ML OPTIMIZATION METHODS FOR APS-U COMMISSIONING

NIKITA KUKLEV (FNAL/ANL)

LOUIS EMERY (ANL)
HAIRONG SHANG (ANL)
MICHAEL BORLAND (ANL)
YINE SUN (ANL)
YOSHITERU HIDAKA (NSLSII, BNL)
GUIMEI WANG (NSLSII, BNL)

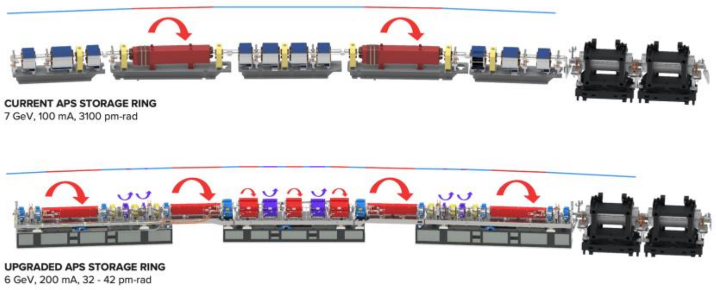
U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY

OPTIMIZATION @ APS-U

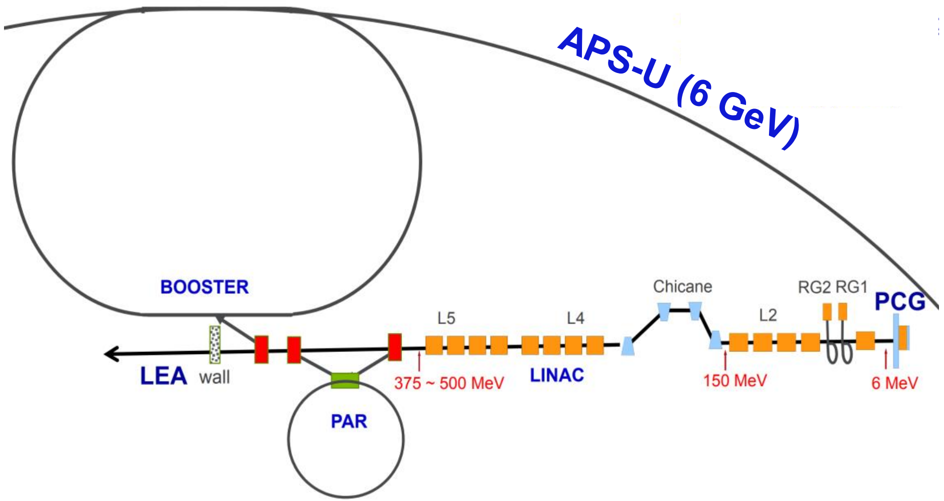
APS just finished a major upgrade (APS-U)

- Extremely compressed timeline
- New storage ring
- New diagnostics and injection scheme



Very successful overall

…but some painful lessons



(NON-PARAMETRIC) ACCELERATOR OPTIMIZATION

Many methods:

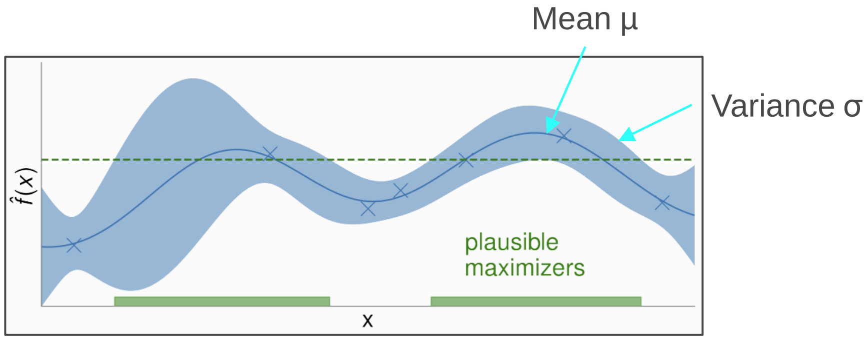
- **Humans**: Scientists, Engineers, Operators
- **Classic methods**: Nelder-Mead Simplex, RCDS, Extremum Seeking
- **Artificial Intelligence**: Genetic Algorithm, Particle Swarm
  - **Machine Learning**: Bayesian Optimization
    - Our preference for '**one-shot**' or '**sampling-efficiency-critical**' applications
    - **Deep Learning**: Reinforcement Learning, Deep Kernel Learning
      - Also a good choice, especially for stabilization

BAYESIAN OPTIMIZATION LOOP

INPUTS $\mathbf{x}$

Set new inputs

MACHINE BLACK BOX



$\mathbf{y} = f(\mathbf{x}) + \varepsilon$

Measure Outputs

BO predicts the most promising places to look, taking into account the uncertainty and expected mean value of GP model

Make next guess

$$\text{UCB}(\mathbf{x}) = \mu(\mathbf{x}) + \sqrt{\beta} * \sigma(\mathbf{x})$$

MAXIMIZE ACQUISITION FUNC.

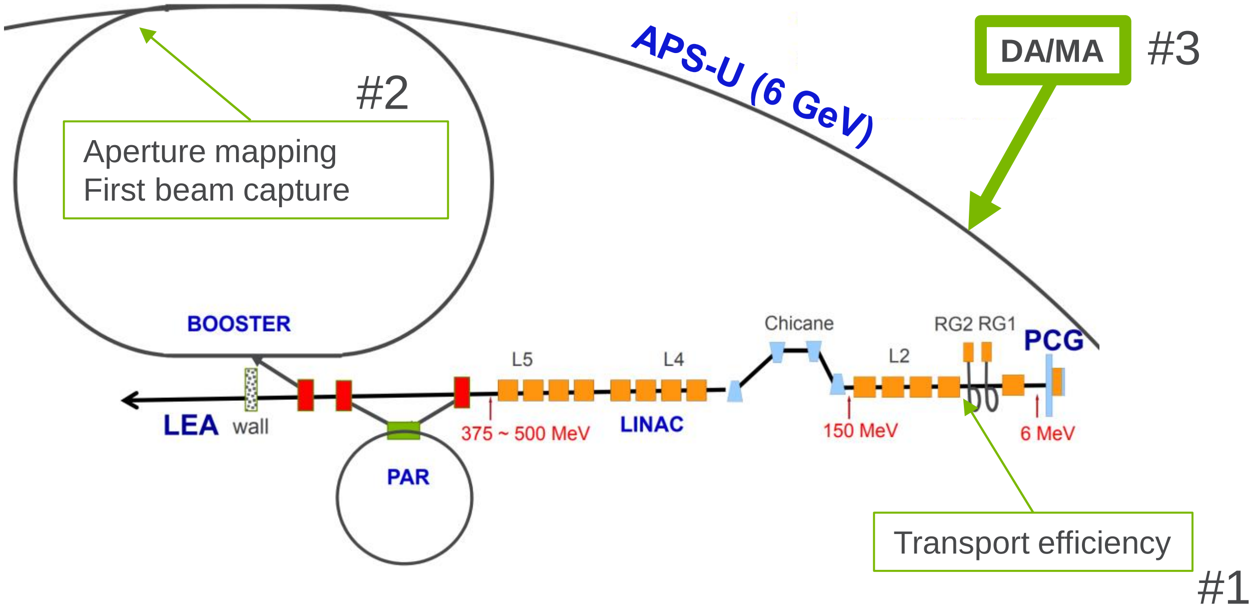
$$\text{p}(\mathbf{f} \mid \mathbf{y}, \mathbf{x})$$

Fit model to data

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

GAUSSIAN PROCESS (GP) MODEL

APS-U OPTIMIZATION EXAMPLES



EXAMPLE 1: LINAC GUN TUNING

Linac efficiency tune-up is a standard task

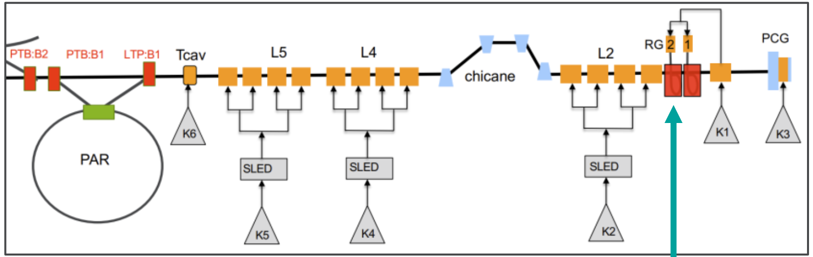
Want:

- Low beam losses (high current)
- High injection efficiency into PAR

Knobs:

- Quadrupoles
- Correctors

(start at source, move down the beamline when possible)



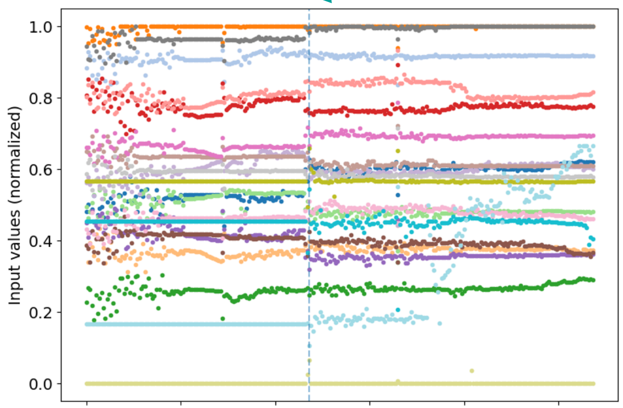
EXP

EXAMPLE 1: LINAC GUN TUNING

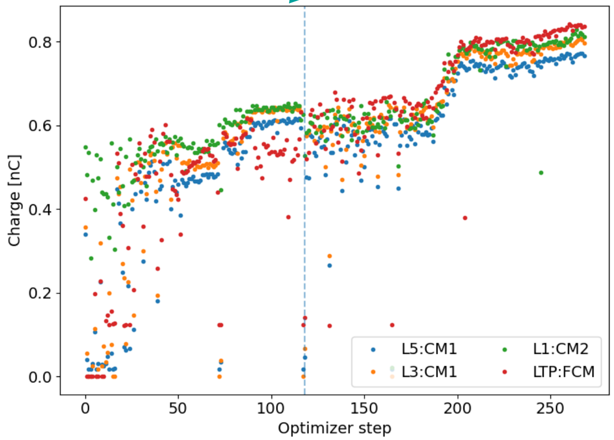
BO reliably manages to find a solution even with 20+ parameters

(not a perfect cure – can't fix broken equipment, which is often source of poor efficiency)

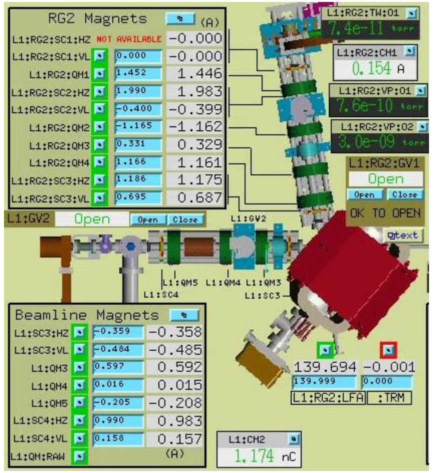
Dynamic variable and objective addition!



