Toward higher brightness and polarization of hadron beams: Digital-Twin-based autonomous control of BNL's hadron accelerator chain

a continuation of

Higher RHIC Polarization by physics informed Bayesian Learning

Georg Hoffstaetter de Torquat for the EIC-BeamAI collaboration

Long term goal: improve EIC physics by ML/AI

Brookhaven National Lab is constructing a 4km long accelerator complex to study basic nuclear physics, e.g.,

- Where do protons get their spin from?
- How did cosmic events produce the isotope distribution?
- How do gluons hold nuclei together

Designated the most pressing next NP project by DOE.
The largest accelerator project in the US today.

Probably the most complex accelerator ever built:

- Polarized protons and electrons.
- Beam cooling (Rf, e, and photon based)
- Superconducting RF acceleration
- Superconducting magnets

DE-FOA-0002875 : ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING FOR AUTONOMOUS OPTIMIZATION AND CONTROL OF ACCELERATORS AND DETECTORS

Title: Beam polarization increase in the BNL hadron injectors through physics-informed Bayesian Learning

Collaborators: BNL, Cornell, SLAC, JLAB, RPI

Budget: $1.5M, 09/01/2023 to 8/31/2025

Funding through DOE-NP DE SC-0024287, contr.# 2023-BNL-AD060-FUND

Funding officer Manouchehr Farkhondeh

FOA requested topic:

- Address the challenges of autonomous control and experimentation
- Efficiency of operation of accelerators and scientific instruments

New NOFO (DE-FOA-0003458, Artificial Intelligence and Machine Learning Applied to Nuclear Science and Technology) for a continuation proposal.

Brookhaven National Laboratory CLASSE Cornell Laboratory for Accelerator-based ScienceS &Education

Description of the project and the current status

- Description of RHICH pre-accelerators
- Topics that can improve polarization:
  - Lower emittance
  - Better timing
  - Lower spin-orbit resonance strengths

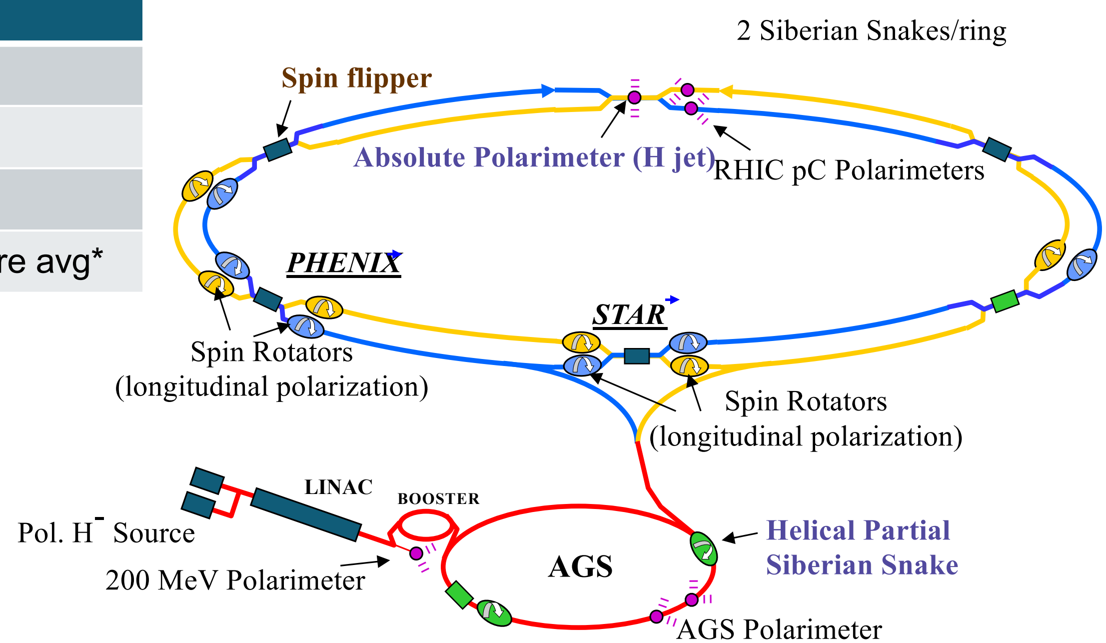
- New project (2025-27) extends to brighter hadron beams and to optimizing the polarized source and linac.

The polarized proton accelerator chain

RHIC Polarized Beam Complex

	Max tot. Energy [GeV]	Pol. At Max Energy [%]	Polarimeter
Source+Linac	1.1	82-84	
Booster	2.5	~80-84	
AGS	23.8	67-70	p-Carbon
RHIC	255	55-60	Jet, full store avg*

* Includes both ramp loss and store decay



	Relative Ramp Polarization Loss (Run 17, full run avg)
AGS	17 %
RHIC	8 %

Characteristics of polarization optimizations

1. Optimal parameter settings are hard to find, and the optimum is difficult to maintain.

2. The data to optimize on has significant uncertainties.

3. Good, approximate models of the accelerator exist.

4. A history of much data is available.

Is this type of problem suitable for Machine Learning?

Why would ML be better suited than other optimizers and feedbacks?

Polarized collider performance vs. beam intensity

Collider luminosity, $\mathcal{L}$

$$\mathcal{L} \propto \frac{N^2}{\varepsilon}$$

N = intensity/ bunch
ε = tran. emittance

Polarized collider figure of merit (for polarization P):

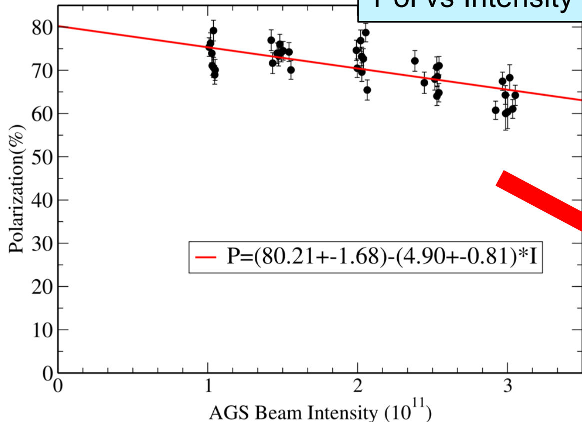
$$\text{FoM} = \begin{cases} \mathcal{L}\, P^2 & \text{transverse spin} \\ \mathcal{L}\, P^4 & \text{longitudinal spin} \end{cases}$$

Since both emittance and polarization degrade with intensity figure of merit decreases rapidly

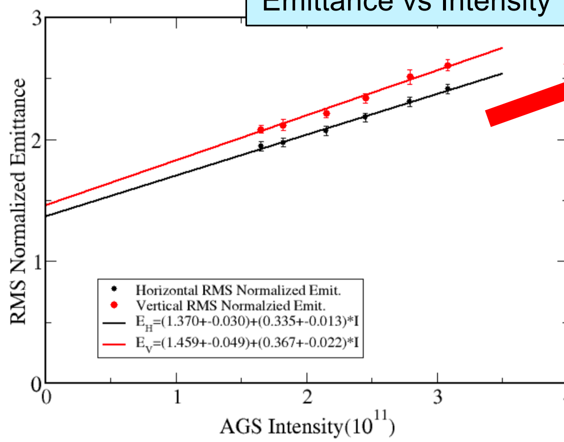
FoM dependence on intensity closer to linear in N than quadratic.

AGS extraction

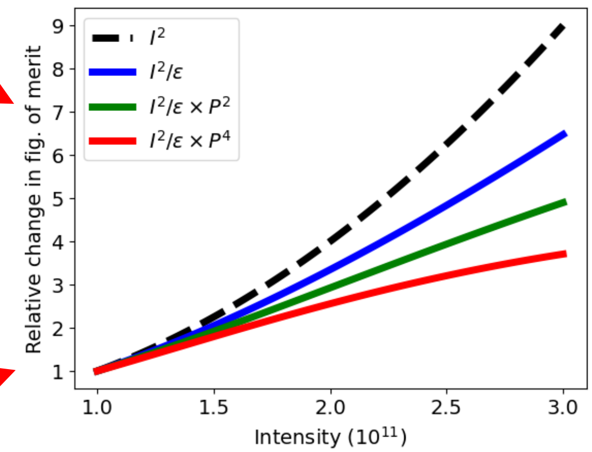
Pol vs Intensity



Emittance vs Intensity



Polarized beam collider FOM



Impact of intensity increase on FoM given emittance and polarization dependence at AGS extraction

Emittance reduction ➔ less depolarization

To reduce and maintain emittances we

- optimize Linac to Booster transfer
- optimize Booster to AGS transfer
- correct optics and orbit in Booster and AGS
- use orbit responses to calibrate models of Booster and AGS.
- split bunches in the Booster for space charge reduction and re-merge them at AGS top energy.

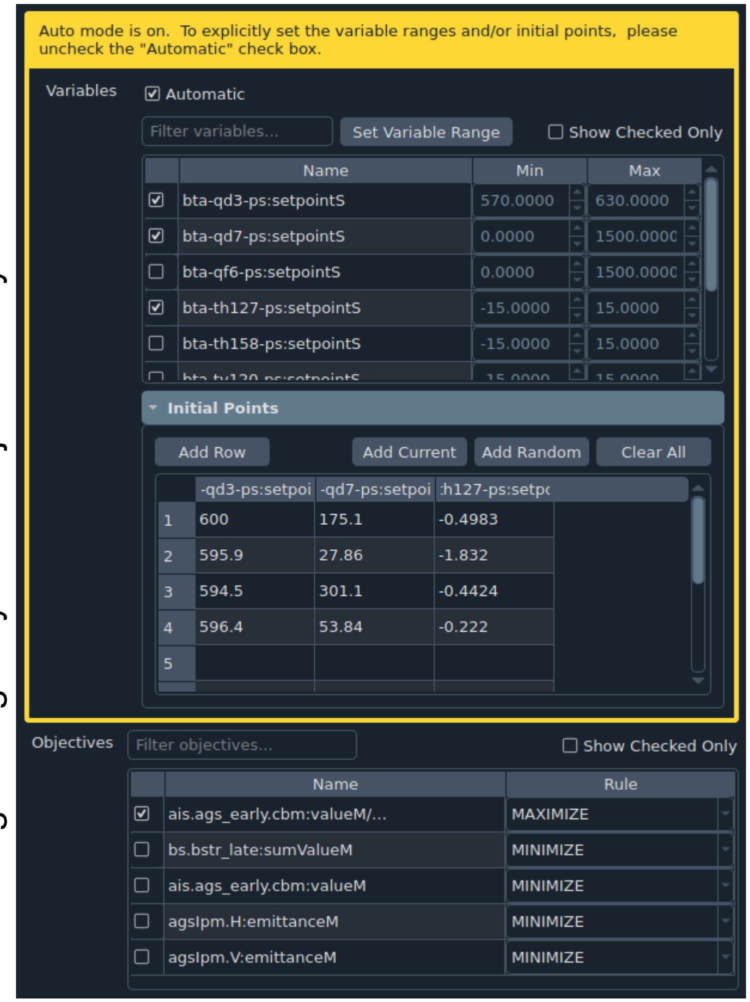
Developed so far, for the control room …

- Bayesian Optimization of Booster injection, available in the control room
- Bayesian Optimization of AGS injection, available in the control room
- Reinforcement Learning routine of AGS bunch merges, partially tested

This allows maintaining optimal emittances automatically.

