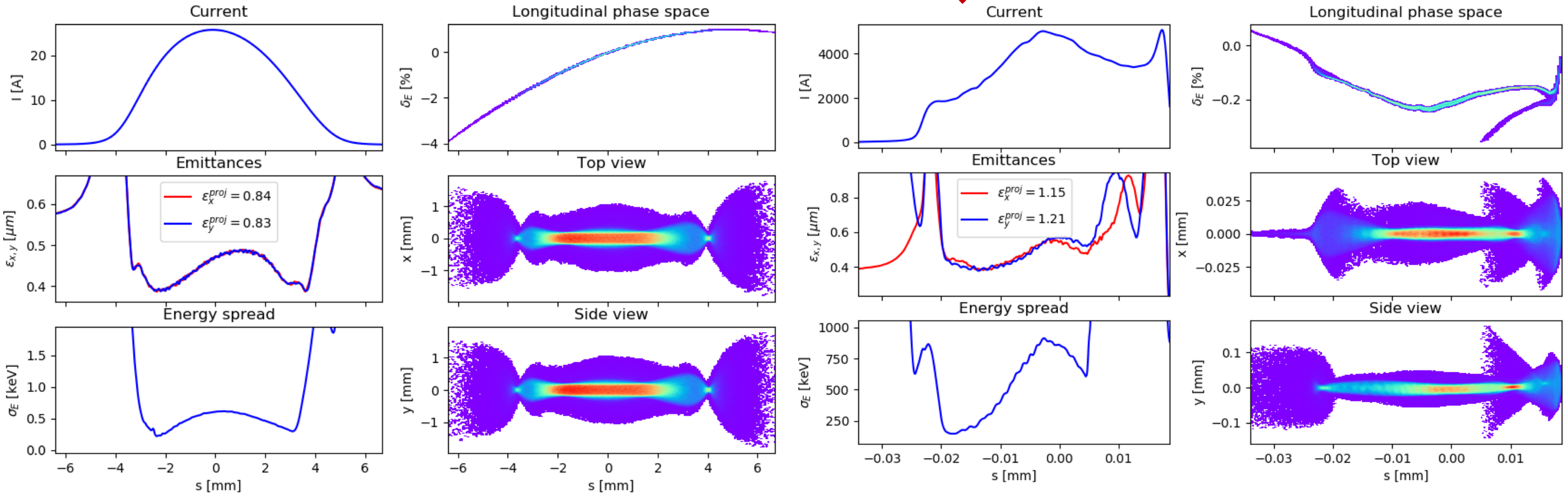
S2E simulation. OCELOT multiphysics simulation toolkit. Motivation



Beam dynamics for 500 pC, 5 kA



- projected x-emittance growth by 37%
- projected y-emittance growth by 44%

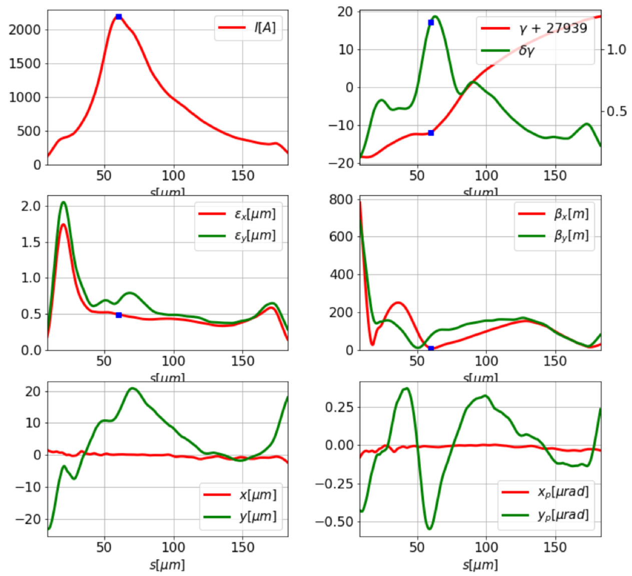


M Dohlus, S.Tomin and I. Zagorodnov, “Shaping the Future of the European XFEL: Options for the SASE4/5 Tunnels“, December 6, 2018

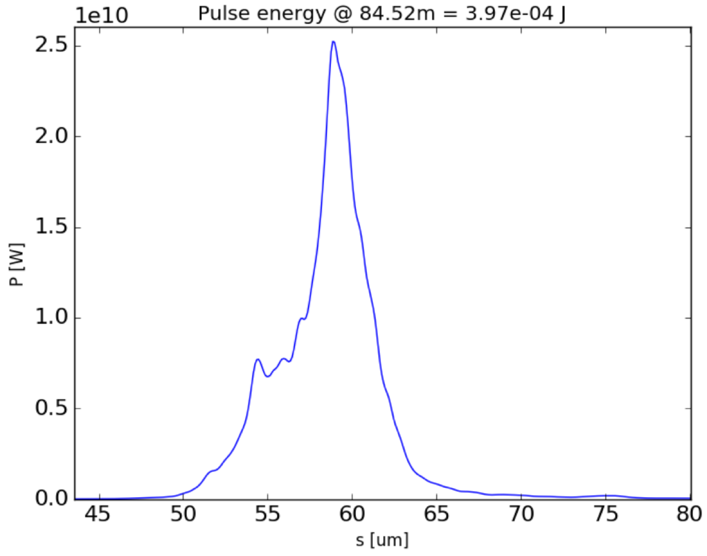
OCELOT toolkit: S2E in control room (beta version)

- Reading quads and cavities settings and measured beta-functions
- Tracking 200000 particles with CSR, SC, wakes through all machine up to undulator section
- Total time calculation 20 mins
- FEL power Estimator (Ming Xie parametrization). – 0.4 mJ
- In reality, we had 1 mJ with nonlinear undulator tapering
- Genesis can be used, as well

Beam slice parameters



FEL power estimation with Ming Xie

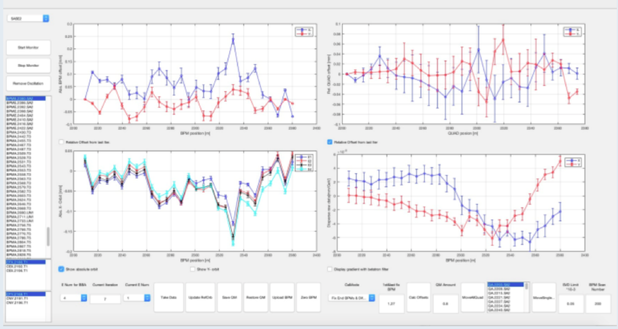


Optimization of accelerator performance

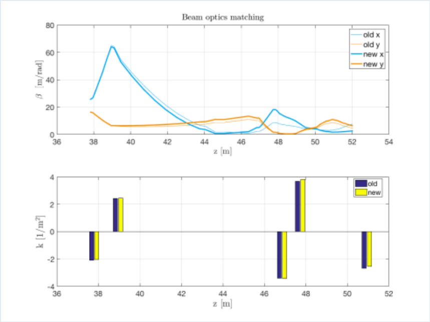
- Modern Free Electron Lasers are complex facilities with hundreds of free tuning parameters
  - Bunch compression, orbit, beam optics, gun optimization, undulator gaps, phase-shifters, etc
- Even when the main accelerator systems work well, manual fine-tuning is necessary to get the best performance and this is time expensive
- More automation of the tuning procedures is needed