INPUTS

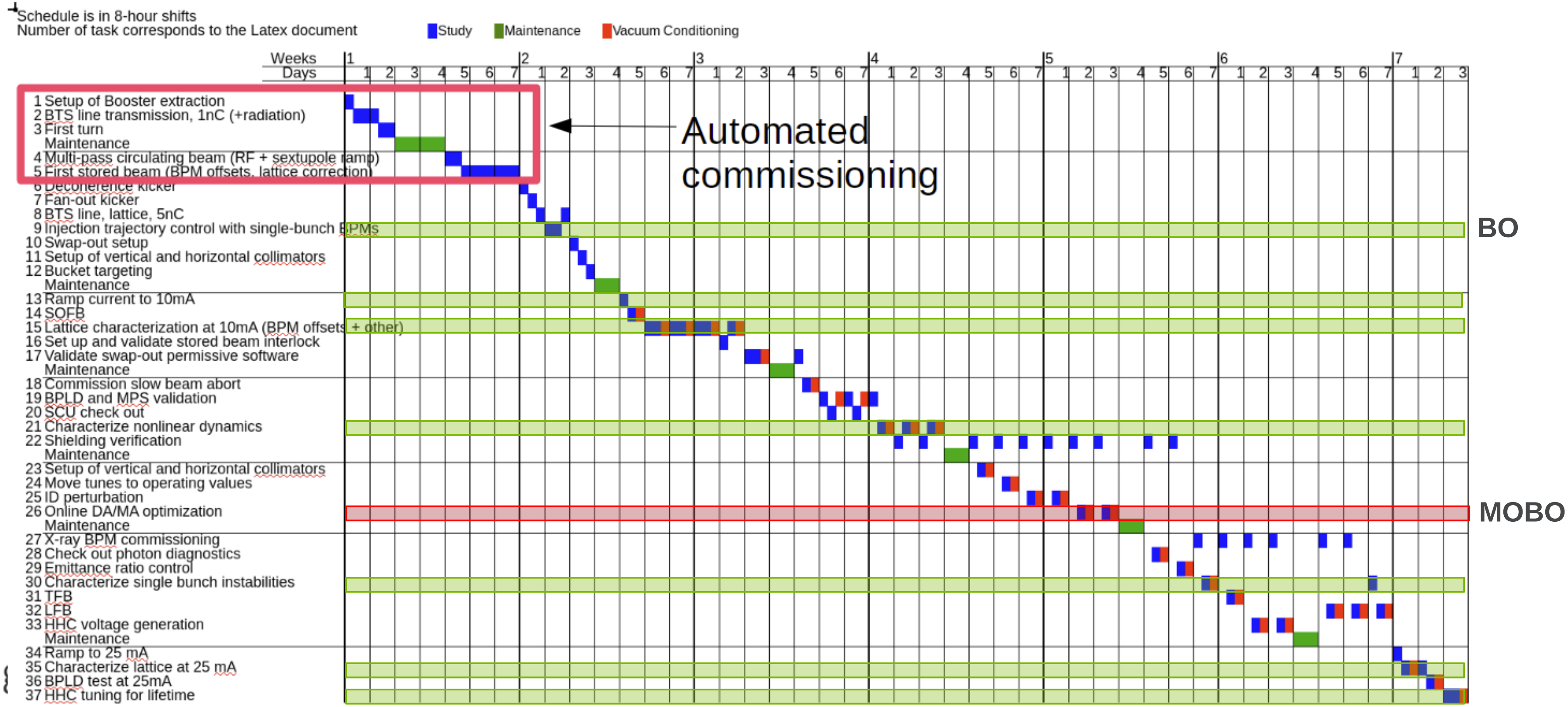


OUTPUTS



EXP

APS-U COMMISSIONING PLAN



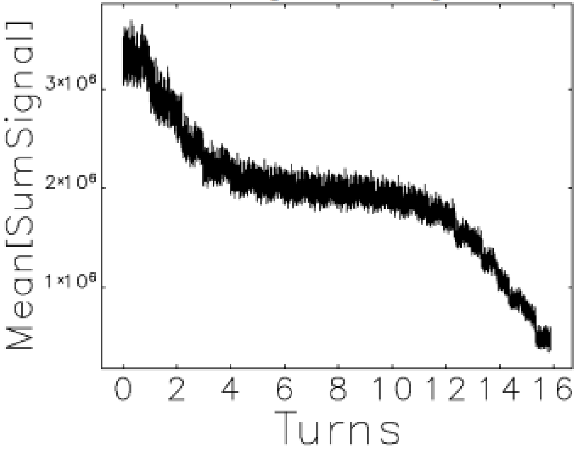
U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

APS-U COMMISSIONING REALITY

First days of APS-U commissioning were hard

- Reversed magnet polarity in transfer line (history of positrons)
- First-turn threading corrected trajectory well but could not result in second turn (simulations predict 5-10 turns)
- Large orbit bumps around septum (> 2 mm) allowed to get ~5 turns
- Tweak fest ensued, resulted in 15 turns

BPM sum signal showing 16 turns

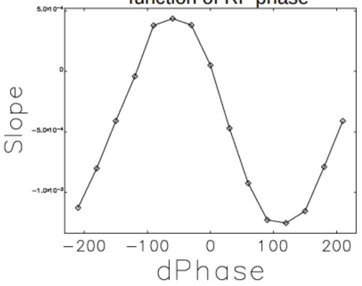


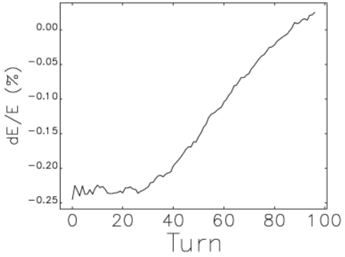
APS-U COMMISSIONING REALITY

Wild optimization found condition with”almost” stored beam

- Turned on RF, scanned and set rf phase – resulted in 80 turns
- Turned on sextupoles
- Bayesian optimization of everything resulted in 40 μA with lifetime of 4 sec
  - BTS exit trajectory and quads
  - Position and angle closed orbit bump at SR septum
  - SR Tunes
  - The optimization still wanted large orbit at septum
- Tunes are 1 unit off, and lifetime is too short and could not be increased

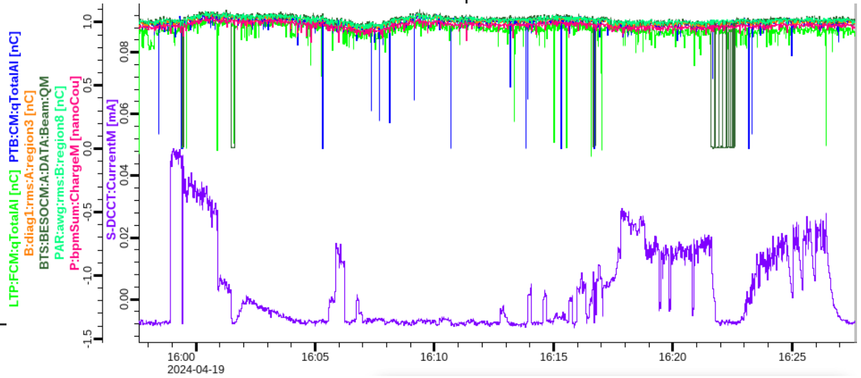
Turn-by-turn dE/E slope as a function of RF phase





Beam energy measurement using trajectory works well

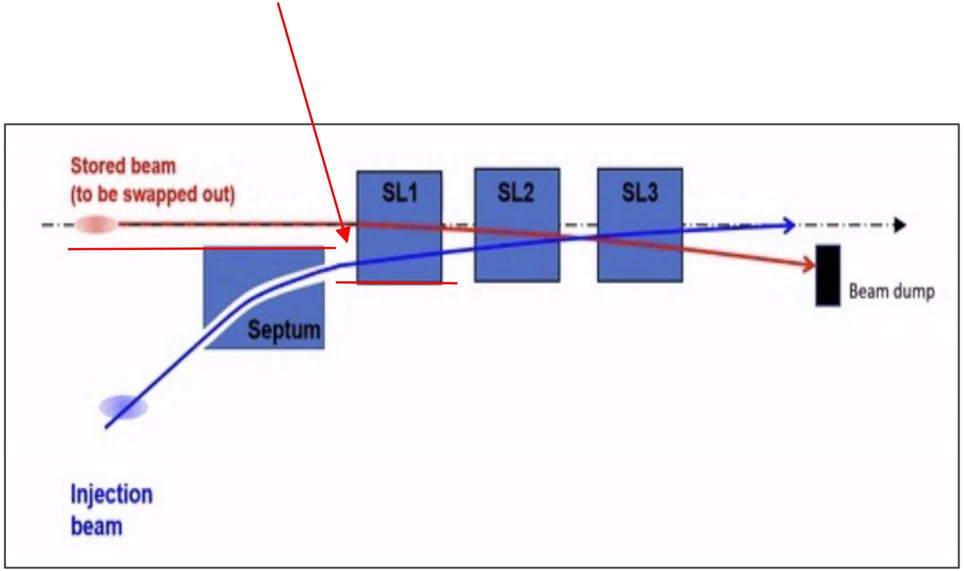
…later reconditioning and running auto commissioning slowly got beam stored



V. Sajaev

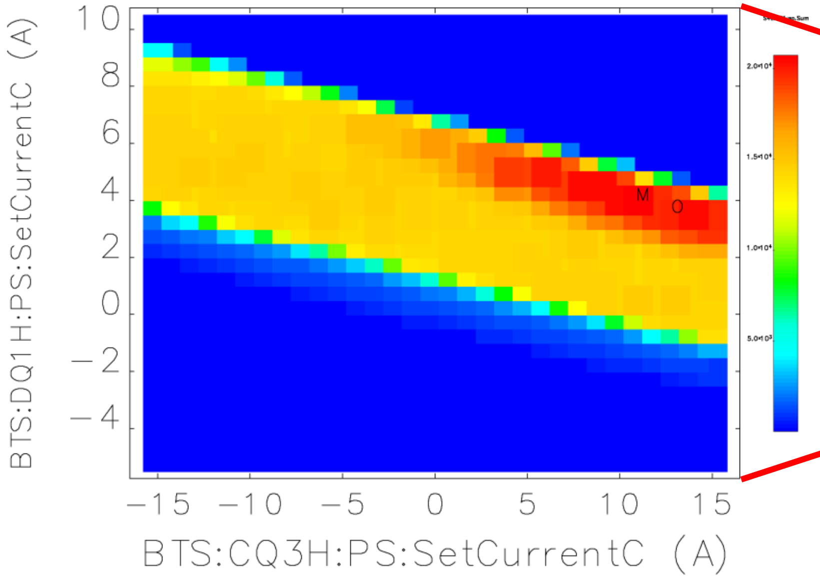
EXAMPLE 2: BAYESIAN EXPLORATION OF INJECTION APERTURES

Problem: apertures of injected beam and stored beam did not agree with model…a lot

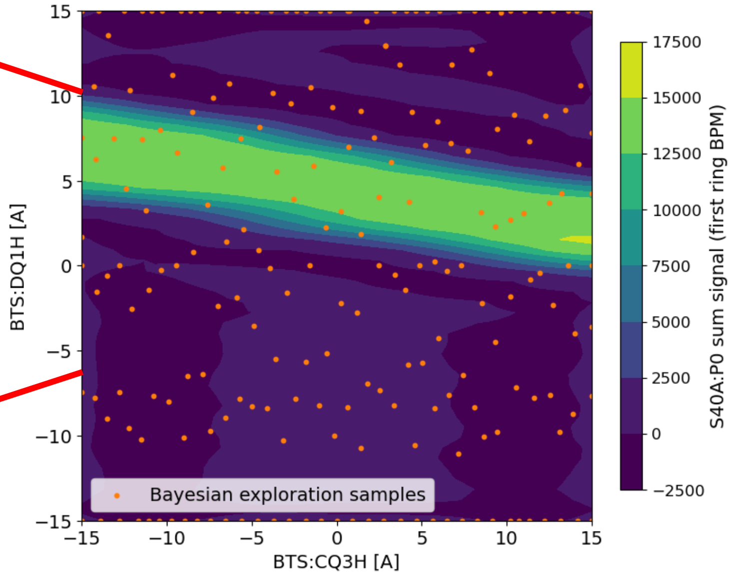


EXAMPLE 2: BAYESIAN EXPLORATION OF INJECTION APERTURES

- Used grid scans to map apertures
- Demonstrated that same data can be inferred quickly with Bayesian exploration



