# Machine Learning in Heavy-ion Accelerators

**Yue Hao**
Professor of Physics
Facility for Rare Isotope Beams
Department of Physics and Astronomy
Michigan State University

This material is based upon work supported by the U.S. Department of Energy Office of Science under Cooperative Agreement DE-SC0000661, the State of Michigan and Michigan State University. Michigan State University operates FRIB as a DOE Office of Science National User Facility in support of the mission of the Office of Nuclear Physics.

# AI/ML for accelerators

Accelerator Design

Construction

Commissioning

Operation

Maintenance

AI/ML can find hidden data structures and build surrogate models for:

- Machine tuning
- Anomaly Detection
- Tomography
- Model predictive control

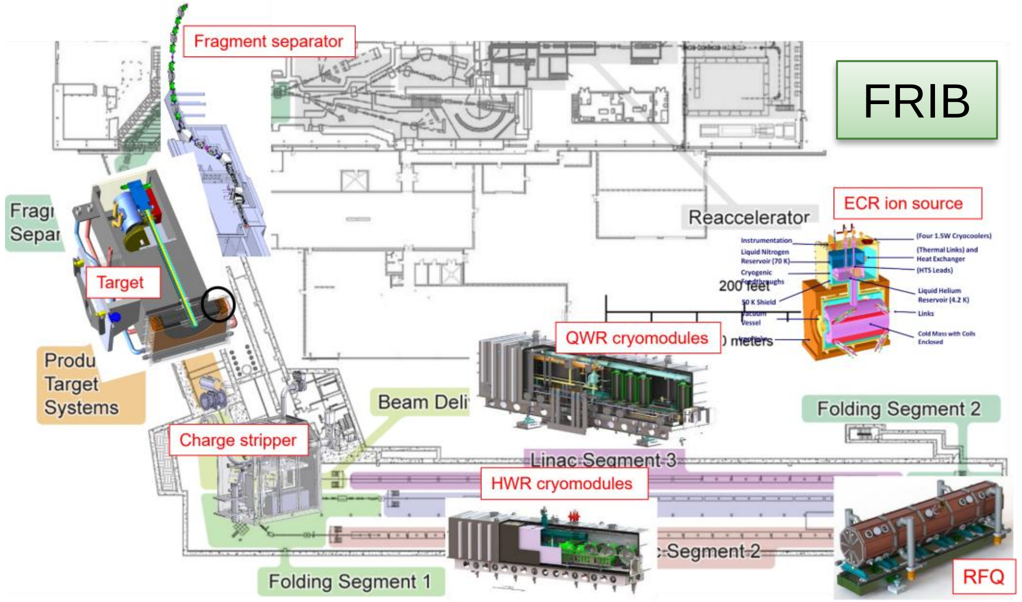
GP+BO

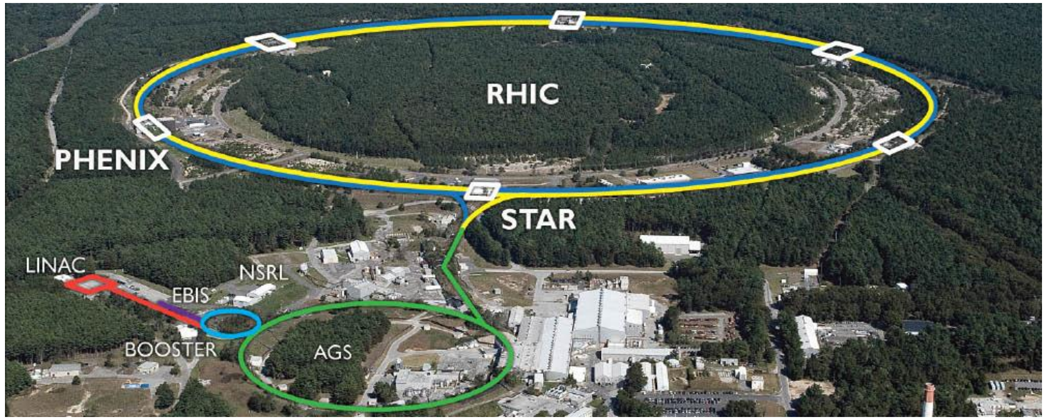
xNN

RL

Auto-Encoder

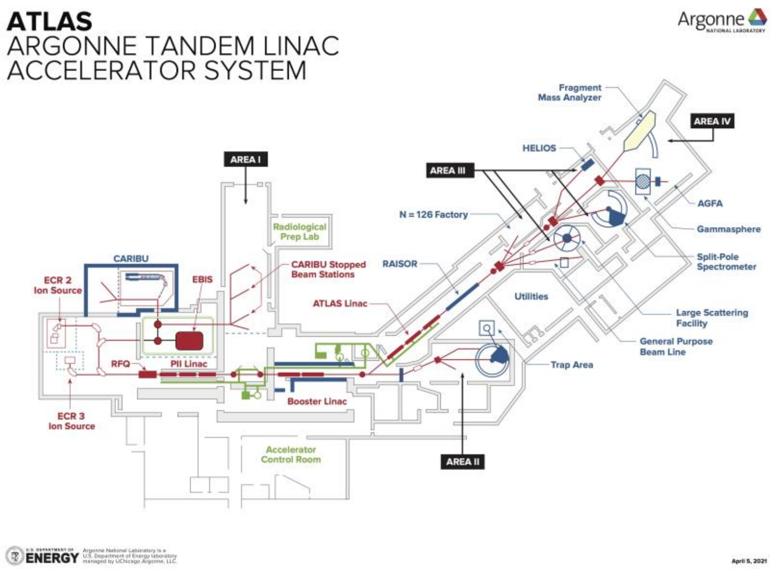

# Heavy-ion Accelerators




Goal of NP Heavy-ion Accelerator

- High power primary beam (FRIB, ATLAS)
