Higher RHIC polarization by Physics-informed Bayesian Learning

Preparation for a ML / AI Proposal to DOE-NP

Georg Hoffstaetter

Collider-Accelerator Department, BNL and Cornell University

Georg.Hoffstaetter@cornell.edu

RHIC MAC December 14, 2022

@BrookhavenLab

DE-FOA-0002875 : ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING FOR AUTONOMOUS OPTIMIZATION AND CONTROL OF ACCELERATORS AND DETECTORS
Budget: $16M, maximum grant $2M, duration 2 years.

Requested topics:

- Efficiently extract critical and strategic information from large complex data sets
- Address the challenges of autonomous control and experimentation
- Efficiency of operation of accelerators and scientific instruments
- AI for data reduction of large experimental data

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

- Gaussian Process (GP) Bayesian Optimization (BO) and physics informed learning.
- When is ML/AI better for accelerator operations than other feedbacks and optimizers?
- 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
- Collaborations: BNL, Cornell, SBU, SLAC, JLAB, RPI

Optimizers for different applications

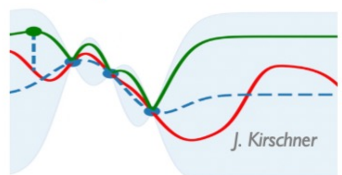
less ← assumed knowledge of machine → more

Model-Free Optimization

Observe performance change after a setting adjustment

→ estimate direction or apply heuristics toward improvement

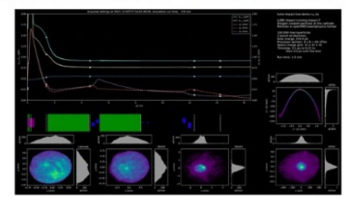
Model-guided Optimization



Update a model at each step

→ use model to help select the next point

Global Modeling + Feed-forward Corrections



Make fast system model

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

gradient descent
simplex
ES

Bayesian optimization
reinforcement learning

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.

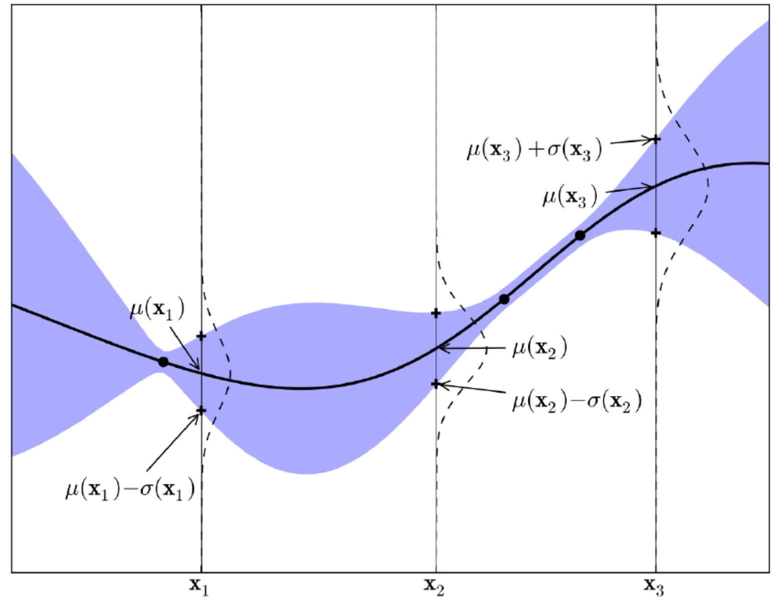
Is this type of problem suitable for Machine Learning?

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 $\sigma^2(x_i|X) = k(x_i,x_i) - k(x_i,X)K^{-1}k(X,x_i)$