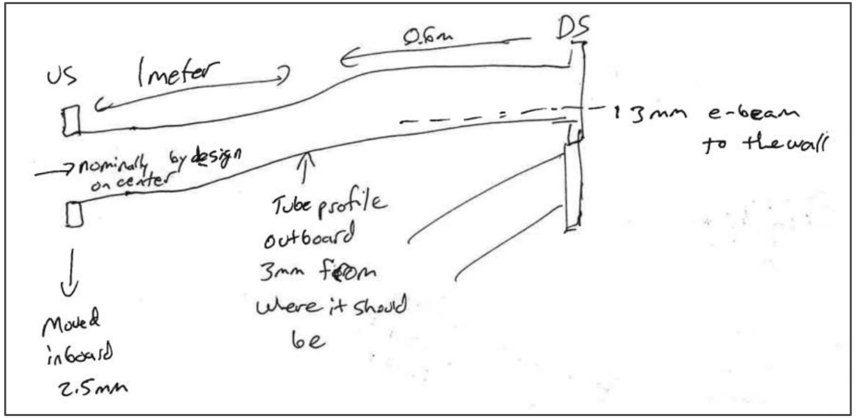
32x32 = 1024 points grid scan



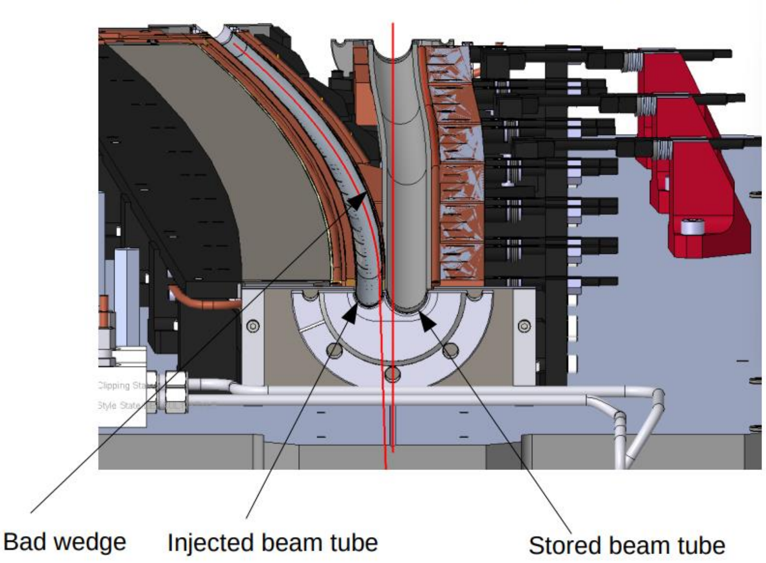
~150 samples using BE

EXAMPLE 2: BAYESIAN EXPLORATION OF INJECTION APERTURES

Cause: septum was 2mm off and bent…moved until proper fix



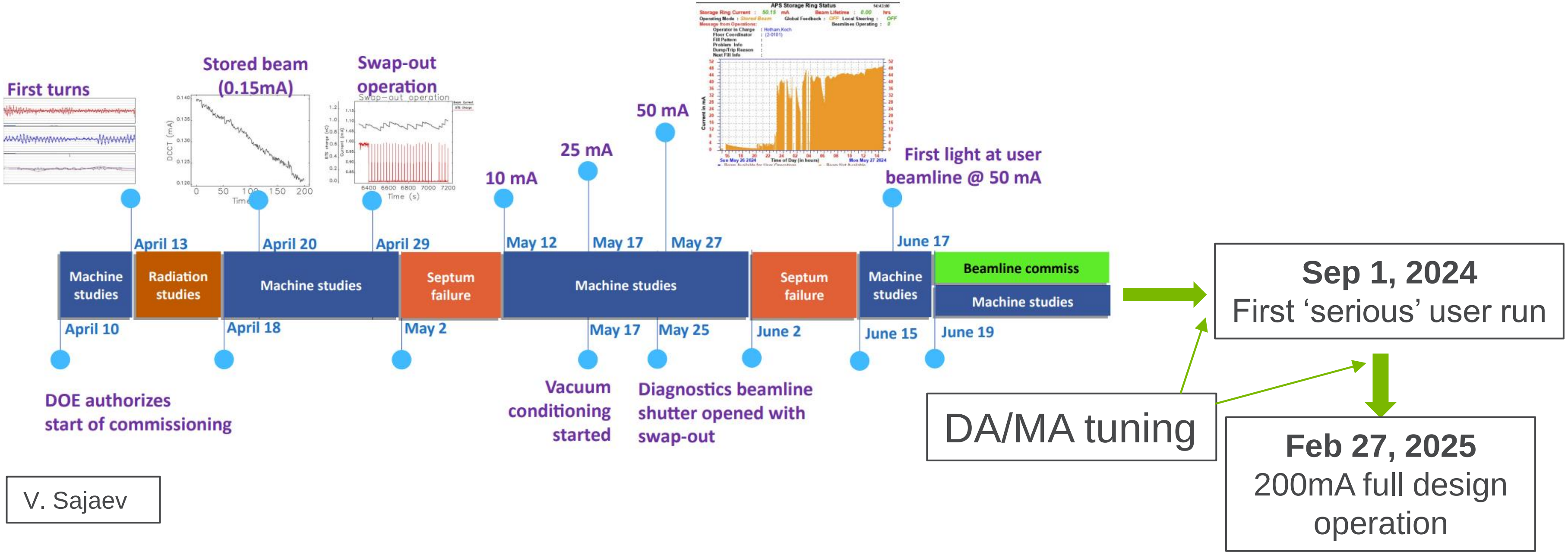
Horizontal cross-section of the septum magnet



APS-U COMMISSIONING REALITY CONTINUED

After beam was established, commissioning proceeded well but had many hardware issues

- Sporadic use of ML tools as required, especially for injection trajectory efficiency



V. Sajaev

LANDSCAPE OF 'ADVANCED' ML

Modern BO optimization and modelling tools are **turnkey**

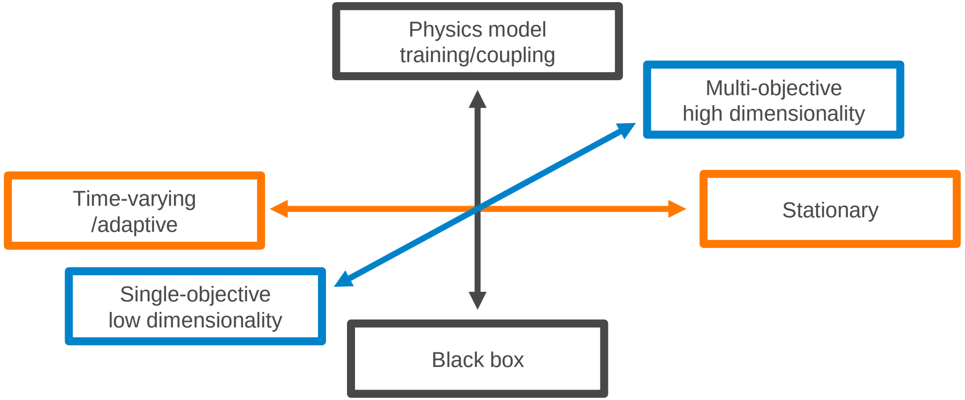
Near guaranteed success for:

- Single-objective smooth function
- <=~40 parameters
- Time-invariant / no hysteresis

"some ML code" option: use BoTorch

"no ML code" option: use Xopt (in our case APSopt)

Where are the future improvements?



LANDSCAPE OF 'ADVANCED' ML

What is "advanced" ML?

Problem/algorithm complexity

- High dimensionality
- Multi-objective
- Decoupled sampling
- Strict safety constraints
- Reformulations like BAX

Physics integration

- Prior mean
- Hyperparameter/kernel pretraining
- Simulation-guided optimization

Experimental execution

- Fast measurements
- Control of sys/stat noise
- Device calibration and compensation

EXAMPLE 3: DA/MA TUNING

No single best LMA/DA solution:

- Objectives fight each other, beamline constraints, etc.

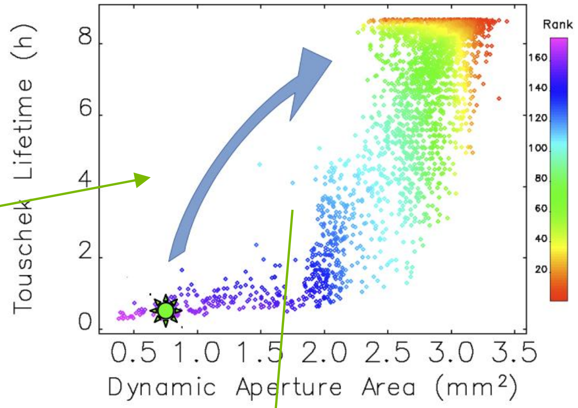
Real machine will be one of random 'seeds'

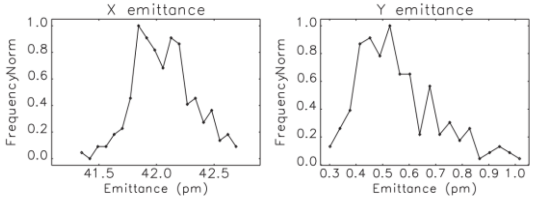
- Need to adjust to get best performance

How to tune?

Require **multi-objective BO** tuning to find best tradeoffs

- Harder than single objective, need **lots** of beam time
- Online use demonstrated previously with MOGA, MGGPO
- APS-U **too complex** to be handled by these methods



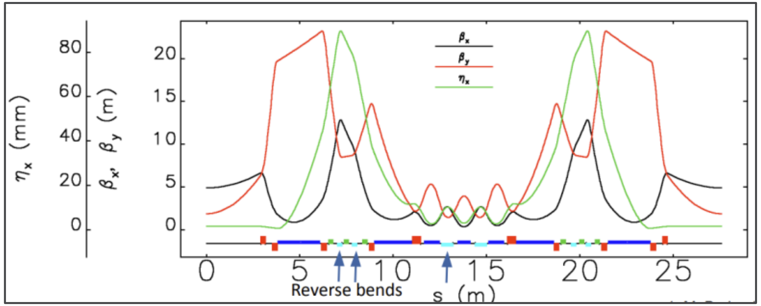


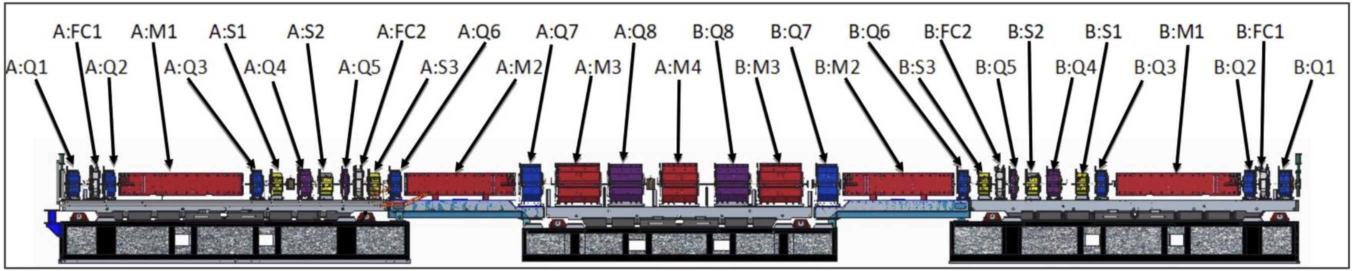
M. Borland et al., HEP GARD ABP: Workshop #2, WG 4 (2020)
M. Borland et al., J. Synchrotron Rad. 21 (2014).
Y. Sun, NAPAC2016 (WEPOB15)
Y. Sun, NAPAC2016 (WEPOB12)
V. Sajaev, PRAB 22 (2019)

[1] https://indico.cern.ch/event/1326603/contributions/5779742/
[2] N. Kuklev at al, ML-enhanced commissioning of the APS-U accelerator complex, https://doi.org/10.18429/JACoW-IPAC2024-TUPS50

DA/MA ML: APS-U REQUIREMENTS

- APS-U lattice is more sensitive than 3rd gen light sources
  - Complex interplay of nonlinear effects
  - **12D-24D =** 12 sextupoles per double sector + 12 in S39/S40 after injection straight



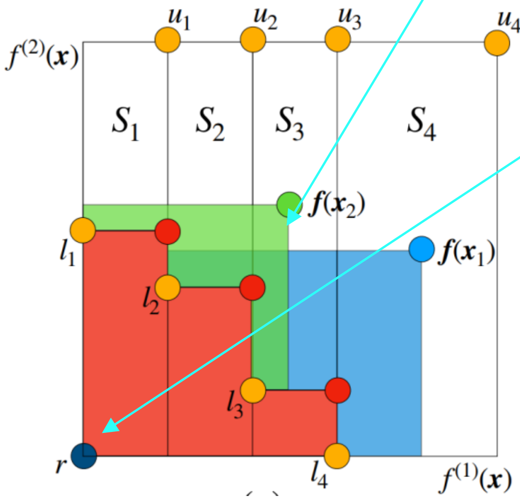


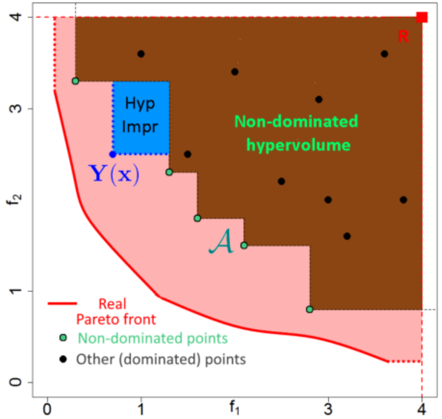
[1] arXiv:2006.05078
[2] arXiv:2105.08195
[3] PhysRevAccelBeams.27.084801

DA/MA ML: MOBO ALGORITHM

BoTorch

- GP models are fitted to each objective independently
- Acquisition function seeks to maximize **hypervolume** relative to **reference point**





https://arxiv.org/pdf/1809.10482.pdf

- Recent progress in ML frameworks has made this computation **tractable** for online optimization in high dimensional spaces (~seconds) [1,2]

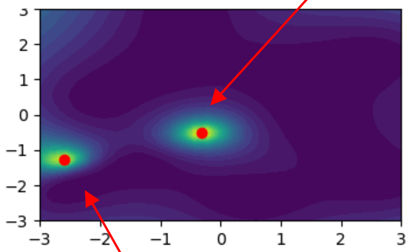
DA/MA ML: RECENT MOBO IMPROVEMENTS

Prior mean: simulation data can be incorporated as a bias

- Using model variance smoothly blends online and offline data
- Promising priors explored first, online data can ‘override’ locally