These were not yet ready for the last RHIC run.

Through Badger by Leve Hajdu and Ryan Russel



Brookhaven National Laboratory

CLASSE Cornell Laboratory for Accelerator-based ScienceS and Education

Booster injection

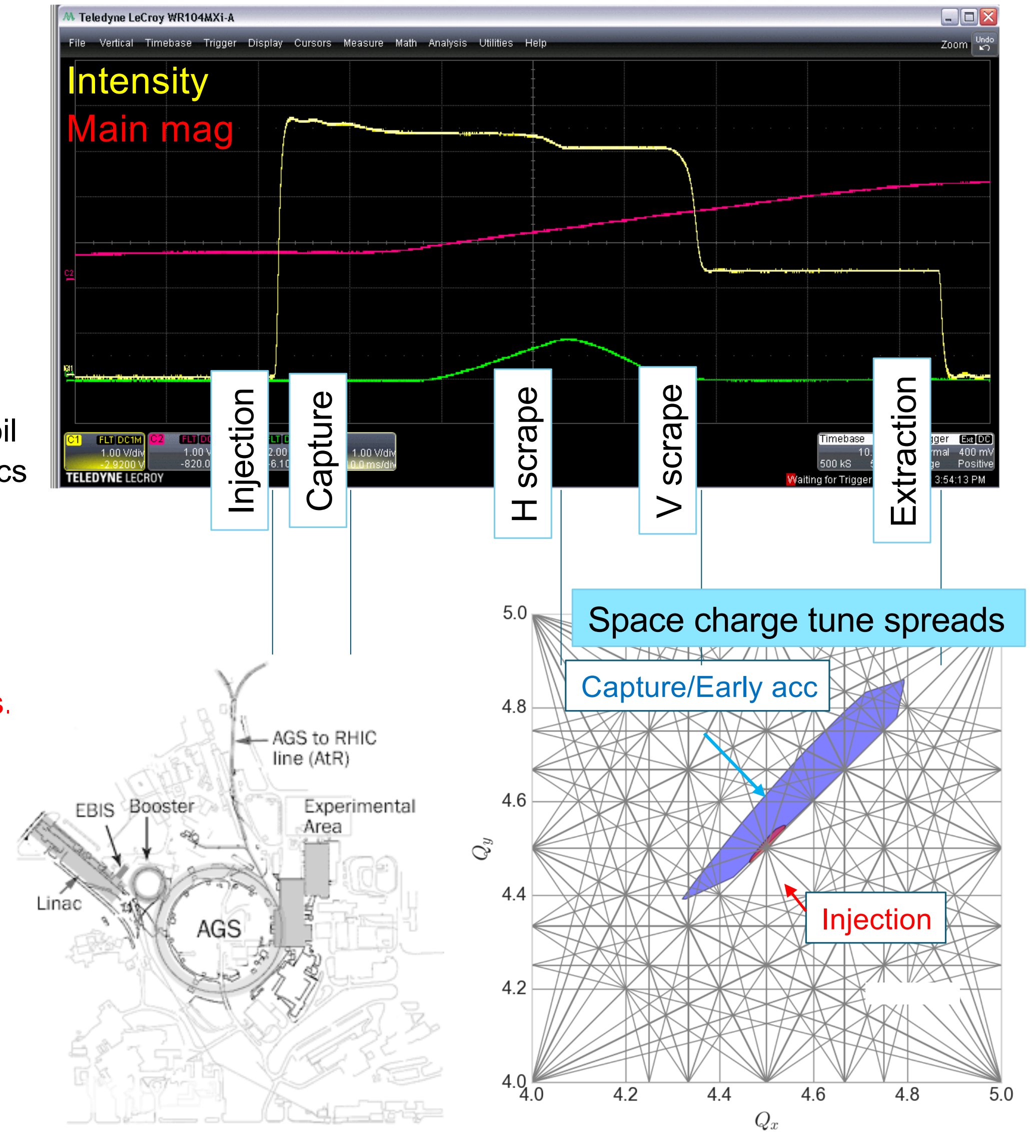
Booster injection/early acceleration process sets maximum beam brightness for rest of acceleration though RHIC

Goal: Get most current past fixed V and H scrapers.

Model building (complex): Injection is complex, incl. ionization foil with 300 turn scattering, H and V scraping, space-charge dynamics an acceleration.

Optimization: Many ”knobs” are available, incl. transfer line magnets, puls length vs. height, RF capture parameters, Booster orbit and optics.
So far, our Bayesian Optimization only uses transfer line magnets.

Instrumentation (complex):
- WCM, BPMs don’t work until after capture
- No transverse profile monitor in Booster
- Scraping efficiency as proxy for brightness optimization
- Emittance only measurable in the extraction line via multiwire



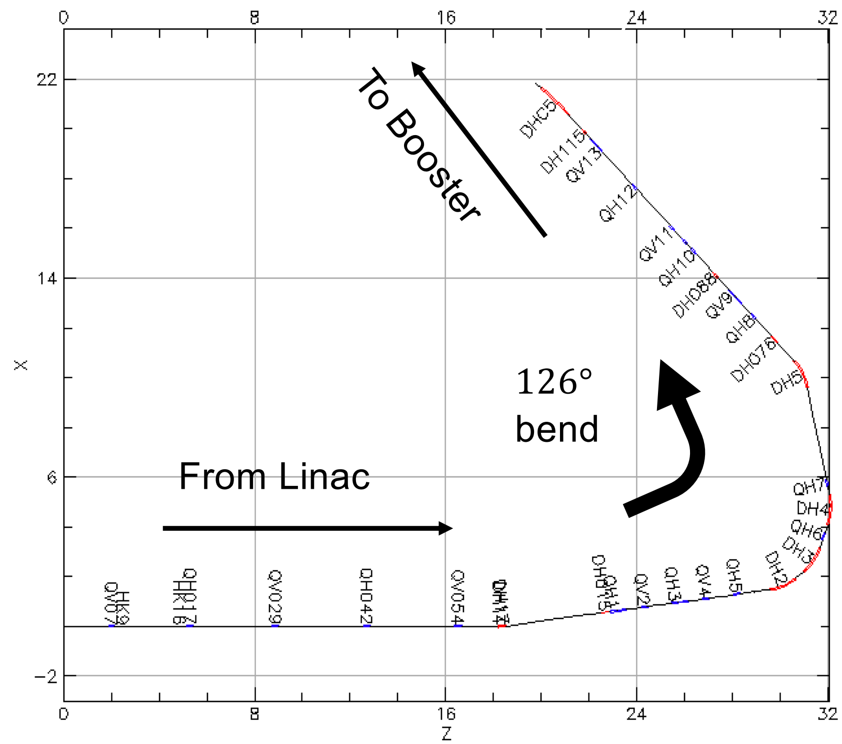
Model building for Booster injection

Booster injection process sets maximum beam brightness for rest of acceleration through RHIC

- Transfer line, including complex injection magnet modeled.
- Ionization foil and 300 turn scattering modeled
- Acceleration to H and V scrapers modeled
- Acceleration modeled, not yet with space charge

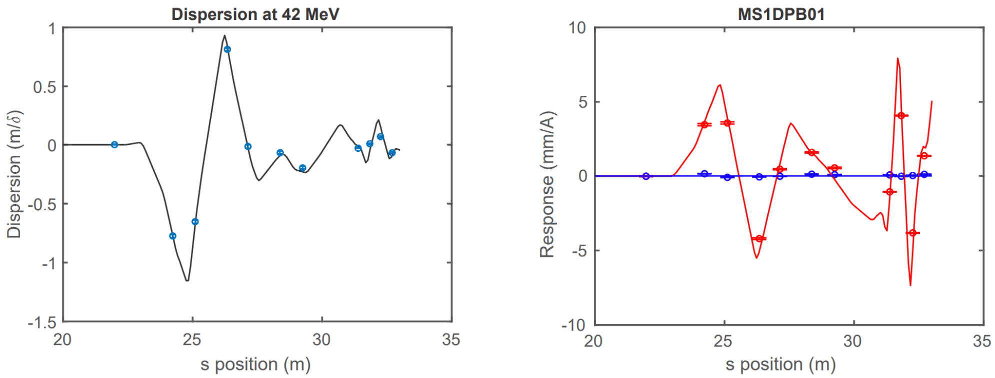
Goals:
(a) set up a digital twin to streamline operations

(b) make Bayesian Optimization physics informed by the model.



Digital Twin for hadron injector sections

A Digital Twin is a bi-directional connection between an accelerator's physics model and its control system.