- High Luminosity (RHIC)
- Various ion species (ALL)
- high availability/Reliability (ALL)



# Outline

- A brief survey of the AI/ML applications in heavy ion accelerator
  - AI/ML based Machine tuning in FRIB, ATLAS and RHIC
    - Various of Bayesian Optimization applications
  - Retrieving beam profiles with AI/ML method
    - Connecting with beam loss
    - Towards Tomography of phase space
- Summary

# Machine Tuning

# AI/ML Methods for Machine Tuning

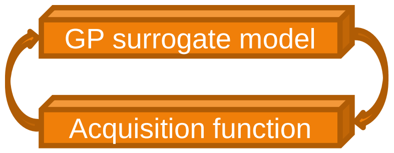
- (model independent) Reinforcement Learning (RL) vs Gaussian Process + Bayesian Optimization (BO)
  - RL can adapt to time variation (e.g. drift), and scales well to large data but sample *in*efficient.


  - BO with Gaussian Process is very sample efficient but is for static problem and scales terribly to large data

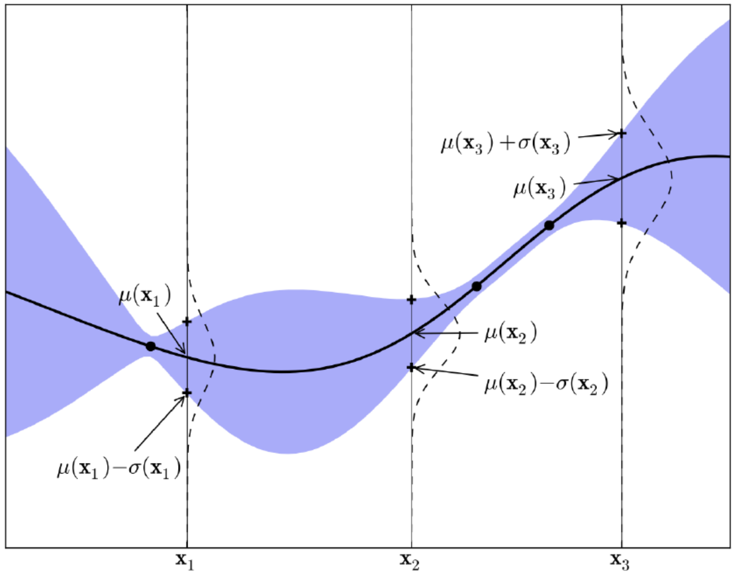
$$p(f\,|\,\mathbf{x},\mathcal{D},\boldsymbol{\theta}) = \mathcal{N}(f\,|\,\mu(\mathbf{x}),\sigma^2(\mathbf{x}))$$