Zoo of High Level Control tools

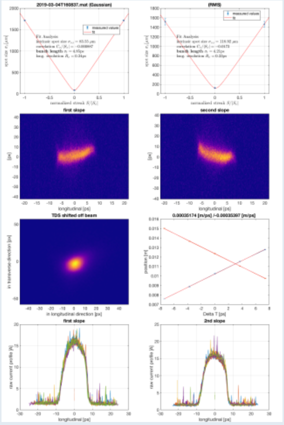
Beam Based Alignment



Beam matching

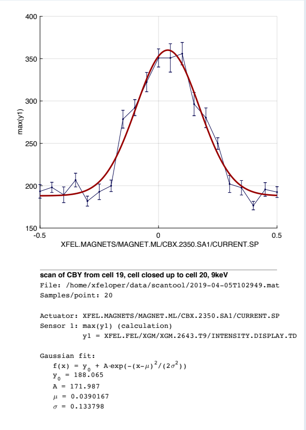


Longitudinal profile measurement



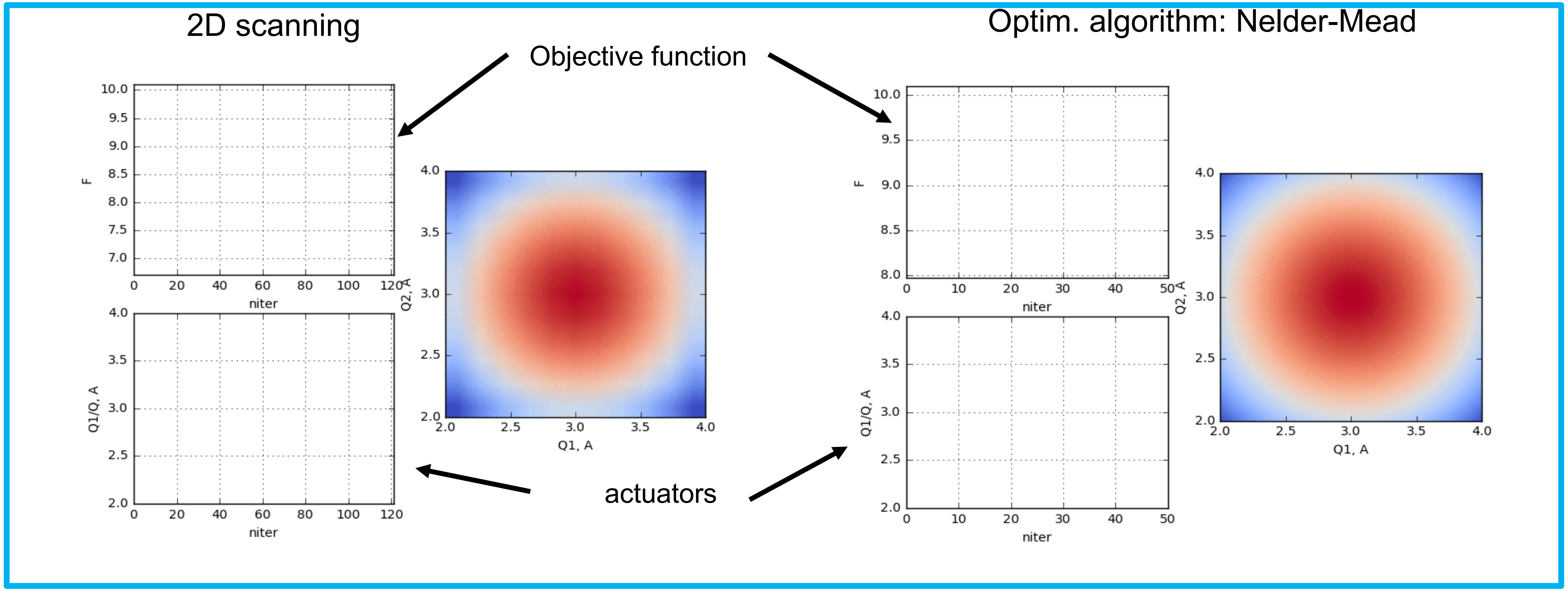
Orbit correction

Scan tool



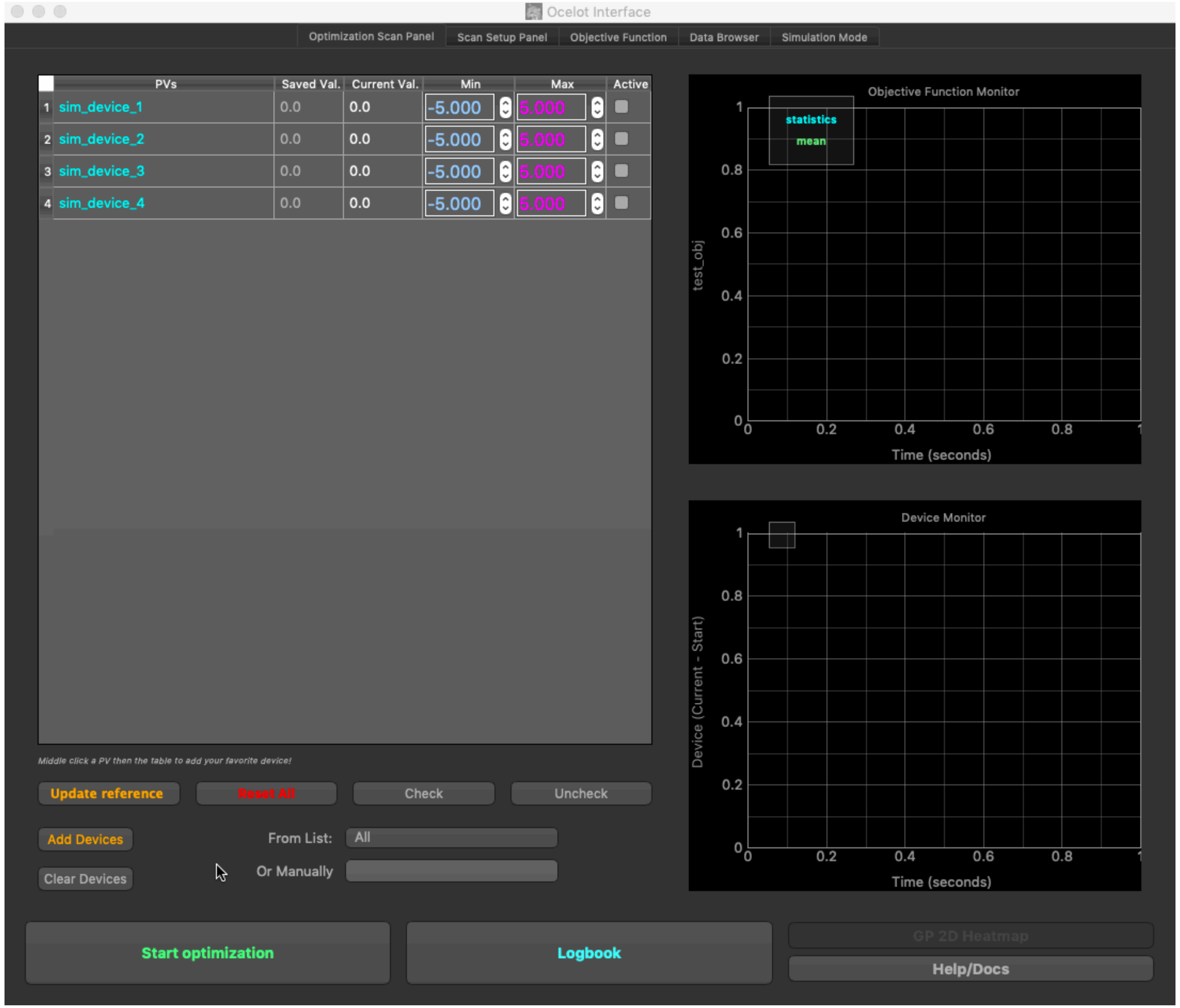
Optimization of accelerator performance. Optimization algorithm

- Optimization algorithms are faster than scanning
- Optimization methods can be model-independent or model-dependent