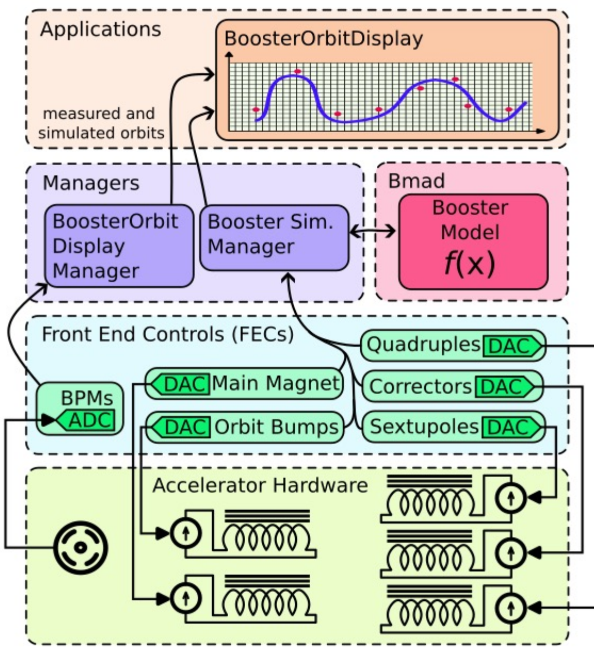


Example digital-twins for CBETA: combine Bmad with EPICS bidirectionally.

- Bmad ➔ control system: DT results are displayed by the control system, just like measured accelerator data.
- Control system ➔ DT: Power supply settings automatically load into the physics model.

Great for continuous comparison of operations and model.
Great for offline development of operations procedures.
Great for virtual diagnostics.

- Additional benefit: Neural network can be trained to predict slow to simulate beam behavior in operations time, e.g. space charge dynamics.
- ML control routines always have the up-to-date physics model available.



DT currently being prepared for the Booster.

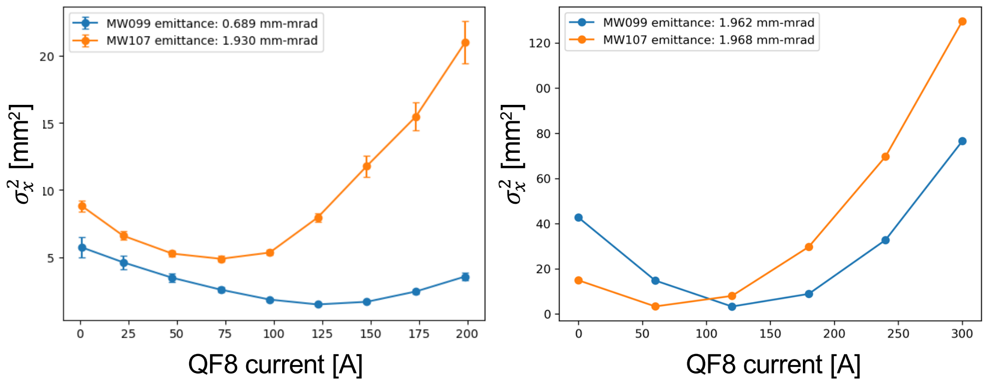
Speed up of BO with physics information

BO of emittance minimization already works, but it could be faster with model information.

- To model injection into the booster, the beam's phase space distribution in the LtB line needs to be known.
- While a NN can be trained to determine the beam's phase space distribution from tomography, the current diagnostics does not permit to resolve x-y coupling.
- Polarized proton beam has such coupling because it is created in a solenoid field.

➔ We will use skew quads in the booster and tilted multi wire detectors to resolve x-y coupling.

➔ Then our BO can be extended by a physics informed model.

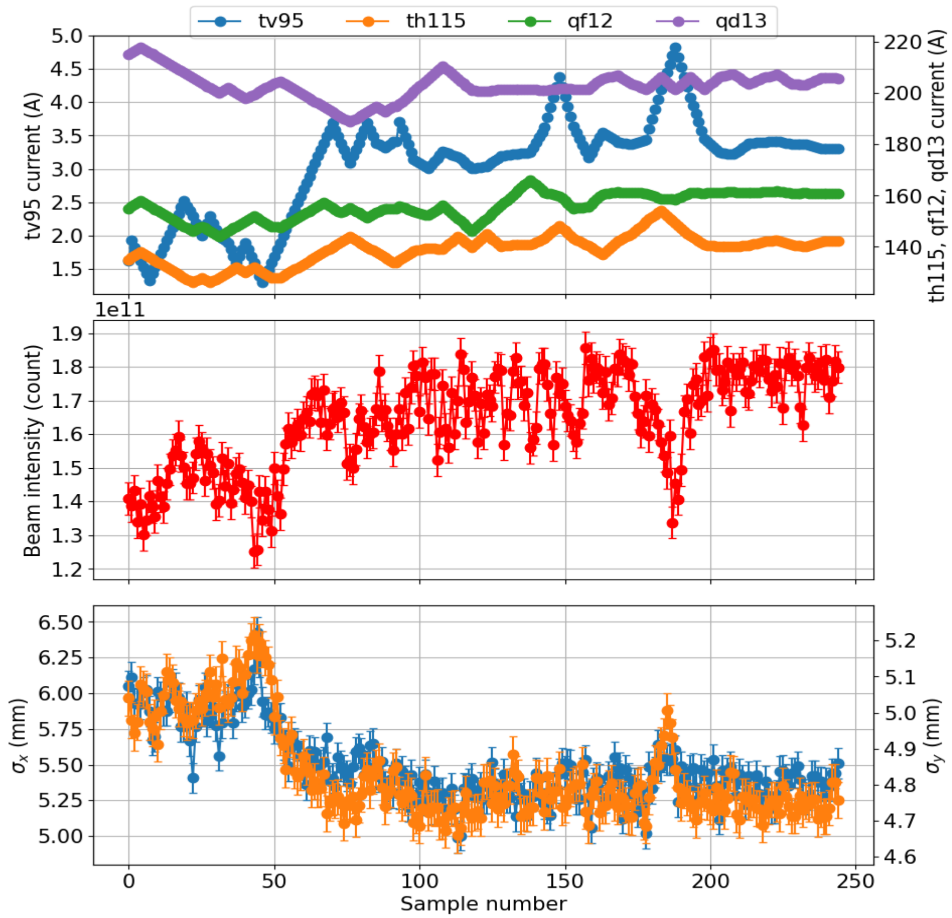


Simulated (left) and measured (right) quadrupole scan results for horizontal quad QF8 observed at two multi-wires (MW099, MW107) in the LtB line.

➔ The x/y projected emittances change along the transfer line, i.e., coupling needs to be considered.

Result: Automatic BO for Booster injection

- Controls: Power supply currents of two correctors and two quadrupoles at the end of the LtB line
- Beam size decrease in both planes in the BtA line in correspondence with intensity increase



Bayesian optimization of the Booster injection process.

Top: power supply currents of two correctors (tv95, th115) and two quadrupoles (qf12, qd13) in the LtB line.

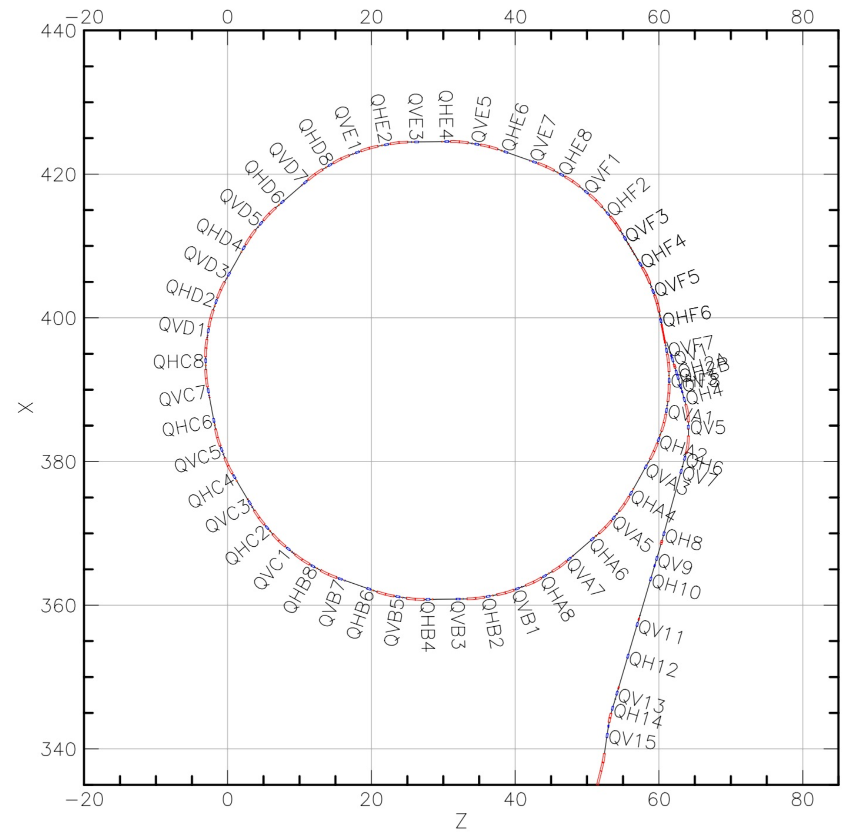
Middle: beam intensity after Booster injection, scaping, and acceleration.

Bottom: Beam size measurements in the BtA line during Bayesian optimization.

Control system: This Bayesian Optimization is now available as a control system application to operators.

