Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

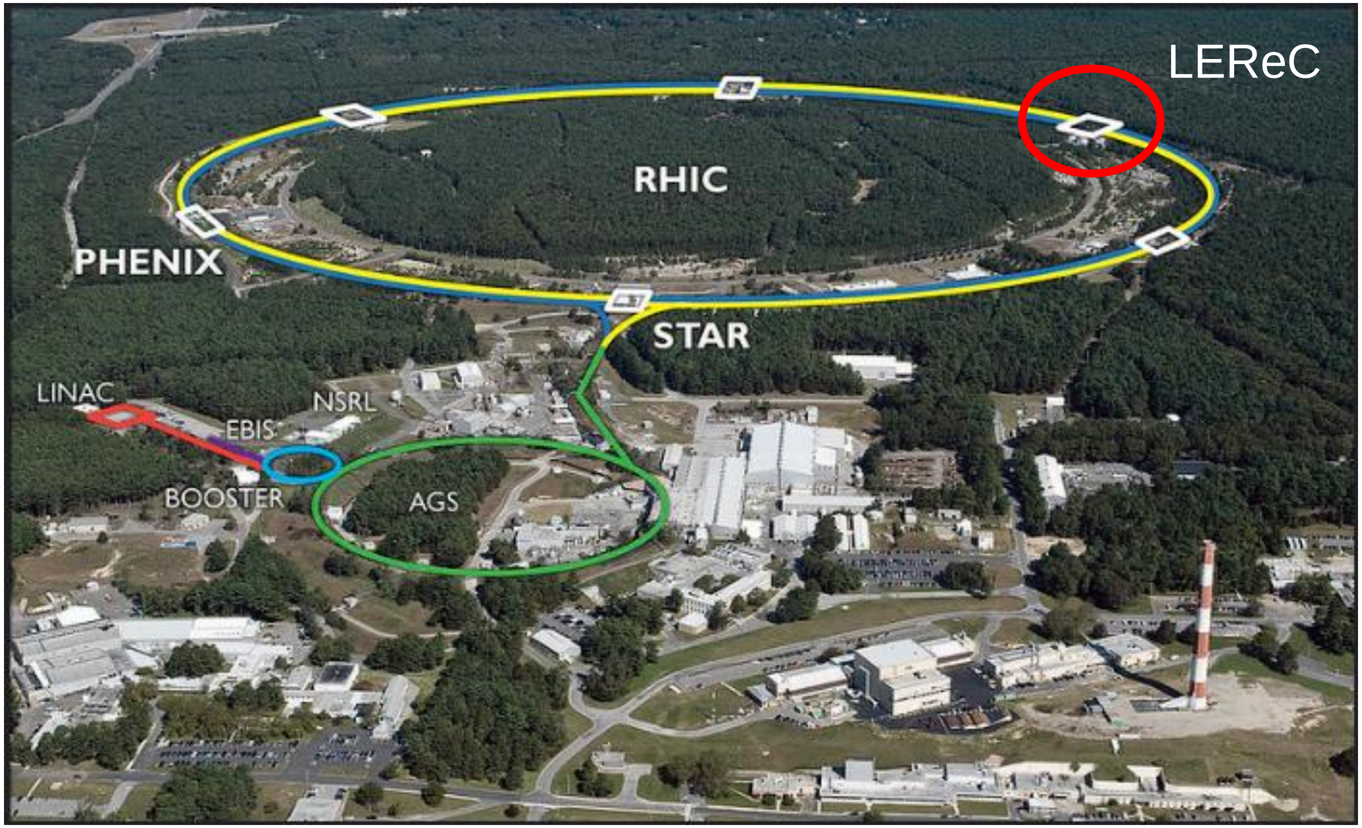
Applying Bayesian Optimization to Achieve Optimum Cooling at the Low Energy RHIC Electron Cooling System

Yuan Gao, Weijian (Lucy) Lin, Kevin Brown, Xiaofeng Gu, Georg Hoffstaetter, John Morris, Sergei Seletskiy, Vincent Schoefer

11/02/2022

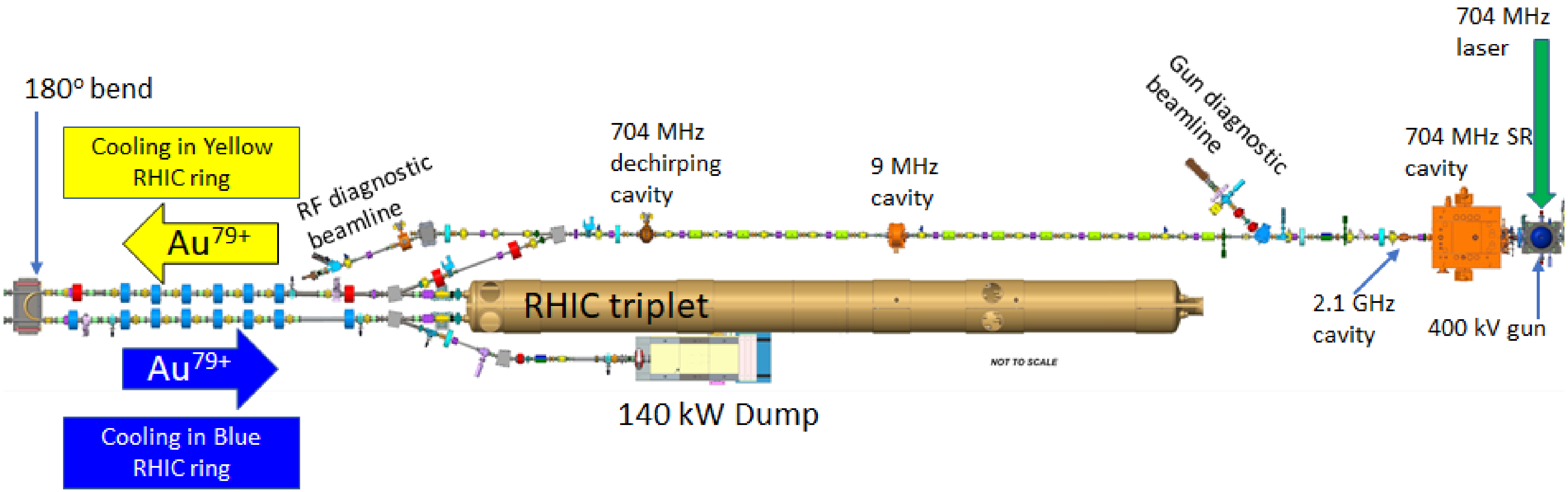
@BrookhavenLab

Relativistic Heavy Ion Collider



- Two 3.8 km counter-rotating supper-conducting rings;
- Six Interaction Regions (IR), LEReC is at IR2;

LEReC System Overview



- LEReC is used to increase the luminosity, it was successfully improved the luminosity multifold in 2020 and 2021 runs;
- 704 MHz e-bunches (grouped into 9 MHz macro-bunches) are produced from the photocathode and accelerated in the SRF cavity to the designed energy (1.6 MeV, 2 MeV);
- Those e-bunches are delivered to the cooling sections (20 meter), where they co-travel with ion bunches.