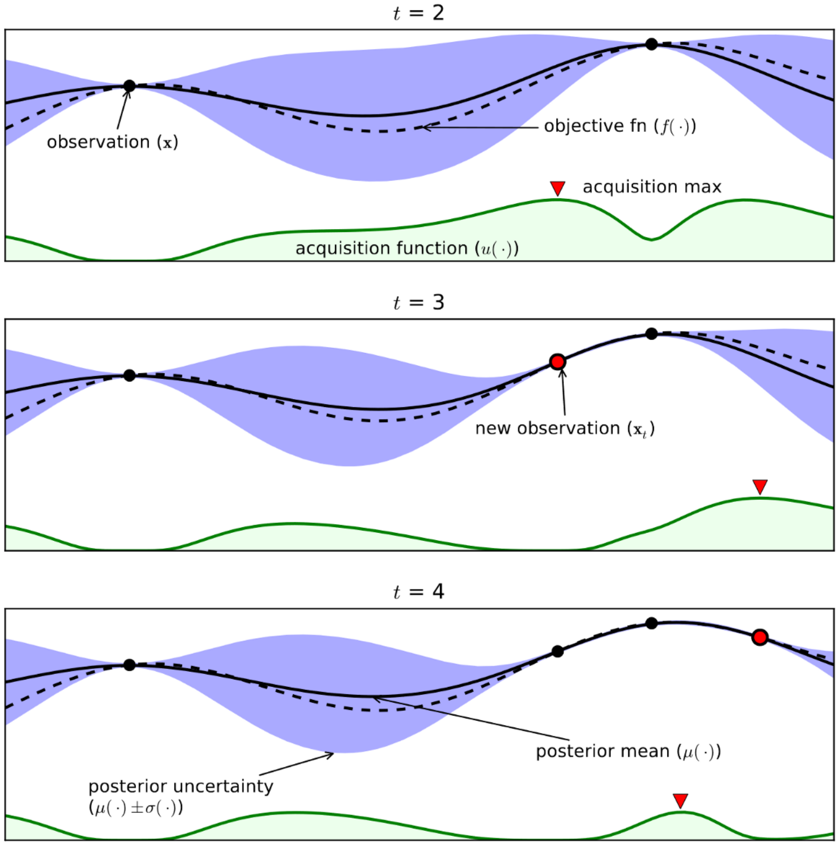
Acquisition Function

[2]

- Guide how input space should be explored during optimization

- Combine predicted mean and variance from Gaussian Process model
  - Probability Improvement (PI)
  - Expected Improvement (EI)
  - **Upper Confidence Bound (UCB)**

$$\mathbf{UCB}(\boldsymbol{x}) = \boldsymbol{\mu}(\boldsymbol{x}) + \boldsymbol{\kappa\sigma}(\boldsymbol{x})$$

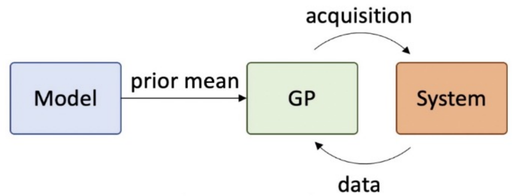


Merit of physics-informed optimization

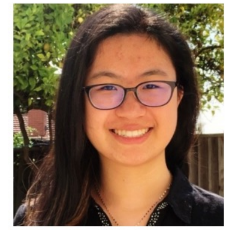
Neural Network System Models + Bayesian Optimization

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

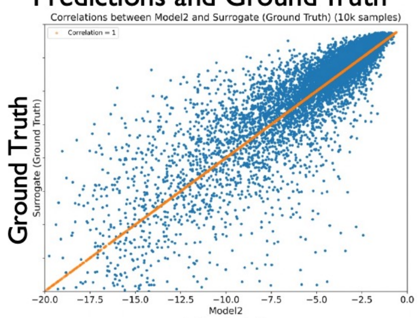


Used the LCLS injector surrogate model for prototyping
variables: *solenoid, 2 corrector quads, 6 matching quads*
objective: *minimize emittance and matching parameter*

