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

Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

ML / AI proposal supported by DOE-NP

Georg Hoffstaetter de Torquat

Collider-Accelerator Department, BNL and Cornell University

Georg.Hoffstaetter@cornell.edu

A collaboration of BNL, Cornell, TJNAF, SLAC, RPI

@BrookhavenLab

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

Desired result: higher proton polarization

- What high-impact operational challenge can be addressed by MI/AI?
  ➔ Polarized protons.
- From the source to high energy RHIC experiments, 20% polarization is lost.
- Polarized luminosity for longitudinal collisions scales with P^4, i.e., a factor of 2 reduction!
- The proton polarization chain depends on a hose of delicate accelerator settings form Linac to the Booster, the AGS, and the RHIC ramp.
- Even 5% more polarization would be a significant achievement.

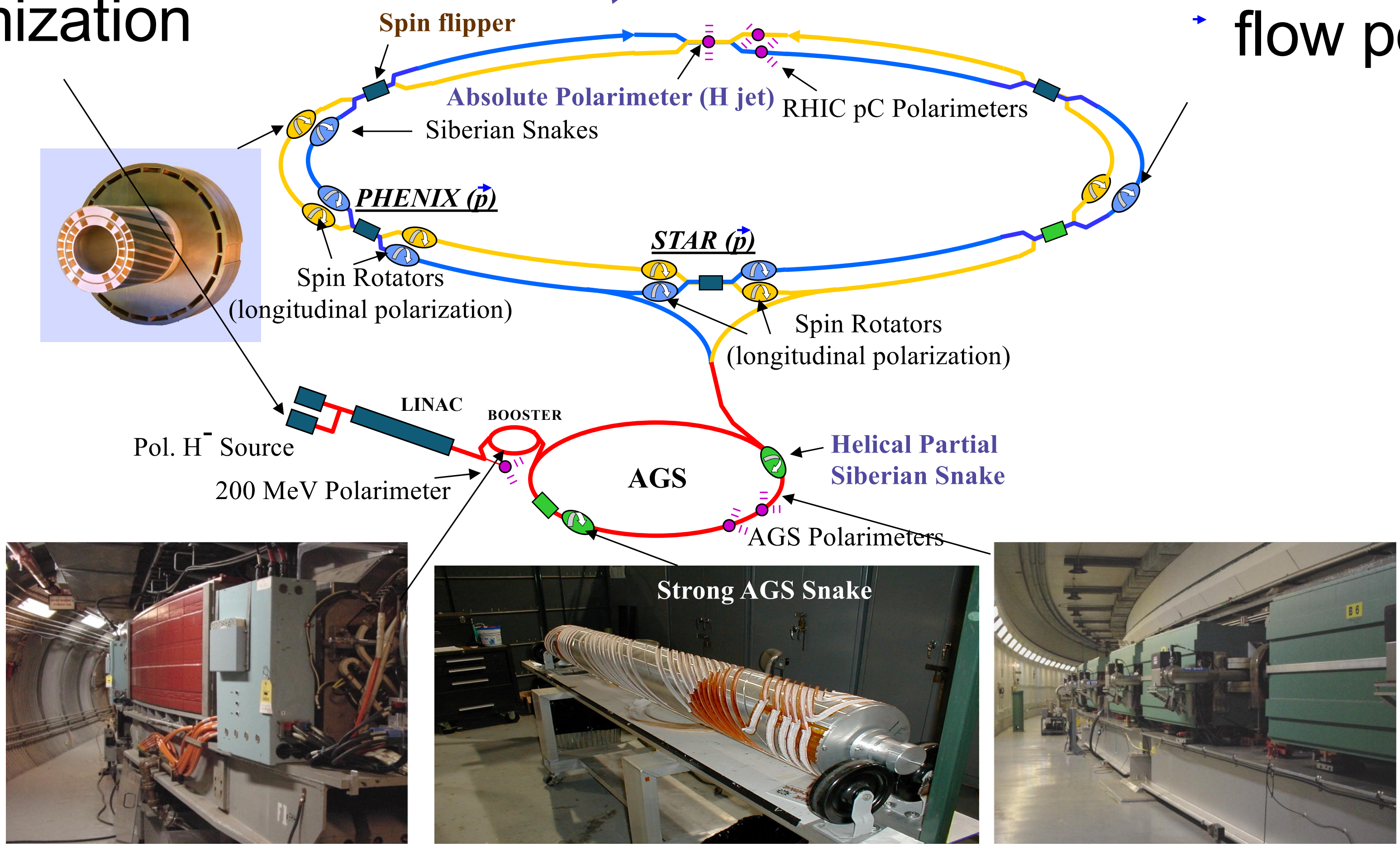
Outline

- Objective of proposed work: higher proton polarization in RHIC and the EIC.
- Polarized-proton acceleration chain.
- Potential avenues toward higher proton polarization.
- (1) Emittance reduction
- (2) More accurate timing of timed elements
- (3) Reduction of resonance driving terms
- Gaussian Process (GP) Bayesian Optimization (BO) and physics informed learning.
- When is ML/AI better for accelerator operations than other feedbacks and optimizers?
- Progress report
- Plans

The polarized proton accelerator chain

Linac – Booster – AGS Optimization

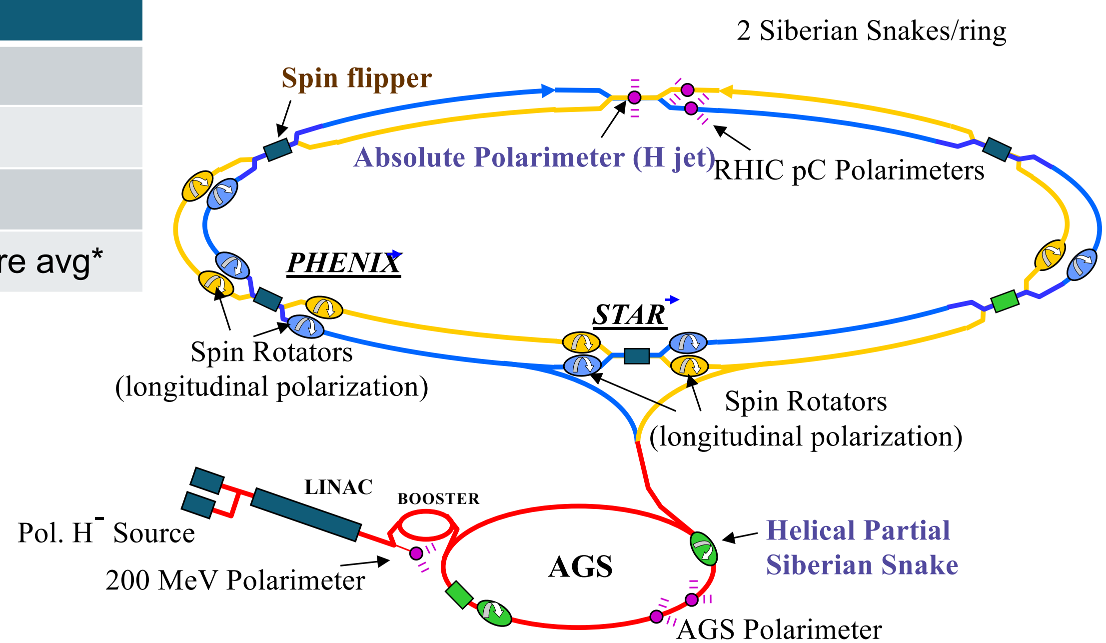
Readily available, large data flow possible



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 %

Topics that can improve polarization

- (1) Emittance reduction
- (2) More accurate timing of tune jumps
- (3) Reduction of resonance driving terms

Optimizers for different applications

less ← assumed knowledge of machine → more

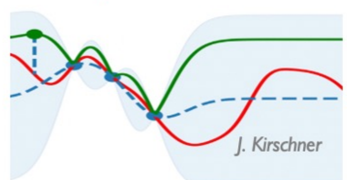
Model-Free Optimization

Observe performance change after a setting adjustment

→ estimate direction or apply heuristics toward improvement

gradient descent
simplex
ES

Model-guided Optimization

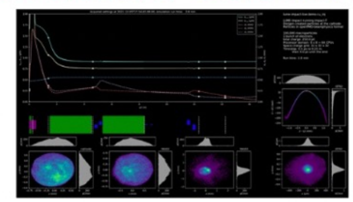


Update a model at each step

→ use model to help select the next point

Bayesian optimization
reinforcement learning

Global Modeling + Feed-forward Corrections



Make fast system model

→ provide initial guess (i.e. warm start) for settings or fast compensation

ML system models +
inverse models

Courtesy Auralee Edelen

Characteristics of involved 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. Models of the accelerator exist.
4. A history of much data is available and can be stored.

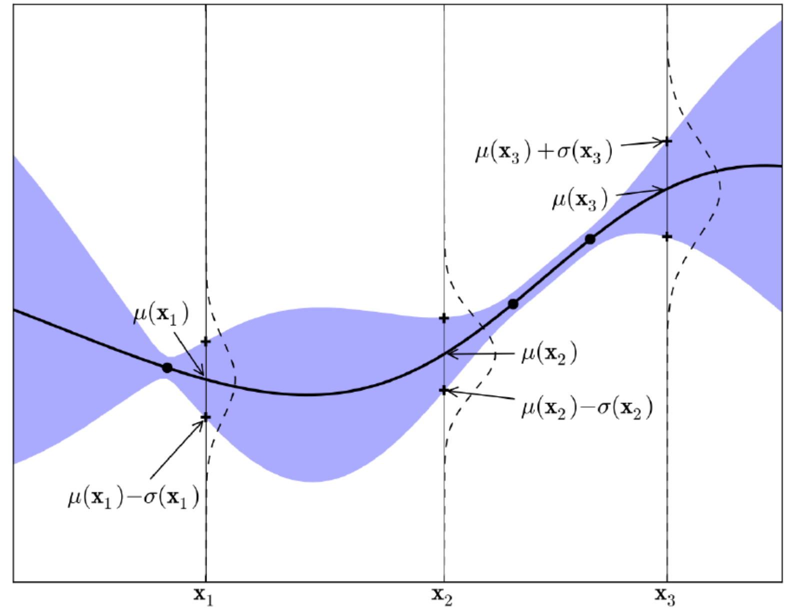
Is this type of problem suitable for Machine Learning?