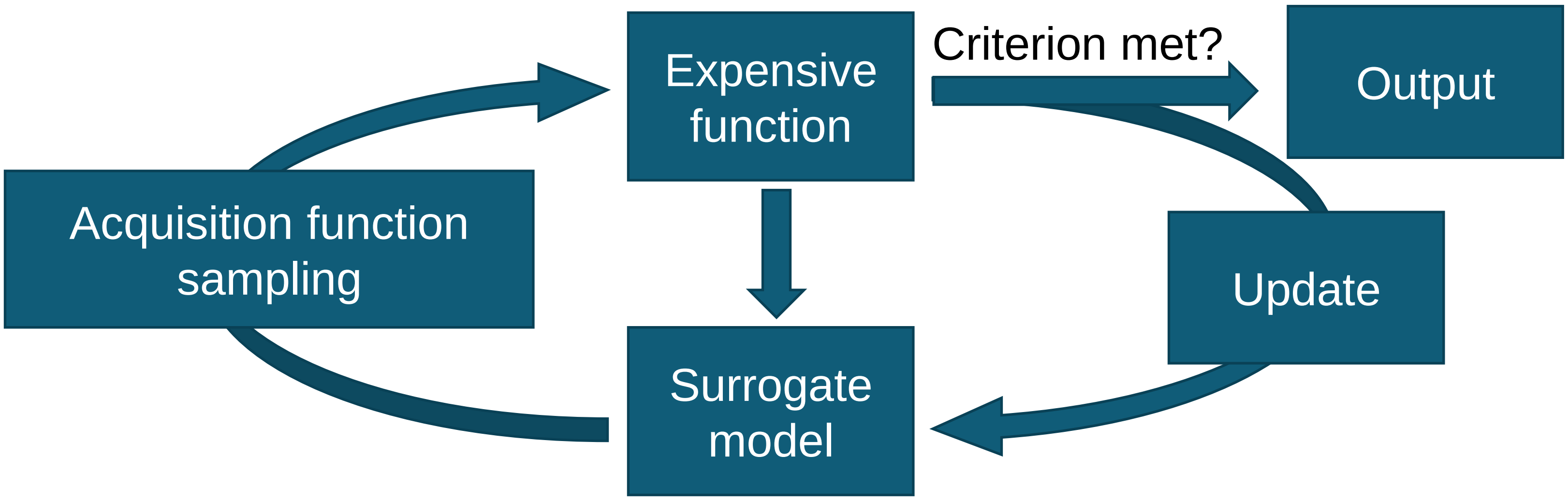
Motivations

- BPM Measurement errors;
- An independent way to optimize the cooling performance.

Method

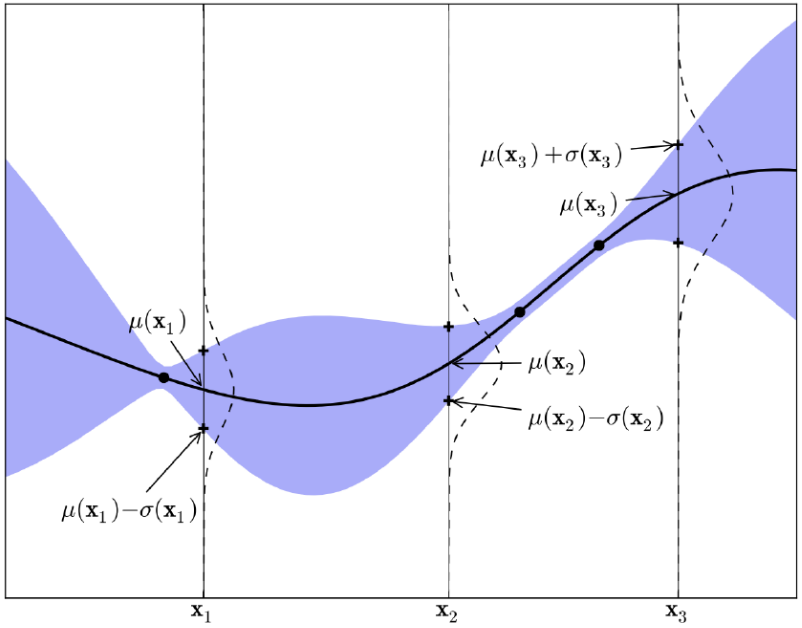
- Bayesian Optimization (BO): a powerful tool for finding the extrema of objective functions that are expensive to evaluate;
- It is called Bayesian because it uses the famous “Bayes’ theorem”.

$$P(f|\mathcal{D}_{1:t}) \propto P(\mathcal{D}_{1:t}|f)P(f)$$



Gaussian Process

[Brochu et al, 2010]



- A probability distribution over possible functions that fit a set of points

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

- The kernel function $k(x_i, x_j)$ describes how closely two points are related.

- The function value at a new sample point x_{t+1} follows $\mathcal{N}\left(\mu_t(\mathbf{x}_{t+1}), \sigma_t^2(\mathbf{x}_{t+1})\right)$ where

$$\begin{aligned}\mu_t(\mathbf{x}_{t+1}) &= \mathbf{k}^T\mathbf{K}^{-1}\mathbf{f}_{1:t}\\ \sigma_t^2(\mathbf{x}_{t+1}) &= k(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) - \mathbf{k}^T\mathbf{K}^{-1}\mathbf{k}\end{aligned}$$

and the covariance matrix $\mathrm{K} = k(X, X) = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_t) \\ \vdots & \ddots & \vdots \\ k(x_t, x_1) & \cdots & k(x_t, x_t) \end{bmatrix}$