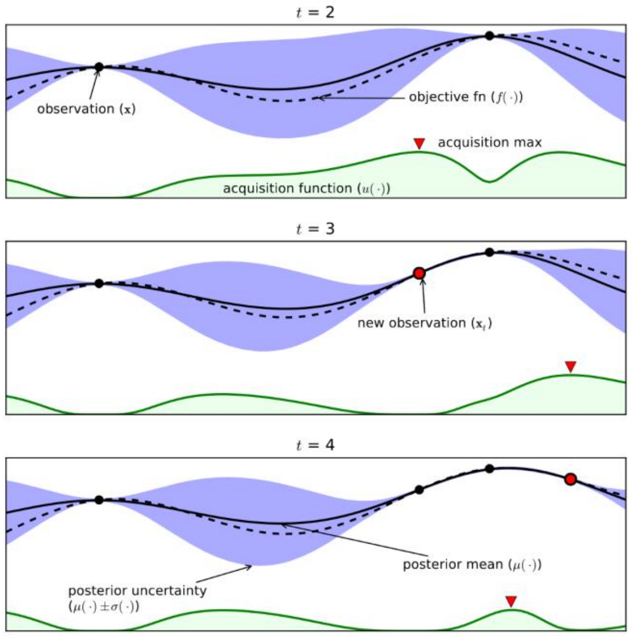
$$\mu(\mathbf{x}\,|\,\mathcal{D},\boldsymbol{\theta}) = m(\mathbf{x}) + \boldsymbol{\kappa}(\mathbf{x},X)K^{-1}(\mathbf{f}-\mathbf{m})$$

$$\sigma^2(\mathbf{x}\,|\,\mathcal{D},\boldsymbol{\theta}) = \kappa(\mathbf{x},\mathbf{x}) - \boldsymbol{\kappa}(\mathbf{x},X)^\top K^{-1}\kappa(X,\mathbf{x})$$