BtA Transfer Line Structure in Bmad

- Lattice can be divided into branches connected with forks to simulate connection to a transfer line
- Require documented coordinates for elements to construct correct geometry
- Beam parameters from the end of one branch is automatically inherited by the start of downstream branch → continuous tracking
- BtA universe with three branches
  - 1st branch: Booster ring with extraction bumps
  - 2nd branch: Extraction line from F2 to F6 septum with F3 kicker on
  - 3rd branch: BtA transfer line



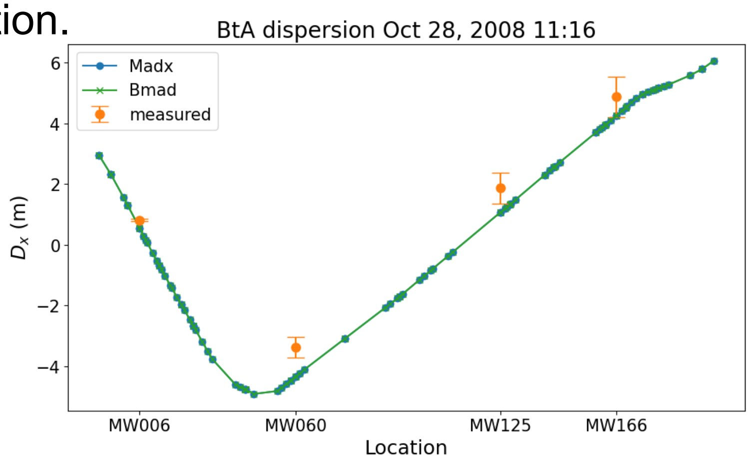
BtA modeling and data comparison

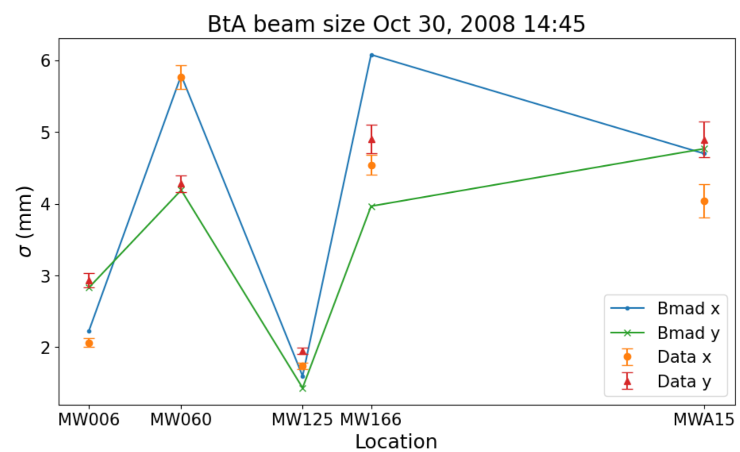
Goals:

(a) set up a digital twin to streamline operations

(b) make Bayesian Optimization physics informed by the model.

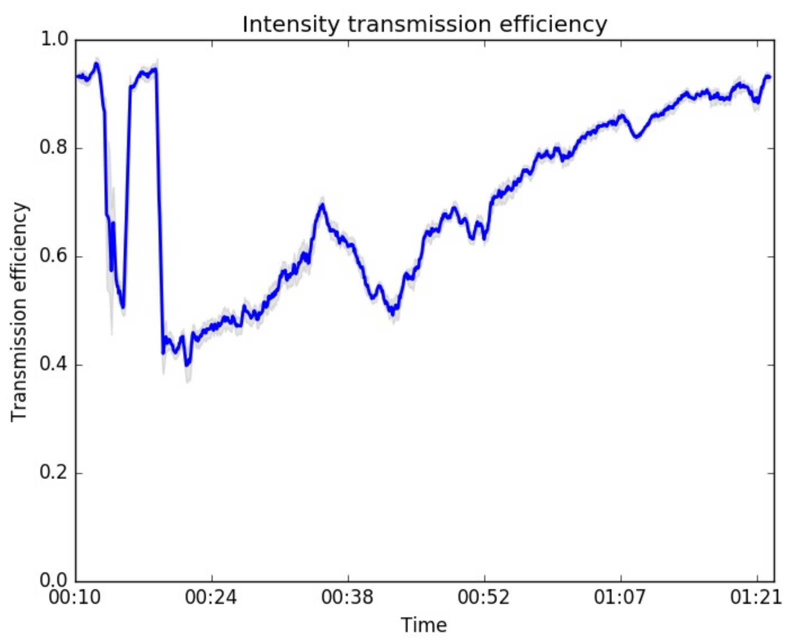
- Bmad tracking leads to horizontal dispersion matching measurements
- Beam size values from bunch tracking show agreements for upstream multi-wire measurements, disagreement downstream needs further investigation
- BO of emittance minimization already works, but it could be faster with model information.

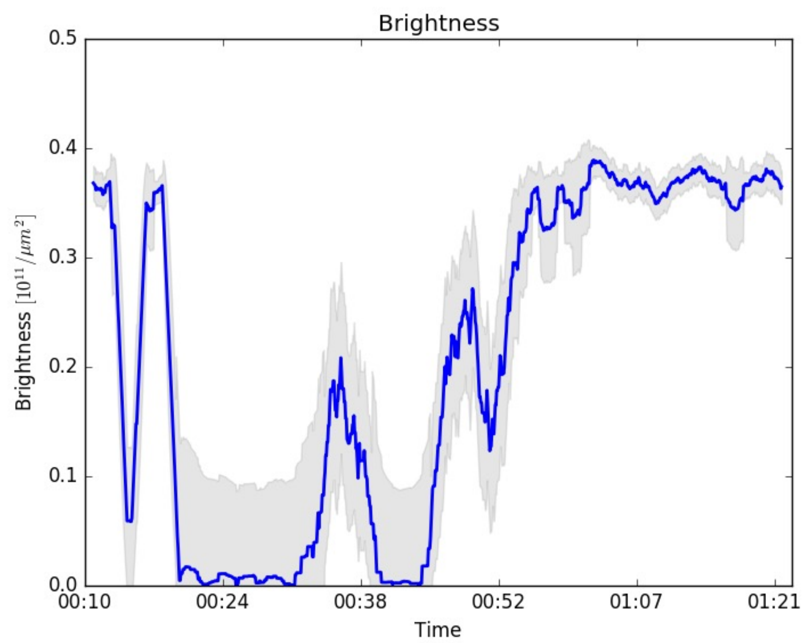




Result: Automatic BO for AGS injection

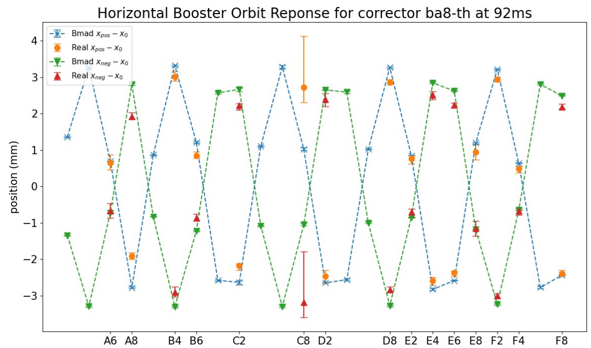
Algorithm efficiently found settings that were different, but at least as good as the previously optimized ones, automatically maintain the AGS injection at optimal performance without human intervention.

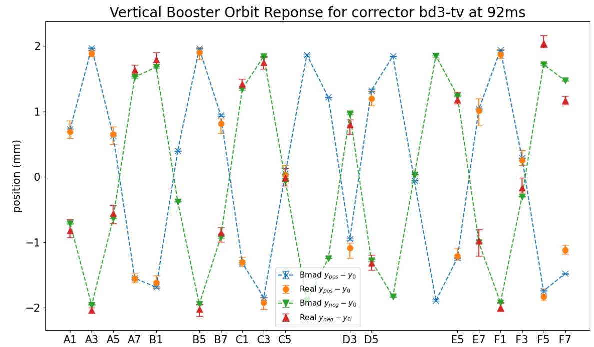




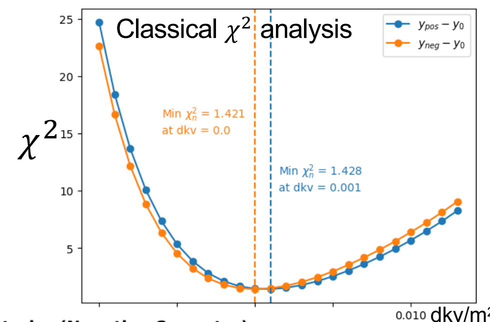
➔ Optimization of current while observing the brightness.

Uncertainty Quantification from orbit responses in the Booster





Orbit response data can be used to find and quantify unknown parameters (e.g., power supply scaling factors, magnet misalignment etc.) in real accelerators,
by Lucy Lin (from C-AD) and Nathan Urban (from CSI)

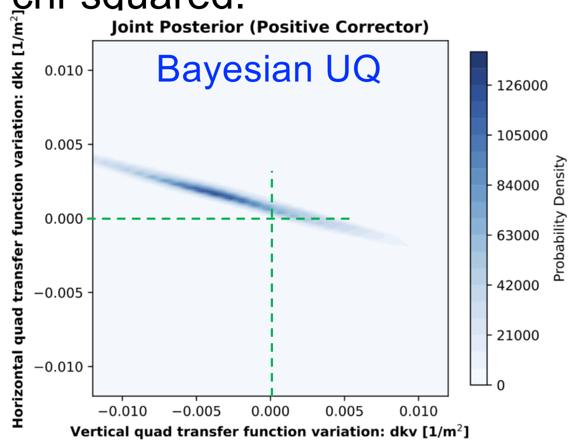


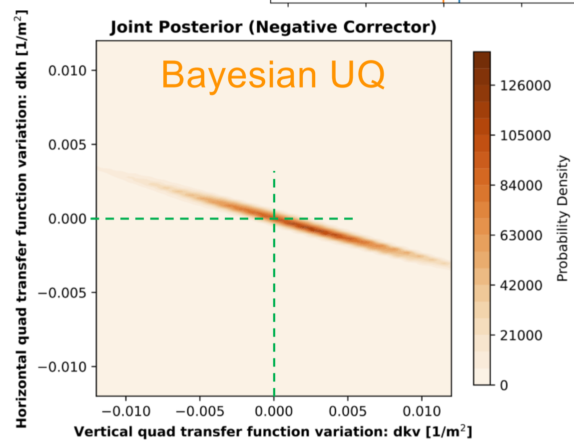
➔ Good agreements between Booster data and Bmad model are reached, with small discrepancies between model and measurement (within 1 mm)

➔ chi-squared/DF = 1.4 for model-experiment. Reasons are analyzed by
(a) Least square fitting to reduce chi-squared.
(b) Uncertainty Quantification.

➔ The main power supply transfer functions (a) do not reduce χ^2, (b) their UQ is consistent with 0

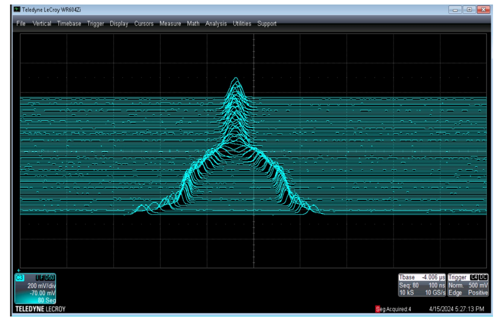
➔ Other error sources are being analyzed.