Brookhaven National Laboratory

CLASSE Cornell Laboratory for Accelerator-based ScienceS & Education

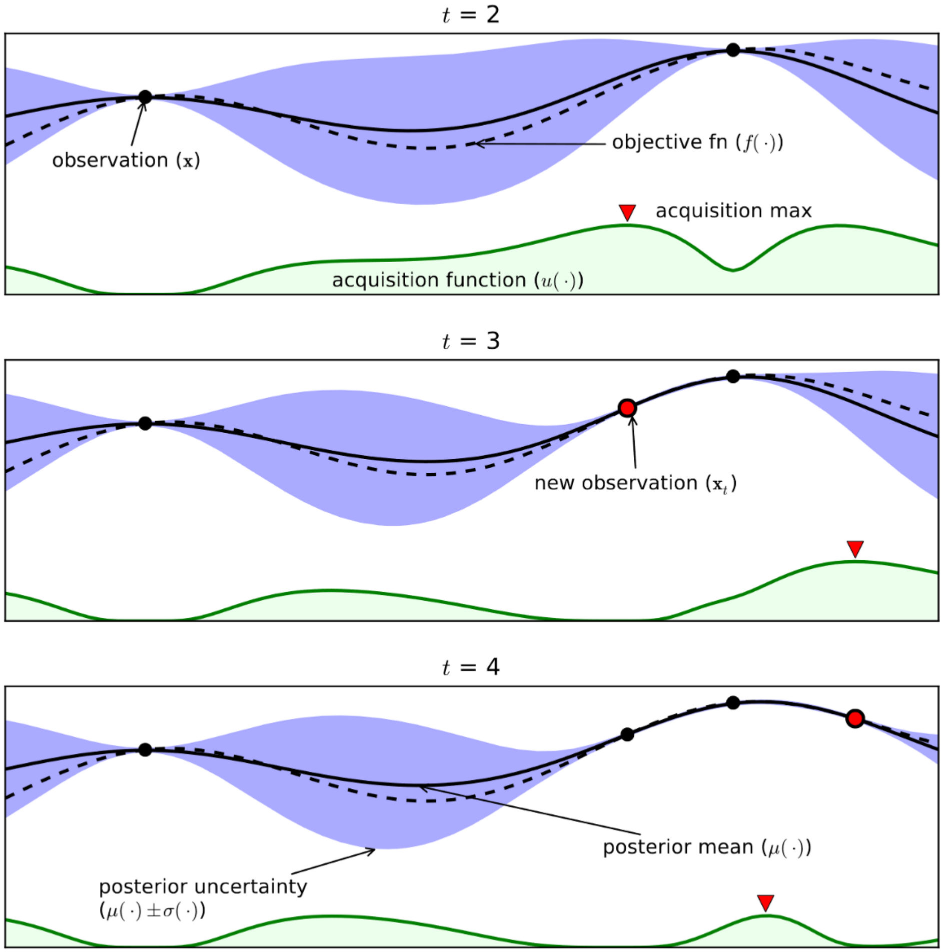
Acquisition Function

- Guide how input space should be explored during optimization;
  - Probability Improvement (PI)
  - Expected Improvement (EI)
  - Upper Confidence Bound (UCB)

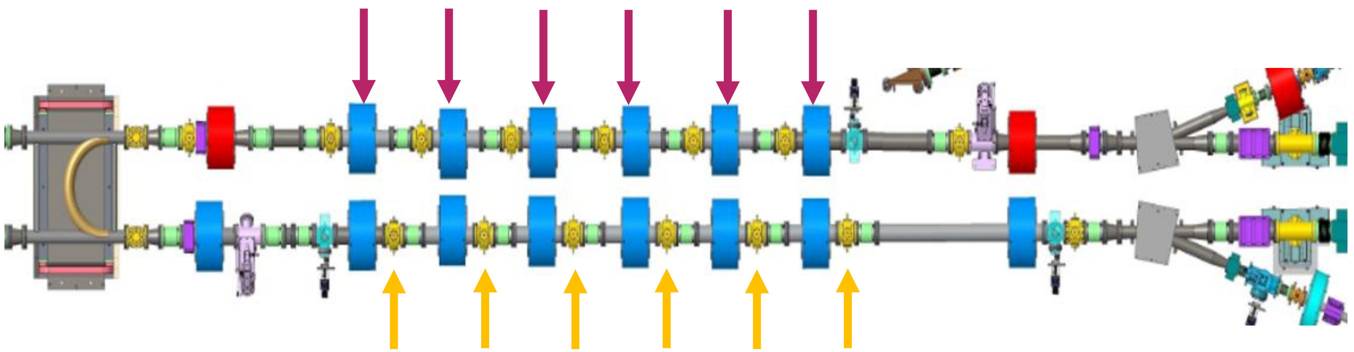
- A combination between predicted mean and variance;

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

[Brochu et al, 2010]



Experiment Settings



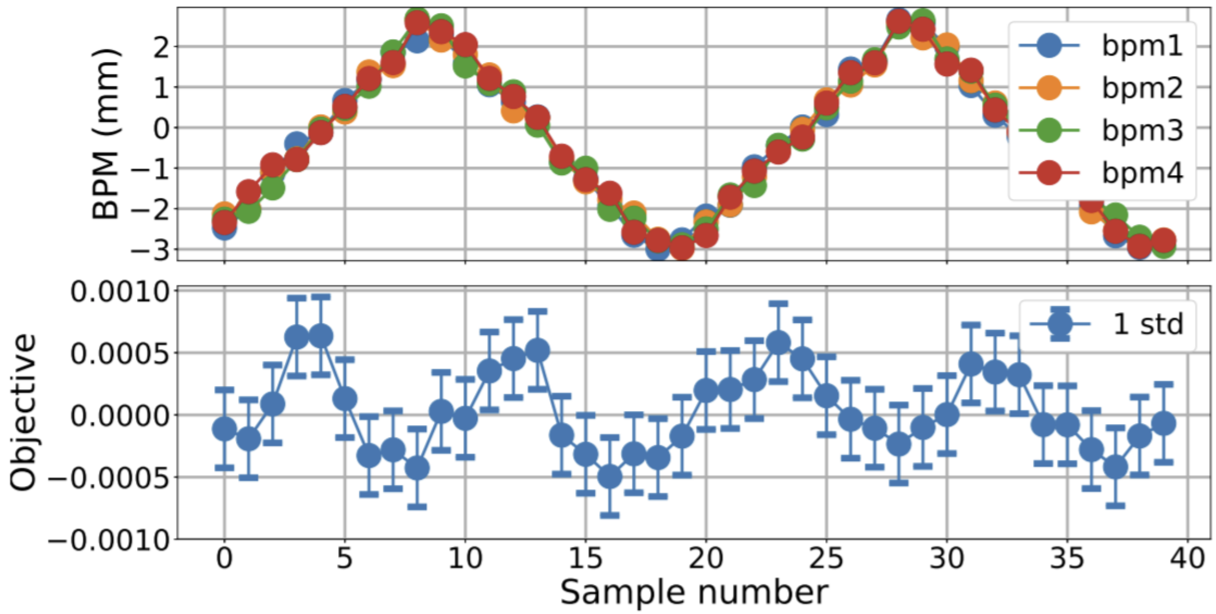
The <u>Goal</u> is to use BO to tune electron trajectories to maximize the ion cooling rate.