$$P(H|M) = \frac{P(M|H)P(H)}{P(M)} \sim P(M|H)P(H)$$

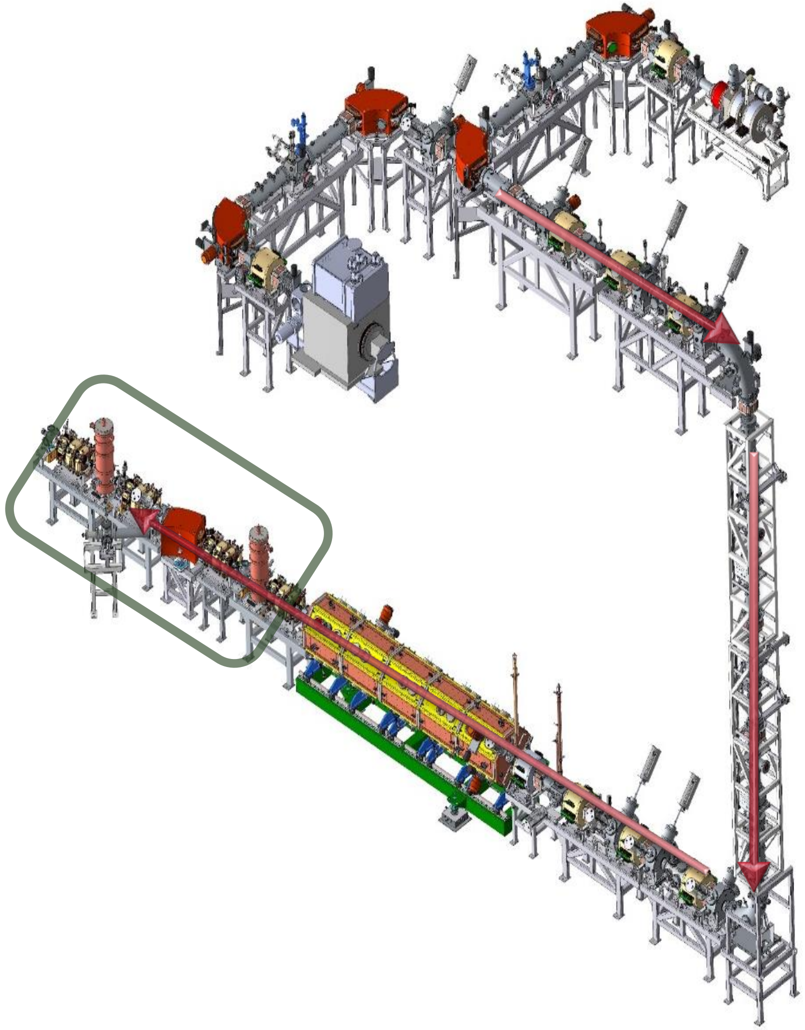



@ Brochu et al, 2010

# FRIB Beam Tuning: Problem

- Goals
  - Minimization 3D beam centroid deviations at 3 MEBT BPMs
  - Maximize beam current ratio $\frac{I_{\text{afterRFQ}}}{I_{\text{beforeRFQ}}}$ at two BCMs
  - Maximize beam current at two BCMs and 3 BPMs
- 6 Decision Knobs:
  - electric currents for magnetic correctors
- Goal Budget: <10 min
- Cost
  - $2\ sec$ for BPM reading
  - $1A/sec$ (max $\pm5A$) for electric current ramping of correctors
  - Additional $15\ sec$ for the electric current polarity change



# FRIB Beam Tuning: Strategy

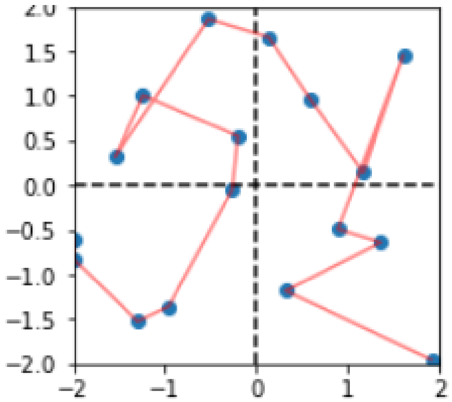
- Asynchronous evaluation
  - Evaluate objective of a candidate solution on machine while computing BO step (model training and candidate query with penalization on currently evaluating candidate)
- Penalize and favor
  - On the currently evaluating candidate one need to penalize for asynchronous BO while at the same need to favor to reduce ramping time.
    - » need careful choice of length scale and weight for the penalize / favor
  - Favor current polarity

$$f_{penal} = -C_{penal} e^{-(x-x_{penal})^2/L_{penal}^2}$$

$$f_{favor} = +C_{favor} e^{-(x-x_{favor})^2/L_{favor}^2}$$

$$f_{polarity} = \begin{cases} +C_{polaity} & \text{if } sign(x) = sign(x_{current}) \\ 0 & \text{else} \end{cases}$$

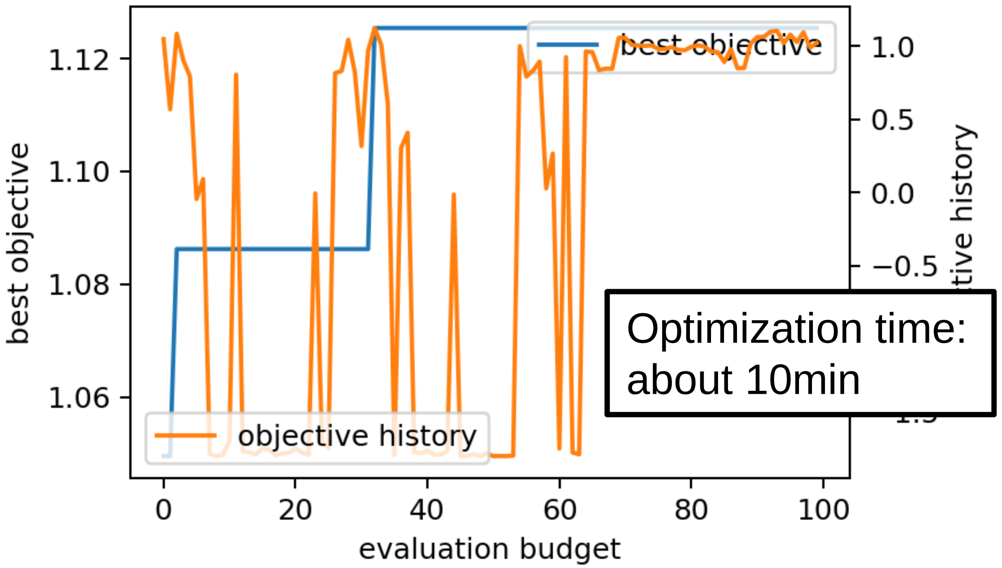
- Minimize ramping path of initial training samples