OCELOT Optimizer

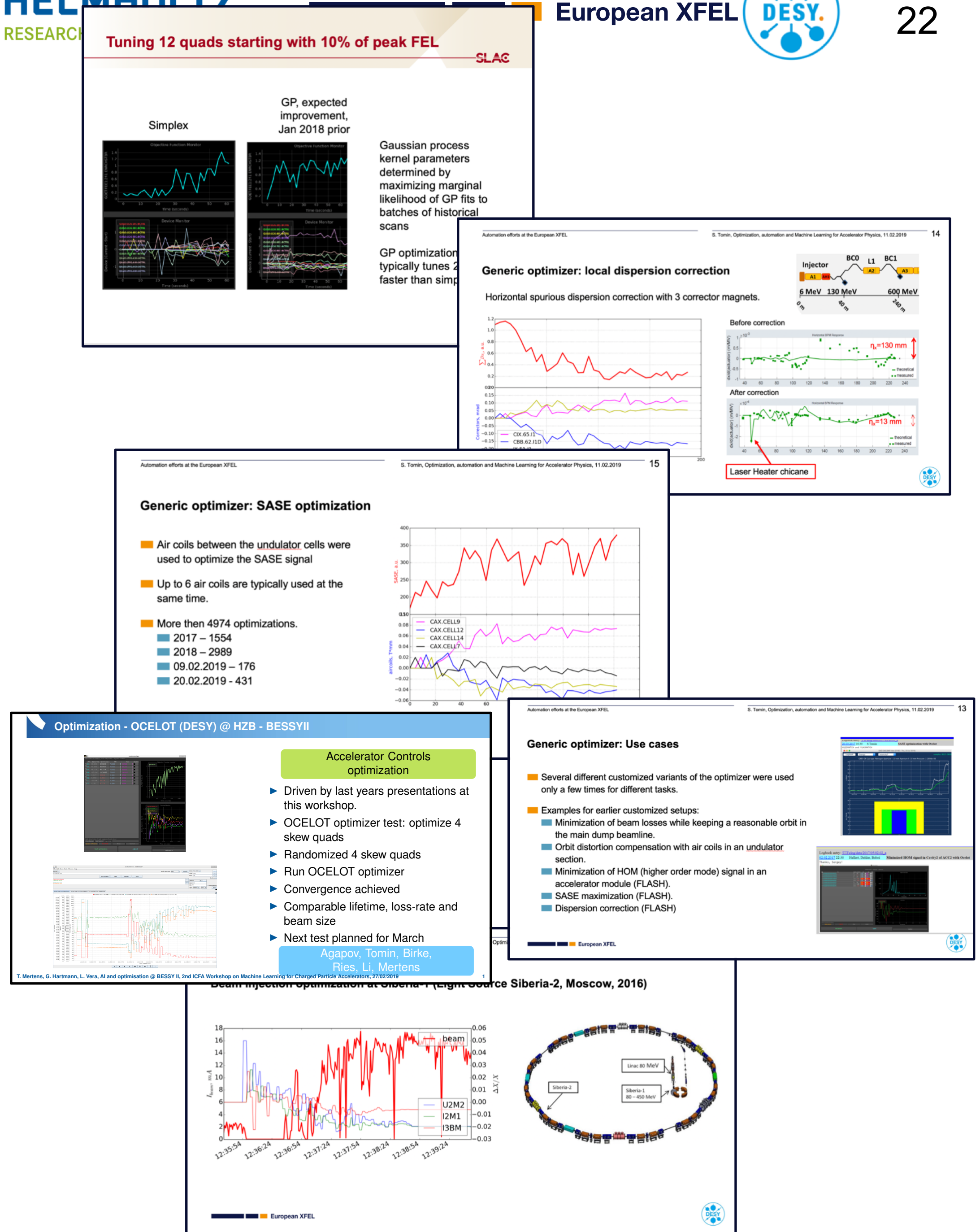
- Optimization algorithms are faster than scanning
- OCELOT optimizer is a flexible platform for optimization:
  - Interchangeable optimization methods
  - GUI
    - ► Add/select device or group of devices
    - ► Craft/modify target function
  - Infrastructure for testing new methods
  - Save/load configs
  - Logging
- Collaboration DESY, EuXFEL, SLAC



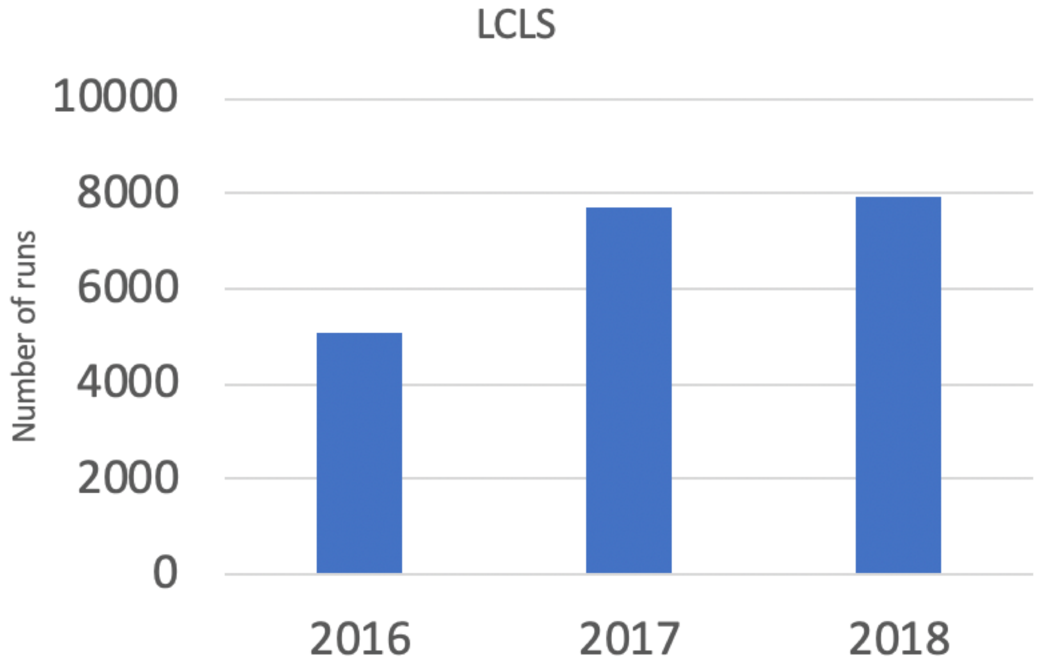
- I. Agapov et al, arXiv:1704.02335
- S. Tomin et al, https://doi.org/10.18429/JACoW-IPAC2017-WEPAB031
- M.W. McIntire et al, DOI:10.18429/JACoW-IPAC2016-WEPOW055

Use cases

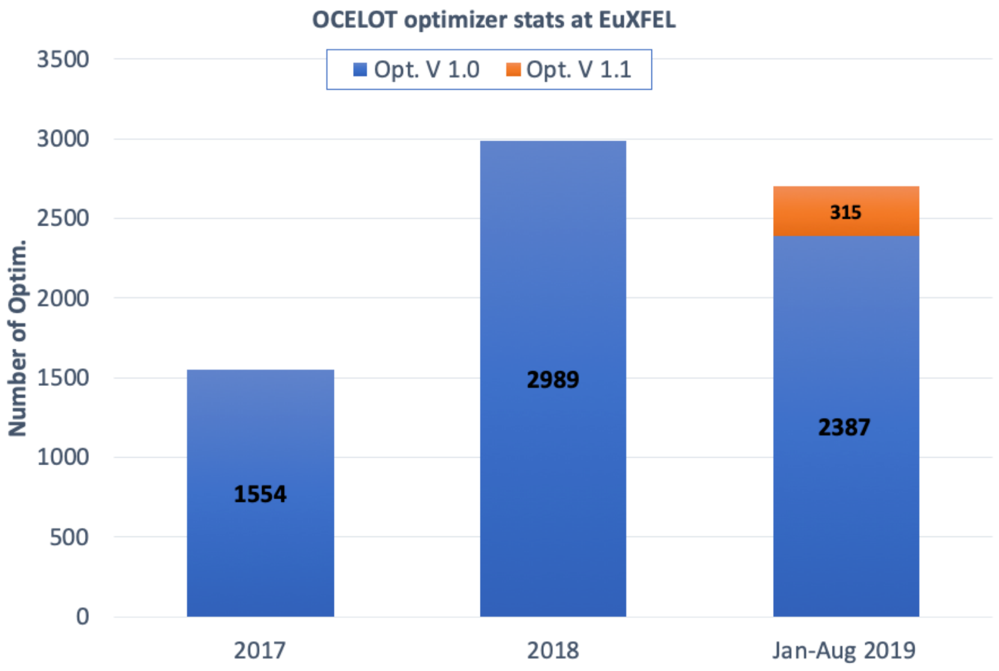
- FEL facilities
  - FEL pulse energy maximization:
    - ► Launch orbit and orbit inside an undulator (EuXFEL)
    - ► Phase-shifters (EuXFEL)
    - ► Orbit in low energy sections (EuXFEL)
    - ► Matching quads (LCLS & EuXFEL)
    - ► RF settings (EuXFEL)
  - Local dispersion correction (EuXFEL & FLASH)
  - HOM signal minimization in cavities (FLASH)
- Storage rings:
  - Injection efficiency optimization (Kurchatov Institute)
  - Beam life time (test at BESSY-II)
  - SPEAR3 (SLAC)