- Ions are assumed in the center position, only the first 4 BPMs are considered;
- Decreasing speed of transverse ion beam size:

$$\lambda = (1/\delta)(d\delta/dt)$$

- Cooling performance is measured by (- λ), a more negative λ means a faster cooling rate;

Initial Sampling



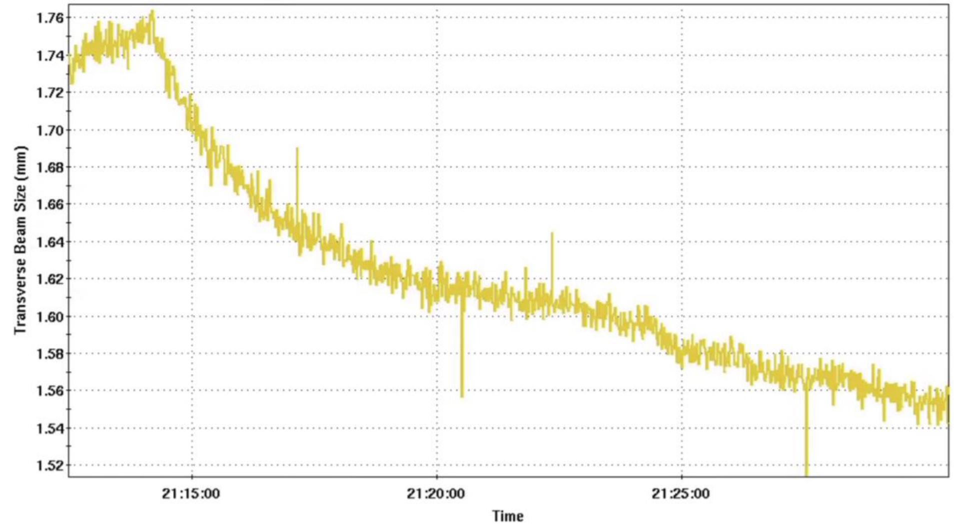
- Input (Top): 4 BPMs, go through the entire [-3, 3] mm range;
- Objective (Bottom): cooling rate (- λ), exhibits a pattern, favors input positions around 0.

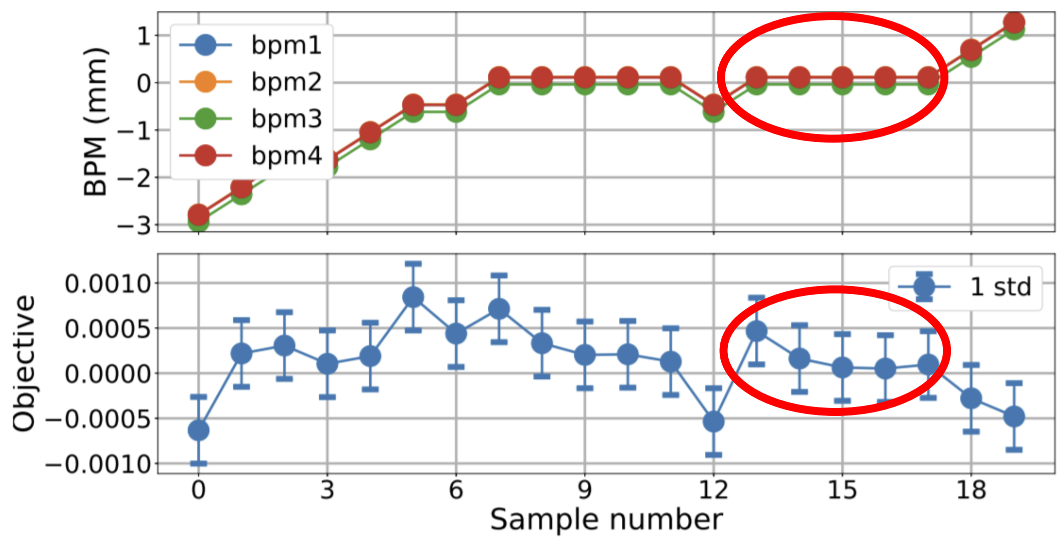
Optimization Strategy in the Presence of Noise

Large noise presents in δ (Top), makes the objective:

- $\lambda = -(1/\delta)(d\delta/dt)$

unstable and unable to converge.