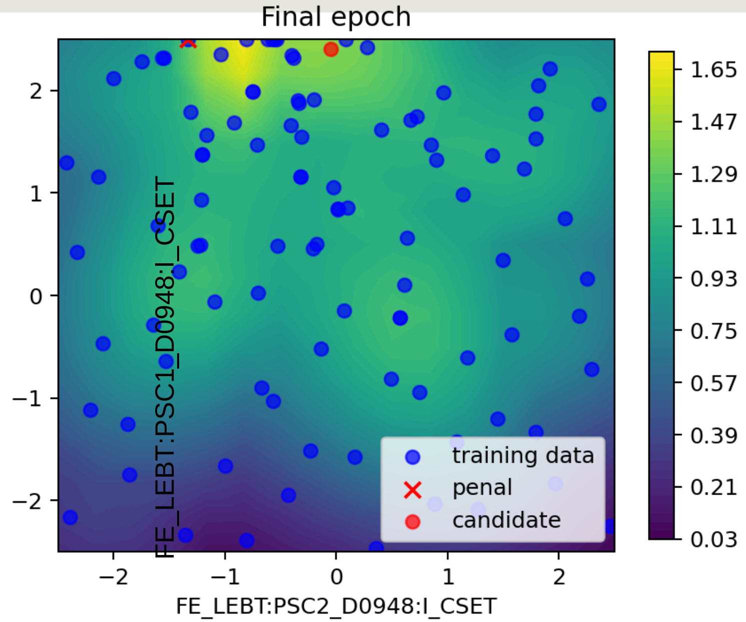

example shows 16 initial samples in 2D decision parameter domain.