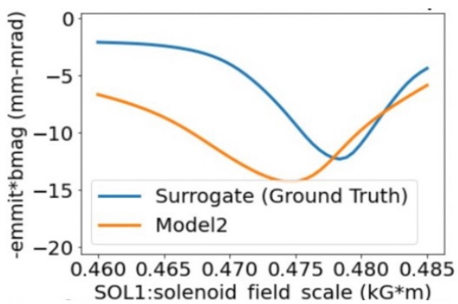


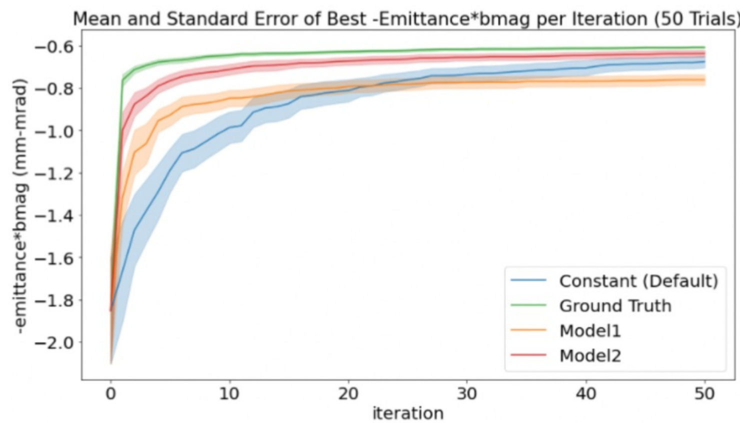
Summer '22 undergrad intern
Connie Xu

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

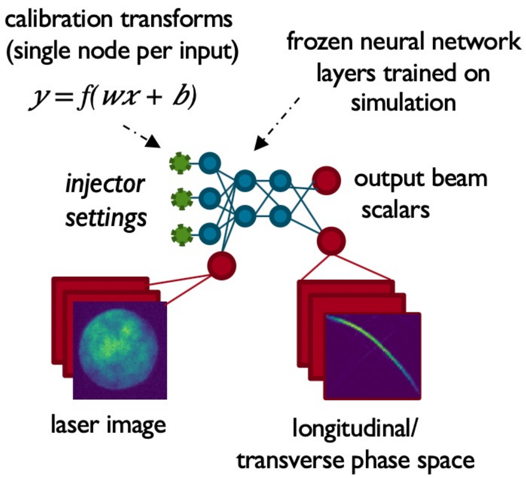
Courtesy
Auralee Edelen

Unknown system parameters

Finding Sources of Error Between Simulations and Measurement

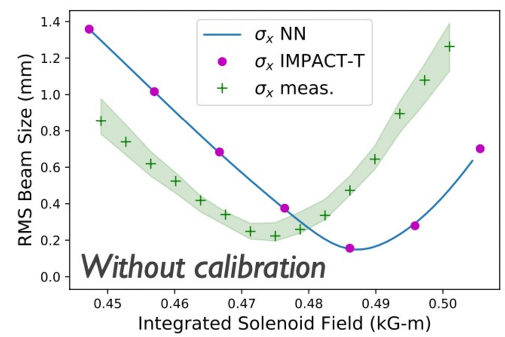
Many non-idealities not included in physics simulations:
- **static error sources** (e.g. magnetic field nonlinearities, physical offsets)
- **time-varying changes** (e.g. temperature-induced phase calibrations)

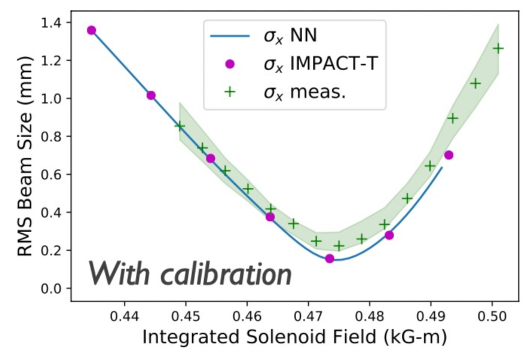
Want to identify these to get ***better understanding of machine*** ***→ fast-executing ML model allows fast / automatic exploration of possible error sources simultaneously***



Inputs
Laser radius
Laser spot sizes
Pulse length
Charge
Solenoid
L0A phase
L0B phase
SQ quad
CQ quad
6 matching quads

Outputs
Beam size (x,y)
Emittance (x,y)
Bunch length





Calibration offset in solenoid strength found automatically with neural network model (trained in simulation, then calibrated to machine)
Example above is simulation-to-machine, but can adapt model over time as well

Courtesy
Auralee Edelen

First studies look promising → current work focuses on examining robustness and extending to larger subsystems

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 (but can use scalarization)	No	No	No	Yes	Yes
Sensitivity to local minima	High (but can use multi-start)	High	High	High	Low	Low (builds a **global** model of f)
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

➔ No derivatives have to be computed.

2. Models of the accelerator exist

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

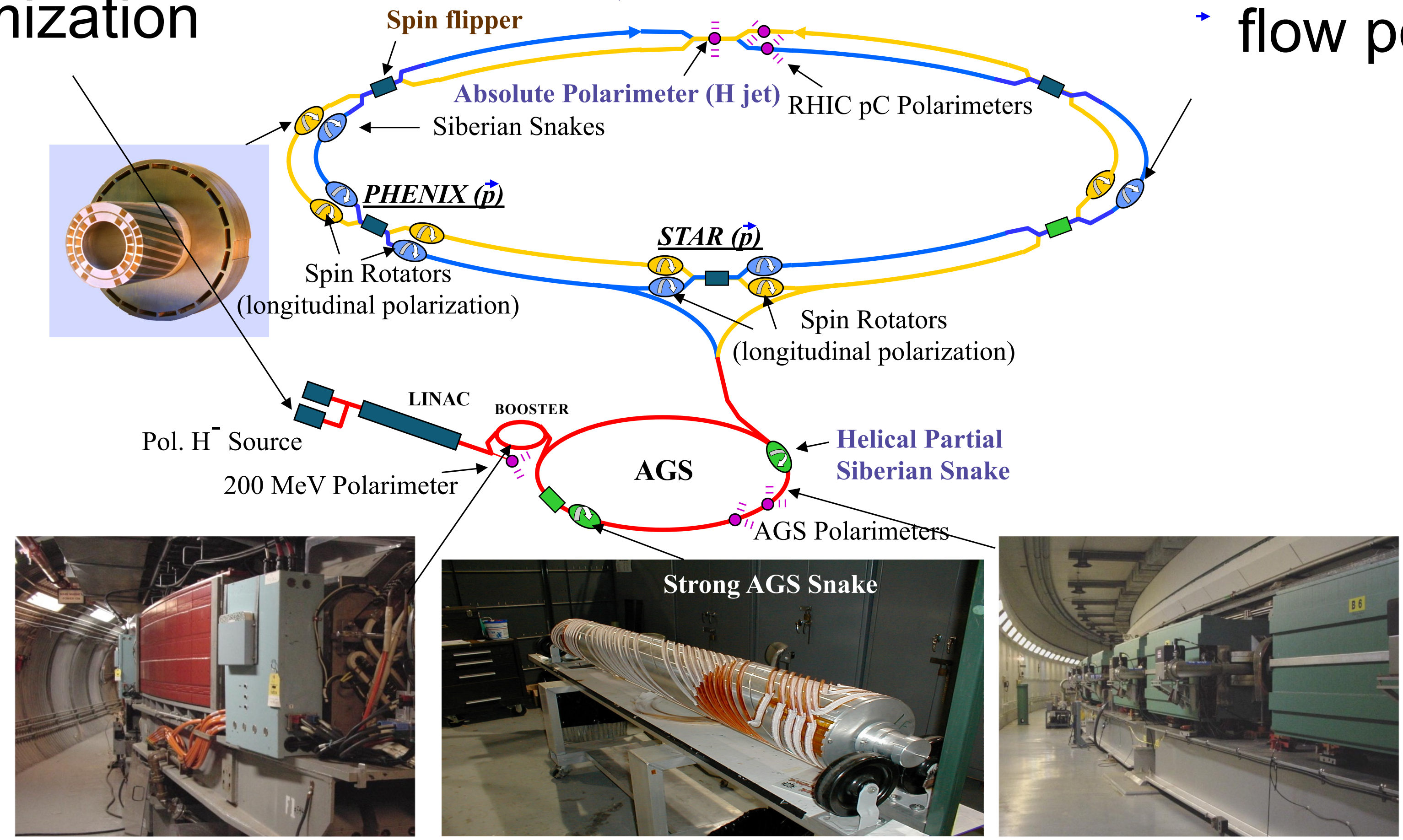
3. A history of much data is available and can be stored