New data overrides prior in nearby area

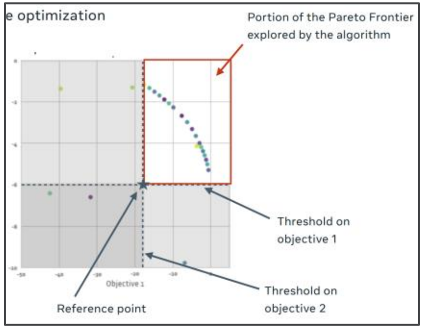
Online GP model impact

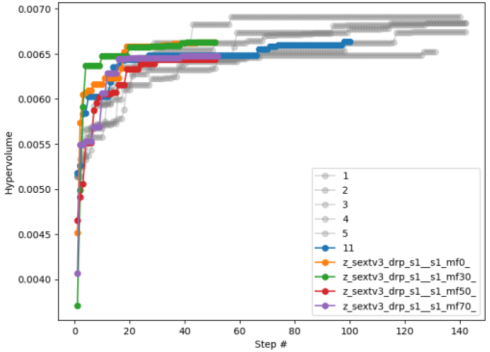


Prior keeps model ‘grounded’ away from new data

Dynamic reference point: dynamic ROI to guide optimizer

- Stabilizes startup, then focuses on region of interest (i.e lifetime >6hrs)



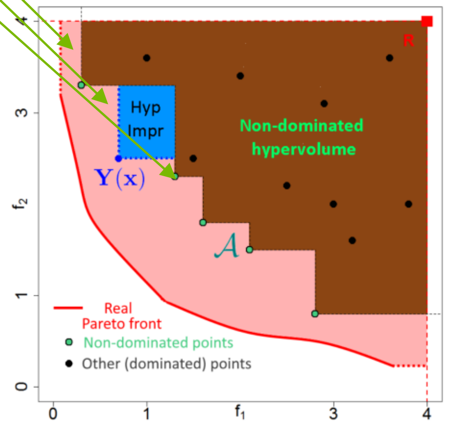


https://indico.bnl.gov/event/16586/contributions/68649/attachments/43703/73634/2021010%20-%20AI4EIC%20workshop%20Ax%20BoTorch%20MOO%20tutorial.pdf

DA/MA ML: RECENT MOBO IMPROVEMENTS

Model-guided numerical refinement: faster and better computation of next point

- Acquisition function optimization starts from finite number of seeds
- Create them as random perturbations of best (pareto front) points within region of interest



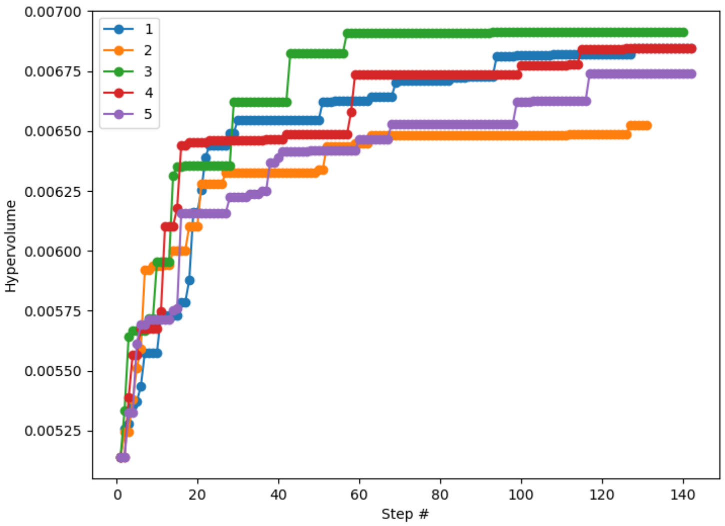
Distributed acquisition function search

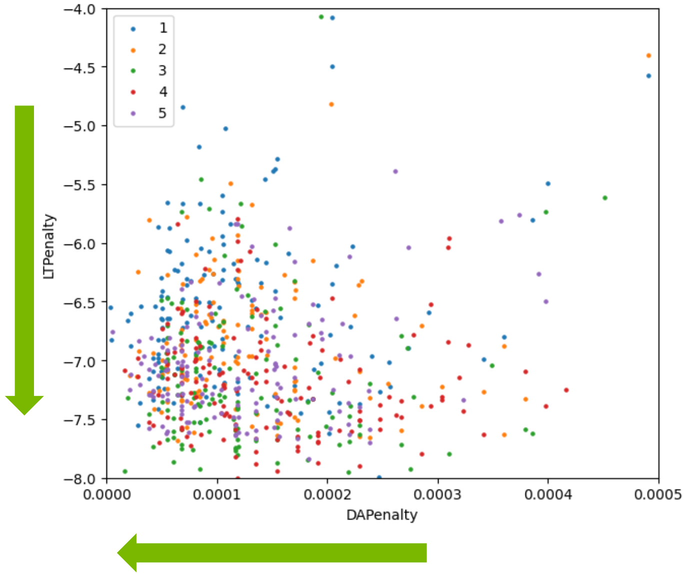
- Improve wall-clock execution time
- Ray for easy job management
- Balances slow downs as number of points grows

DA/MA ML: APS-U SIMULATIONS

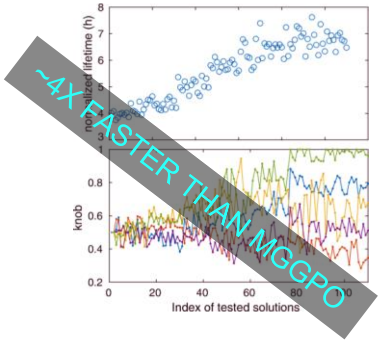
Extensive simulations done in ELEGANT

- Automated optimizer loop on HPC cluster - 5 minutes/point
- Repeatable convergence at expected resolution/noise (see next section)

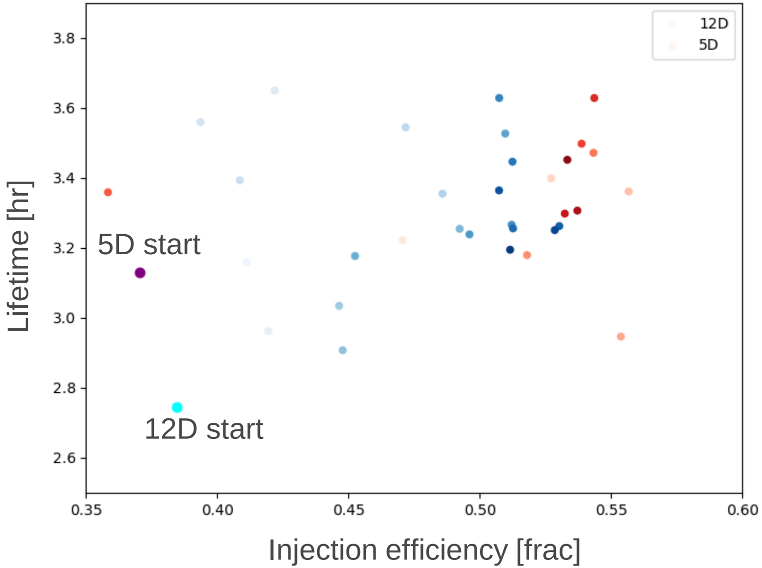




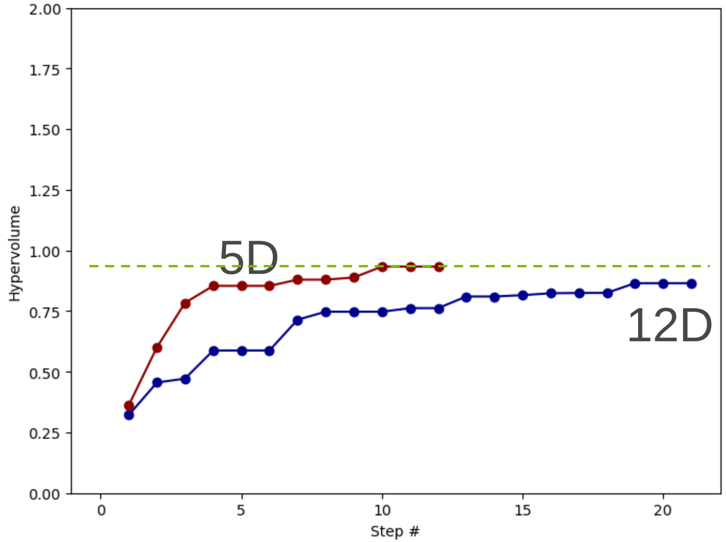
DA/MA ML: EXPERIMENTAL RESULTS IN (OLD) APS



- Test: sextupole optimization in **5D and 12D** (chromaticity null spaces)
  - Expect to recover nominal performance
  - Data consistent, find optimum much faster than previous methods



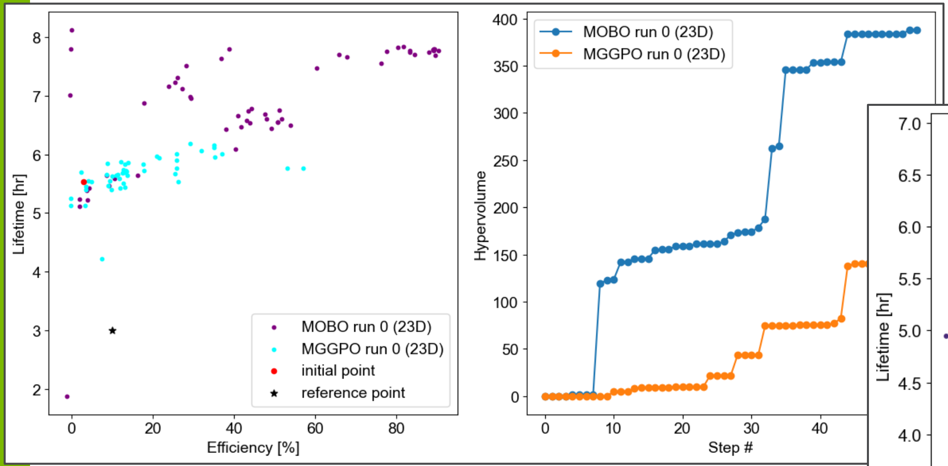
EXP [APS][MAR 2023]



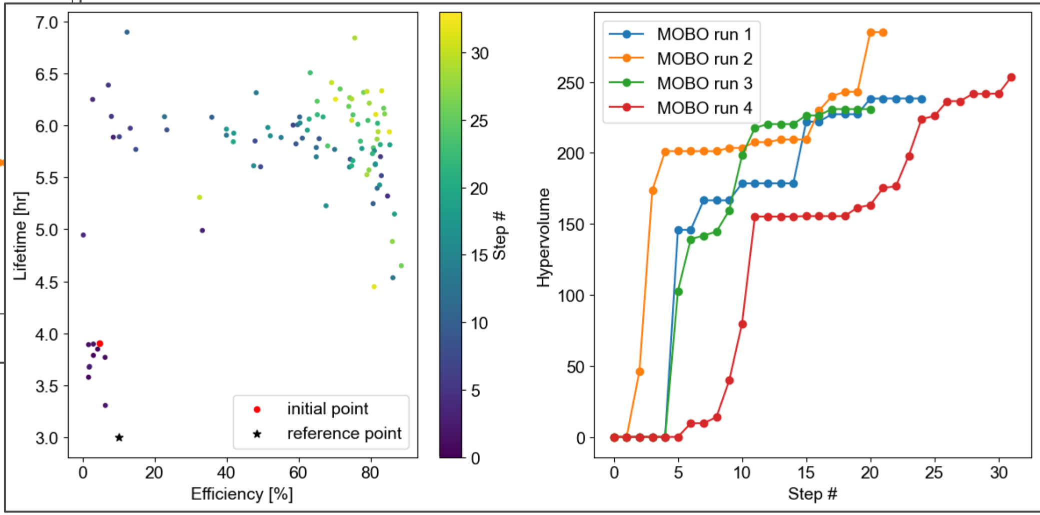
EXP [APS][MAR 2023]

DA/MA ML: EXPERIMENTAL RESULTS IN NSLSII

- Confirmed good performance (improved on NSLSII default!) and repeatable convergence



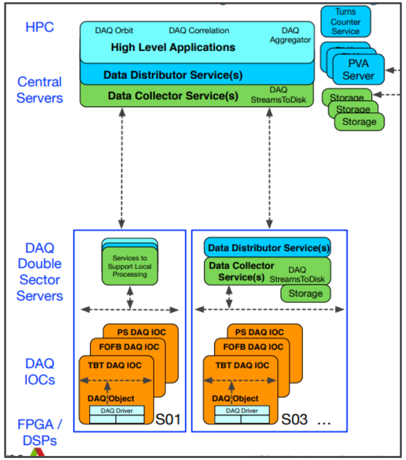
DA/MA in 23D, significant improvement over MGGPO