@ K. Hwang

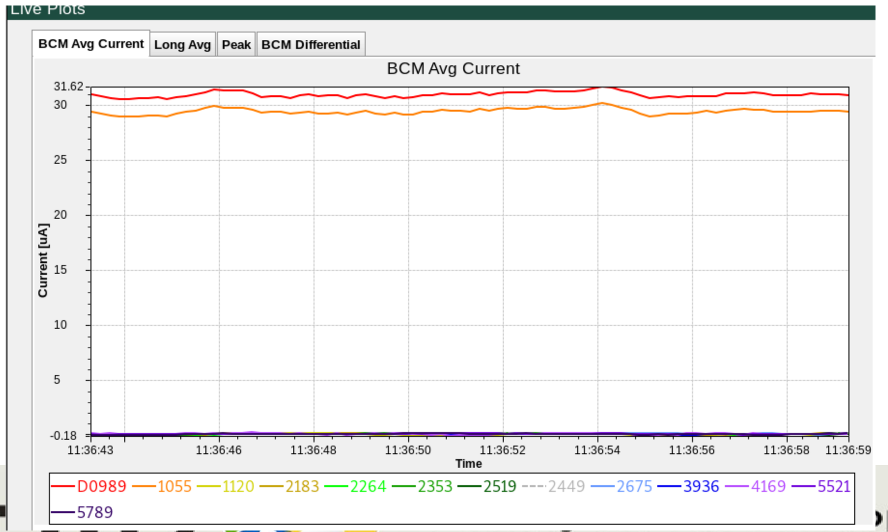
# FRIB Beam Tuning: Results

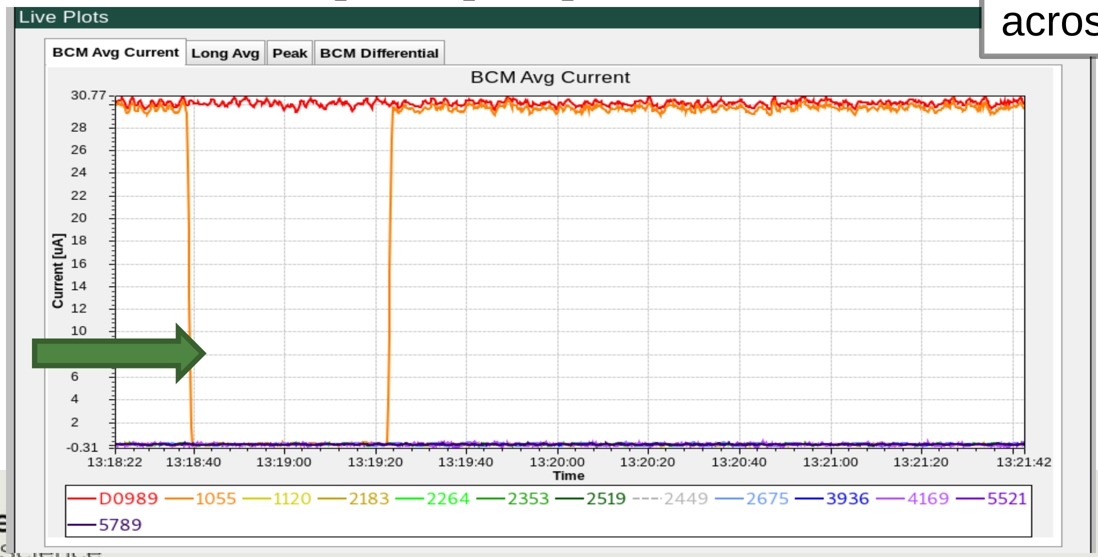



Optimization time:
about 10min




GP mean visualized by projecting maximum across projection axes