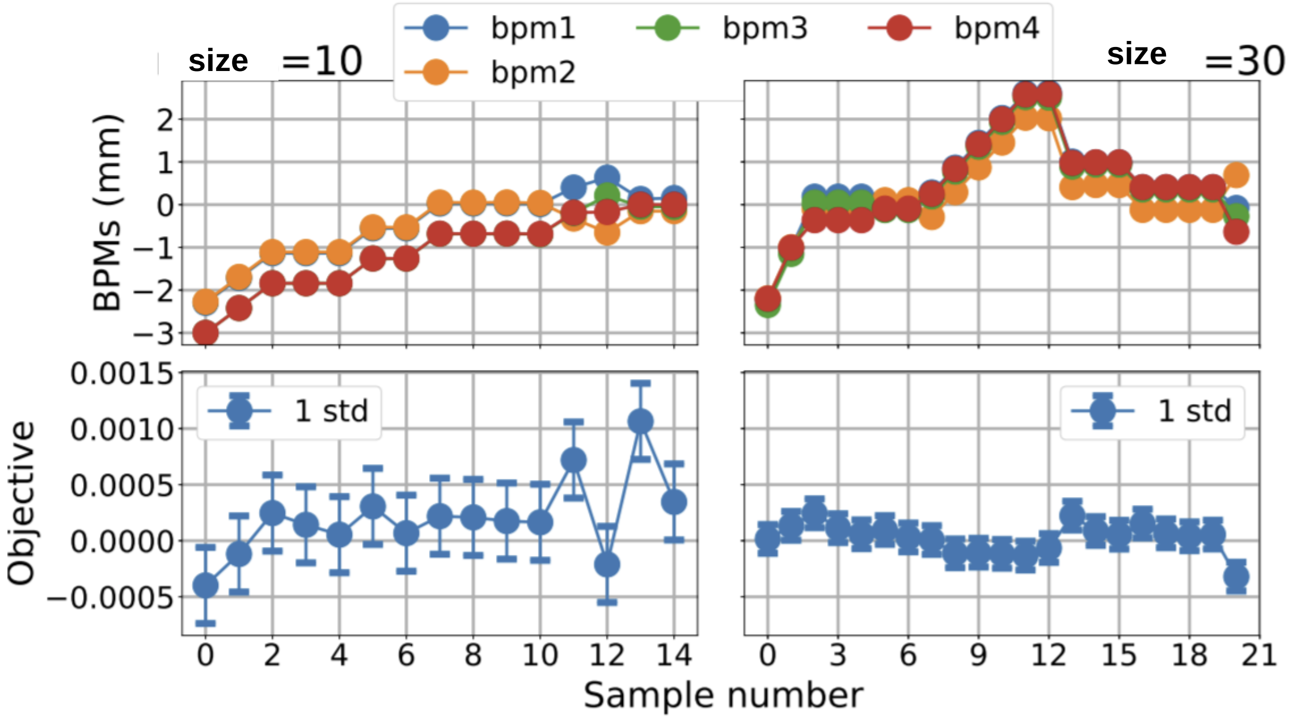




Brookhaven National Laboratory

CLASSE Cornell Laboratory for Accelerator-based ScienceS & Education

Smoothing the Noise

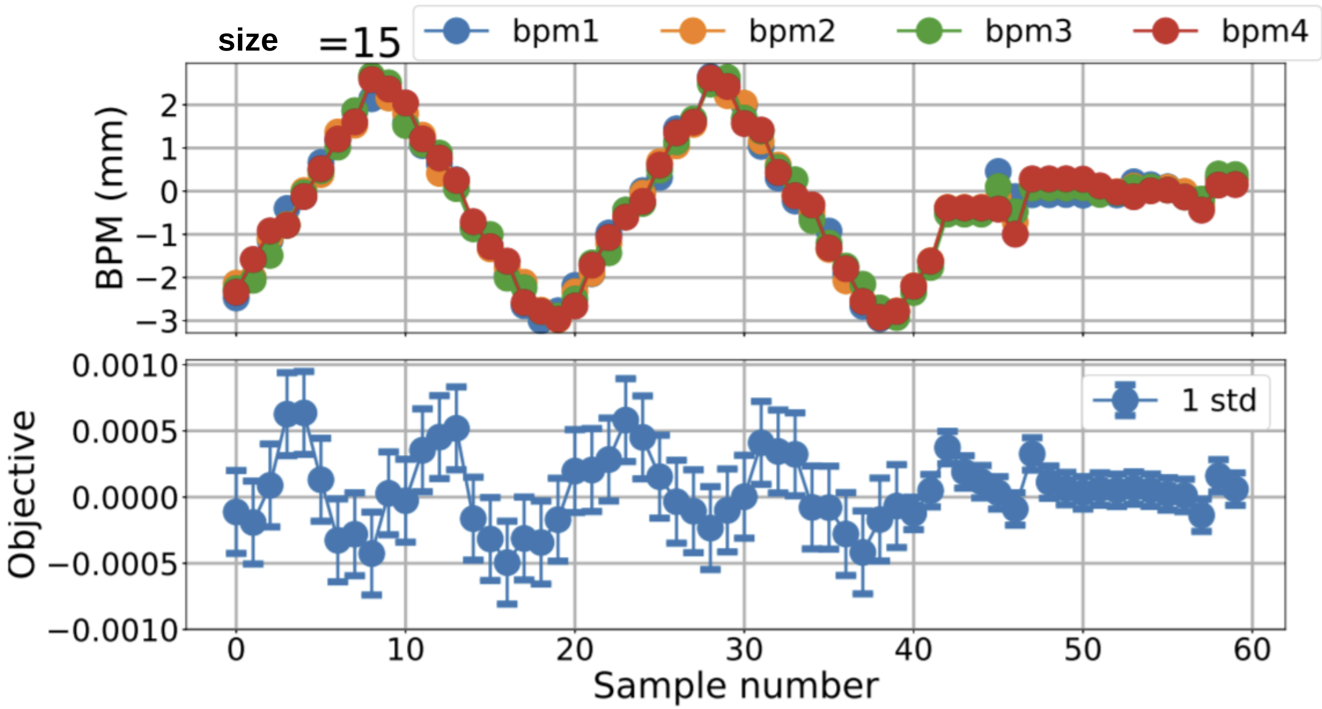


- Use moving average windows, instead of point δ values:

$$\lambda' = (1/\bar{\delta})(\overline{d\delta/dt})$$

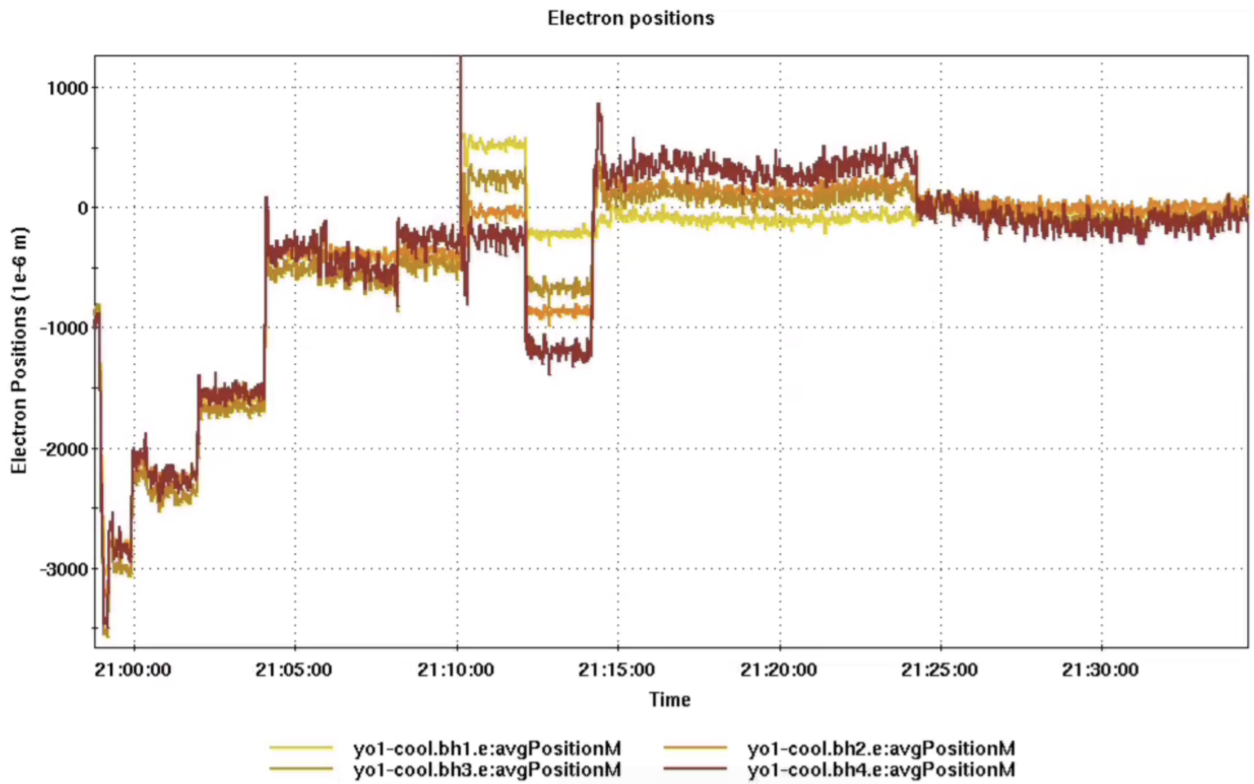
- The window sizes affect how algorithm behaves;

Results



- The final tuning algorithm uses a window size of 15;

Electron Positions Controlled by the BO



- Electron trajectories reported by 4 BPMs;
- The algorithm can tune the trajectories from the farthest points (-3 mm) to the center position and maintain them.