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

Gaussian Process

- GP model built with scikit-learn library
- A probability distribution over possible functions that fit a set of points
- Mean function + Covariance function

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



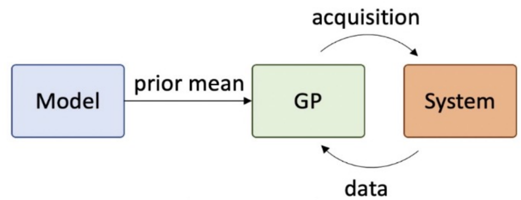
- Kernel: covariance function $k(x_i, x_j)$ of the input variables
- Covariance matrix $\mathrm{K} = k(X,X) = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_t) \\ \vdots & \ddots & \vdots \\ k(x_t,x_1) & \cdots & k(x_t,x_t) \end{bmatrix}$
- At a sample point x_i, Gaussian process returns mean $\mu(x_i|X) = m(x_i) + k(x_i,X)K^{-1}\big(f(X) - m(X)\big)$ and variance $\sigma^2(x_i|X) = k(x_i,x_i) - k(x_i,X)K^{-1}k(X,x_i)$

Merit of physics-informed optimization

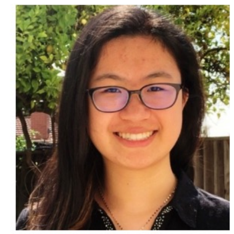
Neural Network System Models + Bayesian Optimization

Combining more expressive models with BO → **important for scaling up to higher-dimensional tuning problems (more variables)**

Good first step from previous work: use neural network system model to provide a prior mean for a GP

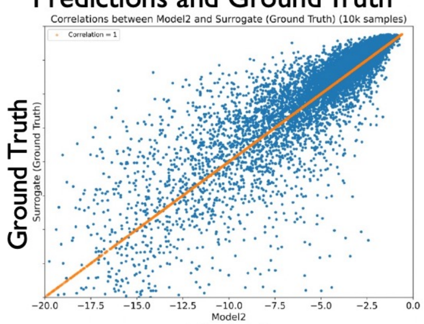


Used the LCLS injector surrogate model for prototyping
variables: *solenoid, 2 corrector quads, 6 matching quads*
objective: *minimize emittance and matching parameter*

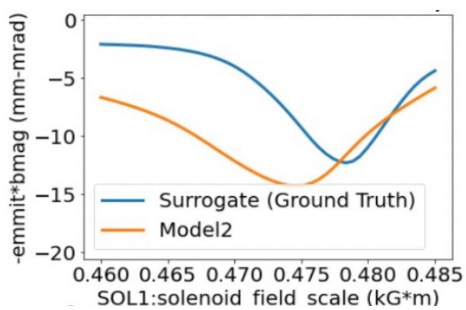


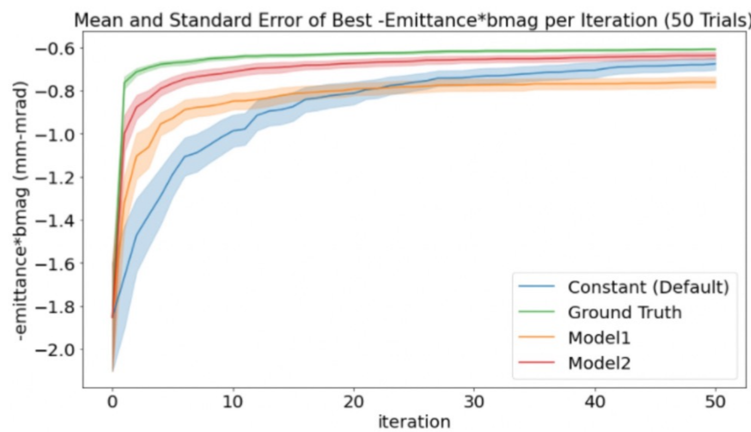
Summer ’22 undergrad intern
Connie Xu

Correlations Between Predictions and Ground Truth



Model 2





Beta = 2.0

Even prior mean models with substantial inaccuracies provide a boost in initial convergence
→ now testing on machine and refining approach

Forthcoming paper at NeurIPS ML for Physical Science

Courtesy
Auralee Edelen

Advantages of Bayesian Optimization

Summary of optimization methods

	Nelder-Mead	Gradient descent	Powell / RCDS	L-BFGS	Genetic algorithm	Bayesian optimization
Sample efficiency	Medium	Medium	Medium/high	Medium/high	Low	High
Computational cost of picking the next point	Low/Medium	Low	Low	Low	Medium (e.g. sorting)	High (esp. in high dimensions)
Multi-objective	No	No	No	No	Yes	Yes
	(but can use scalarization)					
Sensitivity to local minima	High	High	High	High	Low	Low (builds a **global** model of f)
	(but can use multi-start)					
Sensitivity to noise	High	High	High (Powell) Low (RCDS)	High	Medium	Low (can model noise itself)

Summary of optimization methods

	Nelder-Mead	Gradient descent	Powell / RCDS	L-BFGS	Genetic algorithm	Bayesian optimization
Requires to compute or estimate derivatives of f	No	Yes	No	Yes	No	No
Evaluations of f *inherently* done in parallel	No	No	No	No	Yes	No
Hyper-parameters	Initial simplex	Step size: α (+momentum: β)	# fit points Noise level	Accuracy of hessian estimate	• Population size • Mutation rate • Cross-over rate • Number of generations	• Kernel function • Kernel length scales, amplitude • Noise level • Acquisition function

Why is Bayesian Optimization suitable?

1. The data to optimize on has significant uncertainties

➔ Derivatives of measured functions are not required.

2. Models of the accelerator exist

➔ the expected functional form can be included in the function search (Physics-informed learning)

3. A history of much data is available and can be stored

➔ All past data are included to model the function to be optimized.

Note: Reinforcement Learning (RL) can be promising because (a) accelerators have many state variables beyond the optimization objectives, (b) accurate models can reduce the require measurement points of data hungry RL.

➔ Ongoing analysis of BO vs. RL for accelerator control, which will be part of our follow-up proposal.

Topics that can improve polarization

- (1) Emittance reduction
- (2) More accurate timing of tune jumps
- (3) Reduction of resonance driving terms

Emittance reduction ➔ less depolarization

- Optimized Linac to Booster transfer
- Optimized Booster to AGS transfer
- Optics and orbit correction in Booster and AGS
- Beam-based model calibration from orbit responses in Booster and AGS.
- Bunch splitting in the Booster for space charge reduction and bunch re-coalescing at AGS top energy.

Polarized collider performance

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):

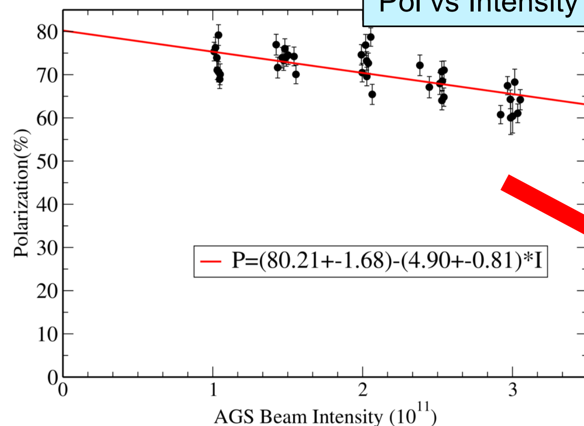
$$\text{FoM} = \begin{cases} \mathcal{L}\, P^2 & \text{transverse spin} \\ \mathcal{L}\, P^4 & \text{longitudinal spin} \end{cases}$$

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

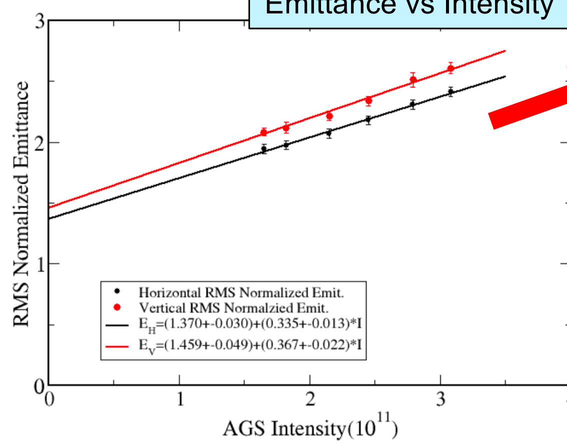
FoM dependence on intensity closer to linear in N than quadratic.