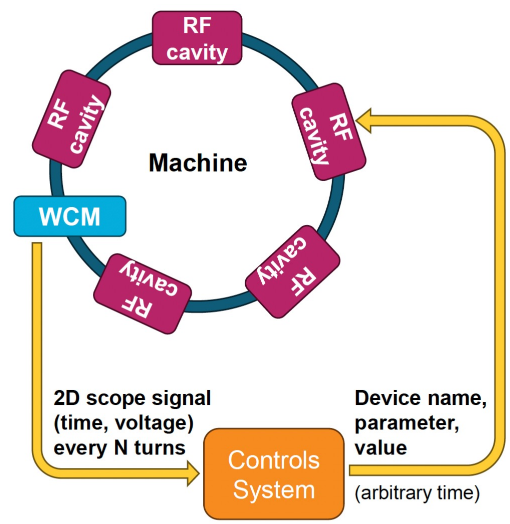
AGS Bunch Merging

- Before transferring to AGS, beam bunch is split into 2 longitudinally to reduce the space charge effect
  -> reduce emittance -> improve polarization
- Bunches are later merged before AGS extraction;
- Requires expert tuning of many parameters:
  - Prone to drift over time;
  - Time consuming;
- Controls: RF voltages, phases
- Goal: Obtain a “good” merged bunch profile:
  - Emittance preservation:
    - No particle lost;
    - Gaussian shape;
    - No “baby” bunches;
  - Stable final bunch profile:
    - Not shifting left to right;
    - Not bouncing up and down;
    - Merged in the center;



Real mountain range data showing 2-to-1 bunch merge in AGS

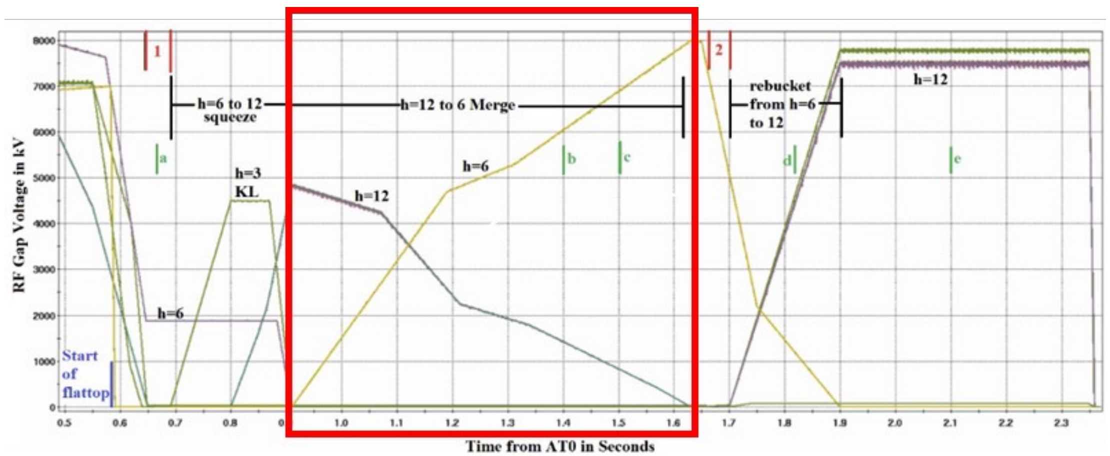
Wall current monitor (WCM) generates voltage vs time signal. Each separated in time by N turns (N accelerator periods)



Cartoon representation of accelerator with WCM, RF cavities (arbitrary number), and input/output

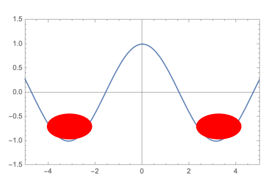
Bunch splitting in Booster / merging in AGS

Splitting in the Booster and merging after AGS accelerator reduces space charge and emittance growth ➔ more polarization

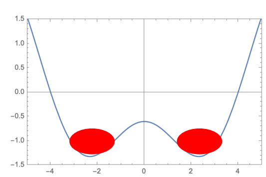


Three RF amplitudes (h=3, 6, 12) in the AGS during bucket manipulation and merging.

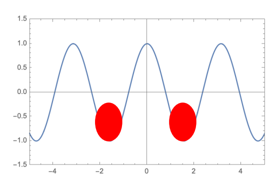
➔ We have set up Reinforcement Learning for the merging section.



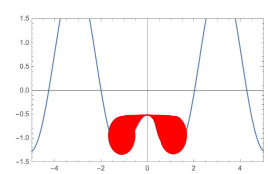
Accelerating RF h=6



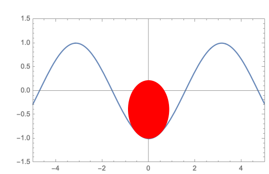
Attracting RF h=3



Close bucketing h=12



Combining h=6



Final bucketing h=6

Reinforcement Learning Tuning test - varying 6 voltage points for each RF system

RF voltages | Bunch traces | Last trace

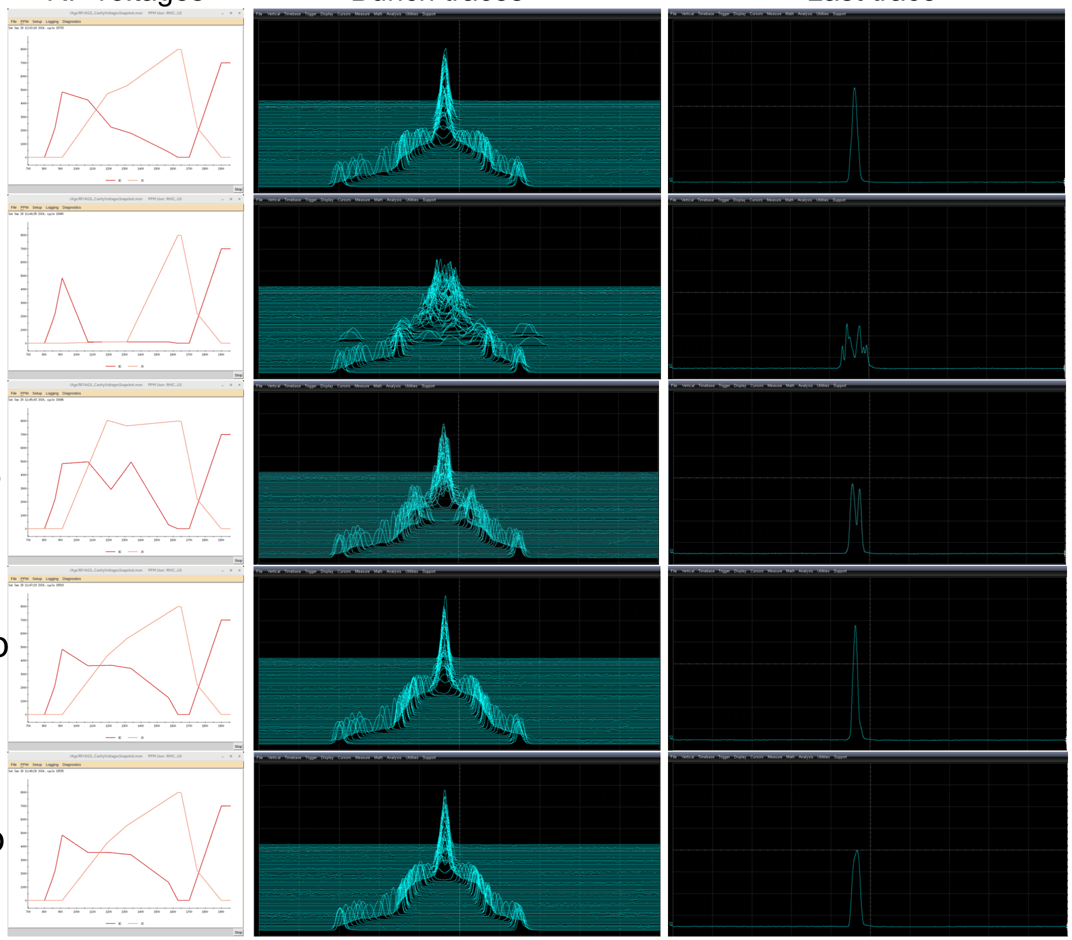
Original

Start

1st step

2nd step

3rd step



Goal: minimize the longitudinal emittance after bunch merging

RF amplitudes as function of time have been optimized in experiments.

Automatic readout of longitudinal emittance not yet available, test used simulated bunch lengths as reward.

Plan: check whether Reinforcement Learning has advantages over BO.

Plan: Include also RF phases as actors and coherent oscillations as state variables.

Determine useful state variables

- measurable
- related to the reward

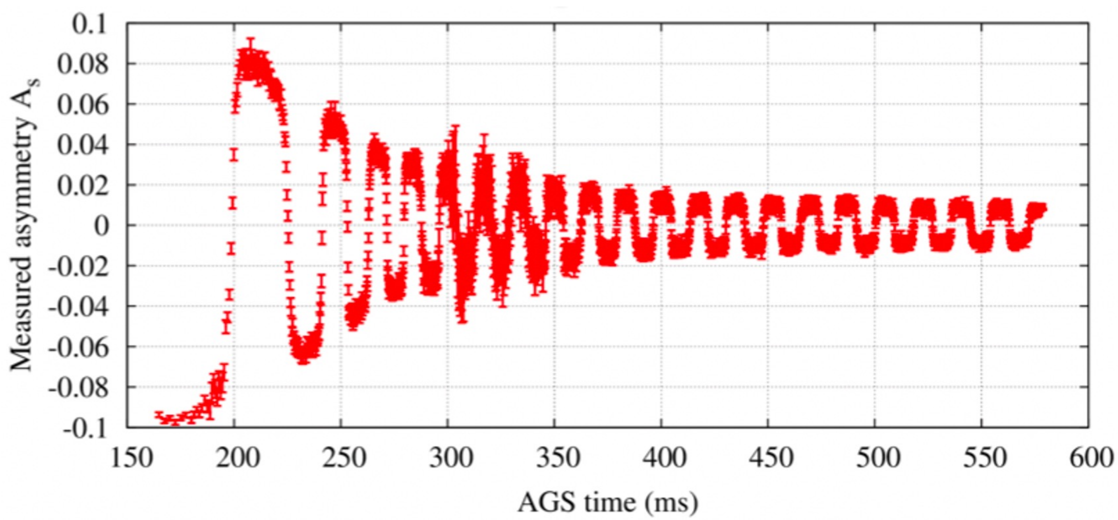
Timing of tune jumps

The G-gamma meter and accurate energy vs. time

(1) Measure the energy by orbit + revolution frequency measurement

(2) Measure of energy by field + revolution frequency measurement

(3) Measure energy by spin flip at every integer spin tune



Combined optimization

➔ measurement with reduced uncertainty at every energy

➔ better timing

➔ higher polarization

Still being worked on. It is less critical with new skew-quad resonance minimization.