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

The polarized proton accelerator chain

Linac – Booster – AGS Optimization

Readily available, large data flow possible



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 Quadrupole calibration from ORM in Booster and AGS.
- Bunch splitting in the Booster for space charge reduction and bunch re-coalescing at AGS top energy.

Linac to Booster transfer

Parameters to vary:

- Transfer line steers
- Main Booster dipol90e field
- Booster beta wave (stop-band quadrupoles) for tune toward ½ and minimum on the foil
- Last two linac phases
- Injection bump elements and their time profile
- Scraper amplitudes

Observables to optimize:

- Transfer efficiency linac ➔ Booster early ramp (2% absolute)
- Emittance from multi wires of the AGS transfer line (5% relative)

Booster to AGS transfer

Parameters to vary:

- Transfer line steerers
- Main AGS dipole field
- AGS RF phase
- Amplitudes of two Injection bumps
- Horizontal orbit in the snakes
- Quadrupole corrections for the snakes
- Injection to accelerator tune change

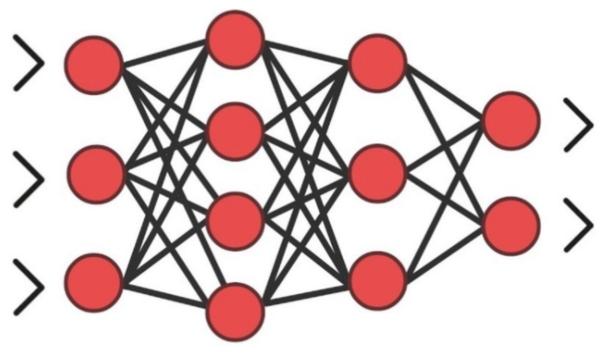
Observables to optimize:

- Transfer efficiency Booster ➔ AGS early ramp (2% absolute)
- Emittance from two IPMs (10% relative)

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

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



Sensitivity studies: error sources

- Sources or error and ranges come from past survey data
- Criteria to quantify & visualize sensitivity:
  - RMS of ORM matrix
  - Beta-beating (vertical & horizontal)

$$\frac{\Delta\beta}{\beta} = \frac{\beta_{measured} - \beta_{model}}{\beta_{model}}$$

Name	Unit	Range
Main magnet roll error	mrad	[-0.5, 0.5]
Main magnet gradient error	m^{-2}	$\pm$ 0.1%
Quadrupole gradient error	m^{-2}	$\pm$ 0.2%
Sextupole offset error	mm	[-8, 8]
Snake magnet roll error	mrad	[-1.5, 1.5]

Where do we put AI/ML?

- ORM will give us
  - BPM and Corrector Anomalies (Trust Analysis)
  - Gradient errors for given conditions
  - Beta-deviations from model
- Dispersion measurements give us
  - BPM Consistency check for given dp/p (BPM Anomalies)
  - Coupling through longitudinal motion (very slow, typically)
- Tune measurements
  - Betatron tune and coupling = destructive measurement in Booster/AGS
  - Tune, Chrom, coupling, emittance, dp/p from RHIC Schottky
- Chromaticity measurements – need to change energy and measure tune
- Orbit Measurements – parasitic = most are time averaged, some turn by turn
- Linear model + small nonlinearities with NN model

Orbit & Optics correction in Booster / AGS

Parameters to vary:

Corrector coils (24 per Booster plane)

Corrector coils (48 per AGS plane)

Observables to optimize:

BPM readings (24 x&y in the Booster) (100um accuracy)

BPM readings (72 x&y in the AGS) (100um for 2mm size at 25GeV)