AGS extraction

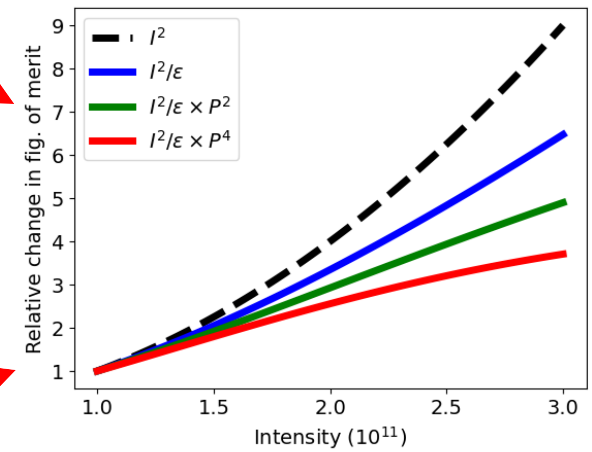
Pol vs Intensity



Emittance vs Intensity



Polarized beam collider FOM



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

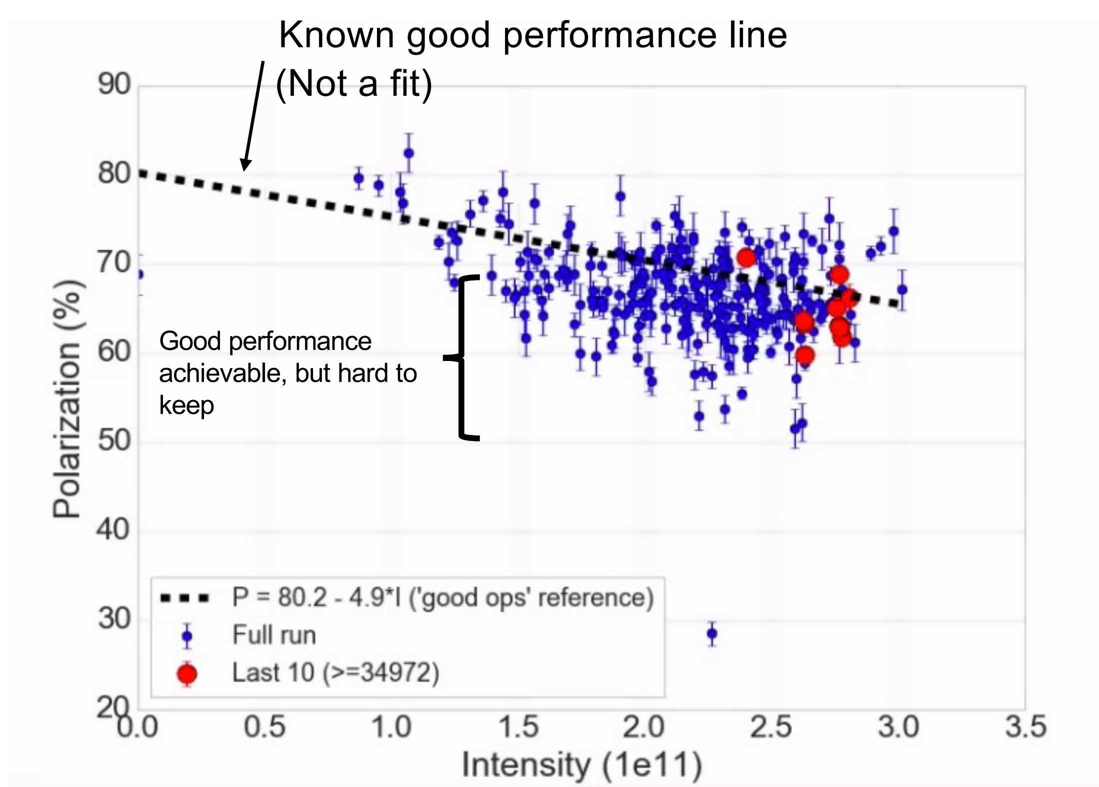
AGS Performance

Highest AGS performance is difficult to achieve ***and*** *maintain*

Value in just holding a known optimum

A combination of maintaining emittances and direct polarization interventions

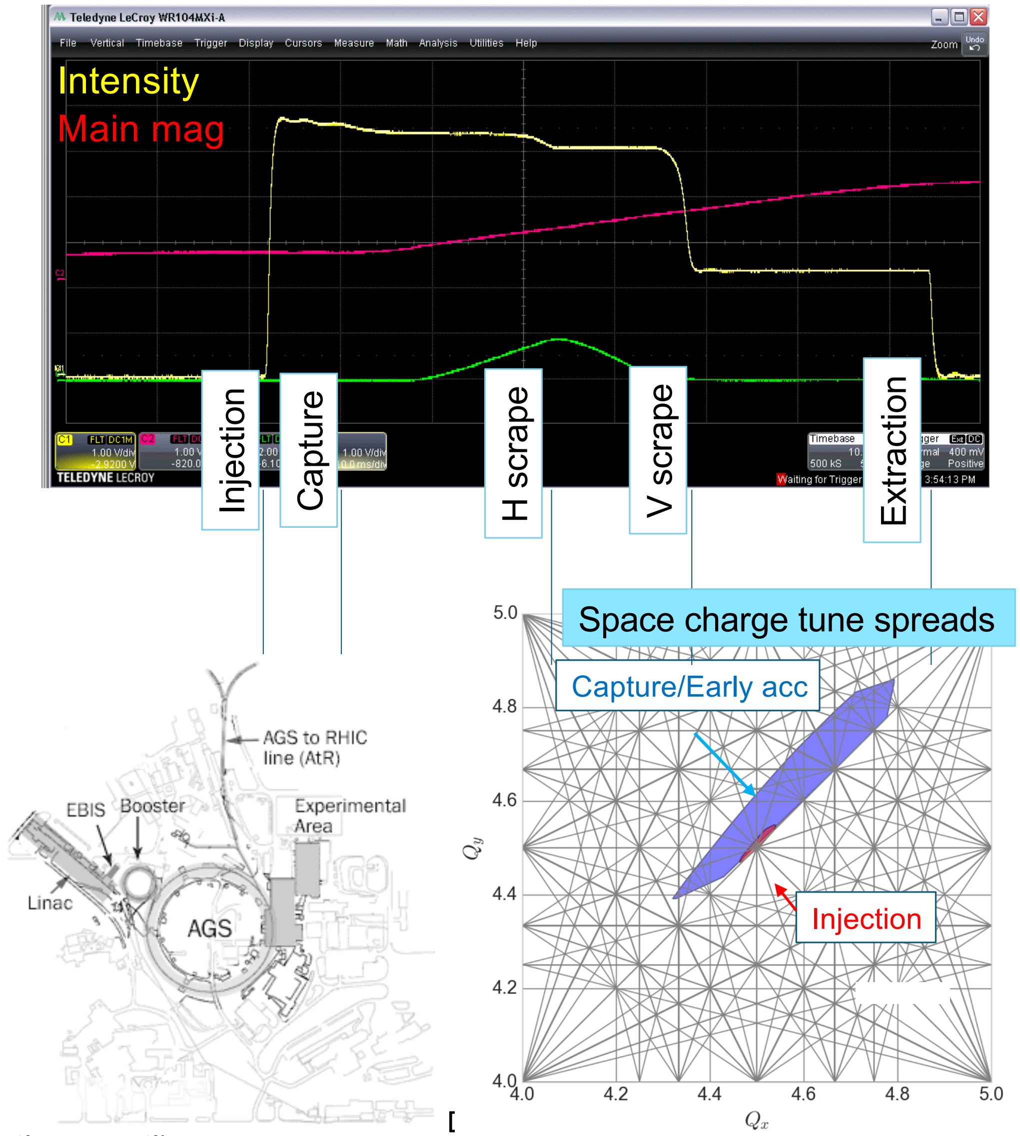
AGS Polarization vs intensity for RHIC fills (Run 24)



Booster injection

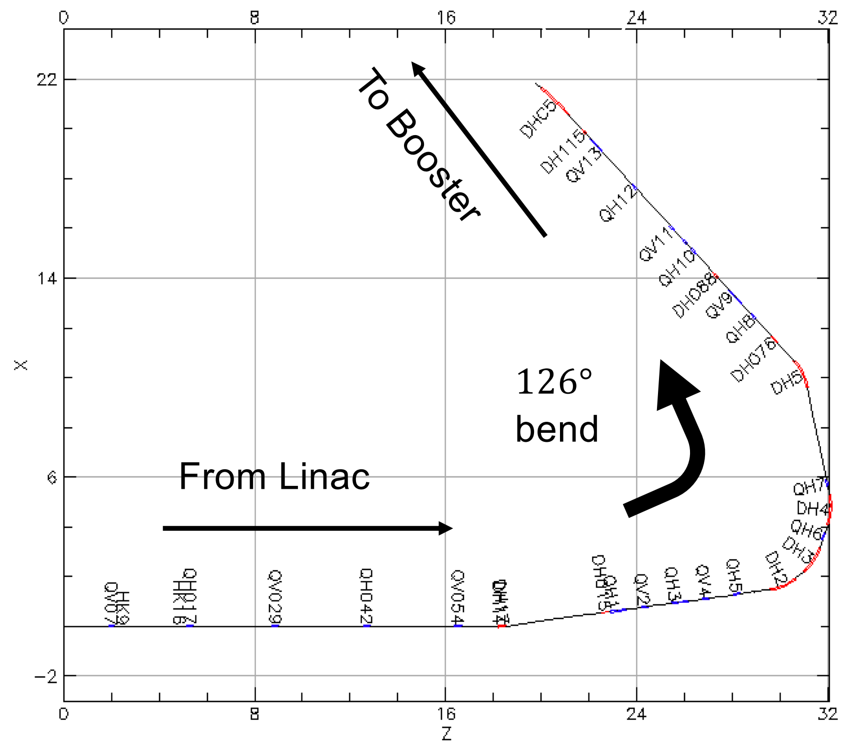
Booster injection/early acceleration process sets maximum beam brightness for rest of acceleration though RHIC

- Many ”knobs”
  - Linac to Booster trajectory/optics matching
  - Optimization of time on foil (Linac pulse length vs height)
  - Linac RF phases affect capture and acceleration efficiencies
  - Booster RF capture rate affects longitudinal emittance (and transverse, via space charge)
  - Booster orbit and optics affect foil scattering, matching and intensity transmission.
  - Betatron ‘stop band’ correctors for intensity, emittance preservation.
- Difficult instrumentation
  - WCM, BPMs don’t work until after capture
  - No transverse profile monitor in Booster
    - Scraping efficiency as proxy
    - Measurable in the extraction line via multiwire
- Difficult model
  - Linac to Booster longitudinal effects
  - Space charge
  - Stripping foil

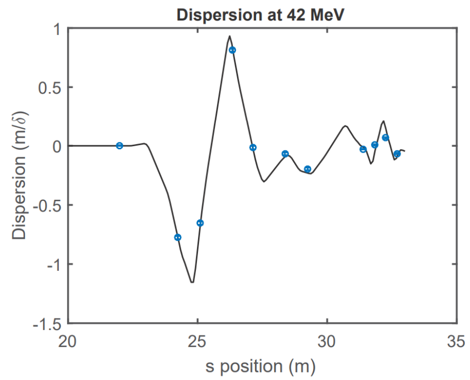


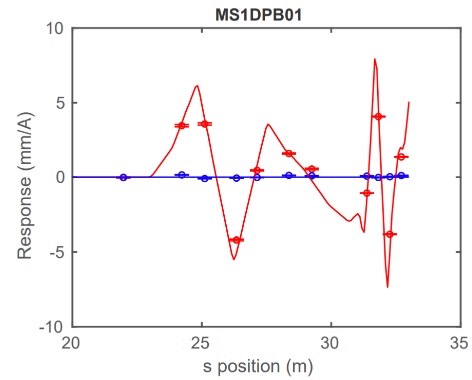
Booter injection

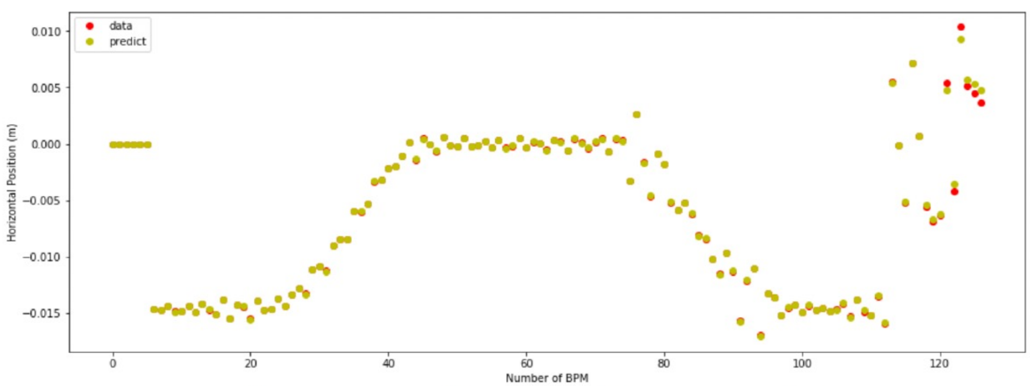
- Booster injection process sets maximum beam brightness for rest of acceleration through RHIC
- Known emittance effect on polarization loss
- Intentional horizontal and vertical scraping reduce emittance to RHIC requirements
- Goal: minimize emittance / maximize beam intensity after scraping
- Controls: Linac to Booster (LtB) transfer line optics
- Method: Bayesian optimization (BO)



Motivation: Digital twin at CBETA





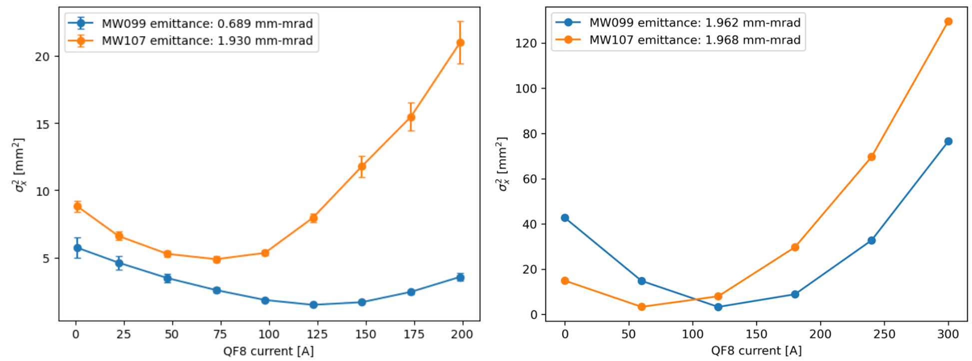


- Had success in building digital-twins for CBETA: combine custom version of Bmad/Tao with EPICS
- CBETA-V: measure beam trajectories and compare to the digital twin in real time on control-system screens
- Neural network can be trained to predict orbit response using Bmad simulation data
- NN model can predict beam behavior due to both linear (correctors) and non-linear (cavity) relationships

Beam in the Linac to Booster Transfer line

- 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.
- X and Y multi wires are not sufficient input for 4-D phase space tomography

➔ 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.