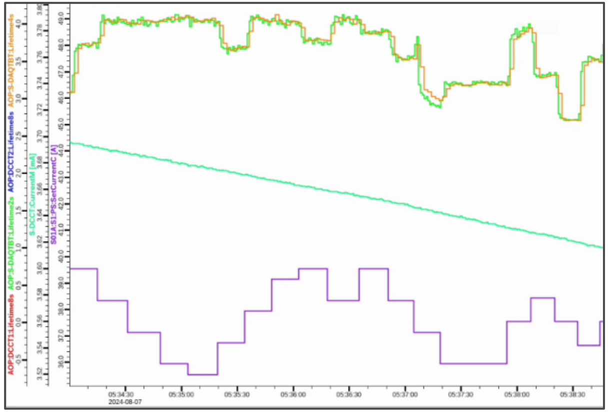
Multi-objective DA/MA in 7D x 5 runs

DA/MA METHODS: MEASURING LIFETIME

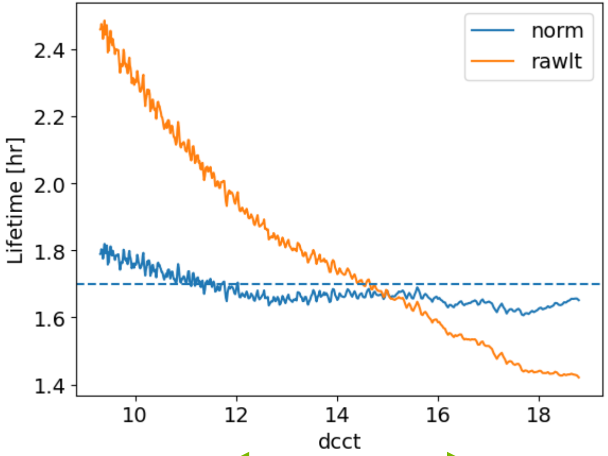
- At low currents, implemented analysis using BPM sum data
  - DAQ streamed 540 TBT sum signals (**0.6GB/s**)
- At high currents, use 10Hz DCCT data (adaptive stopping)
- Normalization:
  - $I^{2/3} = I^{1}$ from Touschek x $I^{-1/3}$ from potential well distortion
  - Experimental calibration revealed other effects



DAQ system PVA readout



Response (top) to sextupole changes (bottom)

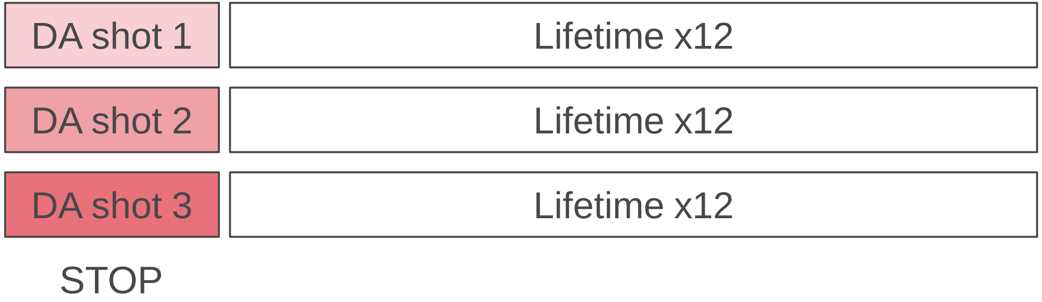


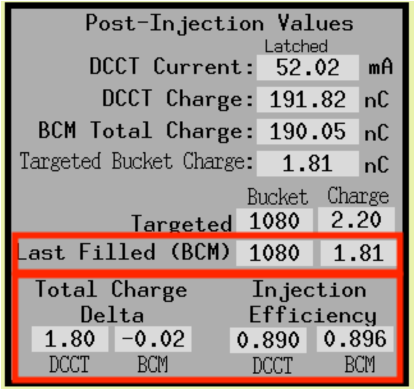
Valid range

DA/MA METHODS: MEASURING DA

DA is approximated by injection efficiency on spoiled trajectory

- Unlike APS, in APS-U can swap-out single bunch (low RF disturbance)
- Use latched measurements to get correct efficiency even at 1Hz injection
- Adaptive stop criterion (3-6 shots) depending on signal-to-noise

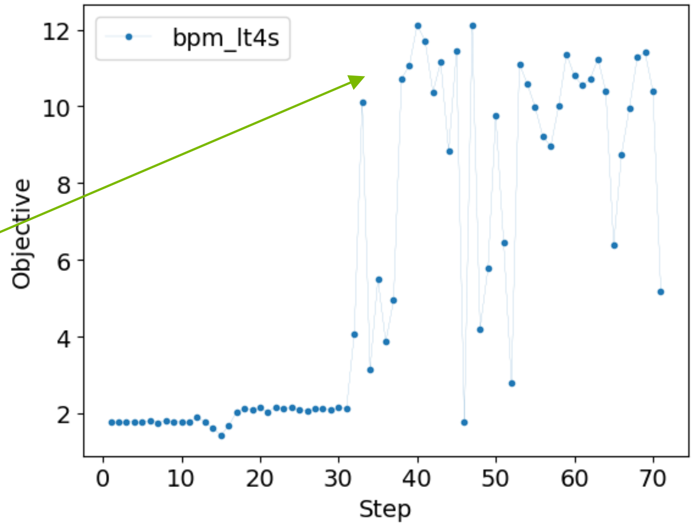




DA/MA: EXPERIMENTAL SETUP

Started nonlinear tuning in August.

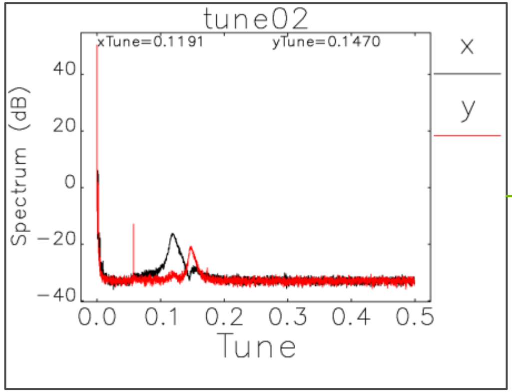
- Updated simulations showed that optimal chromaticity varies a lot (depending on seed)
- Switched from null-space (constant chromaticity) knobs to direct family strengths
- As with APS, optimizer quickly exploited instabilities due to negative chromaticity

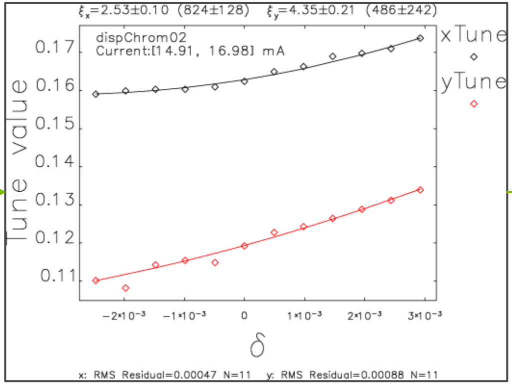


DA/MA: EXPERIMENTAL SETUP

- New solution: use a live simulation model with adjusted lattice
  - Automated measurement and calibration before each study

- “full fidelity” digital twin





Magnet snapshot

Magnet gradients

Adjusted lattice

ELEGANT server process

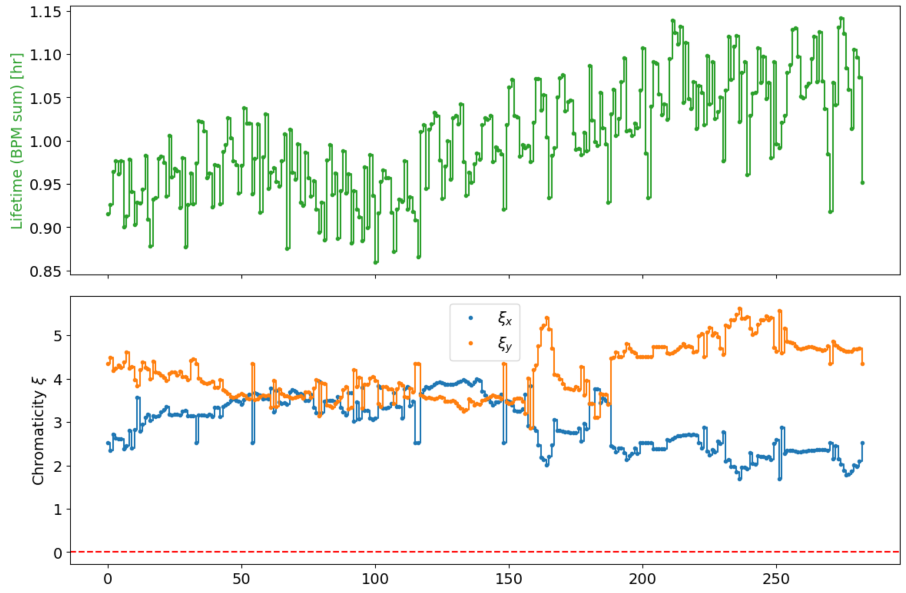
Predicted tune/chromaticity

DA/MA: EXPERIMENTAL RESULTS

For ‘production’ tuning, stayed in 12D.

Because of good alignment, had good efficiency at design sextupole values

	Requirements	As installed
Element to element	30 μm rms	15 μm rms
Module to module	+-100 μm	X: 50 μm rms Y: 35 μm rms



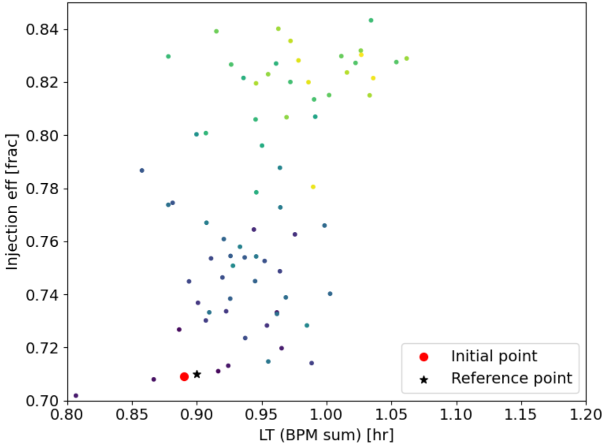
Still, quickly got +15% in lifetime in single-objective run

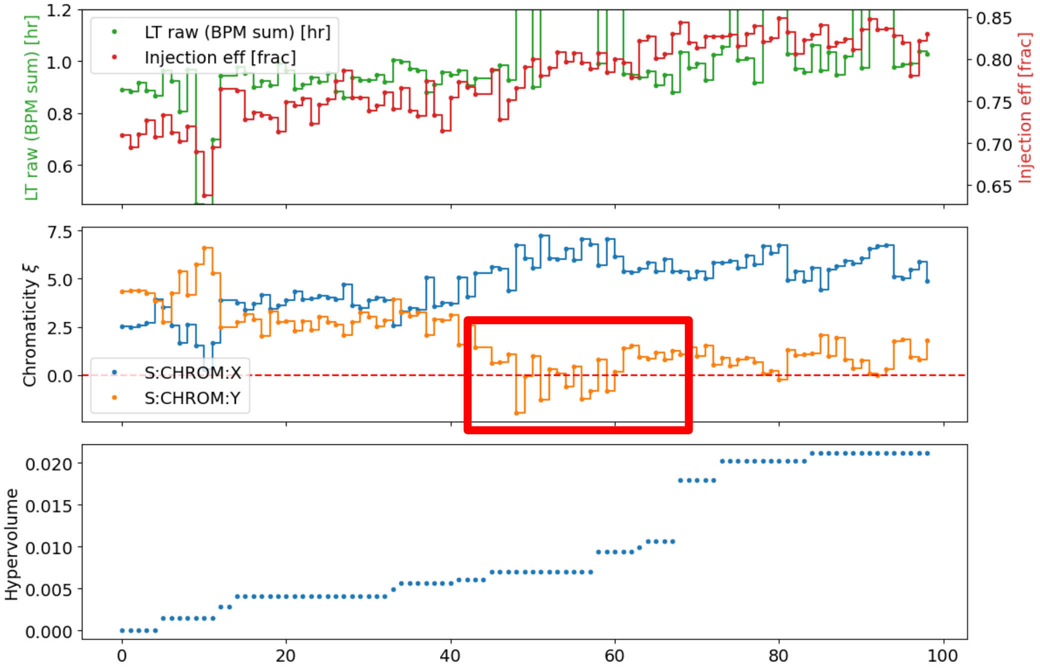
Chromaticity stayed reasonable, did not need constraints

DA/MA: EXPERIMENTAL RESULTS

MOBO performed extremely well too, but quickly found the unstable regions

- Using live model, harshly penalized any –ve chromaticity points
- Obtained same lifetime as previous run AND +12% inj. efficiency





DA/MA: EXPERIMENTAL RESULTS

- To improve sampling speed, integrated live model into the optimizer ('decoupled MOBO')
  - Optimizer proposed point
  - Live model predicts linear optics
  - If violate constraints objective collection skipped
  - Constraint model is updated

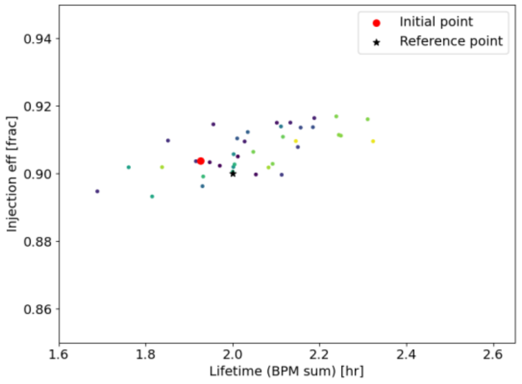
```
Model predicts bad point optics={'nux': 95.1132624825597, 'nuy': 36.1508340446097, 'xix': -2.802417749567054, 'xiy': 6.026185919403116}
```