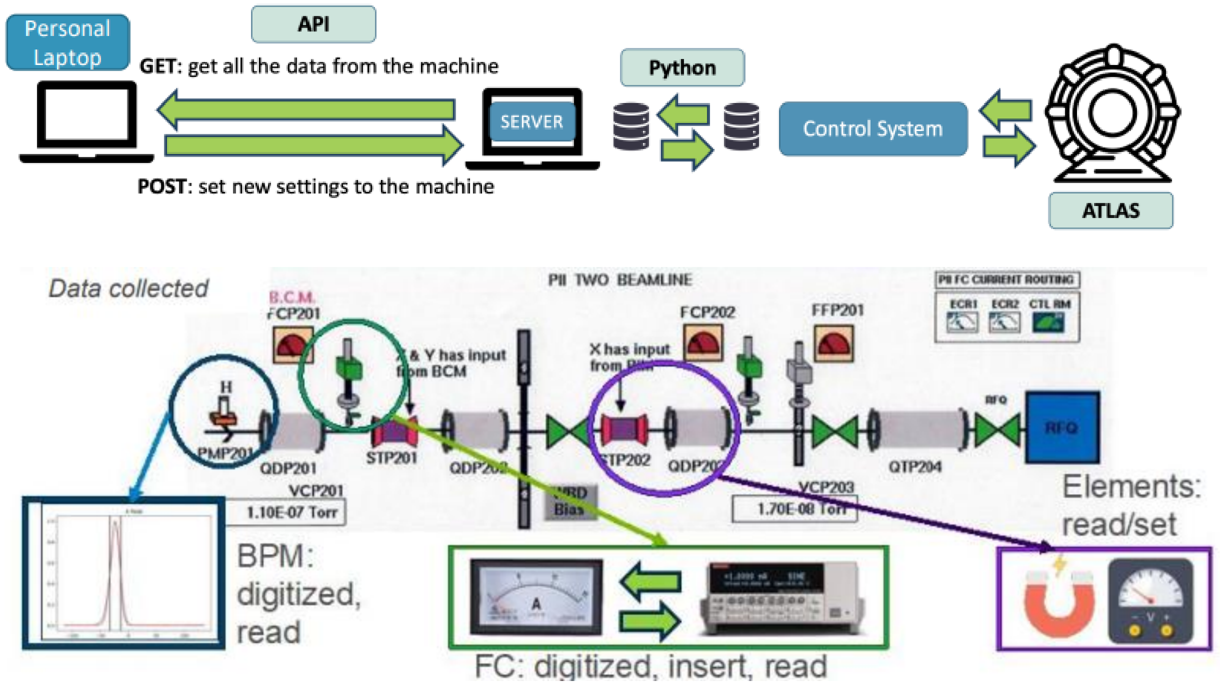


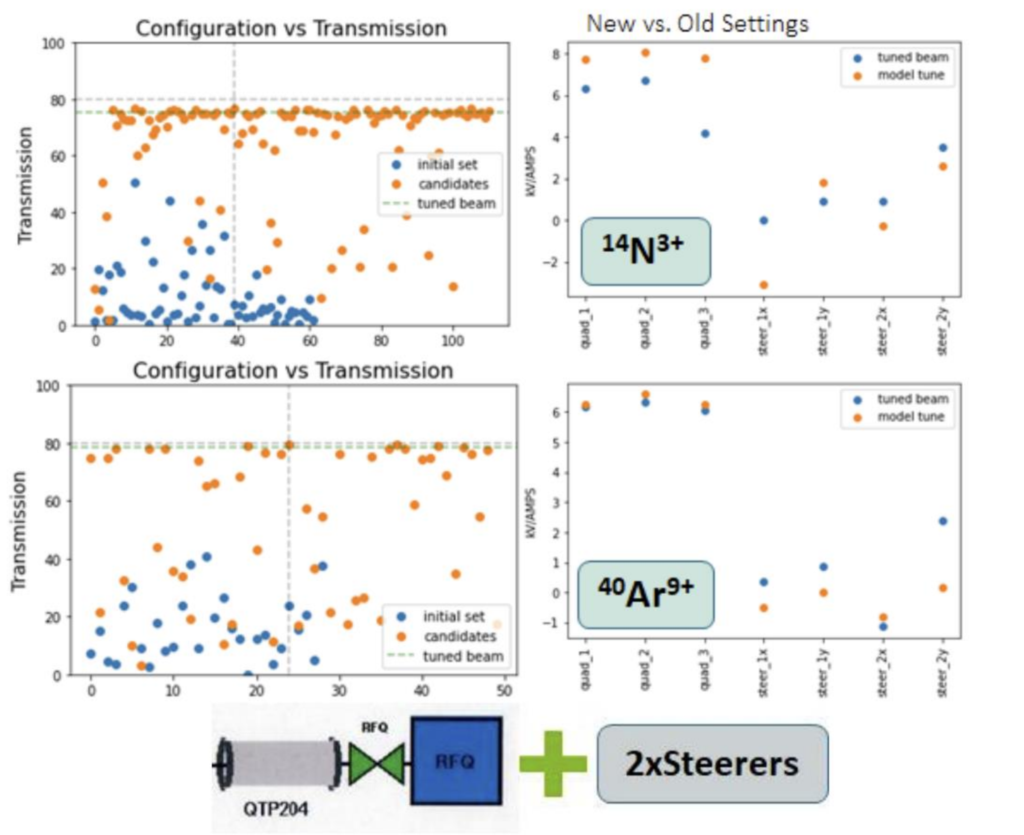

@ K. Hwang

# ATLAS Tuning: Optimize Beam Transmission

- Digitize the Legacy System
  - New Python API for machine tuning
  - Offline modeling: Track code