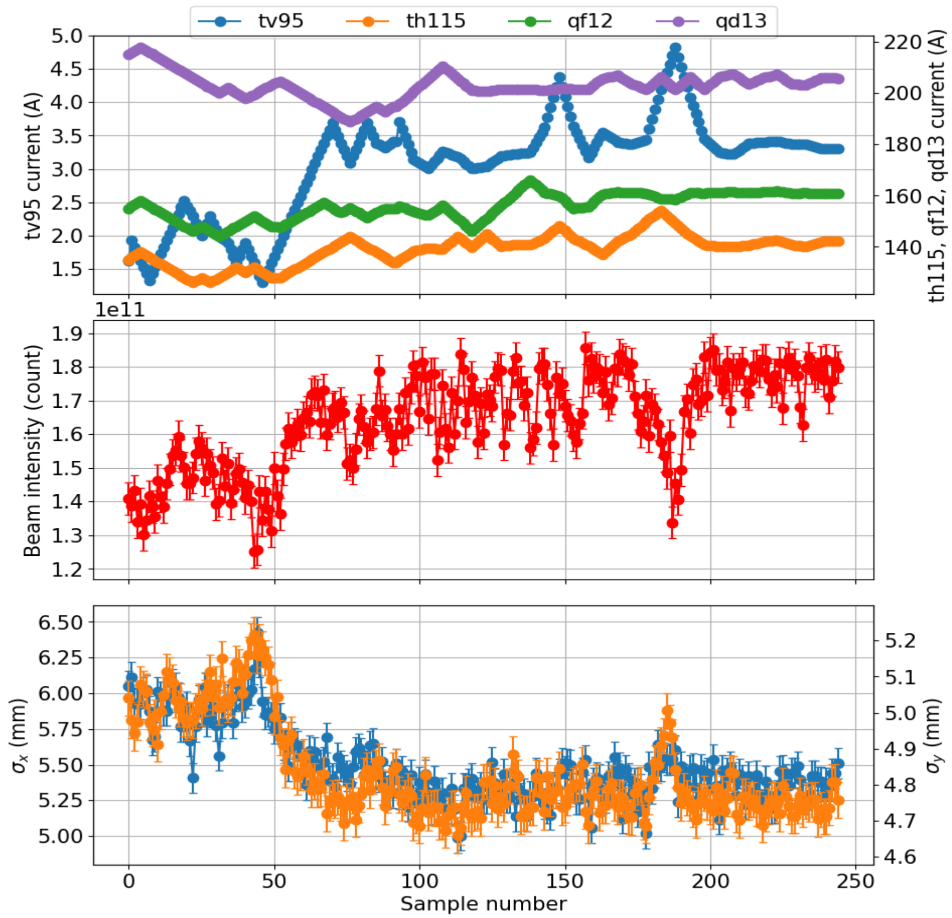


Simulated (left) and measured (right) quadrupole scan results for horizontal quad QF8 observed at two multi-wires (MW099, MW107) in the LtB line.

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

Booster injection: 2 correctors + 2 quadrupoles

- 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.

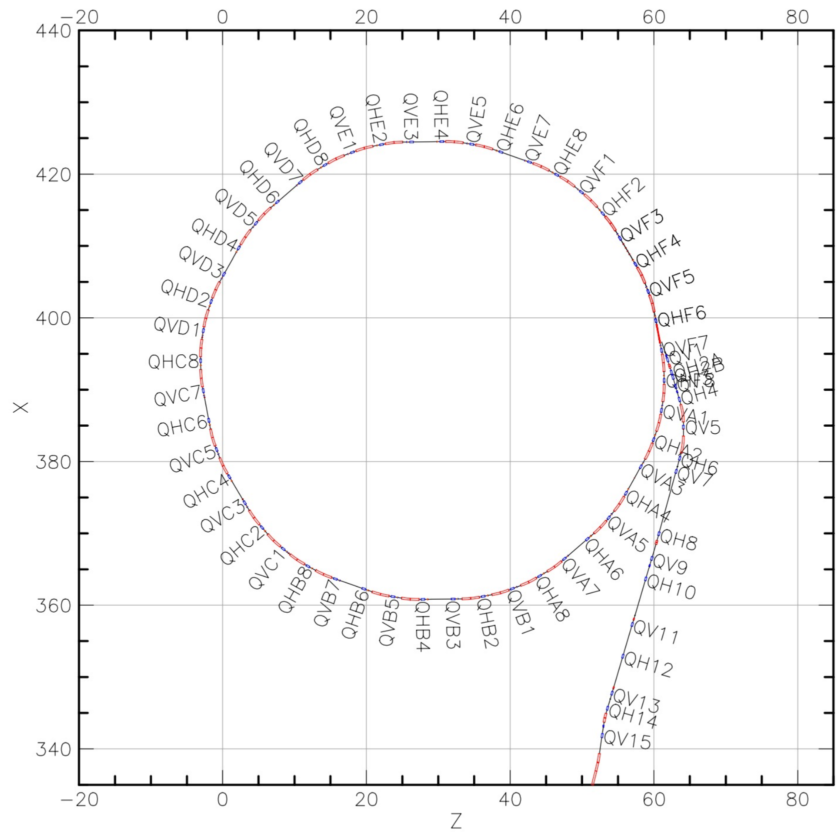
Top: power supply currents of two correctors (tv95, th115) and two quadrupoles (qf12, qd13) in the LtB line.

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

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

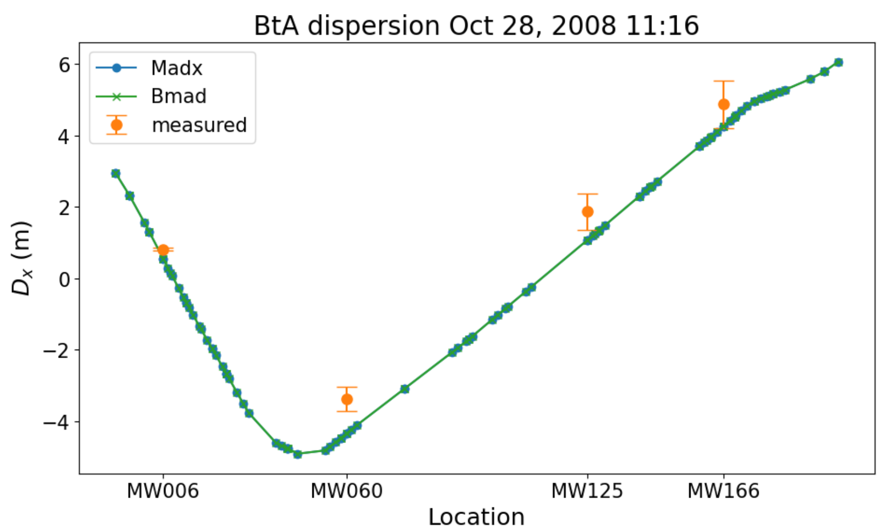
BtA Transfer Line Structure in Bmad

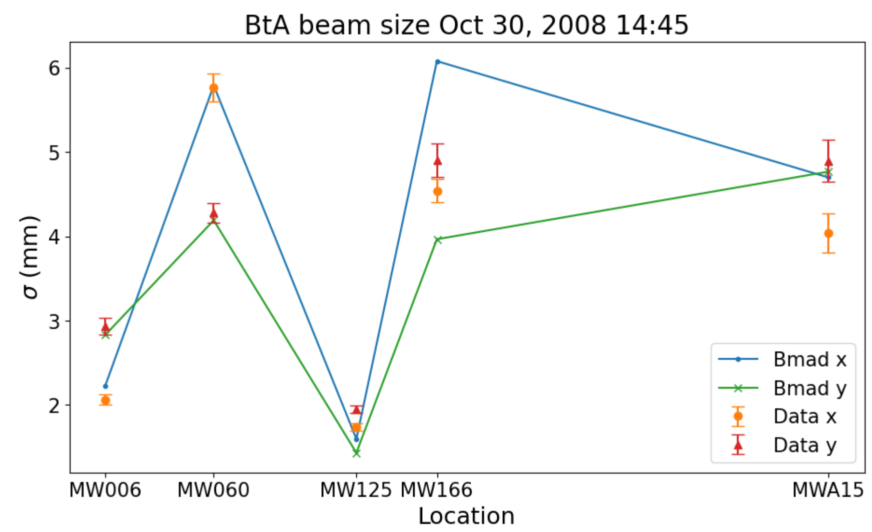
- Lattice can be divided into branches connected with forks to simulate connection to a transfer line
- Require documented coordinates for elements to construct correct geometry
- Beam parameters from the end of one branch is automatically inherited by the start of downstream branch → continuous tracking
- 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