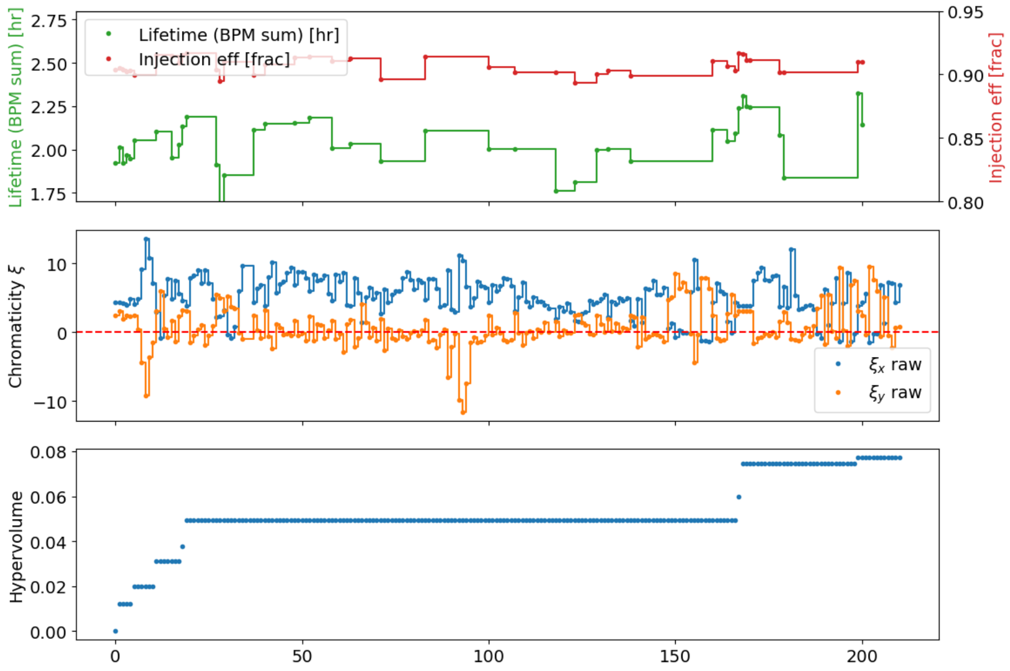
- Why not a prior mean? **Performance** and need for retraining every shift.

DA/MA: EXPERIMENTAL RESULTS

- Tuning was repeated in September, again with good +15% lifetime results
  - (note that many steps are skipped because of live model constraints)

- New solution placed into user operation

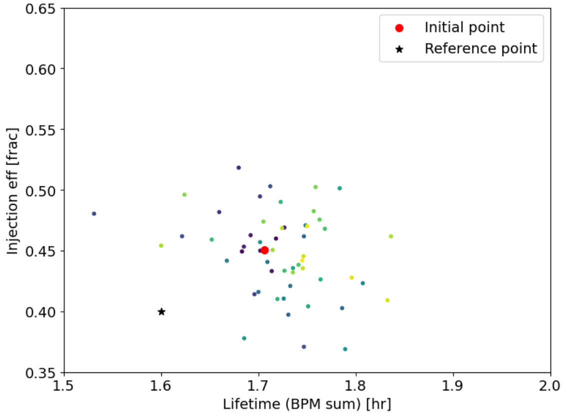




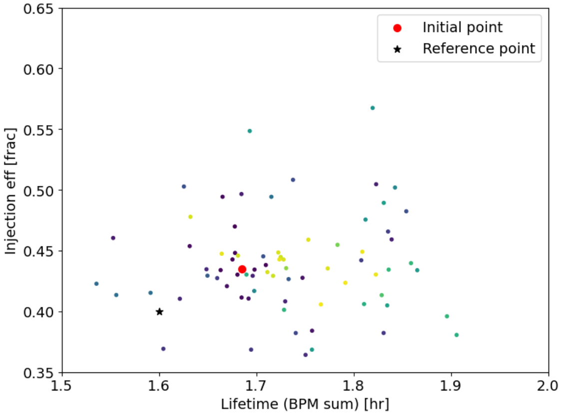
DA/MA: EXPERIMENTAL RESULTS (SPOILED CONFIG)

Had time to do a shift in **intentionally spoiled configuration**

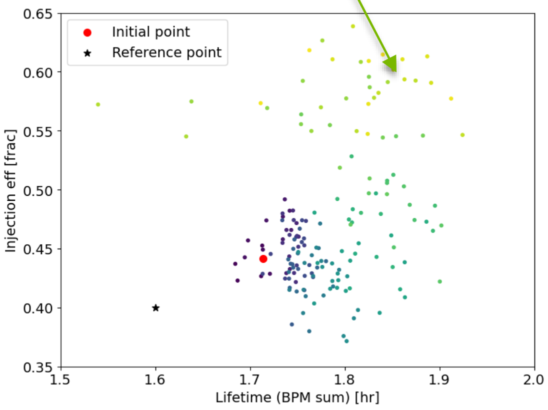
- Showed that full 24D was necessary to get good injection efficiency recovery
- Benchmarked with MGGPO and 12D MOBO – results ranked as expected



MGGPO



MOBO 12D



MOBO 24D

CONCLUSION

ML/AI tools have a variety of uses at light sources and beyond – BO was very helpful

- For APS-U DA/MA online tuning, was the **only feasible option and worked well**

Developed and verified:

- (MO)BO algorithm improvements and user-facing tools (see backup slides)
- Fast and calibrated data collection methods specialized for ML tools
- Integration with live model/digital twin

Results showed sizeable performance improvements in DA and MA

Future plans:

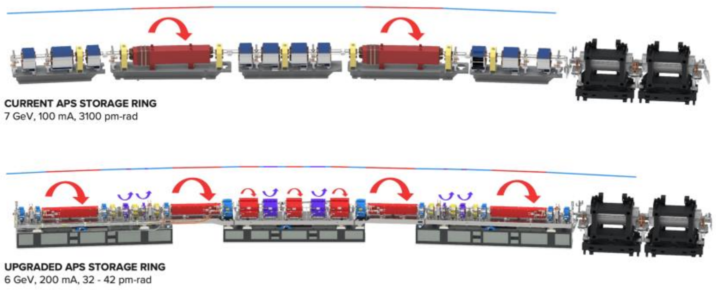
- Running in close-to-production conditions (coupled optics, high current)
- Extension to other machines (example - FNAL booster loss rebalancing)

THANK YOU!

OPTIMIZATION @ APS-U

APS just finished a major upgrade (APS-U)

- Have machines of various age and challenges
- Required customized methods



Linac (≥450 MeV)

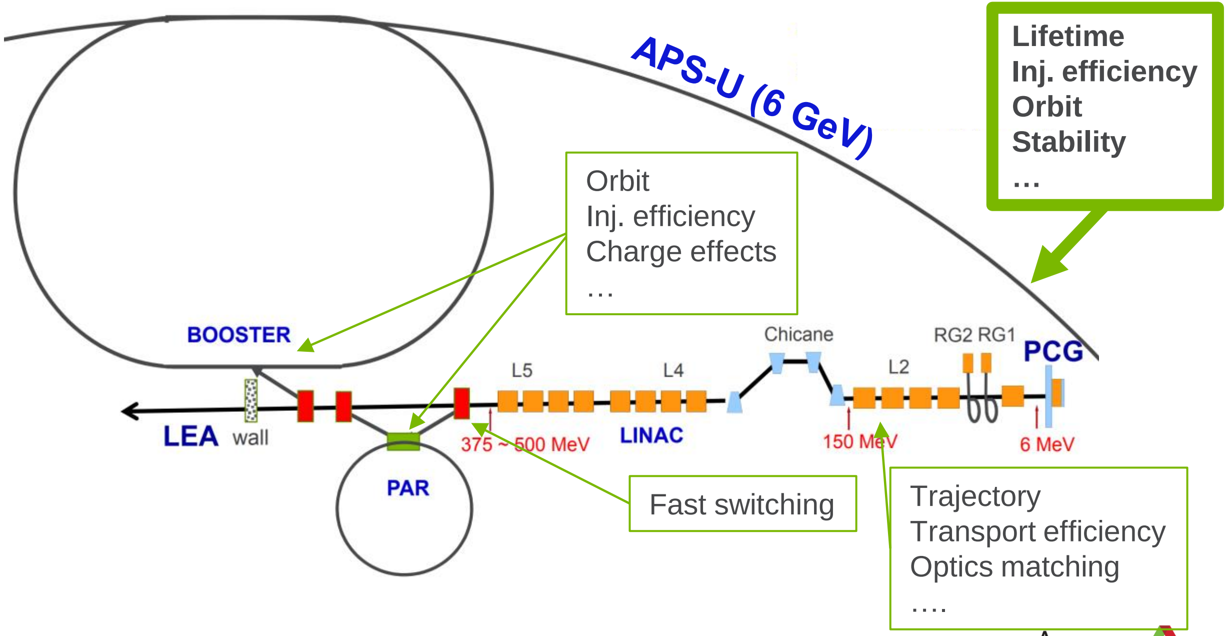
- 1 nC/pulse; 30 Hz rep rate
- Thermionic RF guns: RG1, RG2 (1 hot spare)

Particle accumulator ring (PAR) (≥450 MeV)

- Single bunch; 1-Hz rep rate
- Captures linac pulses in RF1 (9.8 MHz); compresses damped beam in RF12 (117 MHz)

Booster (6 GeV)

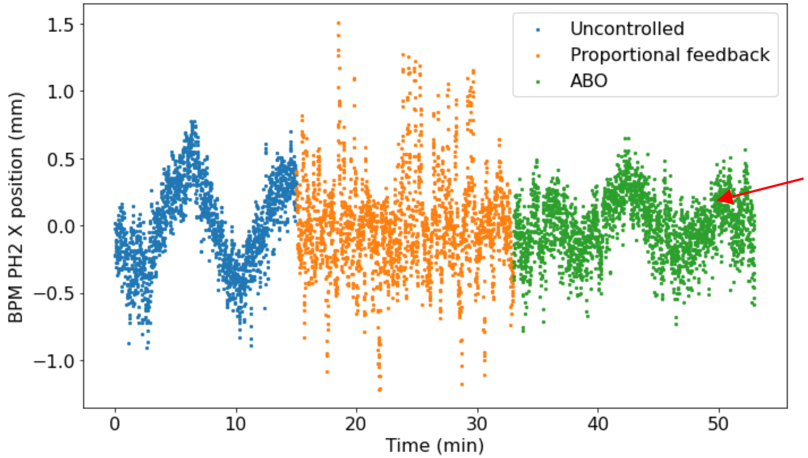
- Single bunch; 1 Hz rep rate
- Magnet ramp linear; RF ramp ~E^4 (352 MHz)
- Natural emittance: 97 nm
- Off-momentum emit (-0.6%): 64 nm



EXAMPLE 2: LINAC ADAPTIVE TUNING FOR DRIFT COMPENSATION

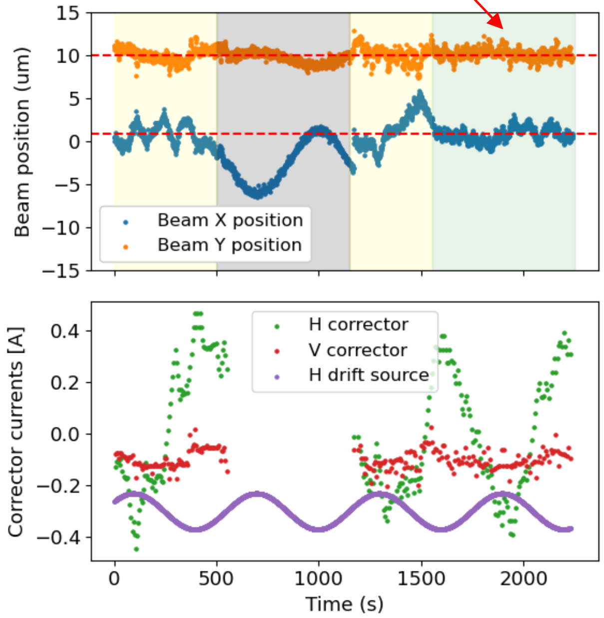
Linac trajectory stabilized through several linear feedback loops

- Due to a variety of reasons, quite noisy
- Successfully demonstrated adaptive (time-dependent) BO as an alternative



(stabilized)