Future Work

- Increase the convergence rate to implement the full control routine on 16 BPMs;

- Physics-model informed GP [1]:
  An alternative way to estimate the kernel function.

- Contextual GP [2]:
  Handle the environmental factors by using separate kernels to model the inputs and contexts.

[1] A. Hanuka, X. Huang, J. Shtalenkova, et al., Physics model-informed gaussian process for online optimization of particle accelerators, Phys. Rev. Accel. Beams 24, 072802 (2021).
[2] A. Krause and C. Ong, Contextual gaussian process bandit optimization, in Advances in Neural Information Processing Systems (NIPS), Vol. 24, (Curran Associates, Inc., 2011).

Data-informed GP, Physics model-informed GP

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

- For convenience, we usually assume the prior mean is the zero-function m(x)=0. A very popular choice for the kernel is the squared exponential function:

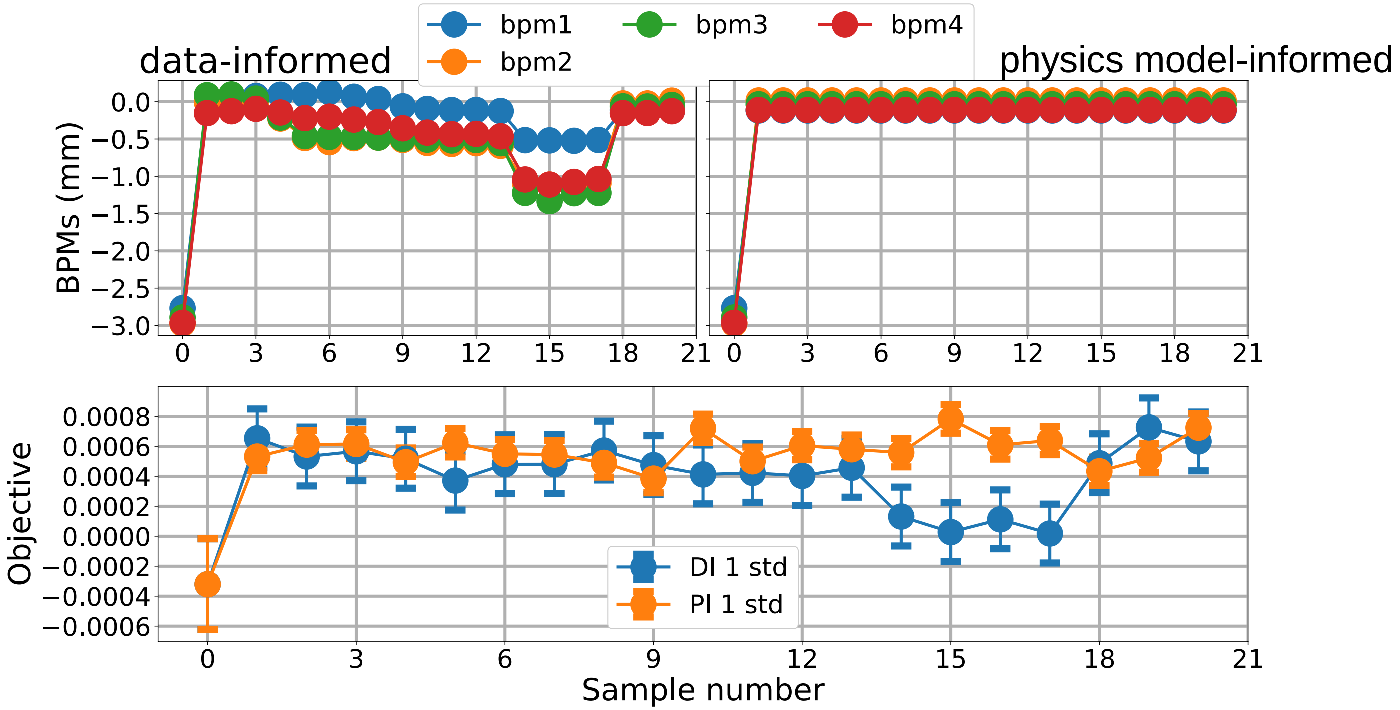
$$k_{\mathrm{RBF}}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^T \Sigma (\boldsymbol{x} - \boldsymbol{x}')\right]$$

- Accurately estimate of the precision matrix ∑ is very important.

- Data-informed GP estimates the Sigma matrix by fitting the data repeatedly.

- Physics model-informed GP, by evaluating the Hessian matrix around the optimal point (could be obtained by physics model/simulation), then calculate the Sigma directly: $\Sigma = -H/2$

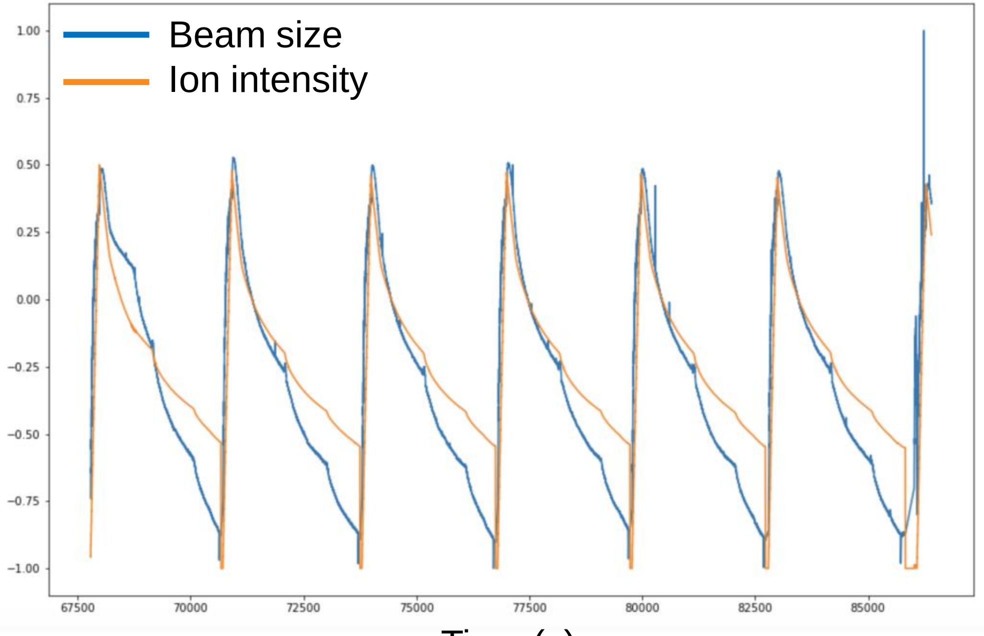
Simulation Comparisons



- Objective function: 4-dimensional Gaussian-like function centered at the origin;
- Physics model-informed GP converges faster and is more stable.