Model building for the AGS

- Proton energy range 2.5 GeV -> 23 GeV
- Polarization preserved using
    - helical dipole snakes
    - + horizontal tune jump
    - Resonance correction in development (would replace tune jump)

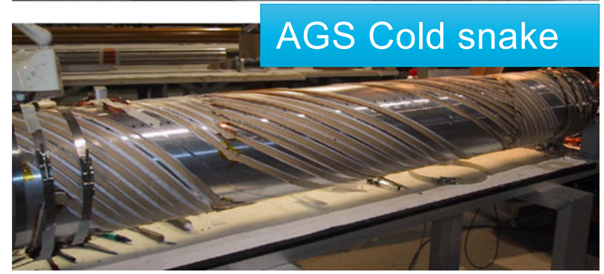
- Requires “near integer” tune
    - Orbit, optics unusually sensitive to errors

- Helical dipoles are complicated magnets
    - Large optical effects at low energy
    - Many related magnetic elements for compensation orbit/optics

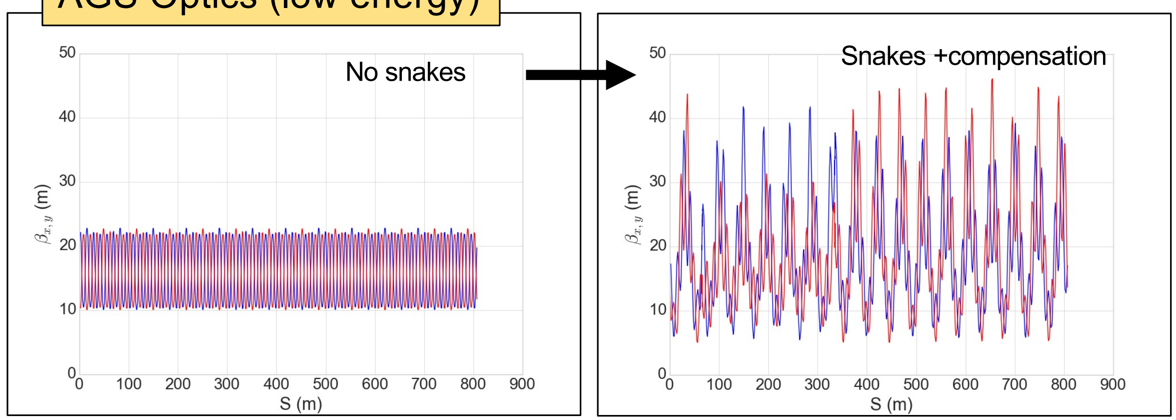
- The complex fields and lattice + high tune requirements are a challenge to modeling (Eiad Hamwi’s work)

AGS Warm snake

AGS Cold snake

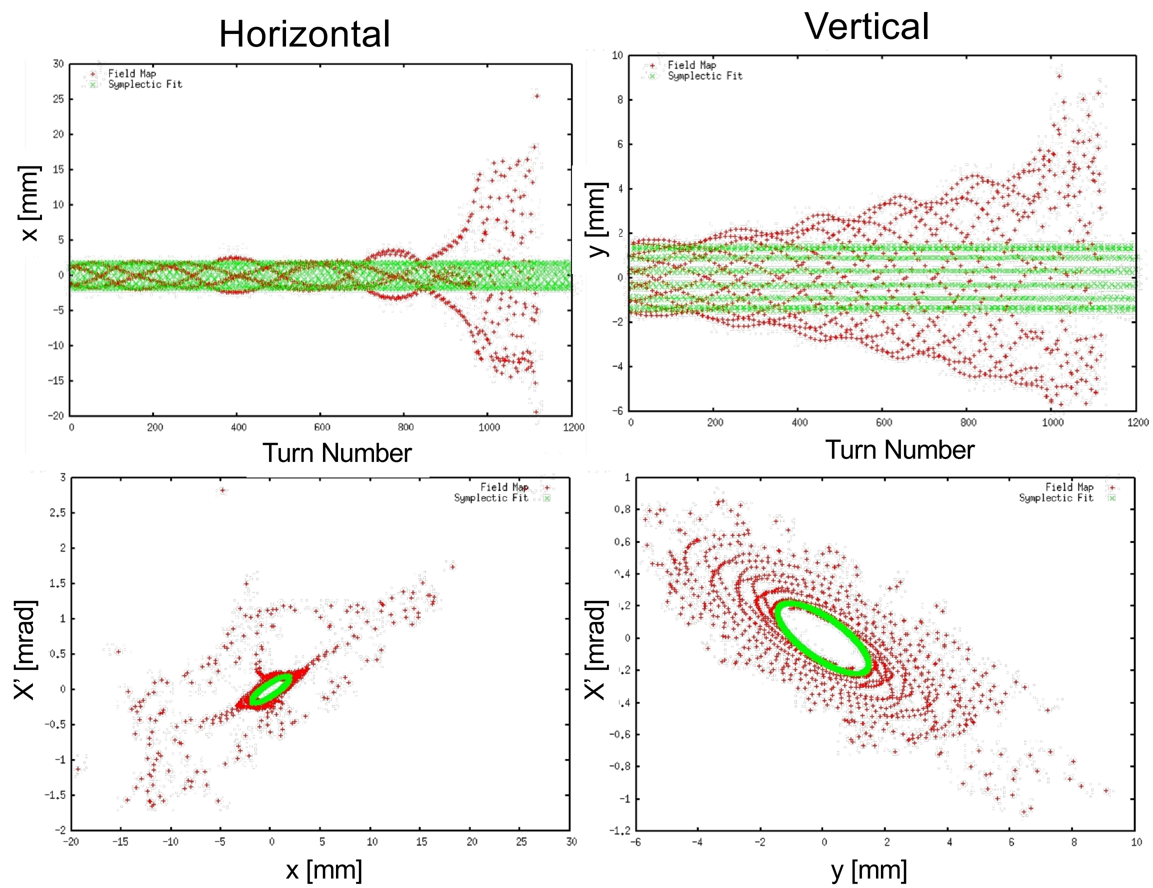


AGS Optics (low energy)



Symplectic AGS Siberian Snake modeling

- AGS Siberian snake field maps violates symplecticity, especially at AGS injection energy
- Symplectic tracking (green) is stable for over 10,000 turns



Reduction of AGS resonance driving terms

Polarization is preserved in the AGS with two partial helical dipole snakes (10% and 6% rotation)

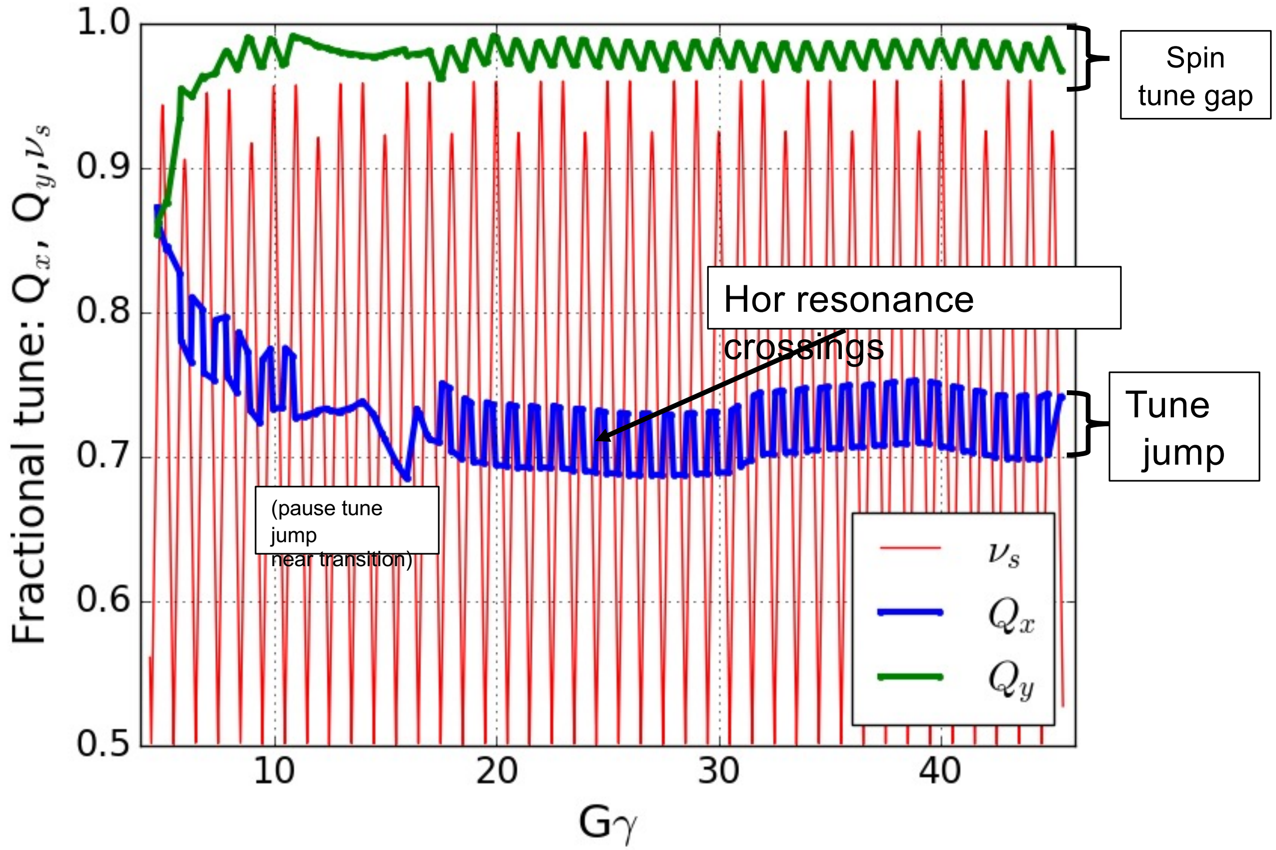
Provides spin tune 'gap' where imperfection and vertical intrinsic resonance condition are never met