Ocelot optimizer stats



The European XFEL

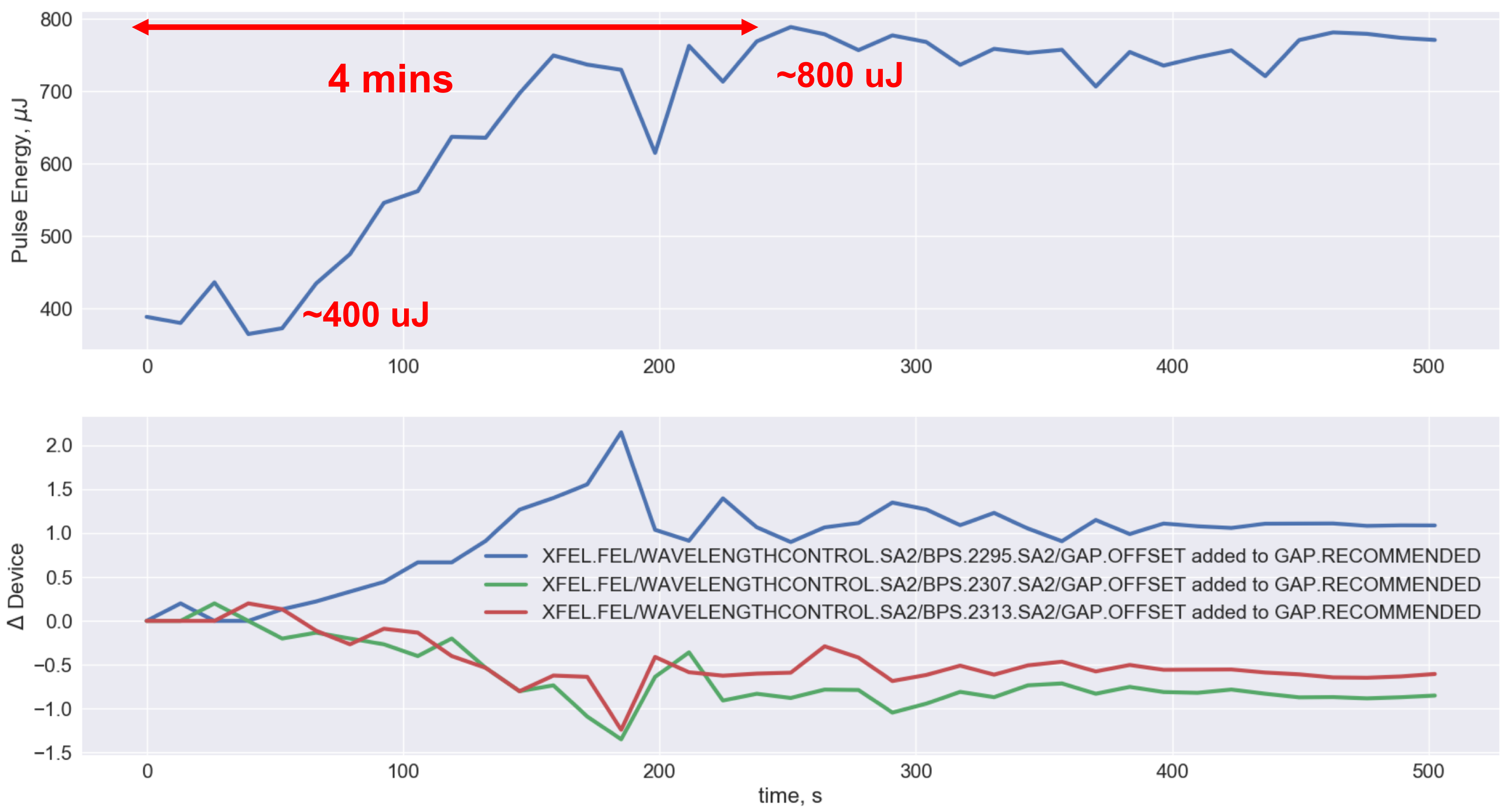


SLAC: Ocelot with Nelder-Mead simplex reduced tuning times on average by 25% to 50% compared to hand tuning

OCELOT Optimizer: Use cases

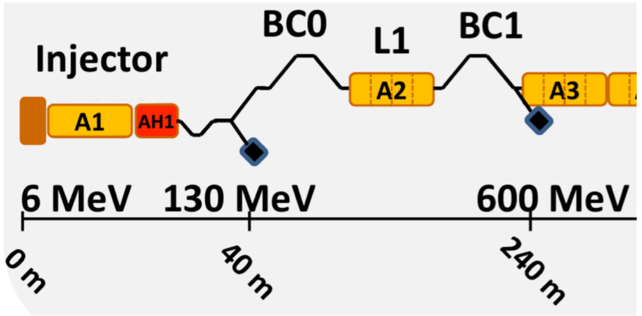
- European XFEL
  - FEL pulse energy maximization:
    - ▶ Orbit inside an undulator
    - ▶ Phase-shifters
    - ▶ Orbit in injector
    - ▶ Matching quads
    - ▶ RF settings

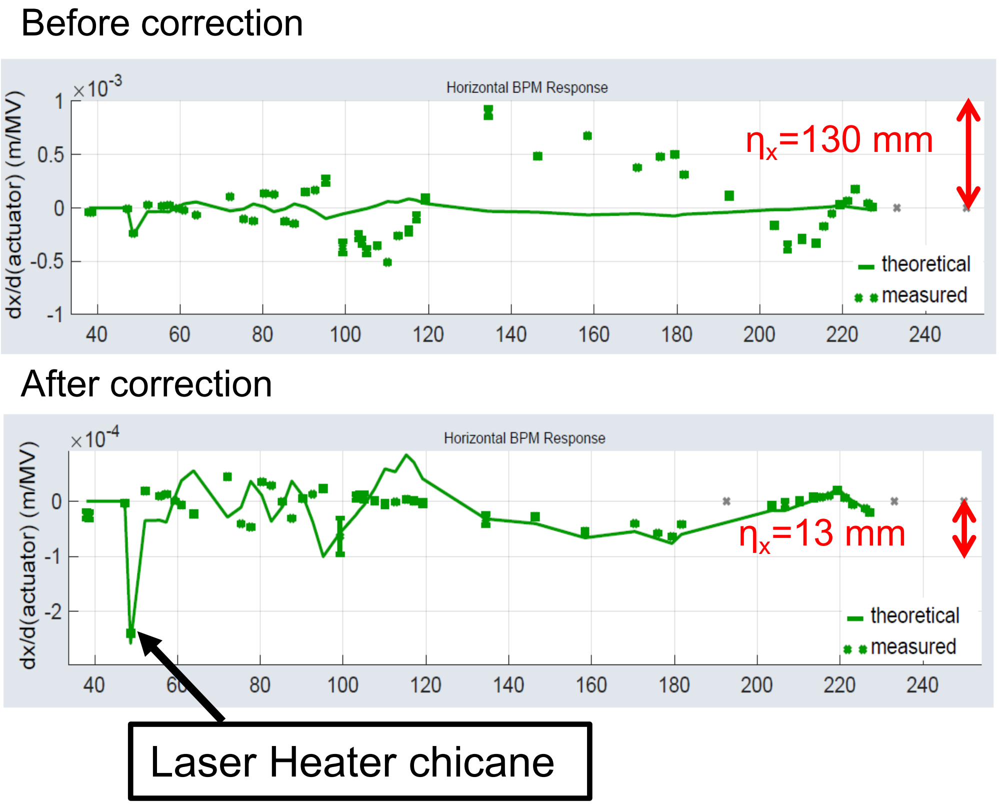
Phase-shifters

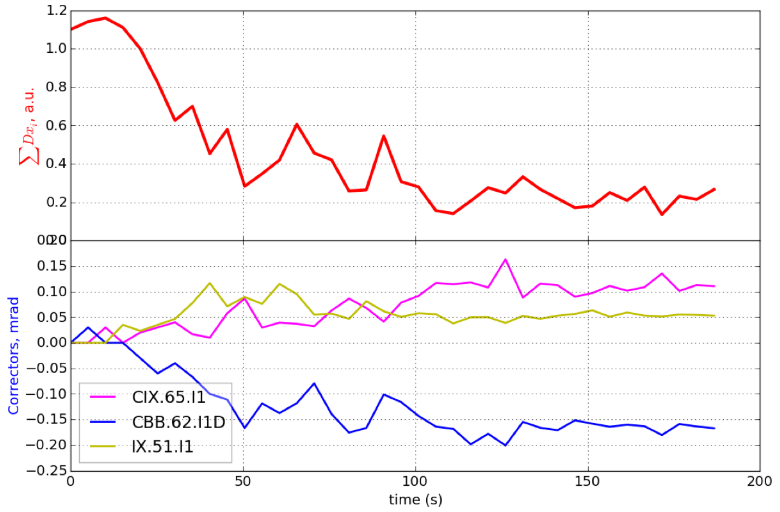


OCELOT Optimizer: Use cases

- European XFEL
  - FEL pulse energy maximization:
    - ▶ Orbit inside an undulator
    - ▶ Phase-shifters
    - ▶ Orbit in injector
    - ▶ Matching quads
    - ▶ RF settings
  - Local dispersion correction in injector