Contextual GP (CGP)

Normalized ion intensity and beam size vs. time



Time (s)

$$k = k_S \otimes k_Z = k_S(s, s') \cdot k_Z(z, z')$$

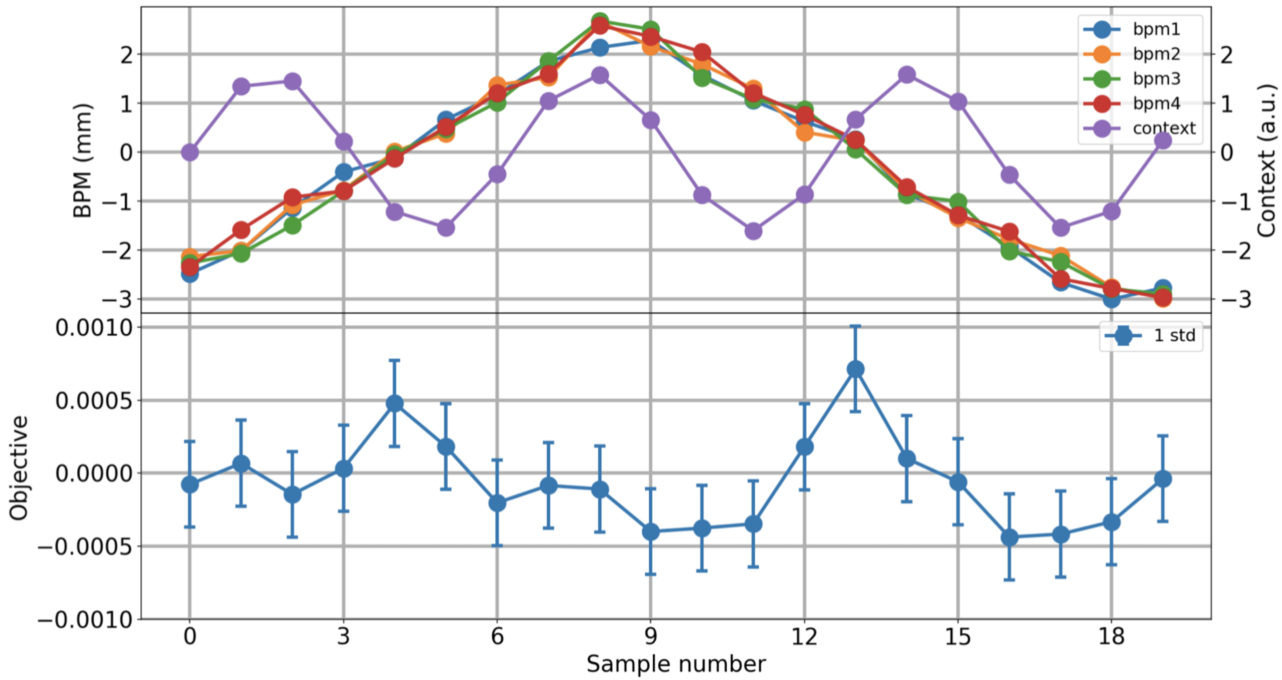
$$k = k_S \oplus k_Z = k_S(s, s') + k_Z(z, z')$$

- Contexts – <u>uncontrollable</u>, <u>varying</u> environmental conditions that affect objective function value;
- In our case – **<u>ion beam intensity</u>** decreases with time and can be treated as an environmental variable;
- Construct a composite kernel – one describes input-specific trend (k_S), the other describes context-specific trend (k_Z):
  - Multiplication
  - Summation