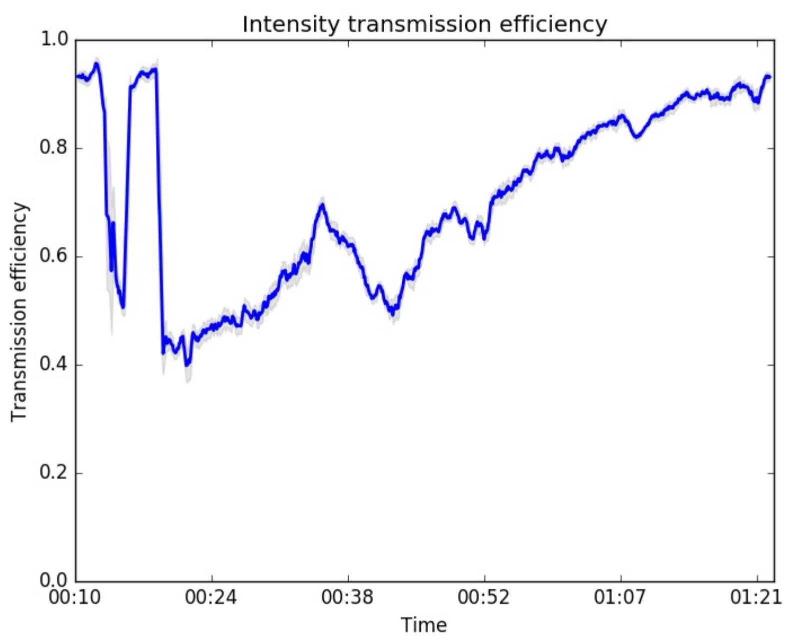
- Bmad tracking leads to horizontal dispersion matching measurements
- Beam size values from bunch tracking show agreements for upstream multi-wire measurements, disagreement downstream needs further investigation

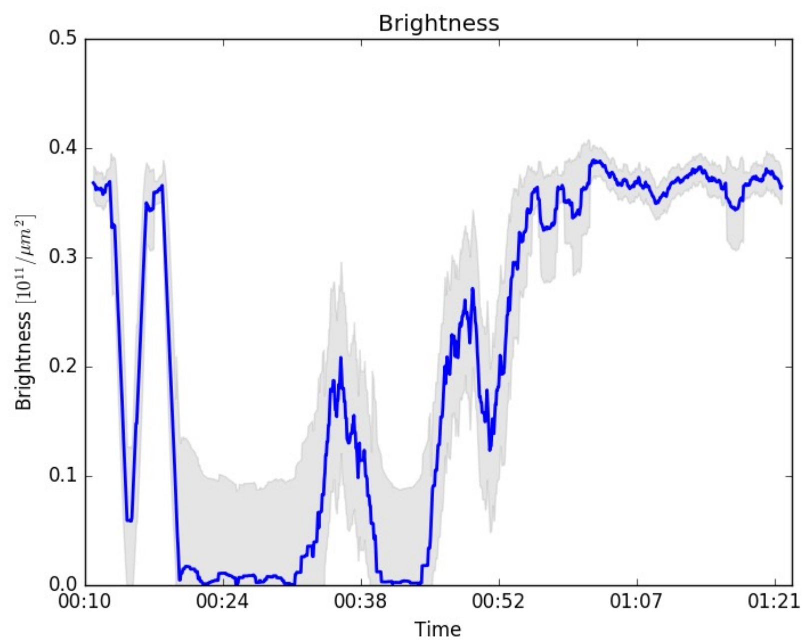




Bayesian Optimized Injection into the AGS

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.

AGS, Polarization and Snakes

- 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)

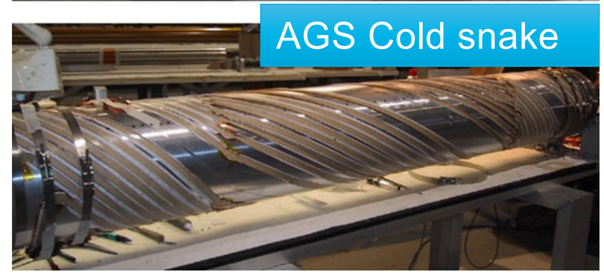
- Requires “near integer” tune
  - Orbit, optics unusually sensitive to errors

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

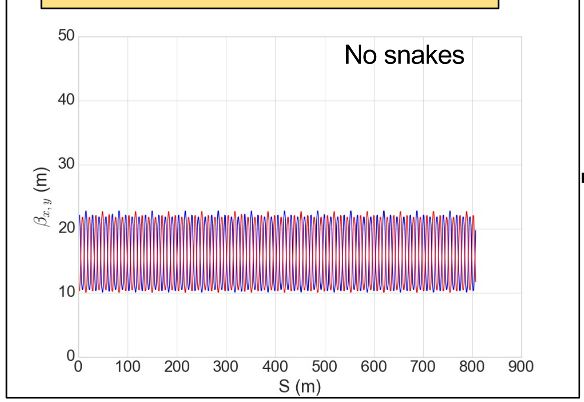
- The complex fields and lattice + high tune requirements are a challenge to modeling (Eiad’s talk)

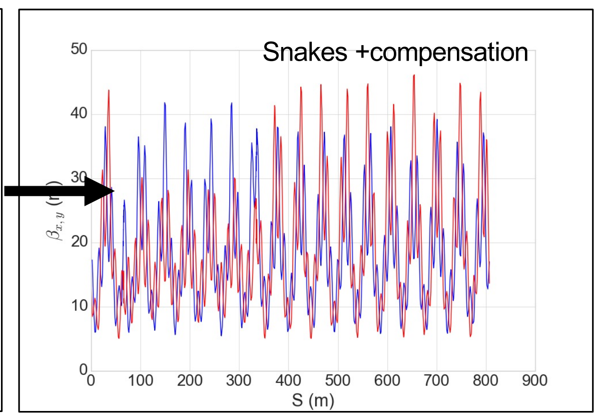
AGS Warm snake

AGS Cold snake



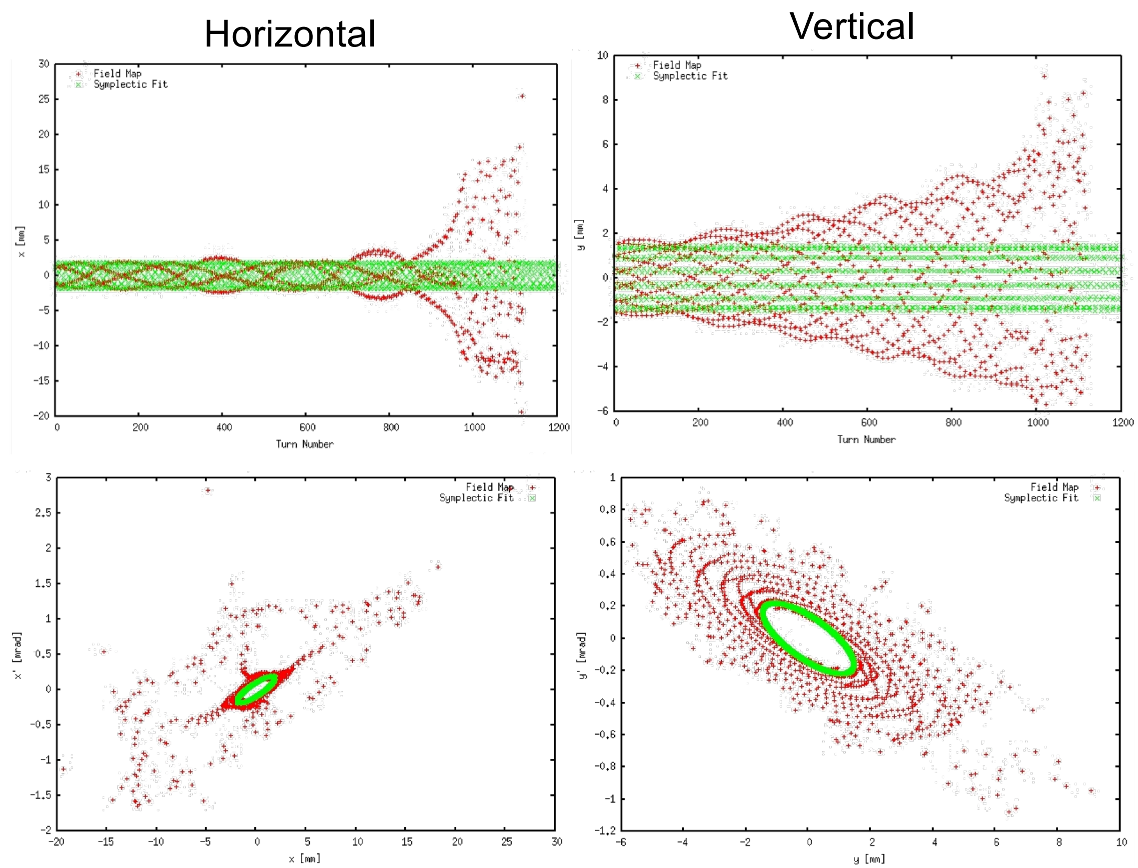
AGS Optics (low energy)





AGS Siberian Snakes modeling

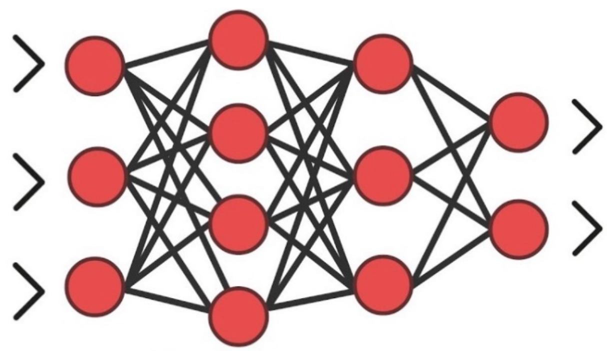
- AGS Siberian snake field maps violates symplecticity, especially at AGS injection energy
- Symplectic tracking (green) is stable for over 10,000 turns

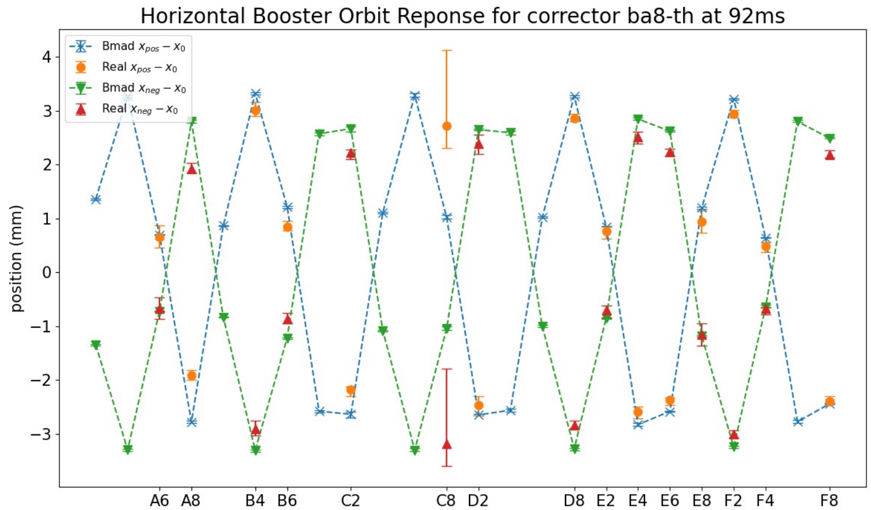


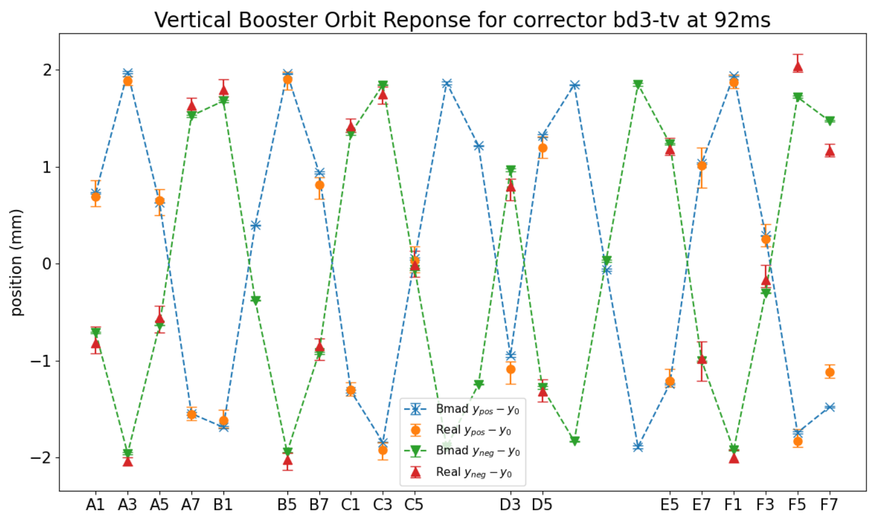
Response Error model for the ORM

- Scan through some common sources of error to see how much ORM changes
- Find relevant parameters to include for building error-detecting model
- **Goal**: establish a neural network that identify error source given a measured ORM