Space-charge emittance increase

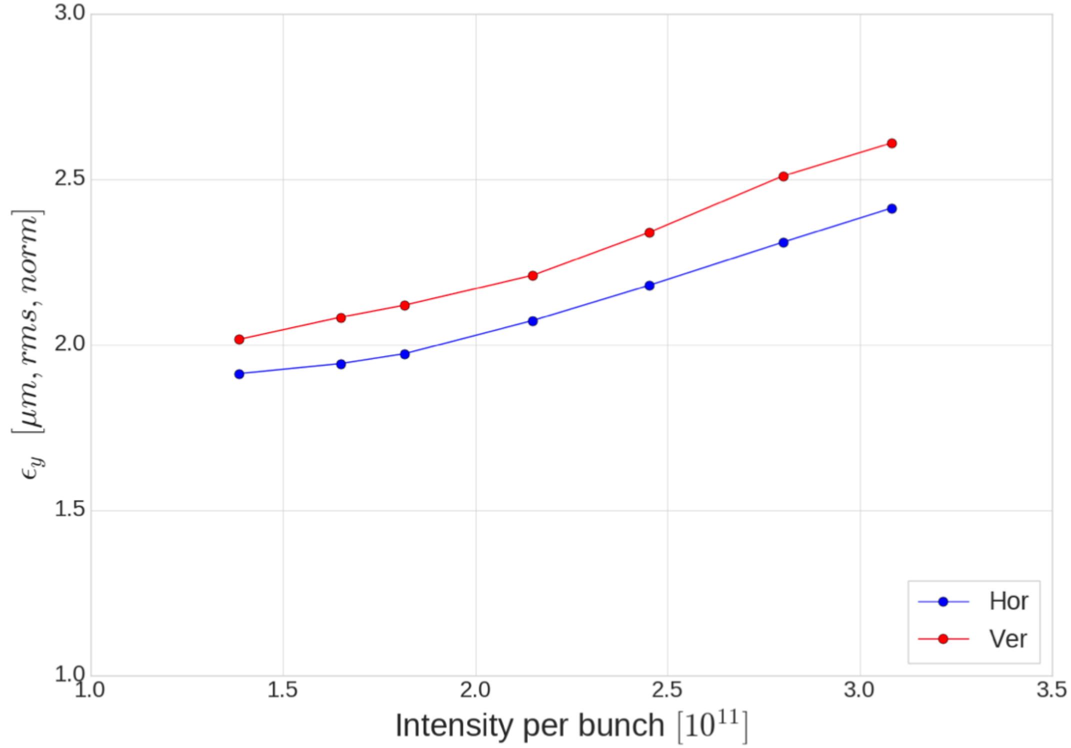
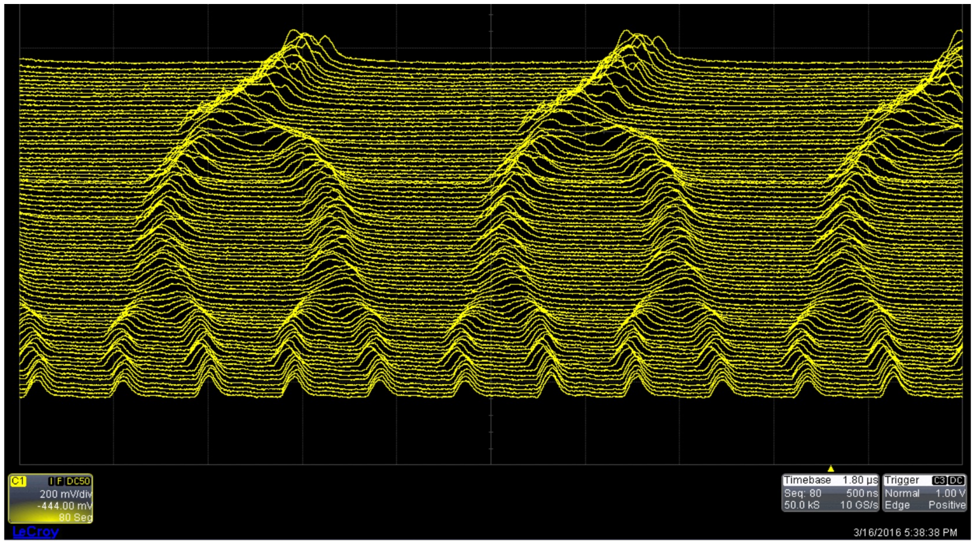