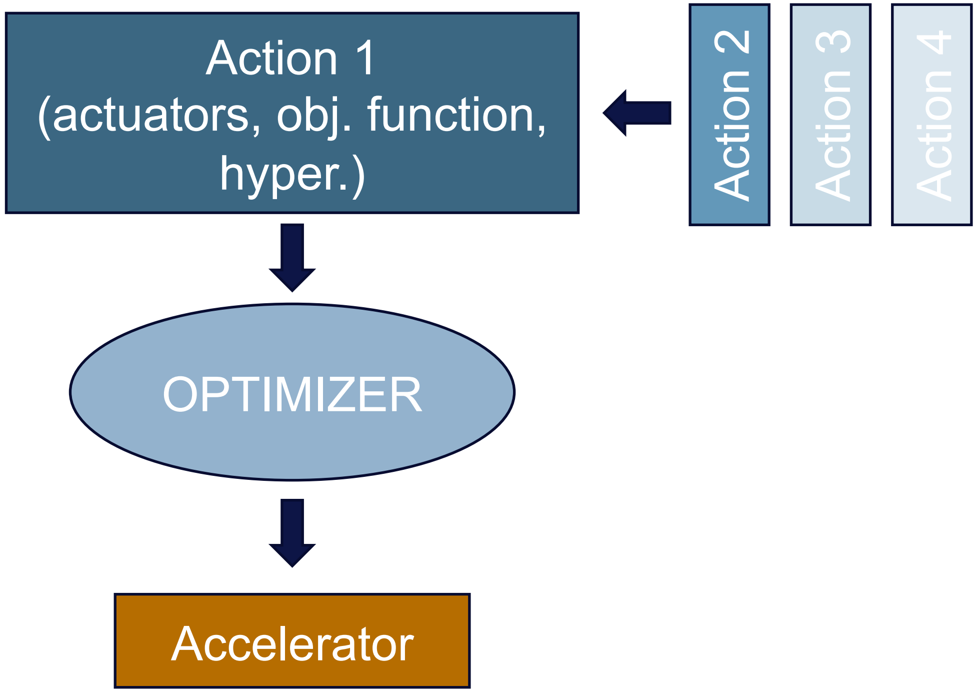




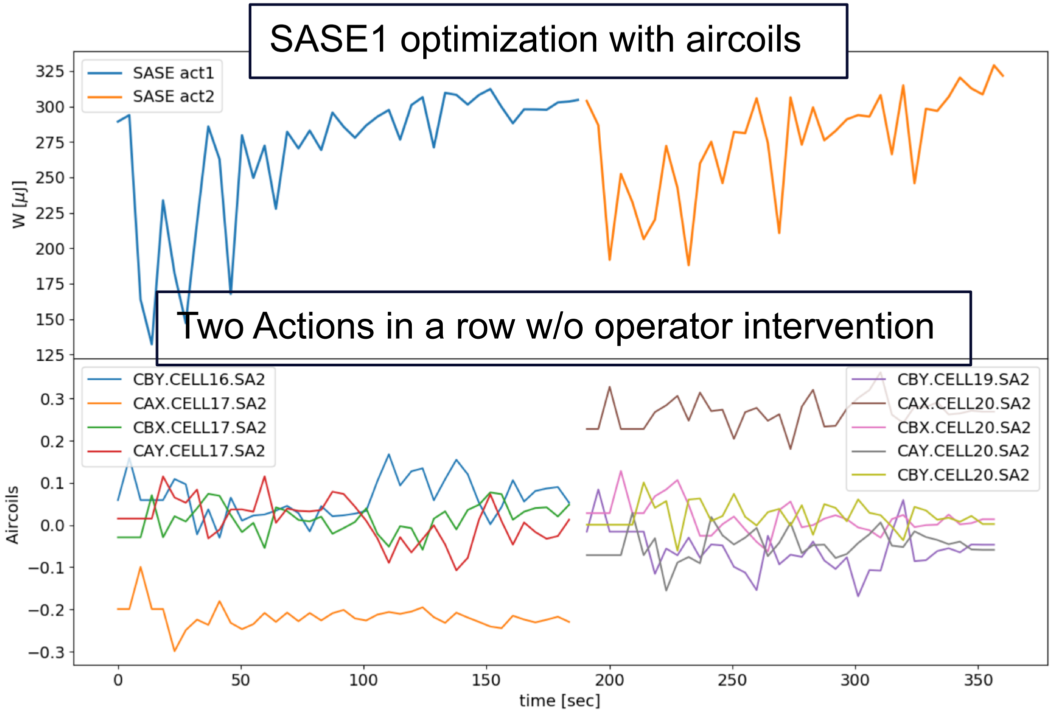


Sequence of optimizations: automatic optimization

- Optimization with small number of actuators (4-6) is more efficient than with many due to noise and slow drifts

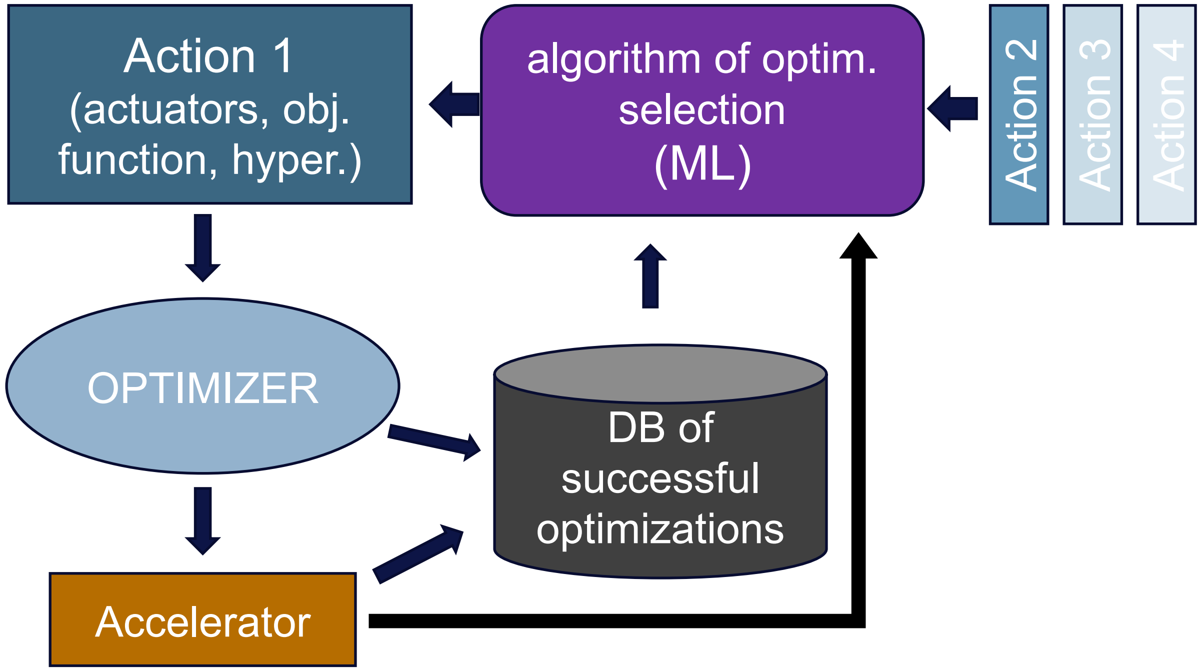


- New version (v1.1) was deployed
- Predefined sequence of optimization
  - without operator intervention
  - Optimizer monitors machine state – paused optimization if necessary.

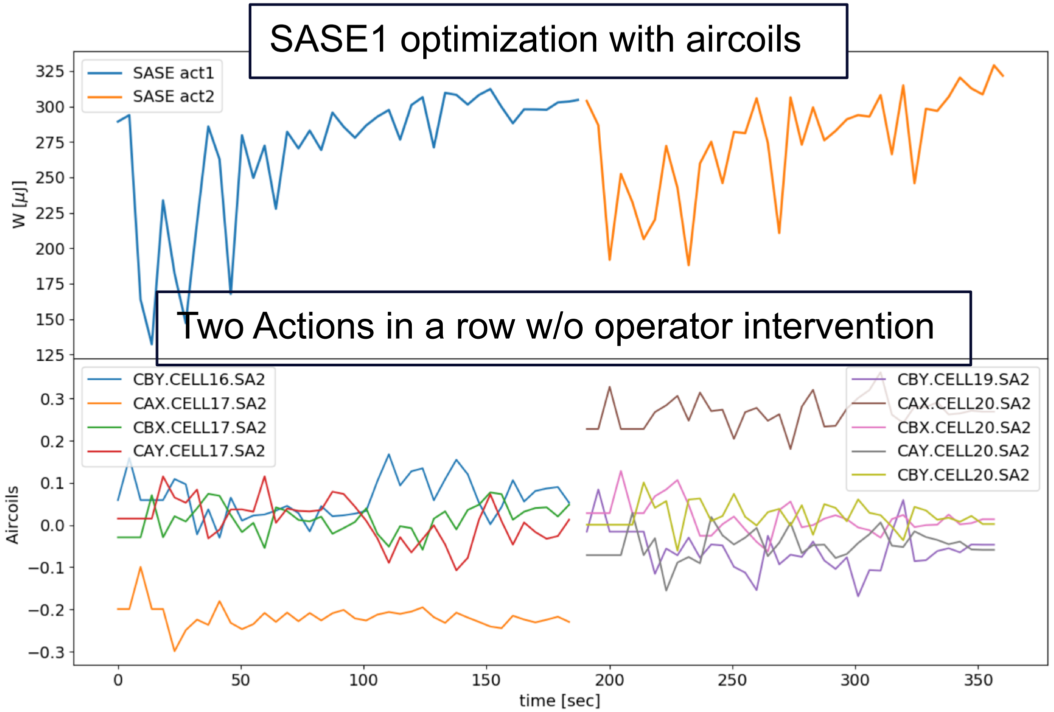


Sequence of optimizations: automatic optimization

- Optimization with small number of actuators (4-6) is more efficient than with many due to noise and slow drifts



- New version (v1.1) was deployed
- Predefined sequence of optimization
  - without operator intervention
  - Optimizer monitors machine state – paused optimization if necessary.