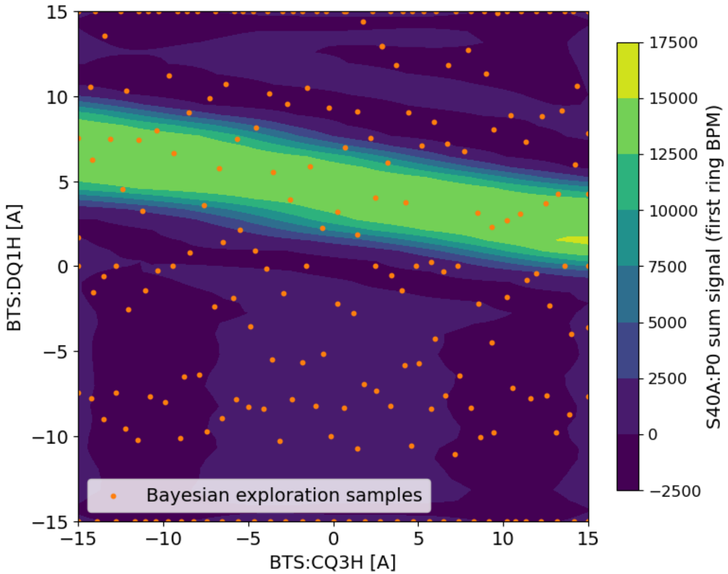
(automatic on/off based on confidence)



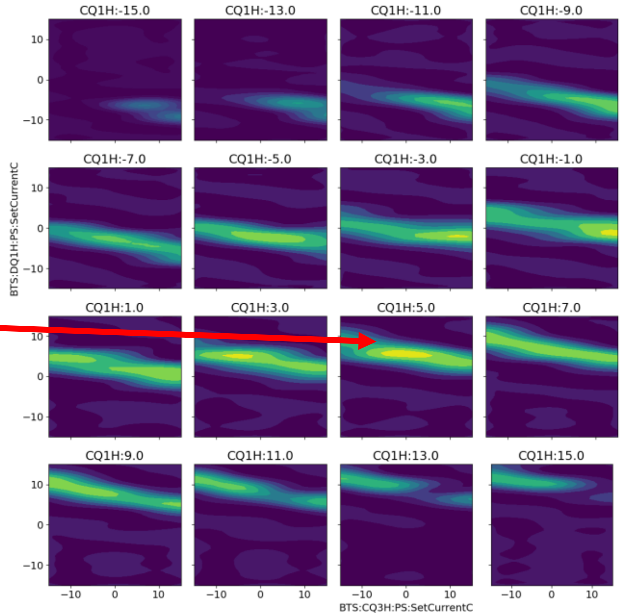
EXAMPLE 3: BAYESIAN EXPLORATION OF INJECTION APERTURES

Adding more dimensions is cheaper with BE

- Confirmed that 3rd corrector was already optimally tuned



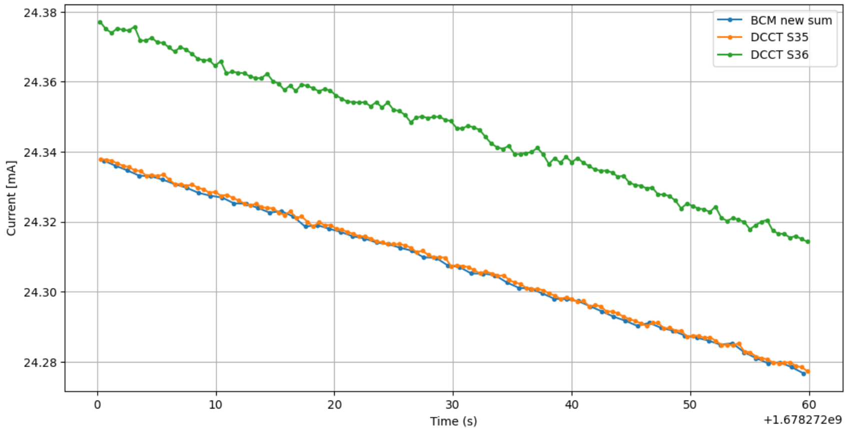
~150 samples using BE

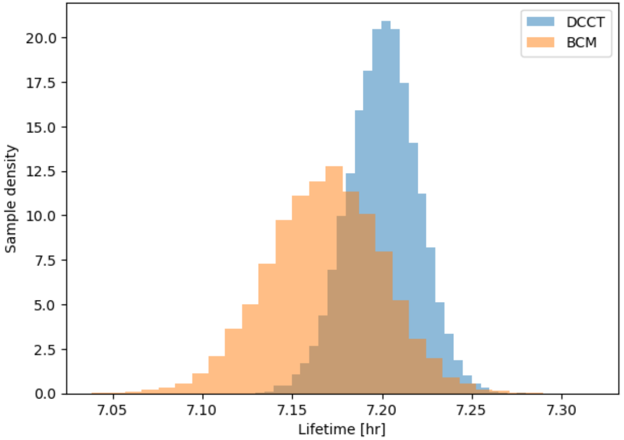


~300 samples using BE

DA/MA METHODS: MEASURING LIFETIME

- DCCT lifetime measurement is most straightforward
- At high currents, 10-20s collection is sufficient for +-0.05hr accuracy





INTEGRATED GUI FOR USERS – APSOPT-GUI

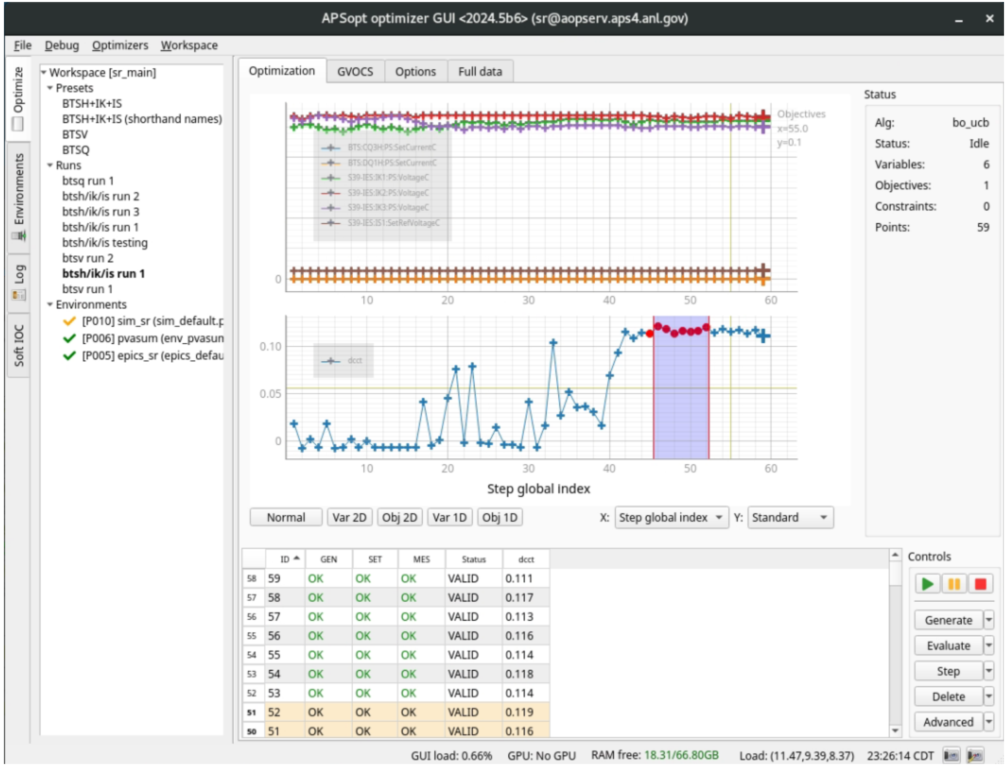
Several goals:

- Expert/R&D and operational modes
- Can be used by operators **without any coding**
- Presets/auto-run for standard tasks for each machine

Modular architecture:

- BO + model viz
- Simplex (for compatibility)
- RL

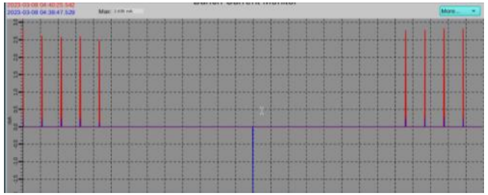
Advanced capabilities based on hands-on experience and needs



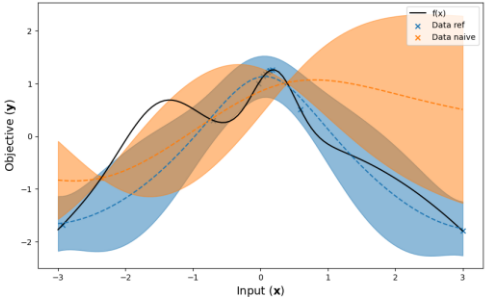
OPTIMIZATION WORKFLOW SUMMARY

Human-in-the-loop guidance

Fast experimental techniques

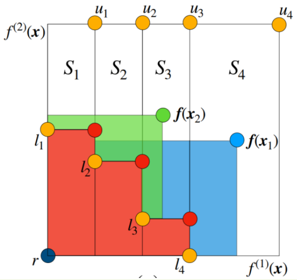


Archive and sim-based priors



Experts

Advanced AI/ML algorithms “APSopt” library

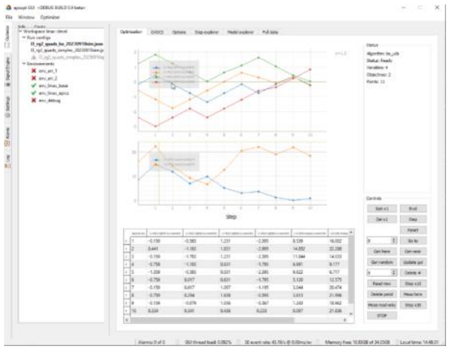


Physics simulations

User’s Manual for elegant

Program Version 2023.4
Advanced Photon Source
Michael Borland

Integrated Optimization Toolkit “APSopt-gui”

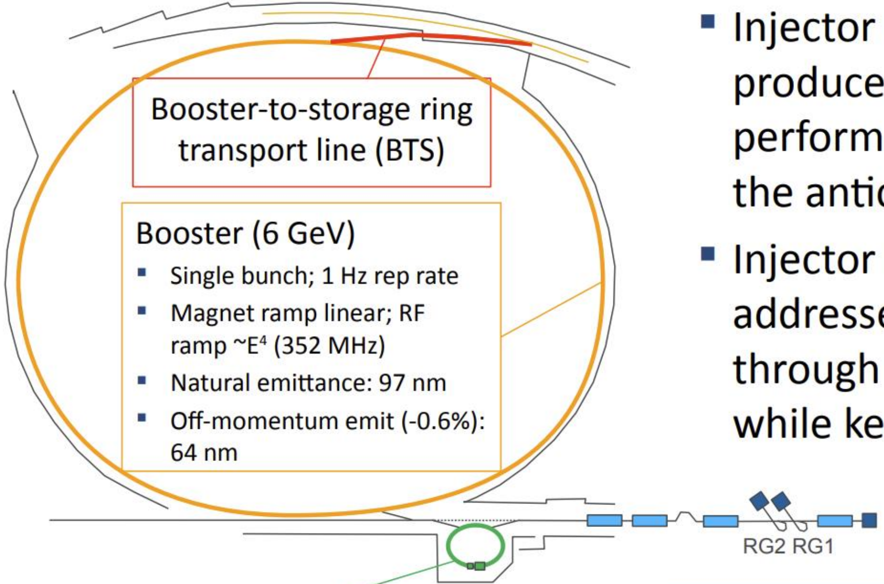


Non-experts

J. Calvey

APS-U Injector Complex

- Injector machines need to reliably produce enough single-bunch charge to perform complete bunch replacement for the anticipated APS-U fill patterns.
- Injector Plan was developed that addresses APS-U injection requirements through upgrades of injector complex, while keeping basic structure.