Brookhaven National Laboratory

CLASSE
Cornell Laboratory for Accelerator-based ScienceS & Education

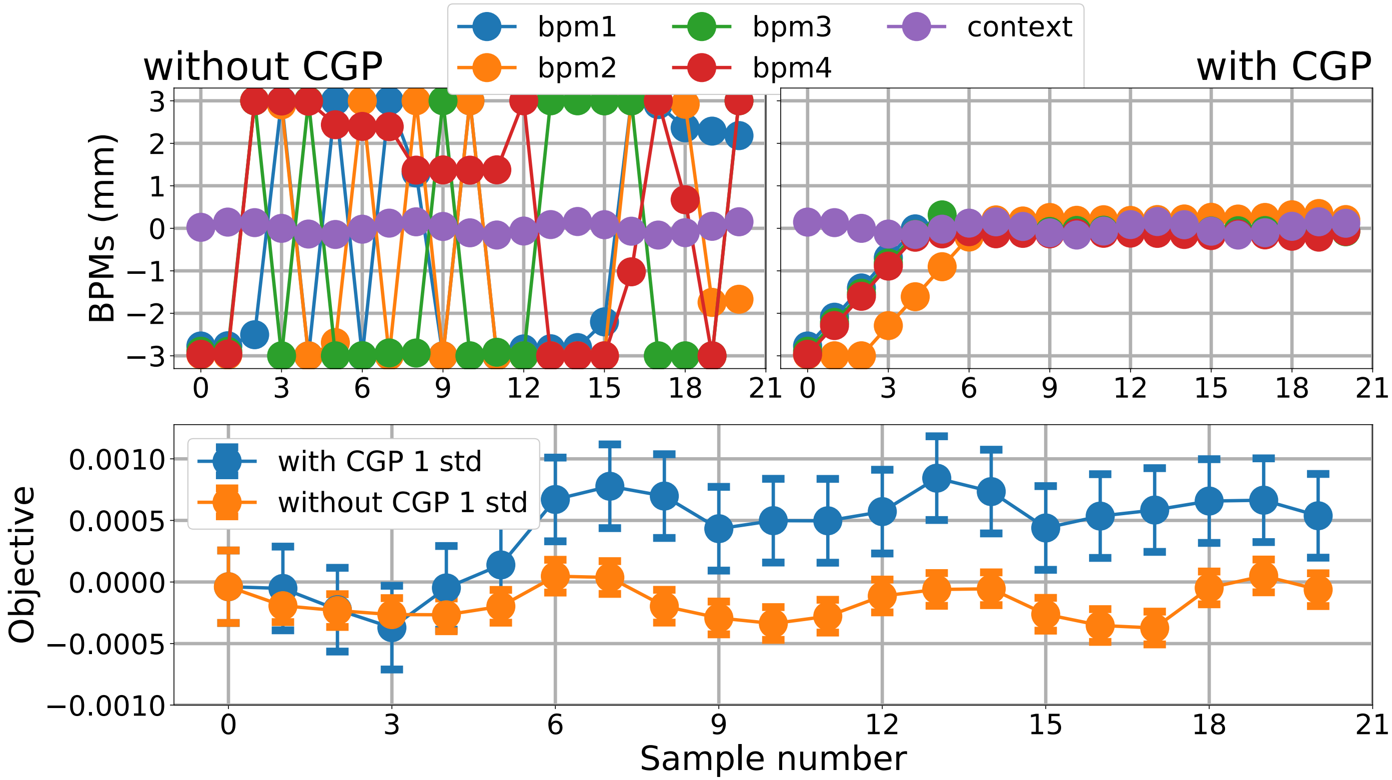
CGP Simulation



- Objective function: 4-dimensional Gaussian-like function centered at the origin plus a sinusoidal function;

- 20 initial samples;

Results comparison: Contextual GP



- Without CGP: algorithm is unable to converge due to the varying context;
- With CGP: algorithm converges in 7 steps and is stable;

Conclusion & Outlook

- The BO method can be very effective in control tasks at accelerator control systems;

- It opens many possibilities of trying different machine learning methods on optimizing performance for control tasks in the RHIC complex, as well as the future EIC.
  - Instrument calibration: Ionization Profile Monitor (IPM) at AGS;
  - Coherent electron Cooling (CeC) experiment at RHIC

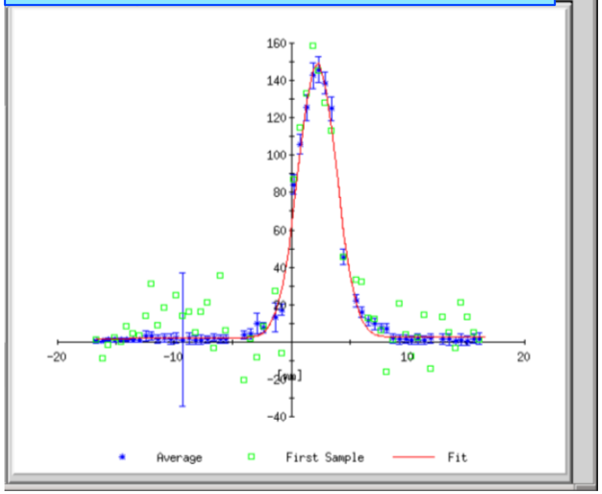
IPM Calibration

- Ionization profile monitor: measures transverse profile of the beam
  - Circulating beam ionizes residual gas in the beampipe;
  - An electric field forces electrons onto a microchannel plate (MCP);
  - Forms a projection of the beam profile;

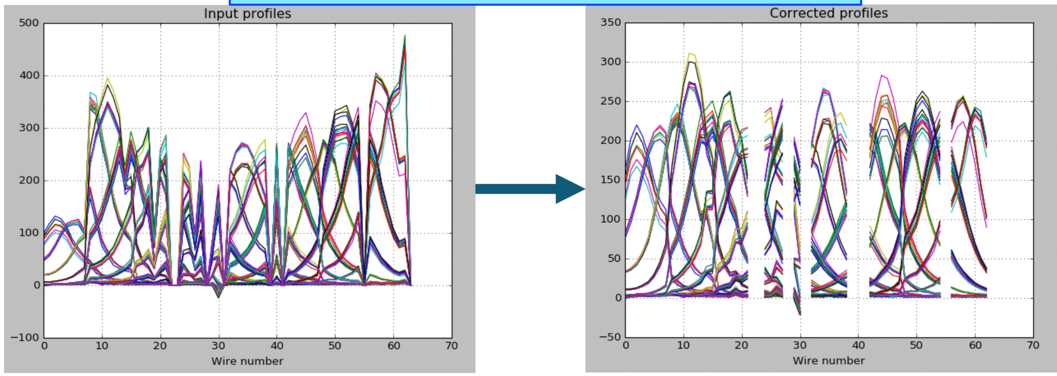
- Beam profile measurement depends on position because of systematic errors in channel gains from
  - Initial channel-to-channel gain variation;
  - Depletion of channel gains over time (systematically faster in region of high beam intensity);
  - Variation in ADC performance;
  - Usually addressed with position scans and offline calibration factors;

- Machine learning/BO opportunities:
  - Confidence intervals for channel gains, profile fit parameters;
  - Identification, imputation of data of ‘bad’ channels;
  - Data assimilation, slow calibration for aging MCPs;

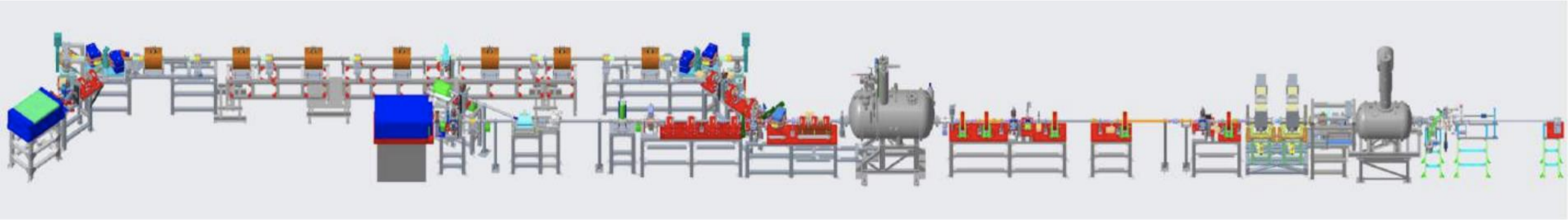
AGS IPM measurement



Position scan for calibration



Improving CeC Operations



- Motivation
  - Tuning of system parameters (i.e. solenoids and trims) are done blindly to obtain desirable beam status
  - Optimization is done by time-consuming genetic algorithm (GA)
- Goal
  - Virtual diagnostics: tuning parameters ↔ YAG screen images
  - Multi-objective optimization: peak current, emittance, energy spread etc.

Thank you !