Model-dependent optimizations

- Model-dependent optimizations use/construct a regression model which can predict result.
  - Mathematical/physical accelerator model
  - Model based on data analysis (statistic, ML)

- Compression scenarios optimization

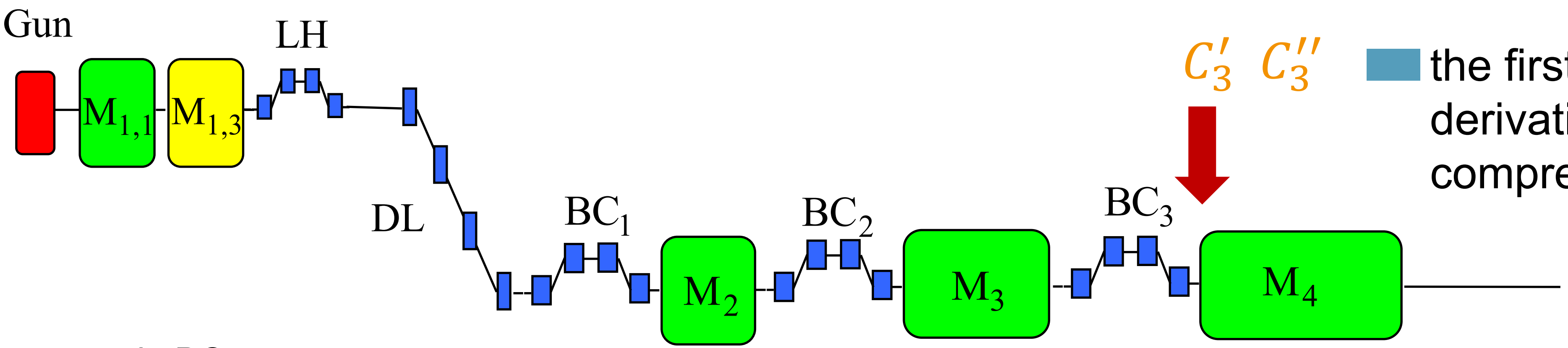
- Adaptive orbit feedback

Charged Particle Beam Dynamics (CPBD) module (linacs, rings) ⟷ **Online beam control**

- Orbit correction
- Adaptive FB
- Optimizer

Bunch compression optimization

Working point (11 parameters of longitudinal beam dynamics)



C_3' C_3''

- the first and the second derivatives of the global compression

- energy in BC:
- Deflection compaction factor:
- compression factor:

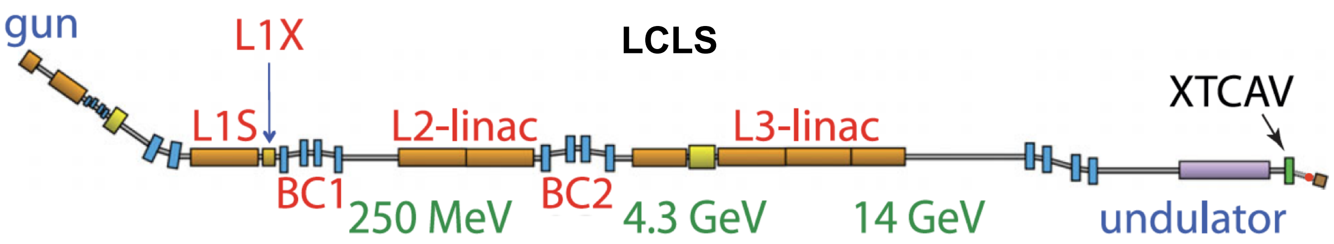
$E_1 = 130$ MeV, $R_{56}^1 = ?$, $C_1 = ?$

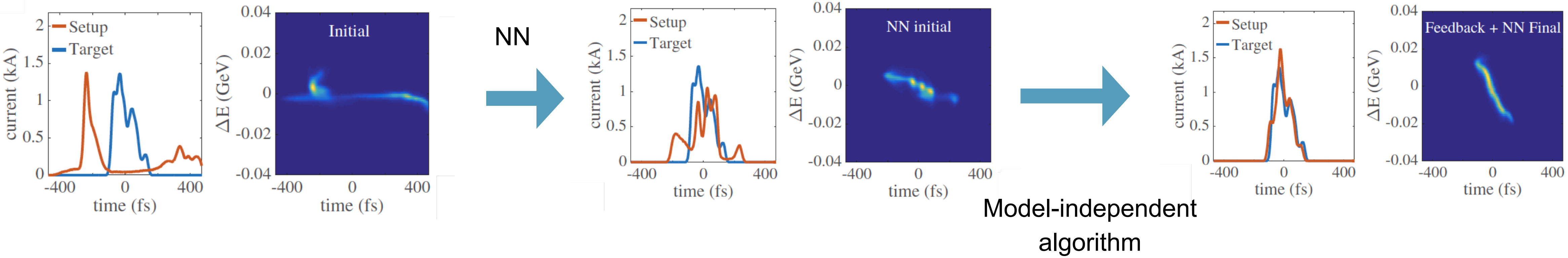
$E_2 = 700$ MeV, $R_{56}^2 = ?$, $C_2 = ?$

$E_3 = 2400$ MeV, $R_{56}^3 = ?$, $C_3 = I_{\mathrm{final}}/I_{\mathrm{gun}}$

- Optimizing working point
  - Theoretical analysis (+ simulations)
  - measurement
  - tuning against the FEL performance

Model-independent and ML methods for control of longitudinal beam dynamics



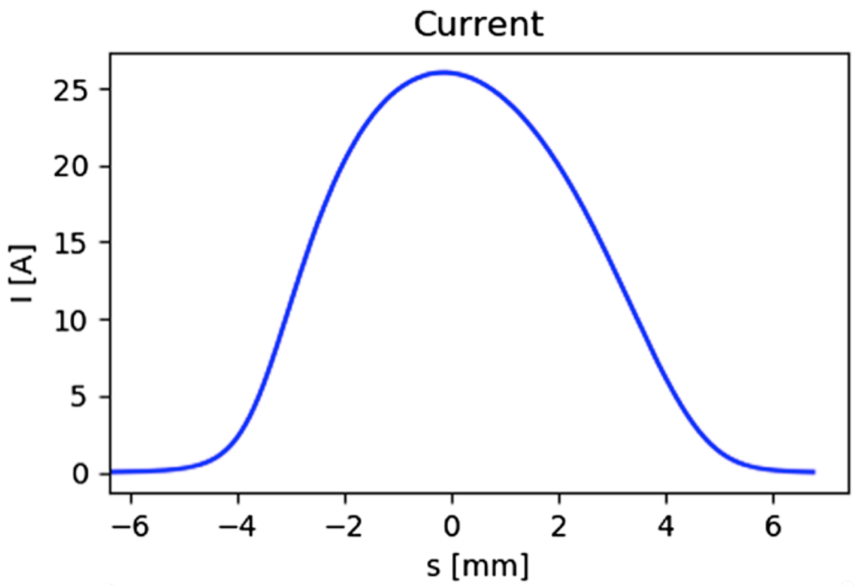


PHYSICAL REVIEW LETTERS **121,** 044801 (2018)

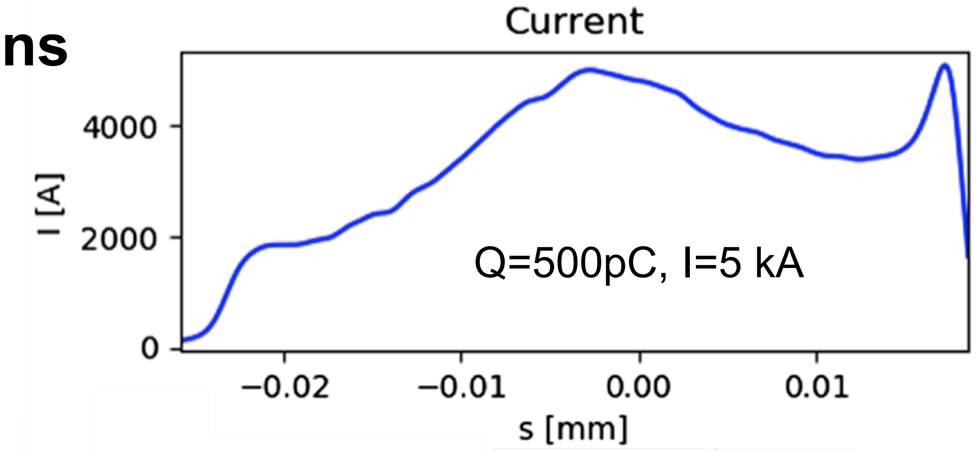
Demonstration of Model-Independent Control of the Longitudinal Phase Space of Electron Beams in the Linac-Coherent Light Source with Femtosecond Resolution