- $\nu_s \neq N$ (full spin flips)
- $\nu_s \neq N +/- Q_y$

Horizontal resonance condition still met

- $\nu_s = N +/- Q_x$
- Horizontal resonance are weak, but many (82 crossings)
- Currently handled with fast tune jump

  $\Delta Q_x = 0.04$, 100 μs



Partial snakes drive horizontal depolarizing resonances

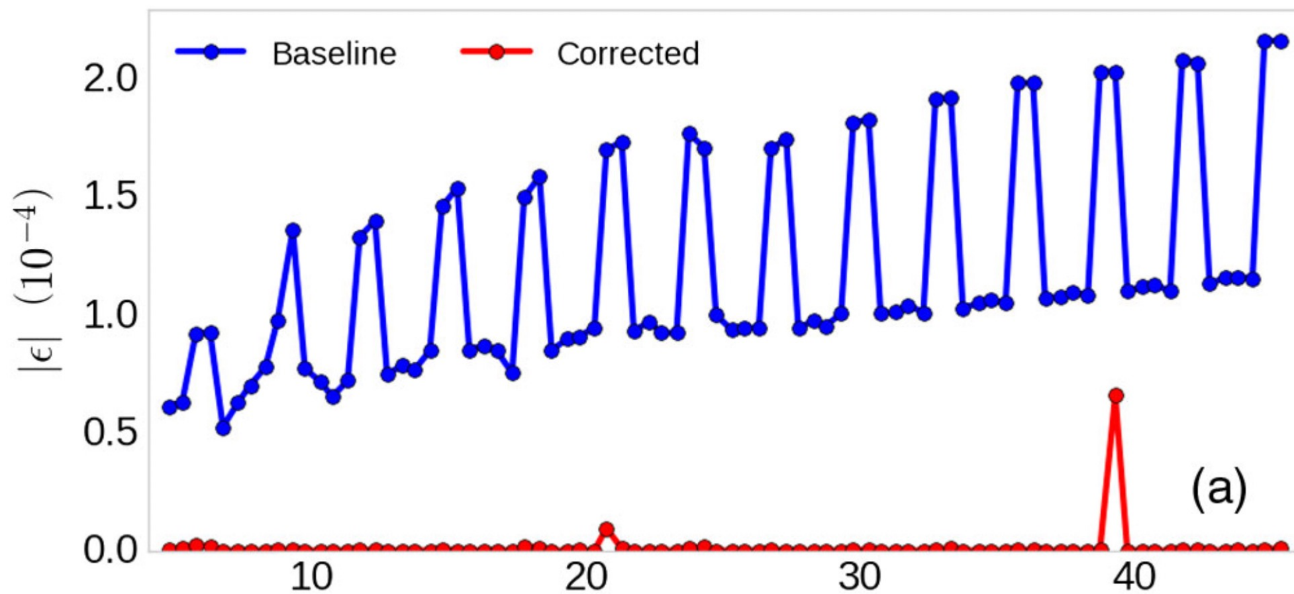
→ Compensate by other coupling elements, e.g., skew quads

Reduction of AGS resonance driving terms

- Two snakes, separated by 1/3 circumference
  - Modulated resonance amplitude highest near Gɣ = 3N (when snakes add constructively)
- Horizontal resonances occur every 4-5 ms at the standard AGS acceleration rate

ML/AI:
Physics informed Learning of the optimal skew quad strength + optimal timing.

Horizontal Resonance Amplitudes in AGS



Brookhaven National Laboratory
CLASSE
Cornell Laboratory for Accelerator-based ScienceS &Education

AGS Spin Resonance Correction Skew Quadrupoles

- A set of 15 pulsed skew quadrupoles, each with an individual power supply
- Designed to excite coupling resonance to compensate the 82 depolarizing resonances associated with horizontal betatron motion in the AGS partial snakes

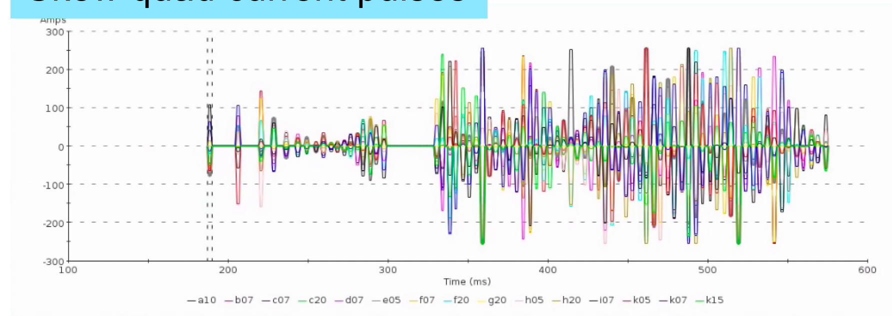
- 15 knobs, 82 different resonances
  - Expected effect is 10-15% gain in polarization
  - A +/-2% measurement takes 5-10 minutes

See presentation by Vincent Schoefer

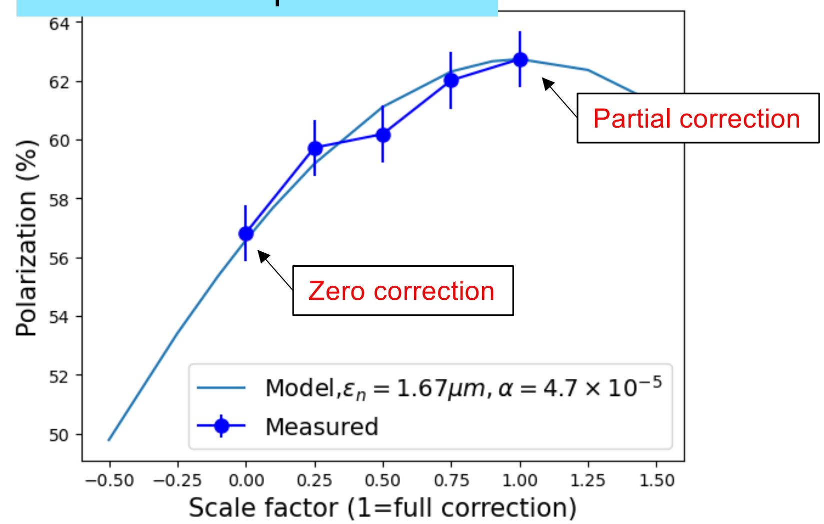
- Run 24: Observation of polarization gain factor (+10%) during acceleration (similar to existing tune jump), with the about the top half the pulses enabled)

- Plans for further improvements (enabling more pulses for 5-10% more gain):
  - Addressing model inaccuracies at low energy
  - Iteration on orbit centering
  - Possible optimizations based on ML methods

Skew quad current pulses



Correction amplitude scan



Bayesian UQ

- Bayesian UQ to probe and quantify sources of simulation error.
- Inputs are *probability distributions*:
  - “**Likelihood**”: distribution of data given params
    normal centered on simulation

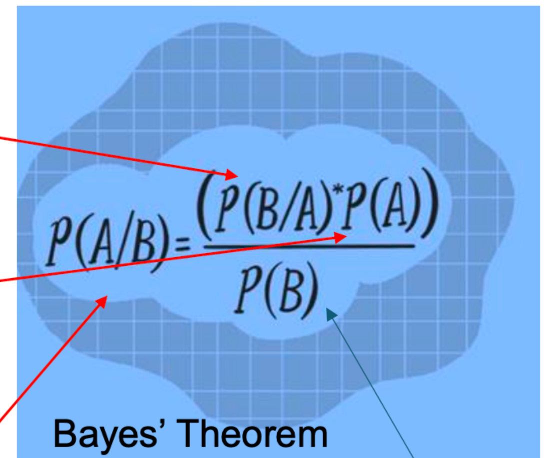
$$\mu = \mathrm{sim}(\mathrm{perturbed};\mathrm{params}) - \mathrm{sim}(\mathrm{unperturbed};\mathrm{params})$$
$$\sigma = \sqrt{2} \times \mathrm{bpm\ err.}$$

  - “**Priors**”: expert knowledge of *parameters*, *e.g.*

Some additive parameter: $\mathrm{additive}_i \sim \mathrm{Normal}(\mu_i, \sigma_i)$

Some multiplicative parameter: $\mathrm{multiplicative}_i \sim \mathrm{LogNormal}(\mu_i, \sigma_i)$

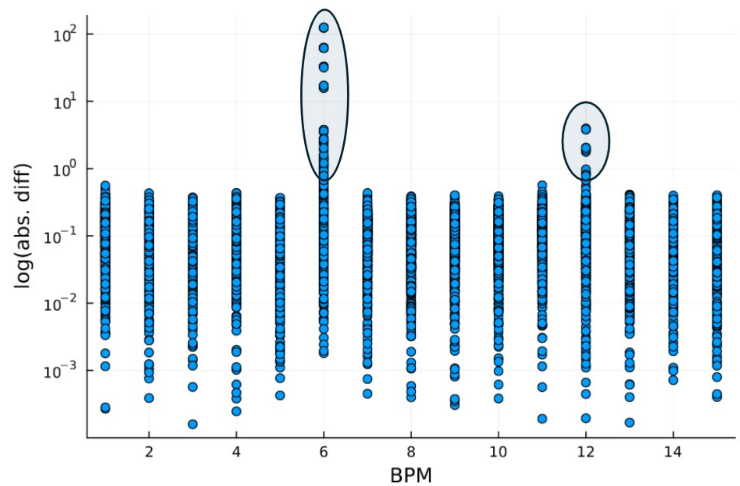
- Output is the “**posterior**”: distribution of the parameters given the data
- Sample using Markov chain Monte Carlo methods via the Julia “Turing” package.