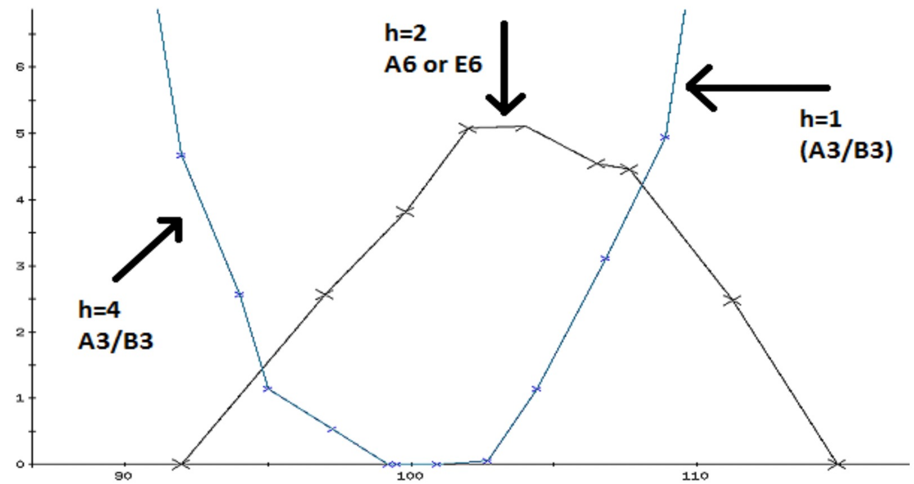
Figure 3.168: Normalized transverse emittances of polarized proton beam at AGS extraction energy ($\gamma = 25.5$) as a function of intensity.

➔ Splitting bunches before AGS acceleration can reduce the emittance.

Bunch splitting / coalescing



Mountain range display of the wall current monitor signal for the 4:2:1 Booster merge used for EBIS Au.



: Rf gap voltages, harmonics, and cavities involved in the standard 4:2:1 Booster merge used for EBIS Au. The x-axis is ms from Bt0 and the y-axis is the voltage reference. The h=2 cavity has 2 gaps, and A3 and B3 have 1 gap. So, since both A3 and B3 are used for h=4 and h=1 the relative voltages here should be correct.

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

Bunch splitting and coalescing

Parameters to vary:

3 RF amplitudes and phases, and their timing

Observables to optimize:

Mountain range width (5% relative)

Mountain range oscillations (10% of a sigma)

Baby-bunch currents (2%)

Emittance in the multi-wire to the AGS (5% relative)

Emittance from two IPMs (10% relative)

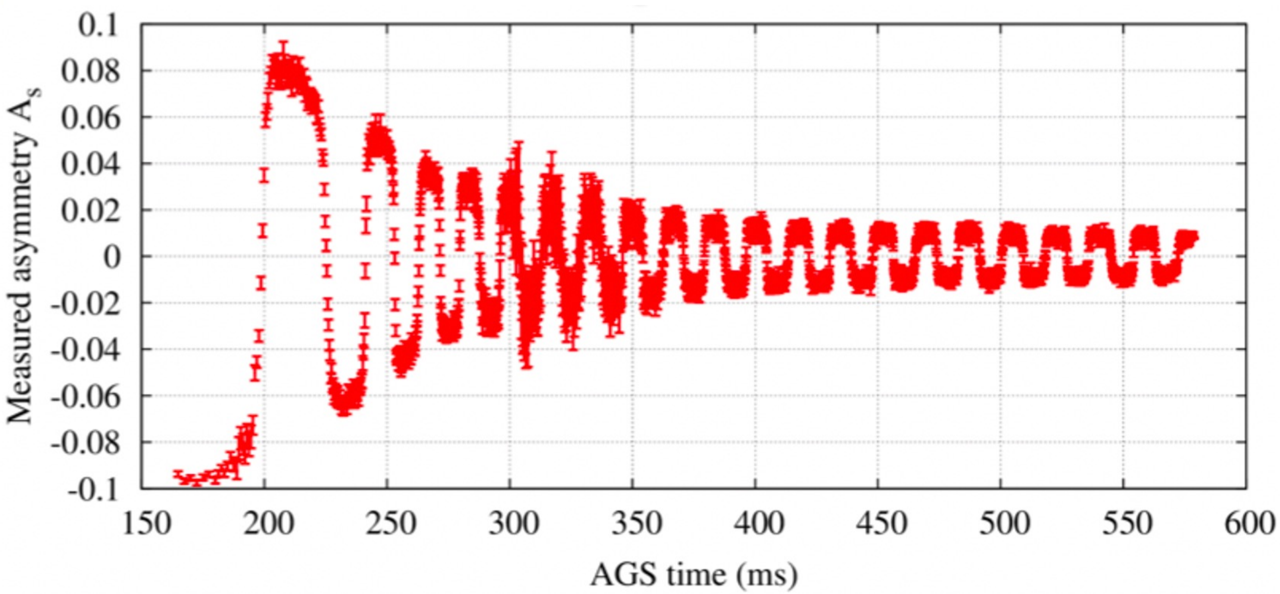
Timing of tune jumps

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), P(t)