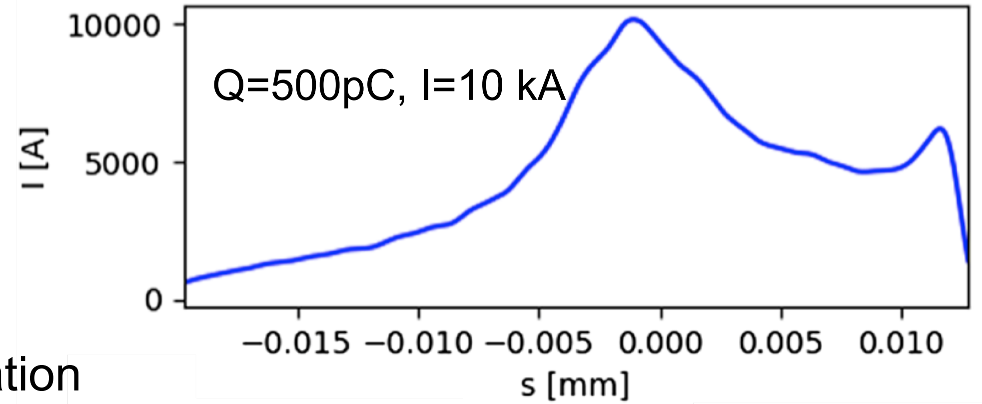
Alexander Scheinker,[1,*] Auralee Edelen,[2] Dorian Bohler,[2] Claudio Emma,[2] and Alberto Lutman[2]
[1]*Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, New Mexico 87545, USA*
[2]*SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA*

Model-dependent compression tuning. Simulations



5 – 10 iterations with CSR, SC, wakes





1. $f_{n-1} = A_x(x_{n-1})$
2. $\Delta f_{n-1} = f_0 - f_{n-1}$
3. $g_n = g_{n-1} + \Delta f_{n-1},$
4. $x_n = A_0^{-1}(g_n)$

$f = A_x(x)$ non-linear transformation of the RF parameters (x) to compression parameters (f)

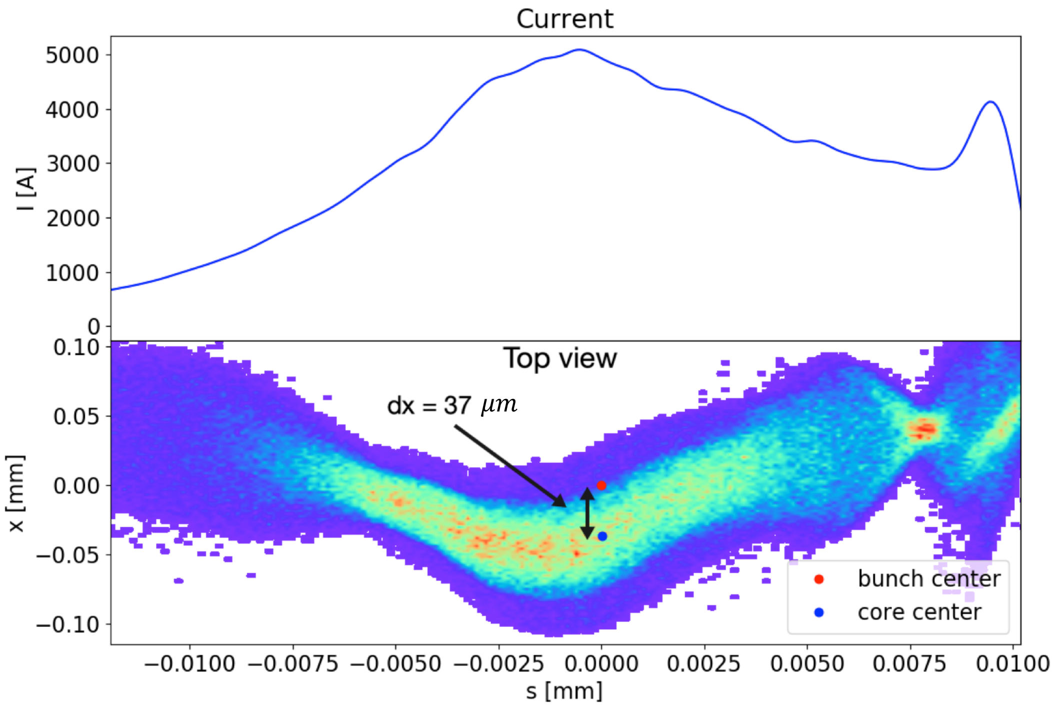
I. Zagorodnov and M. Dohlus, Phys. Rev. ST Accel. Beams 14, 014403 (2011).

I. Zagorodnov, M. Dohlus, and S. Tomin, Phys. Rev. Accel. Beams 22, 024401, 2019.

Adaptive Feedback

- The "Adaptive Feedback” is a statistical optimizer exploiting the orbit jitter and its correlation with a fast FEL intensity signal (shot-to-shot resolution) to optimize the undulator launch orbit
- Correcting the orbit to zero BPM positions does not always mean a straight line for lasing slice

Current and top view of the electron beam (250 pC, 17.5 GeV) in front of SASE2 undulator. Simulation result

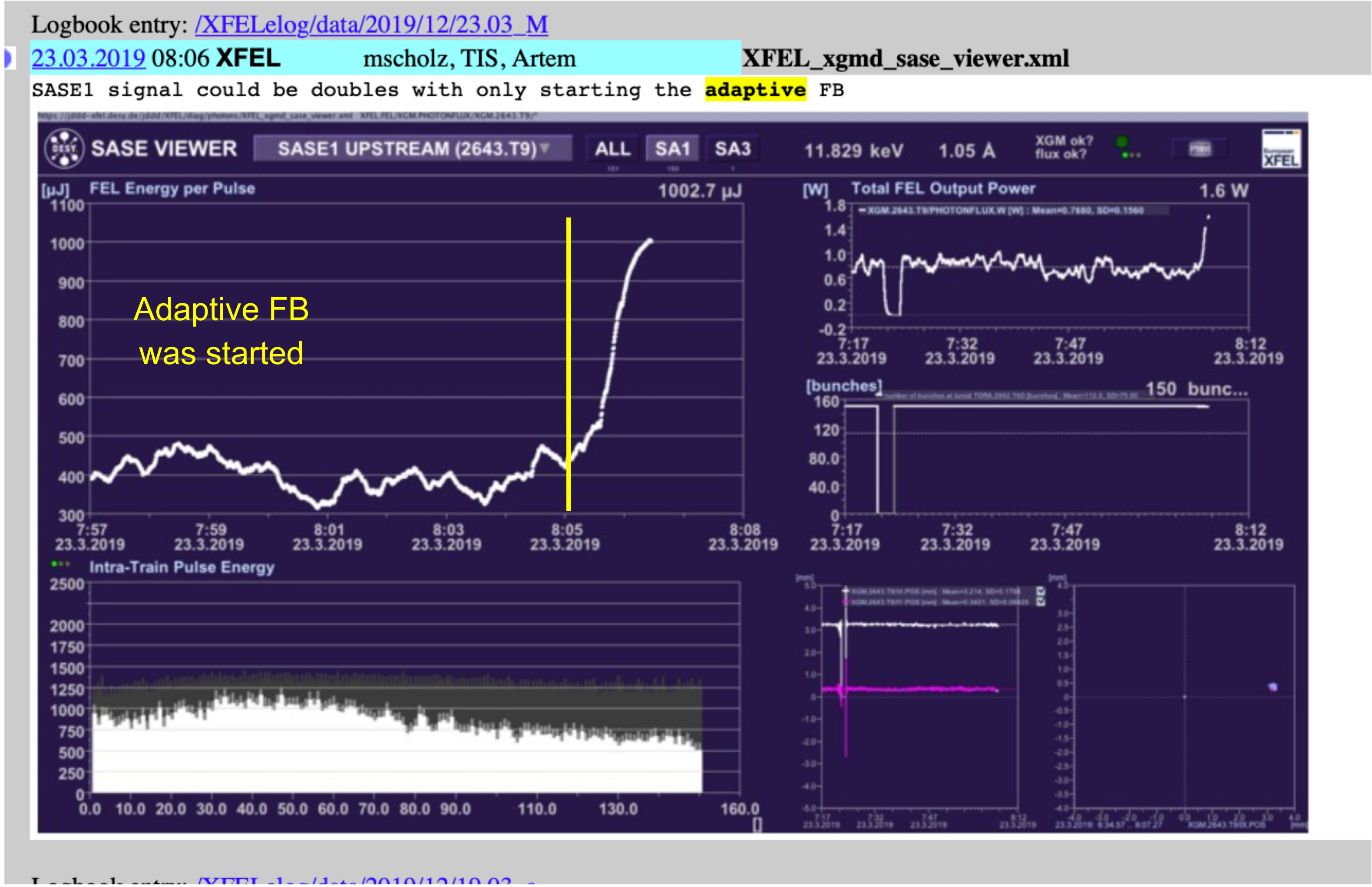


M. Dohlus, S. Tomin, and I.Zagorodnov, “Beam Dynamics at the European XFEL up to SASE4/5”, Workshop „Shaping the Future of the European XFEL: Options for the SASE4/5 Tunnels“