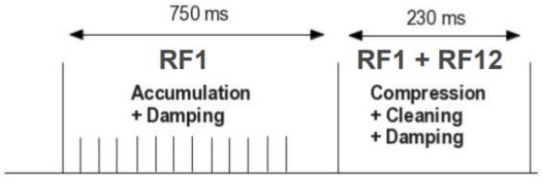


Booster-to-storage ring transport line (BTS)

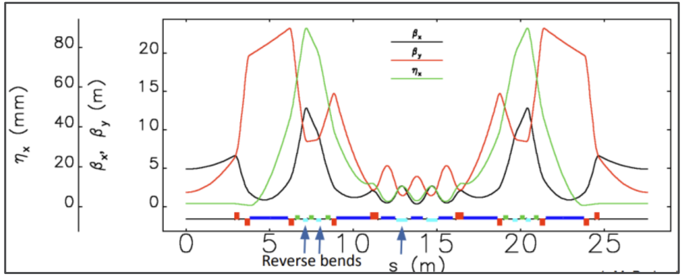
Booster (6 GeV)
- Single bunch; 1 Hz rep rate
- Magnet ramp linear; RF ramp ~E^4 (352 MHz)
- Natural emittance: 97 nm
- Off-momentum emit (-0.6%): 64 nm

Particle accumulator ring (PAR) (≥450 MeV)
- Single bunch; 1-Hz rep rate
- Captures linac pulses in RF1 (9.8 MHz); compresses damped beam in RF12 (117 MHz)

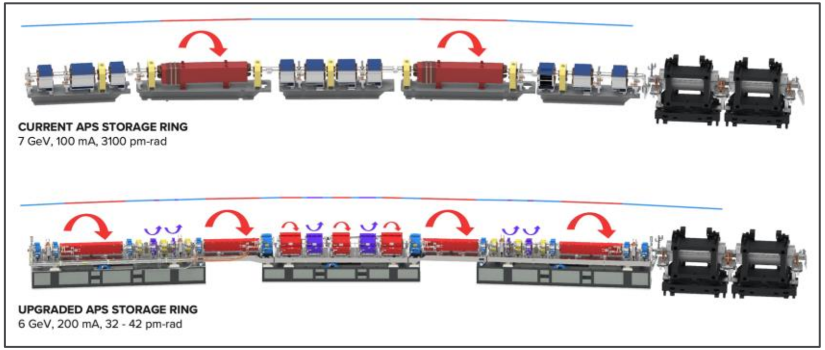
Linac (≥450 MeV)
- 1 nC/pulse; 30 Hz rep rate
- Thermionic RF guns: RG1, RG2 (1 hot spare)

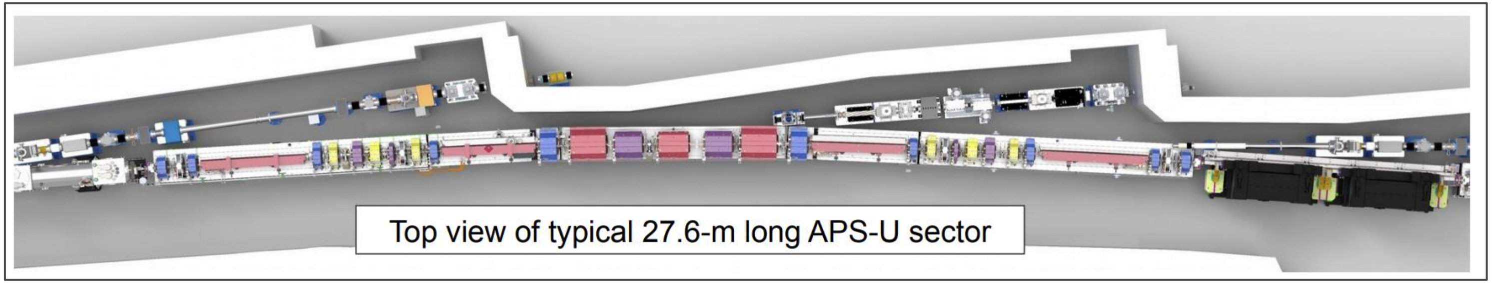


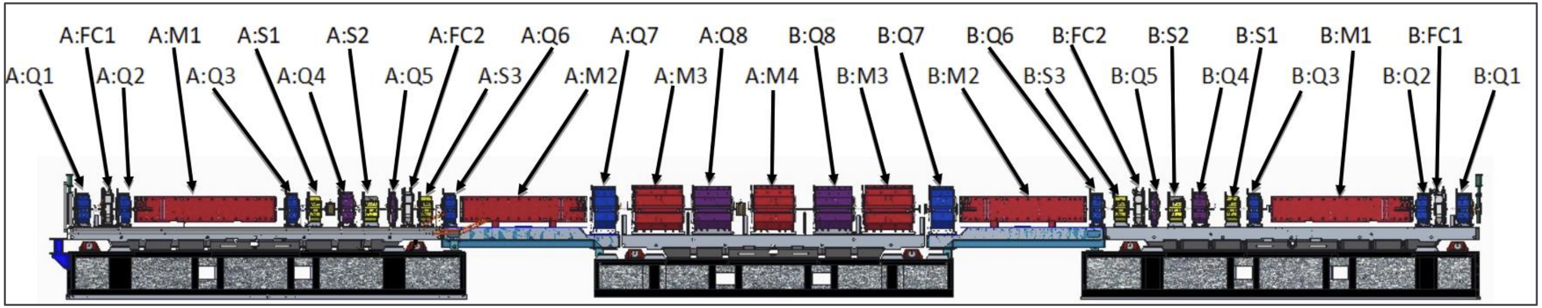
PAR cycle



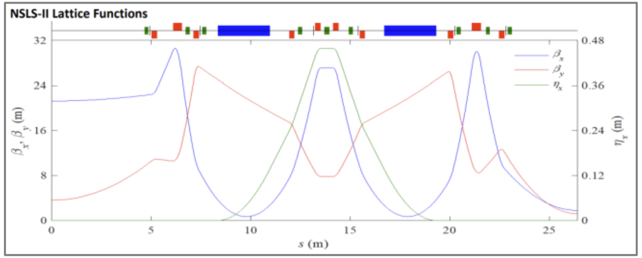
Multi-bend hybrid lattice – rely on
I/3I linear phase advance to cancel resonances





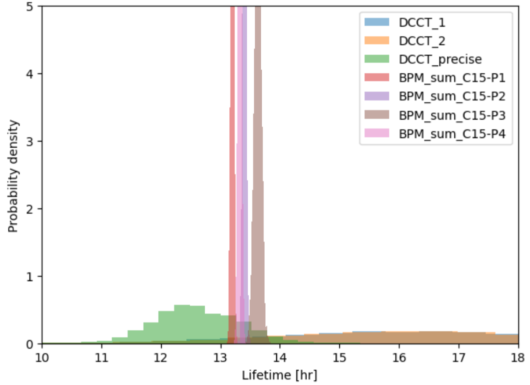


MOBO #2: DA/MA TUNING @ NSLS-II

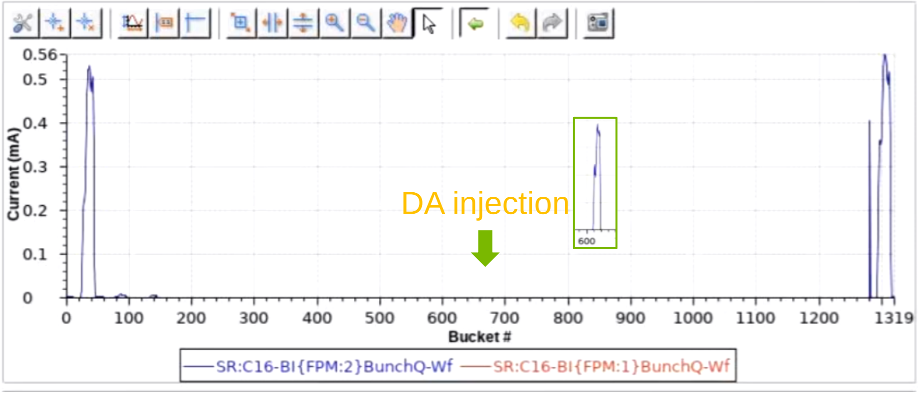


- During downtime, tested at NSLS-II **(big thanks to collaborators!)**
- New facility = new measurement techniques

Lifetime via BPM sum average (~15s)



Special bunch pattern to suppress instabilities/enable BxB feedback



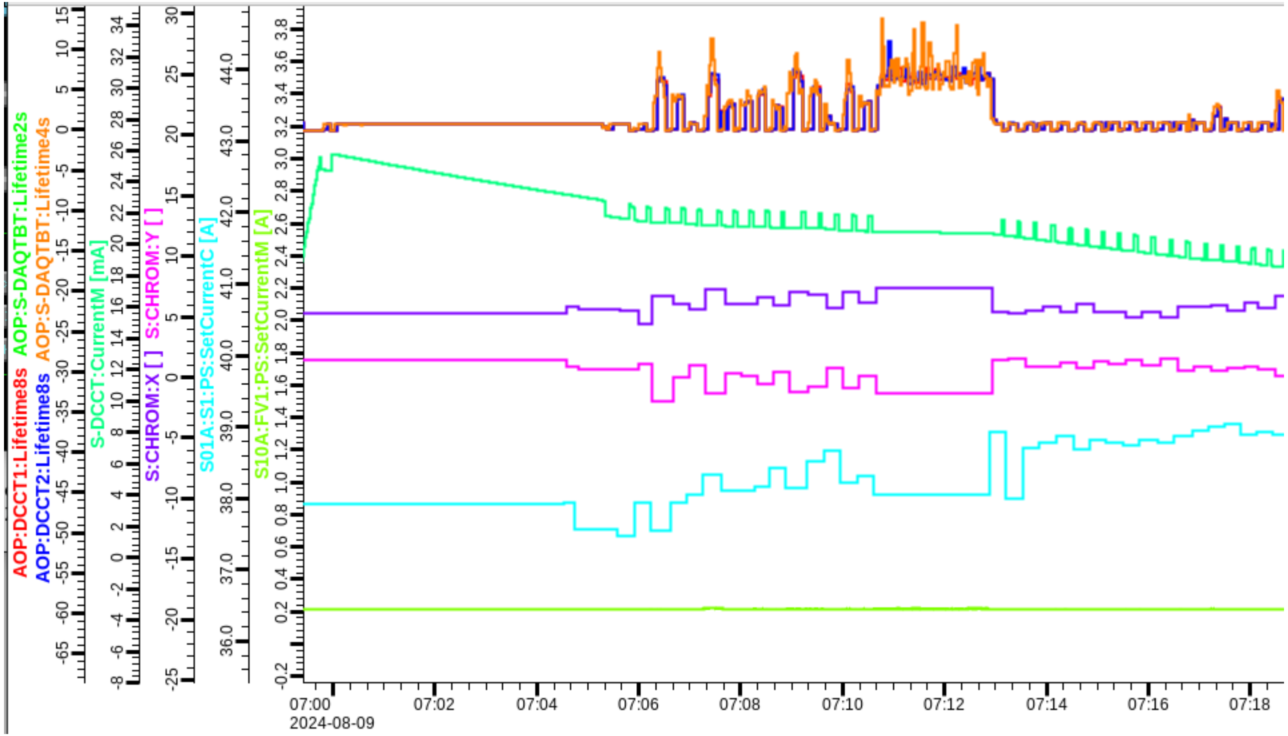
EXP [NSLS][NOV 2023]

EXP [NSLS][NOV 2023]

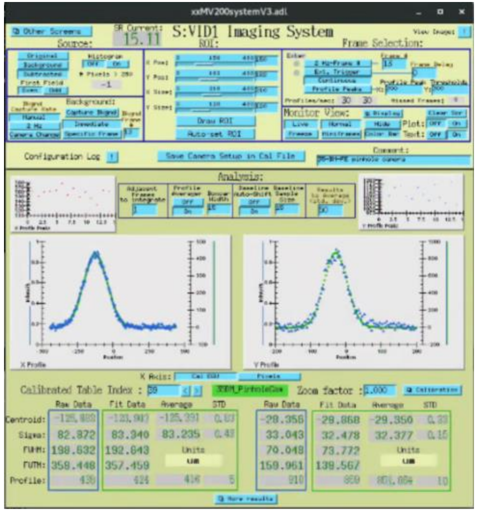
- Injection efficiency from DCCT ratio (~20s)
- Tricks to kick beam with jaws closed/beam position limits enabled + reduce sextupole orbit feed-down

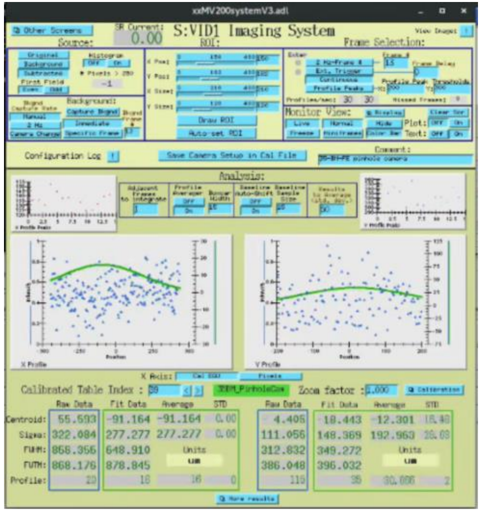
- **Overall implementation/debugging time: ~2 shifts, 15 hours**

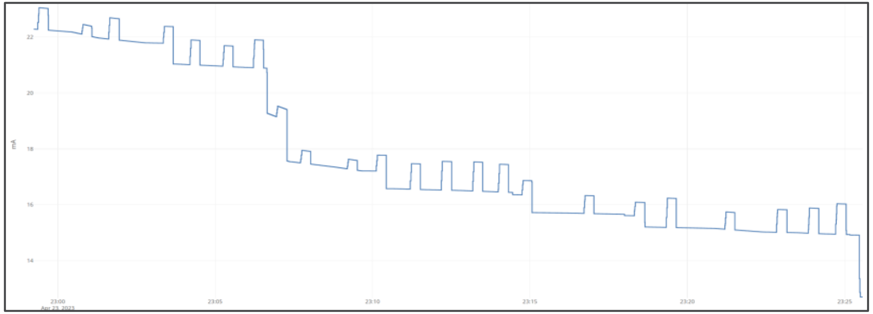
TYPICAL OPTIMIZATION RUN DATA (NOTE LIFETIME SPIKES)



BONUS – RAW SEXTUPOLE INPUTS WITH LARGE RANGE CAUSE MOBO TO CHEAT







MOBO quickly learned to **cheat lifetime by inducing instabilities**