Reduction of AGS resonance driving terms

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

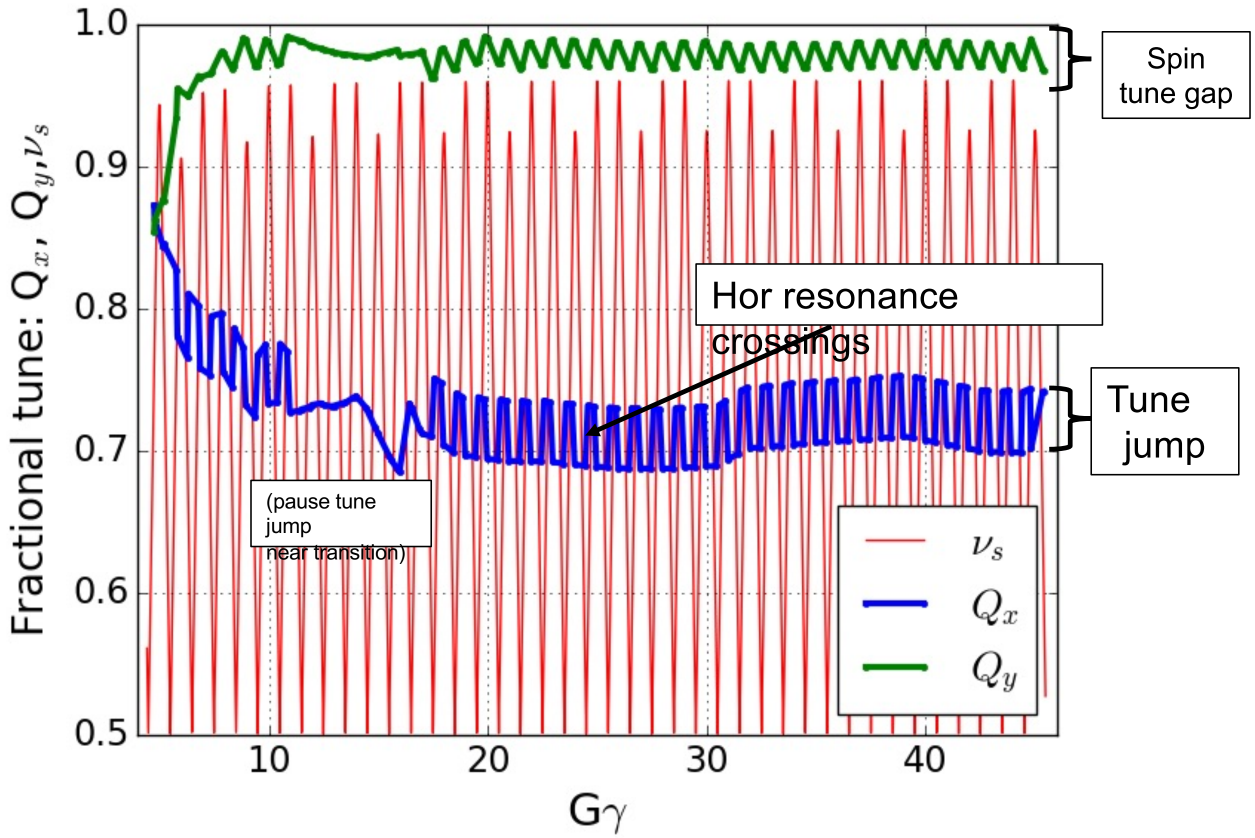
Provides spin tune 'gap' where imperfection and vertical intrinsic resonance condition are never met

- $\nu_s \neq N$ (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

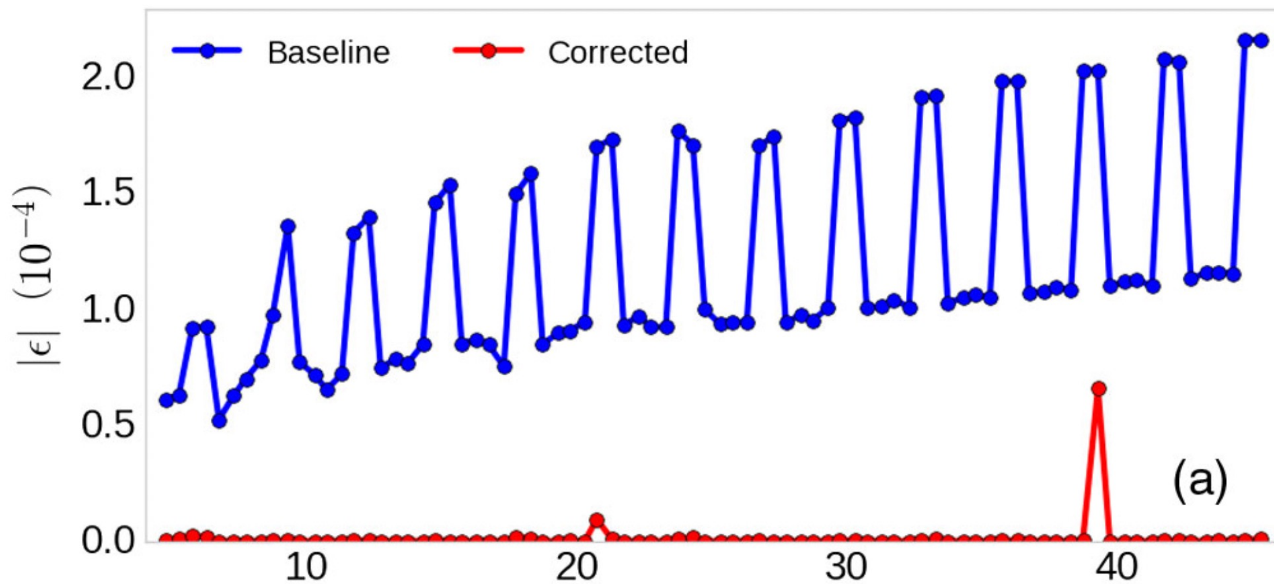
➔ Compensate by other coupling elements, e.g., skew quads