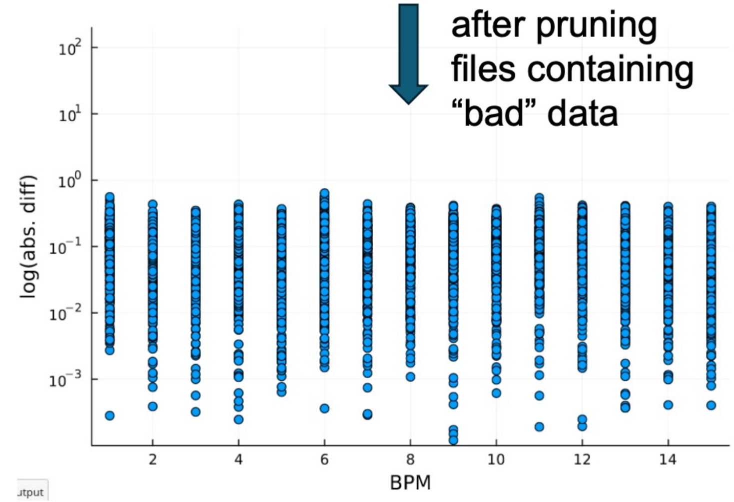
(“Evidence” not needed for MCMC)

Dataset

- The “2022” dataset contains orbit measurements where each corrector in turn is set to -22A, 0A and +22A.
- 3-5 measurements per corrector setting.
- Focus on x-plane orbit. Build orbit responses
  - POSITIVE: +22A – 0A
  - NEGATIVE: -22A – 0A
- Studied fluctuation between meas. of each orbit response over all possible pairings.
  - Subtract simulation result to account for current fluctuations
  - Recenter about mean to focus on spread.
- Found a number of outliers in BPMs 6 and 12 that required pruning.
- (BPM 6 has known issues)



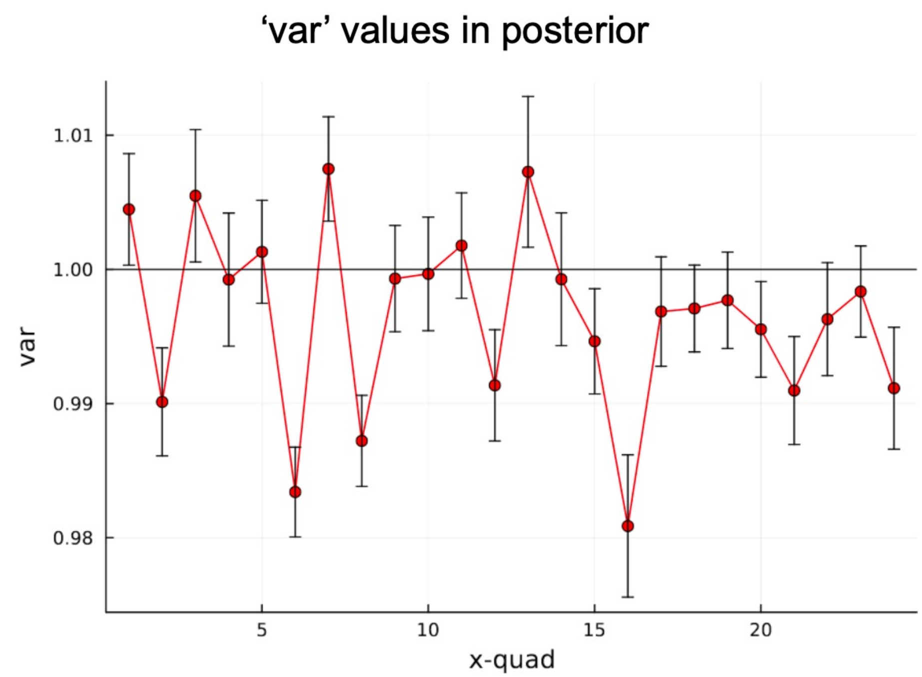
after pruning files containing “bad” data



Brookhaven National Laboratory

Preliminary UQ result

- Performed UQ conditioned on 45 orbit response measurements
  - 1 pairing (after+before) for each perturbed corrector
- Probabilistic model includes prior/likelihood distributions for
  - 'var' values (LogNorm 1.0 ± 0.01)
  - Measurement errors (Normal, σ = sqrt(2) * BPM err ~ 0.21 mm)
- Does not account for
  - Measurement errors on readback currents on static magnets or correctors
  - Any uncertainty on machine characteristics or other internal params (e.g. coefficients in transfer func)
- Observe interesting pattern in output var values, some well distinguished from unity.
- Possible hint for origin of discrepancies?



SciBmad a ML-oriented Toolkits (Libraries)

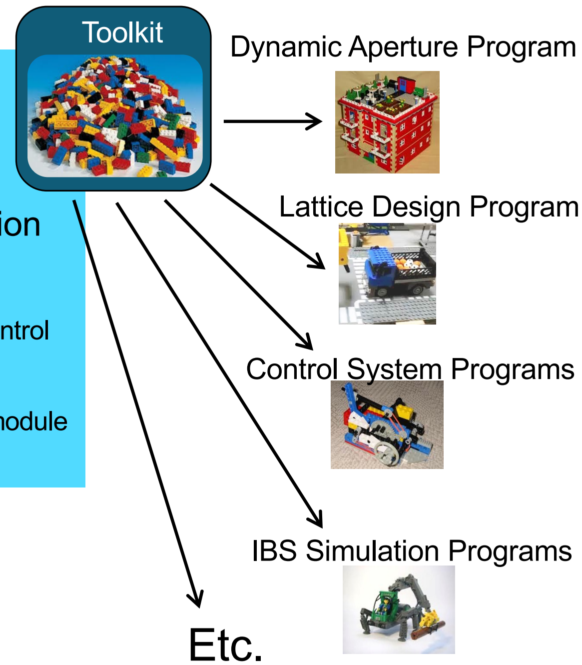
Advantages the toolkit:

Fully differentiable (reverse and forward)

➔ excellent for Neural Network optimizations

➔ Excellent for Bayesian optimization with slope information

- Cuts down on the *time* needed to develop programs.
- Cuts down on programming *errors* (via module reuse).
- Provides a simple mechanism for lattice function calculations from within control system programs.
- *Standardizes* sharing of lattice information between programs.
- Increased *safety*: Modular code provides a firewall. For example, a buggy module introduced into the toolkit will not affect programs that do not use it.



This project is

- funded by DOE-HEP
- has a growing list of collaborators
- has a weekly wise people meetings
- ➔ **is looking for collaborators**

Summary

- DOE-NP funded project for the enhancement of proton polarization using ML/AI. Goal: 5%.
- Bayesian Optimization for **automatically optimizing and maintaining emittances** in Booster and AGS are available in the control room.
- **Reinforcement learning** routings for optimal bunch merging are being developed and have been tested in the control room.
- Models of LtB line, Booster, BtA line, and AGS have been much improved, will be the basis for **differentiable digital twins**.
- **Excellent team** from BNL, Cornell, JLAB, SLAC, FNAL, and RPI has formed, three PhD students have graduated (Bohong Huang, Lucy Lin, Matt Signorelli. Latter two are postdocs at BNL).
- **Reduction of resonance driving terms** already works above transition energy, may use ML belove γ_t.
- **SciBmad, the differentiable ML oriented Beam modeling toolkit** is ready for application
- A **continuation proposal has been accepted** extending to (a) polarized sources and linac, (b) Reinforcement Learning, (c) differentiable Digital Twins, (d) emittances of unpolarized beams.
- The EIC-BeamAI collaboration is extended to the international EIC accelerator collaboration.

Brookhaven National Laboratory CLASSE Cornell Laboratory for Accelerator-based ScienceS &Education

Acknowledgement

Brookhaven National Laboratory

- Kevin Brown, Michael Costanzo, Matt Cowan, Weinin Dai, Bahwin Dhital, Zhihua Dong, Yuan Gao, Levente Hajdu, Kiel Hock, Georg Hoffstaetter de Torquat, Bohang Huang, Natalie Isenberg, Lucy Lin, Chuyu Liu, John Morris, Linh Nguyen, Yihui Ren, Vincent Schoefer, Shruti Tajne, Nathan Urban, Shinjae Yoo, Keith Zeno

CLASSE — Cornell Laboratory for Accelerator-based ScienceS & Education

- Joseph Devlin, Eiad Hamwi, Georg Hoffstaetter de Torquat, Daria Kuzovkova, Lixing Li, David Sagan, Matthew Signorelli, Jonathan Unger, Sophia Yang

Rensselaer

- Yinan Wang

SLAC National Accelerator Laboratory

- Auralee Edelen, Ryan Roussel

Jefferson Lab — Thomas Jefferson National Accelerator Facility

- Armen Kasparian, Malachi Schram, Kishan Rajput

Fermilab

- Tia Miceli

radiasoft

- Nathan Cook, Jon Edelen, Chris Hall

The main goal of the original project (FY 2023- 25)

- Goal: 5% more polarization out of the AGS.

➔ Achieved was about the same polarization without tune quad jumps in the AGS, even though it was only effective above transition.

➔Much promise for even more polarization increase

- ML method developments

➔Beam brightness BO of Booster injection works and is available in the control room

➔Beam brightness BO of AGS injection works and is available in the control room

➔Improved modeling of all parts from end of linac to end of AGS

➔Development of digital twins.

➔Digital twin parameter determination by UQ

The new continuation project extends our methodology to brightness increases for more luminosity and to optimization of the polarized source, including x-y coupled beams.

Annual budget and the total received to date

	Year 1	Year 2	Year 3	Totals
a) Funding allocation				$ 1.5 M
b) Actual cost to date				$ 1.0M

We have underspent because a postdoc (Lucy Lin) who is one of our main contributors could be funded from other sources. We obtained a no-cost extension and are hiring a new post doc on this (Eaid Hamwi).