$$\begin{pmatrix} \Delta\nu_1 \\ \Delta\nu_2 \\ \cdots \\ \Delta\nu_{N-1} \\ \Delta\nu_N \end{pmatrix} = J^{+}_{model} \begin{pmatrix} \Delta R_{11} \\ \Delta R_{12} \\ \cdots \\ \Delta R_{n(m-1)} \\ \Delta R_{nm} \end{pmatrix}$$







Orbit response data in AGS 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
- Good agreements between AGS Booster data and Bmad model are reached, despite some faulty BPMs (i.e., PUEHC8)
- Small discrepancies (within 1 mm) beyond error bars is being investigated
- chi-squared/DF = 1.4 – physics reasons for discrepancy are being sought by **Uncertainty Quantification**.

➔ The main power supply transfer functions are not an explanation. Error sources are being analyzed.

Space-charge emittance increase

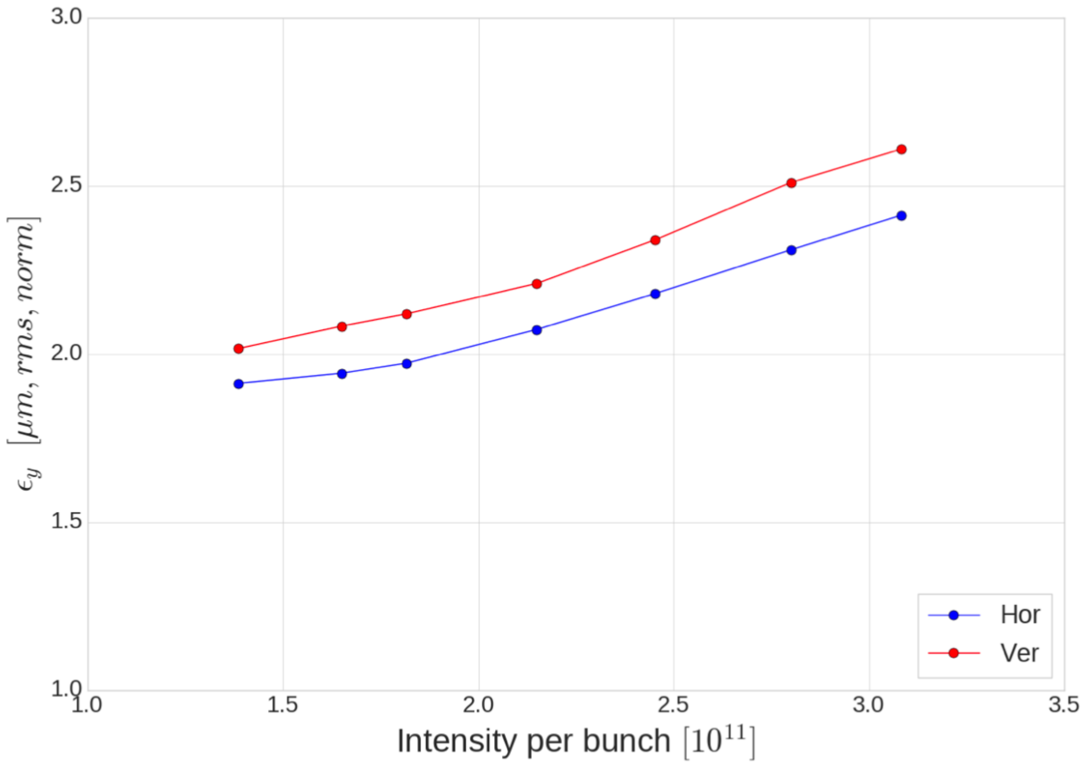
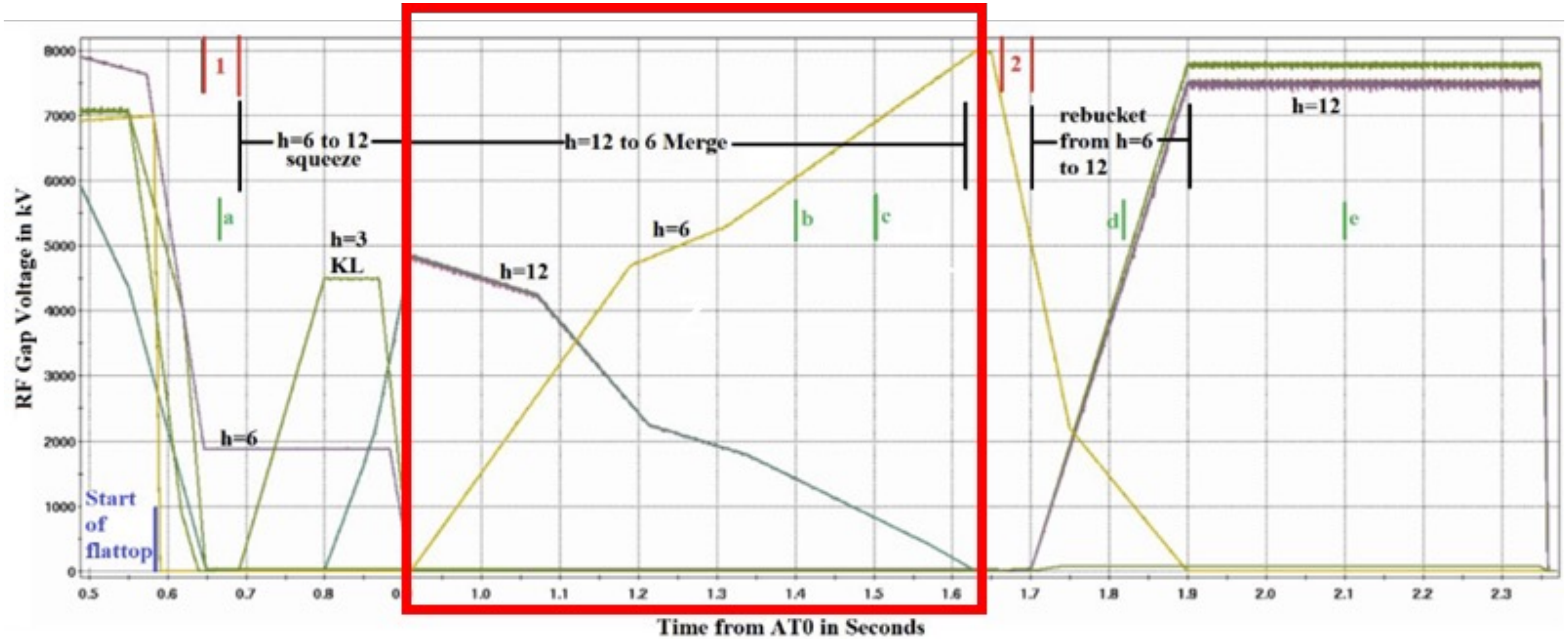


Figure 3.168: Normalized transverse emittances of polarized proton beam at AGS extraction energy ($\gamma = 25.5$) as a function of intensity.

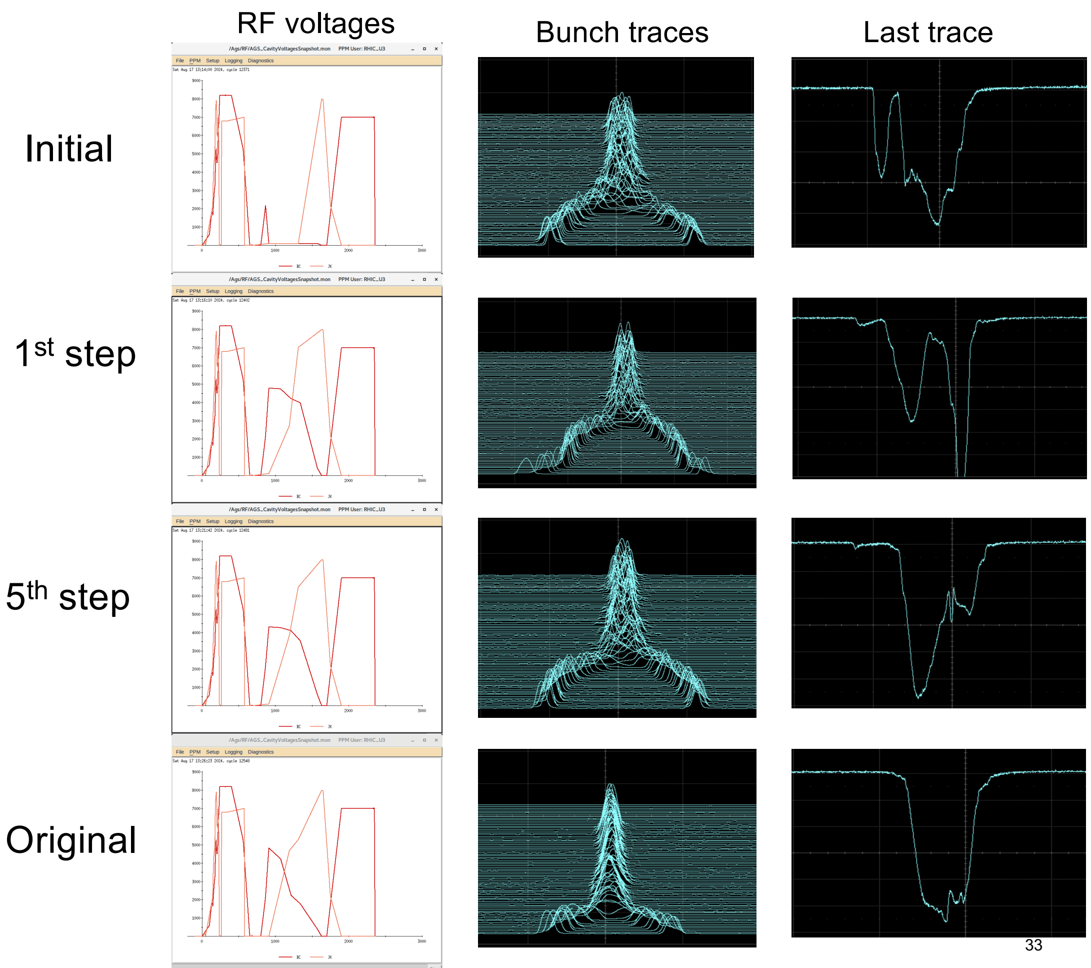
➔ Splitting bunches before AGS acceleration can reduce the emittance.