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



Brookhaven National Laboratory
Georg.Hoffstaetter@cornell.edu

Reduction of resonance strengths

Parameters to vary:

14 Skew quad amplitudes at each of 80 resonances

Timing of skew quad changes

Observables to optimize:

Polarization after the ramp (2% relative)

Polarization at intermediate energies (2% relative)

Personnel involved

Georg Hoffstaetter (C-AD and Cornell) – Accelerator physics
Kevin Brown (C-AD and Stony Brook) – Controls implementation
Vincent Schoefer (C-AD) – Controls implementation
Natalie Isenberg (CSI) – ML with uncertainties
Nathan Urban (CSI) – ML/AI consulting
Yuan Gao (C-AD) – ML applications
Lucy Lin (Cornell) – PhD student
Thomas Robertazzi (Stony Brook) – ML with uncertainties
David Sagan (Cornell) – accelerator modeling
Auralee Edelen (SLAC) – ML/AI consulting
Yinan Wang (RPI) – ML/AI consulting

Summary

- A proposal is being prepared 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 Bayesian Optimization
- Items to be addressed:
- Emittance reduction (orbit, optics, bunch splitting)
- More accurate timing of quadrupole jumps (G-gamma meter)
- Reduction of resonance driving terms (Horizontal spin matching with skew quads)

Publications

ML/AI efforts at BNL/CAD

- B. Huang, C. González-Zacarías, S. Sosa Güitrón, A. Aslam, S. G. Biedron, K. Brown, T. Bolin, *Artificial Intelligence-Assisted Design and Virtual Diagnostic for the Initial Condition of a Storage-Ring-Based Quantum Information System*, IEEE Access, Volume 10, 2022, pp.14350-14358
- Y. Gao, W. Lin, K. A. Brown, X. Gu, G. H. Hoffstaetter, J. Morris, and S. Seletskiy, *Bayesian optimization experiment for trajectory alignment at the low energy RHIC electron cooling system*, Phys. Rev. Accel. Beams 25, 014601 – Published 7 January 2022
- Y. Gao, J. Chen, T. Robertazzi, and K. A. Brown, *Reinforcement learning based schemes to manage client activities in large distributed control systems*, Phys. Rev. Accel. Beams 22, 014601 – Published 2 January 2019
- W. Lin, M. A. Sampson, Y.C. Jing, K. Shih, G. H. Hoffstaetter, J. A. Crittenden, *Simulation Studies and Machine Learning Applications at the Coherent Electron Cooling Experiment at RHIC*, IPAC2022, Bangkok, Thailand
- Y. Gao, K. A. Brown, X. Gu, J. Morris, S. Seletskiy, W. Lin, G. H. Hoffstaetter, J. A. Crittenden, *Experiment Of Bayesian Optimization For Trajectory Alignment At Low Energy RHIC Electron Cooler,* IPAC2022, Bangkok, Thailand
- Y. Gao, K. A. Brown, P. Dyer, S. Seletskiy, H. Zhao, *Applying Machine Learning to Optimization of Cooling Rate at Low Energy RHIC Electron Cooler,* IPAC2021, Campinas, SP, Brazil

Polarized proton beams at BNL

- K. Zeno, An overview of Booster and AGS Polarized Proton Operations during Run 17, BNL-114742-2017-TECH (10/2017)
- K. Zeno, Run 16 Tandem gold performance in the injectors and possible improvement with AGS type 6:3:1 bunch merge in the Booster, C-A/AP/576, (10/2016)
- V. H. Ranjbar, Approximations for crossing two nearby spin resonances, Phys. Rev. ST-AB 18, 014001 (2015)
- Y. Dutheil, L. Ahrens, H. Huang, F. Méot, A. Poblaguev, V. Schoefer, K. Yip, Energy Calibration and Tune Jumps Efficiency in the PP APS, in Proc. PAC2014, Dresden/Germany (2014)
- V. H. Ranjbar, M. Bai, H. Huang, A. Marusic, M. Minty, V. Ptitsyn, Experimental Effects of Orbit on Polarization Loss in RHIC, Proc. IPAC2012, New Orleans/LA (2012)
- V. Schoefer, L. Ahrens, K.A. Brown, J.W. Glenn, H. Huang, Optics Error Measurements in the AGS for Polarized Proton Operation, Proc. PAC2011, New York/NY (2011)
- V. Schoefer, Using betatron coupling to suppress horizontal intrinsic spin resonances driven by partial snakes, Phys. Rev. AB 24, 031001 (2021)
- V. Schoefer, AGS Horizontal Resonance Compensation Overview, Presentation at BNL (2022)

Thank you and Questions?