G. Gaio, M. Lonza, Automatic FEL Optimization at FERMI, Proc. of ICALEPCS2015

Adaptive Feedback

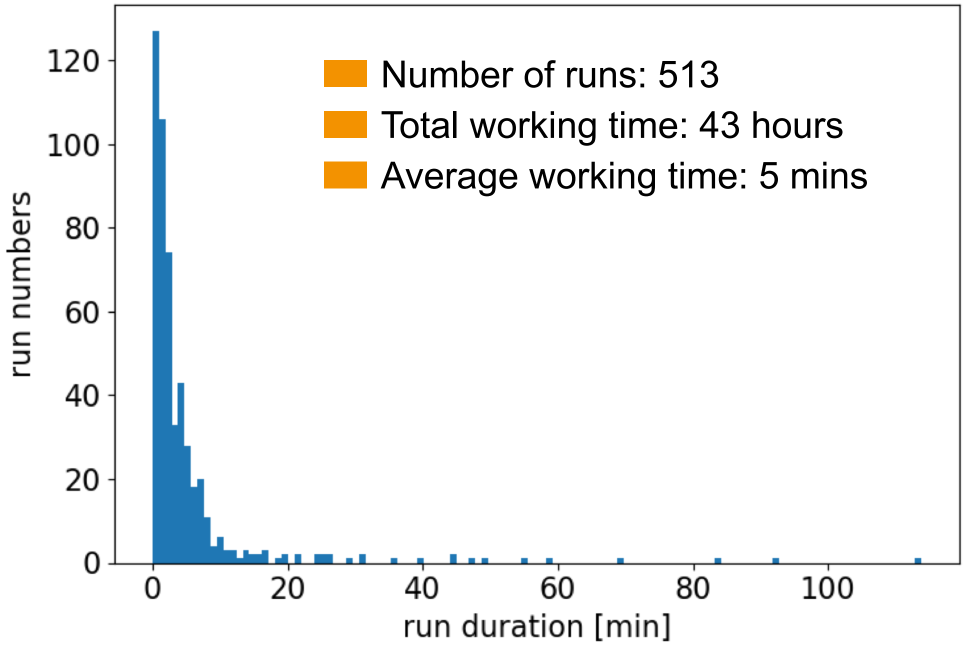
Adaptive Feedback GUI



Adaptive Feedback statistics

- Adaptive Feedback has become one of the main tools for SASE tuning
- In some cases the adaptive feedback is used as an orbit feedback
- The soft X-Ray FEL pulse energy signal is not sensitive to the orbit jitter in the SASE3 undulator.
  - artificially induced orbit changes have to be used to catch correlations
- Active search?

Statistic of the Adaptive Feedback runs from March 11 to April 7, 2019

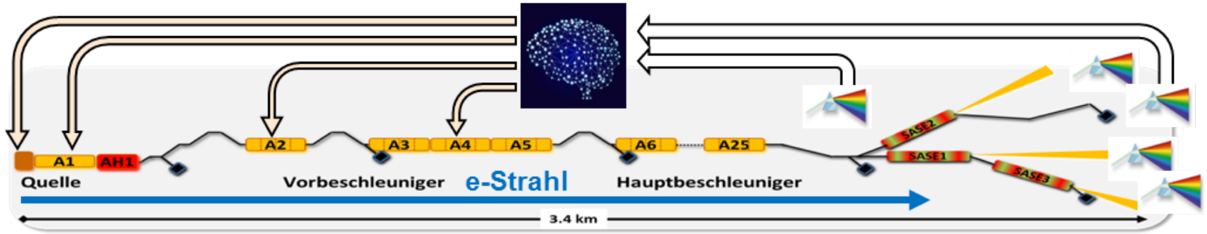


XFEL simulations and optimizations. Summary

- OCELOT is part of DESY/EuXFEL toolkit for beam dynamics and FEL simulations
- Model-free optimization methods were widely used during commissioning of the European XFEL and now it is a part of the daily European XFEL operation
- We can apply more automation such as a sequence of optimization without operator intervention. In the future, ML methods will be used to define the sequence
- A model-based method such as the Adaptive orbit feedback proves extremely useful and more advantageous compared to purely empirical methods.

Applications of Machine Learning in MSK

- Real-time control of photon parameters
  - Train neural network at the to control of photon parameter , e.g., pulse energy, center wavelength, pointing stability, pulse duration and energy chirp
  - System-on-the-chip implementation on the existing hardware architecture



- Quench detection in the RF system:
  - Hybrid anomaly detection based on a model-based nonlinear parity space approach followed by a classification step using support vector machines
  - Factor graph approach to stochastic model-based fault diagnosis: residual generation by a Kalman filter, residual classification by a Gaussian mixture model
  - → Next steps:
    - → Online implementation of developed algorithms
    - → Model extension to further influencing subsystems (synchronization system, cryogenics, vacuum, klystron)

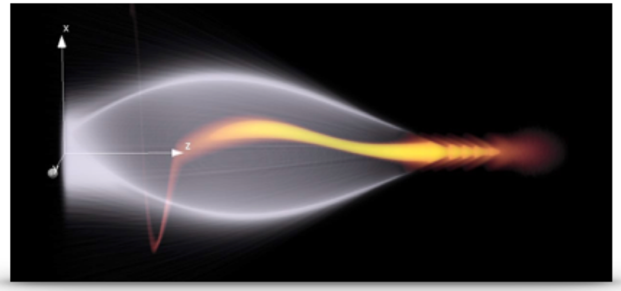
- Predictive maintenance for the laser-based synchronization system:
  - Health- and performance monitoring of components
  - Fault detection and identification for predictive maintenance
  - → First steps:
    - → Data Mining: selection of appropriate data channels (long-time archiving), dimensionality reduction, analysis

Courtesy of Annika Eichler

Plasma accelerator simulations: Accurate and efficient simulations codes

HiPACE

a Highly efficient plasma accelerator emulation



3D quasi-static Particle-In-Cell code

- Allows ~100x speedup for highly-relativistic beams.
- Dynamic time-step adjustment.

Spectral electromagnetic solver

- High performance FFT-solver.
- No numerical Cherenkov radiation.

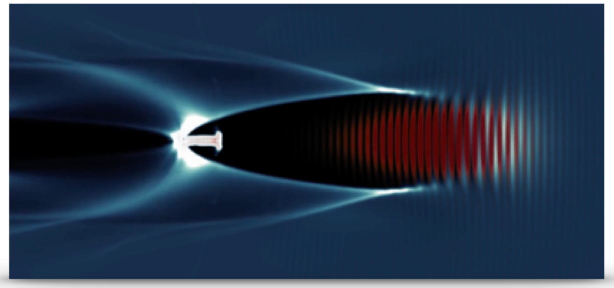
Highly parallel & written in C

- Fully parallelized and well scalable.
- Suitable for high-performance CPU clusters.
- Interfaces with other PIC and particle tracking codes.

T. J. Mehrling et al., Plasma Phys. Control. Fusion 56, 084012 (2014).

FBPIC

Open source electromagnetic Particle-In-Cell code



Quasi-3D geometry

- 3D accuracy at 2D costs.

Spectral electromagnetic solver

- Highest precision
- No numerical artifacts

Boosted frame

- Accelerate simulations by orders of magnitude
- Instability-free solver

Highly parallel & written in Python

- CUDA GPU acceleration
- Multi-GPU domain decomposition

R. Lehe et al., Comput. Phys. Commun. 203, 66-82 (2016).

Courtesy of Alberto M. de la Ossa

thank you for your attention!