Bunch splitting in Booster / merging in AGS



Splitting in the booster and coalescing after AGS accelerator reduces space charge and emittance growth ➔ more polarization

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



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, therefore experimental setup uses simulated bunch lengths as reward.
- Plan: check whether Reinforcement Learning has advantages over BO.
- Plan: Include also RF phases as actors
- Determine useful state variables
  - measurable
  - related to the reward

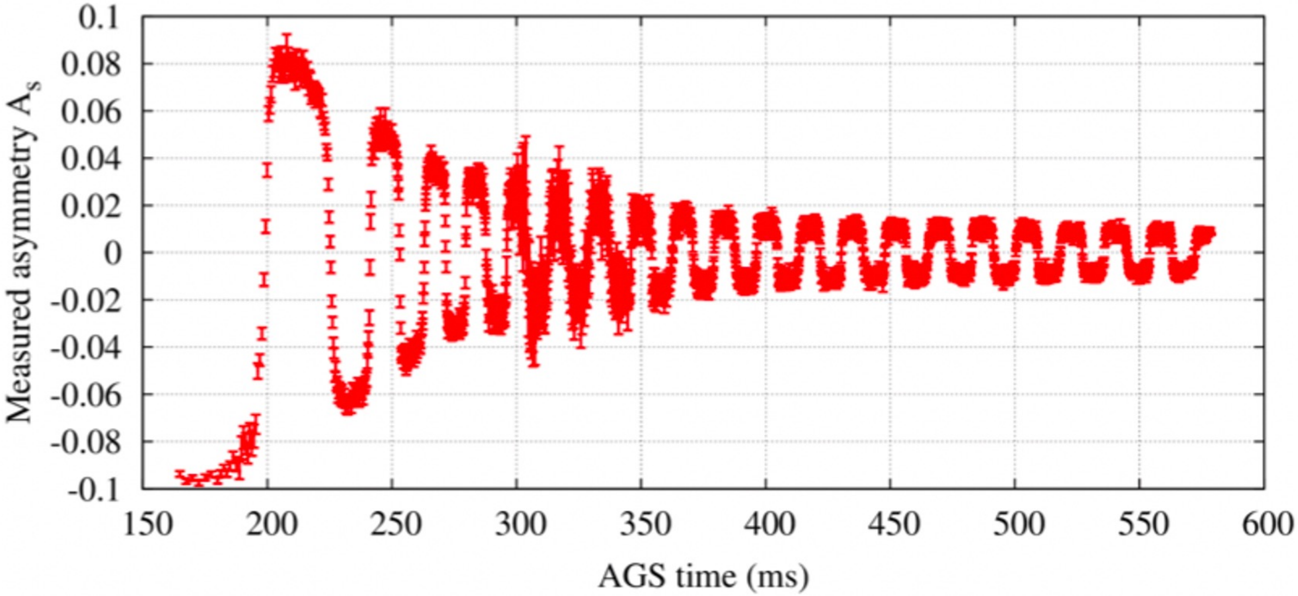
Timing of tune jumps

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

➔ better timing

➔ higher polarization

Improved energy timing

Parameters to vary:

Time profile of the time-jump quadrupoles

Observables to optimize:

Revolution frequency (1.E-6)

Radial offset from BPM readings (20mu average)

Main dipole fields Hall-probe at injection (0.1%) + integrating coil (2%)

E(t) by measure f(t), x(t), B(t)

Measuring Energy (ɣ)

Calibration of the Ggamma meter consists of measuring Gg(B) and Gg(f) at the same times in the cycle and fitting parameters until they agree to sufficient precision

Dedicated calibration ~2 weeks

Essentially an inverse problem with data assimilation, good candidate for uncertainty quantification (how well can we determine these parameters, which is responsible for most variation?

From RF frequency

$$G\gamma = G\frac{1}{\sqrt{1-\frac{1}{c^2}\left(\frac{f}{h}\right)^2(2\pi)^2(R_0+dR)^2}}$$

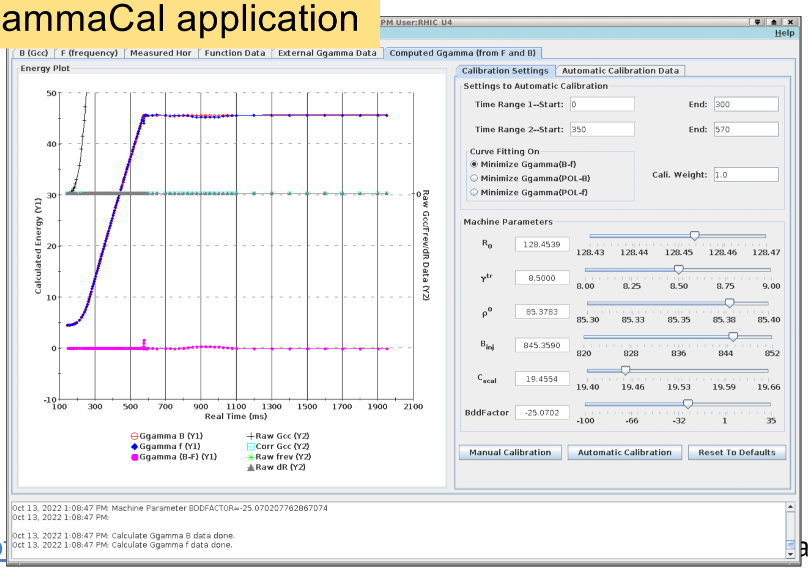
$$G\gamma = G\sqrt{\left[\frac{(1+\gamma_{tr}^2\, dR/R_0)\rho_0 c\,(B_{inj}+B_{clock}/C_{scal})}{M_0}\right]^2+1}$$

From field

In RED:
Measured quantities
f = RF frequency
dR = radial shift from 'zero'
B_{clock} = Field reported by Gauss clock

In BLUE:
Machine parameters (not known to sufficient precision)
γ_{tr} = transition gamma
R0 = true central radius of AGS
ρ_0 = avg bend radius of AGS main magnet
C_{scal} = Gauss clock calibration (gauss/tick)
B_{inj} = Dipole field at injection
Bdfactor = [NOT IN FORMULA] Gauss measurement sensitive to dB/dt (B-dot), not well understood

AgsGgammaCal application



Reduction of AGS resonance driving terms

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

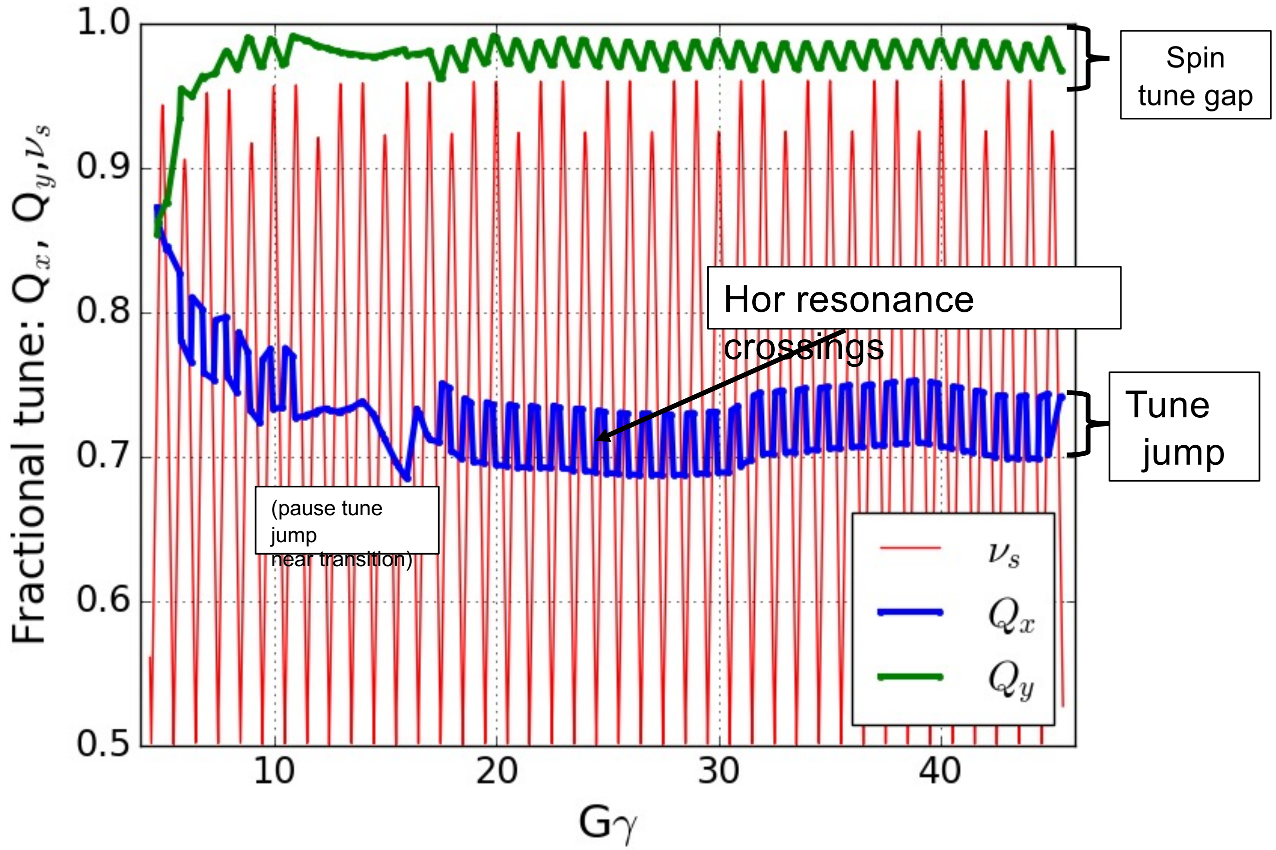
Provides spin tune 'gap' where imperfection and vertical intrinsic resonance condition are never met

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

Horizontal resonance condition still met

- $\nu_s = N \pm Q_x$
- 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

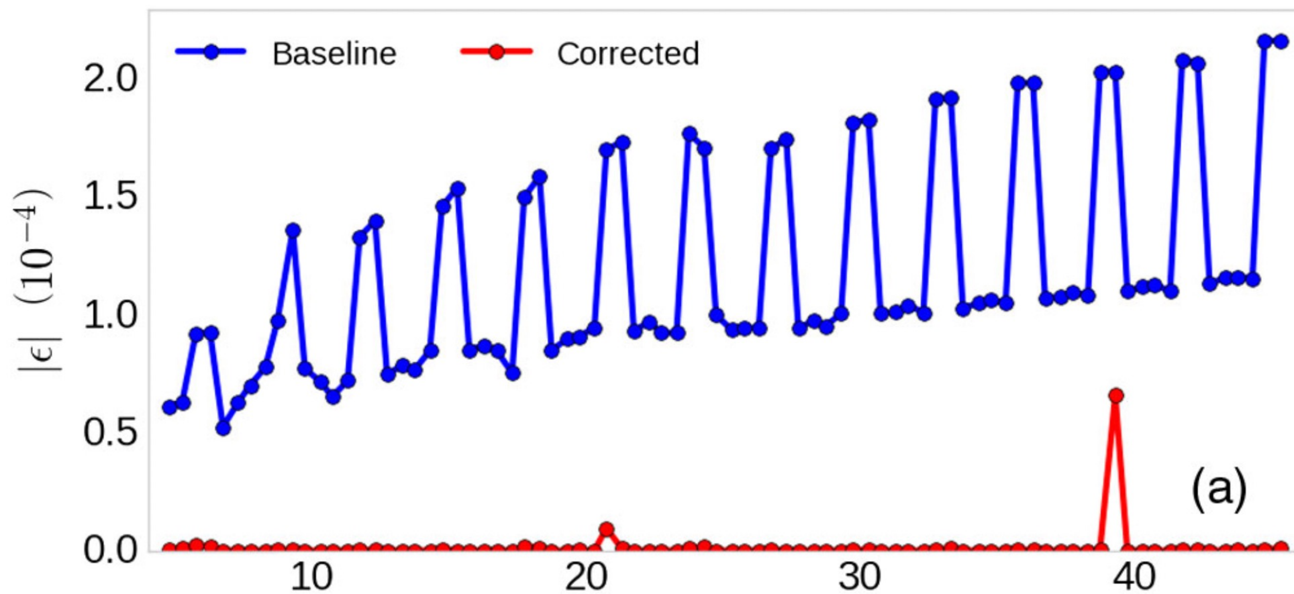
➔ Compensate by other coupling elements, e.g., skew quads

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



AGS Spin Resonance Correction Skew Quadrupoles

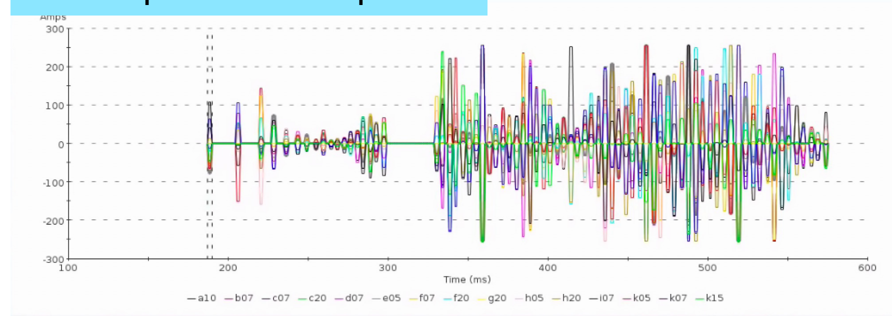
- A set of 15 pulsed skew quadrupoles, each with an individual power supply
- Designed to excite coupling resonance to compensate the 82 depolarizing resonances associated with horizontal betatron motion in the AGS partial snakes

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

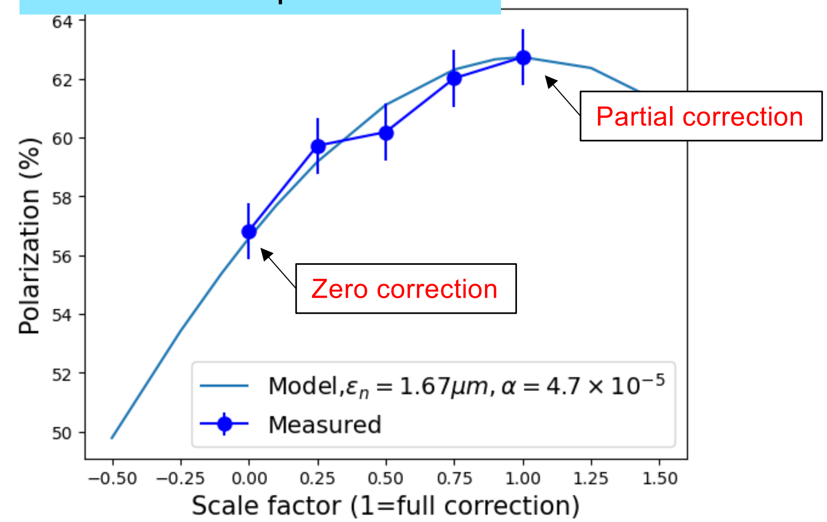
- Run 24: Observation of polarization gain factor (+10%) during acceleration (similar to existing tune jump), with ~half the pulses enabled)

- Further improvements (enabling more pulses, +5-10% gain):
  - Addressing model inaccuracies at low energy
  - Iteration on orbit centering
  - Possible optimizations based on ML methods
    - No solid plan for how to approach this

Skew quad current pulses

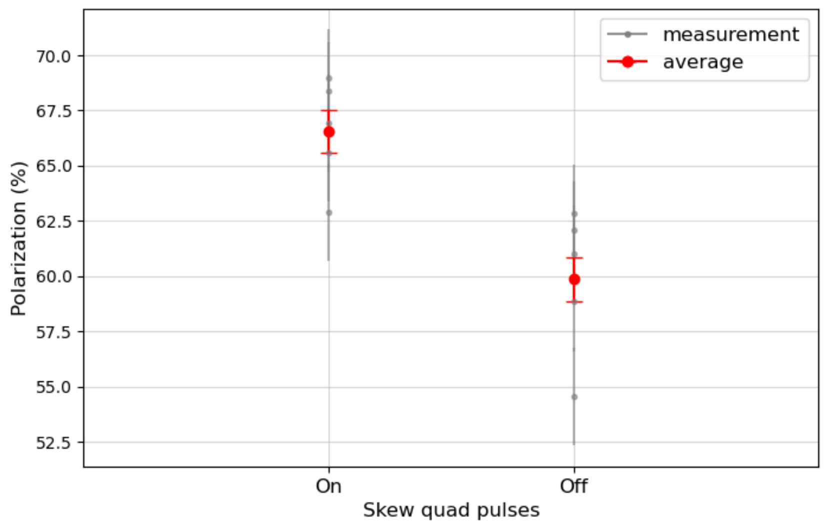


Correction amplitude scan

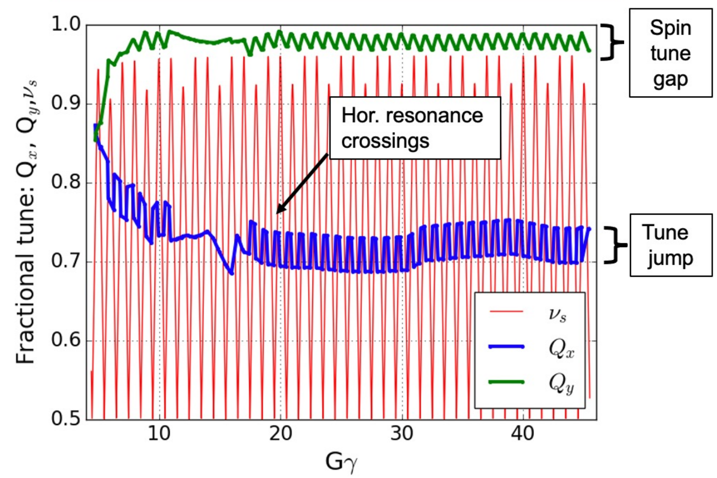


AGS skew quads

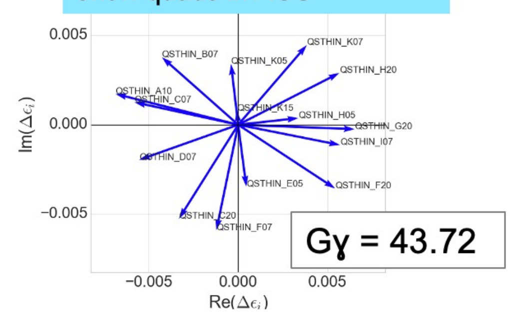
- Partial snakes in the AGS helps avoiding vertical resonances
- Goal: compensate 82 horizontal resonances with 15 pulsed skew quadrupoles
- Satisfactory results for above-transition resonances



Betatron and spin tunes during AGS ramp



Spin resonance terms from skew quads in AGS



Georg.Hoffstaetter@cornell.edu

SciBmad a ML-oriented Toolkits (Libraries)

Advantages the toolkit:

Fully differentiable (reverse and forward)

➔ excellent for Neural Network optimizations

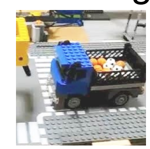
➔ Excellent for Bayesian optimization with slope information

- 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.

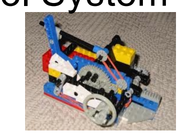
Toolkit

Dynamic Aperture Program

Lattice Design Program



Control System Programs



IBS Simulation Programs

Etc.

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%.

- Several accelerator optimizations can impact polarization.

- These topics are of the type suitable for physics- informed Bayesian Optimization and we are evaluating suitability for Reinforcement Learning.

- Excellent team has formed, items being addressed:
- Emittance reduction (orbit, optics, bunch splitting) already works in the Booster
- Improved model building and programing of digital twins of all parts
- Reduction of resonance driving terms already works above transition energy
- Accelerator studies show the utility of ML for the pre-accelerator chain.

Dominant Participants

BNL: Kevin Brown, Weinin Dai, Bhawin Dhital, Yuan Gao, Levente Hajdu, Kiel Hock, Bohong Huang, Natalie Isenberg, Nguyen Linh, Chuyu Liu, Vincent Schoefer, Nathan Urban

Cornell: Georg Hoffstaetter de Torquat (also BNL), Lucy Lin, Eiad Hamwi, David Sagan, Matt Signorelli

SLAC: Auralee Edelen

JLAB: Malachi Schram, Aarmen Kasparian

RPI: Yinan Wang

Radiasoft: Nathan Cook, Jon Edelen, Chris Hall

Thank you and Questions?