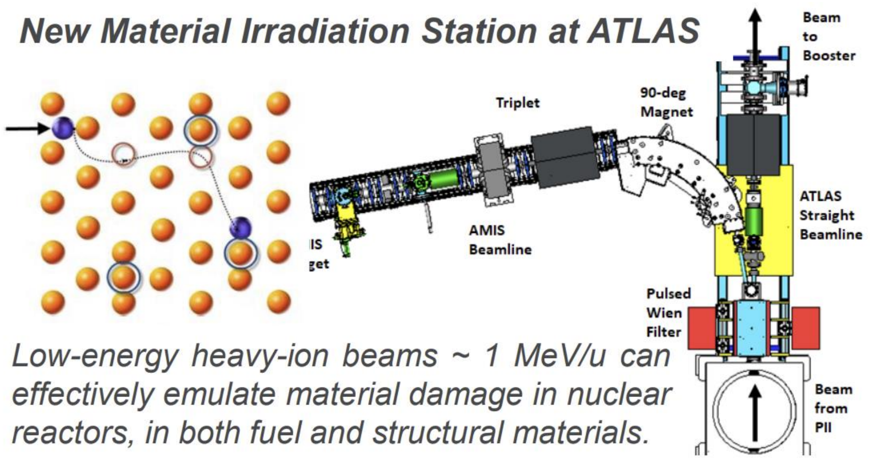

@ B. Mustapha



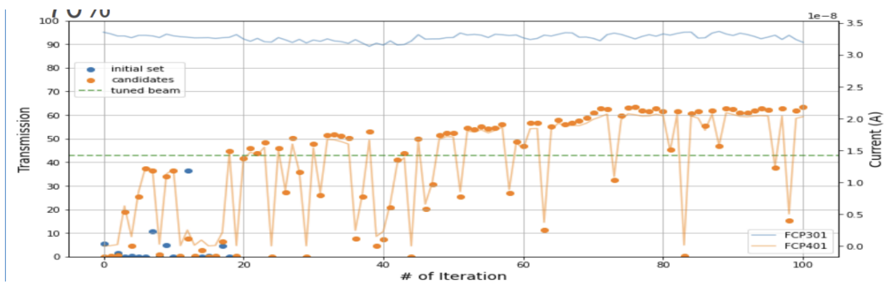

- 7 varied parameters (3 quads + 2 steerers)
- Optimization of beam transmission
- Case of $^{14}N^{3+}$ : 29 historical + 33 random tunes
- Case of $^{40}Ar^{9+}$ : 29 historical tunes

# ATLAS Tuning: Multi-Object BO

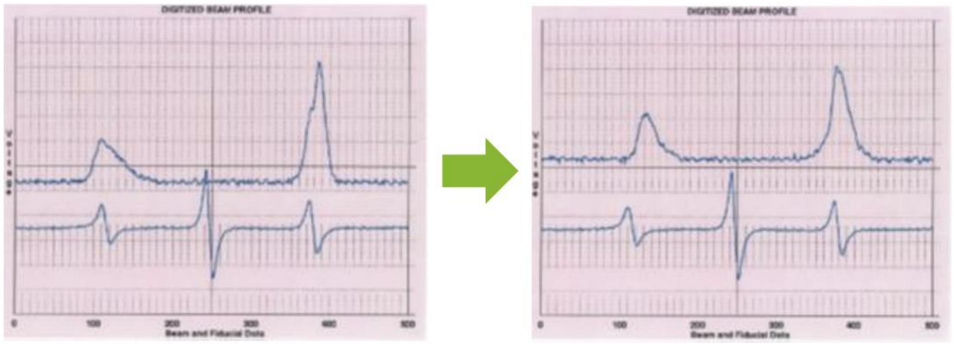
Use Bayesian Optimization for dual objections

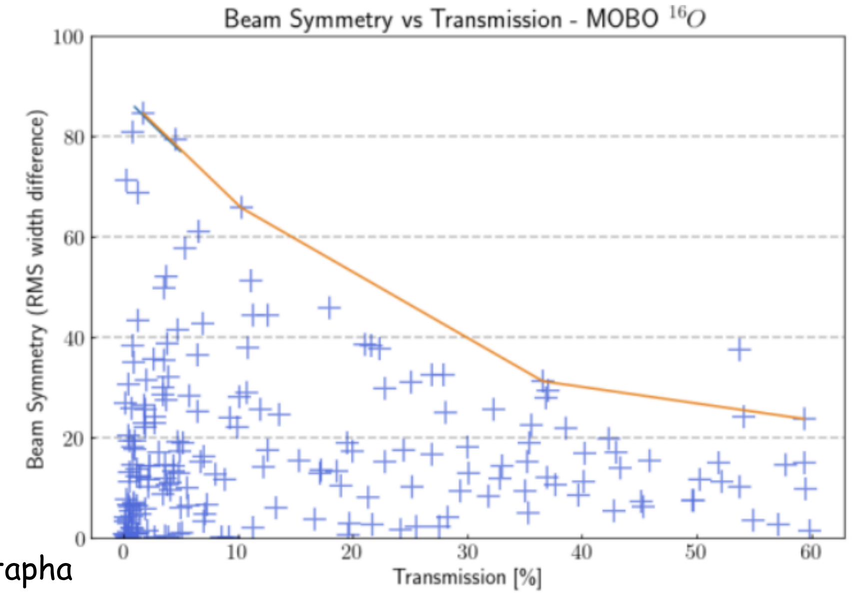



Optimizing transmission




Optimizing Profile, symmetric

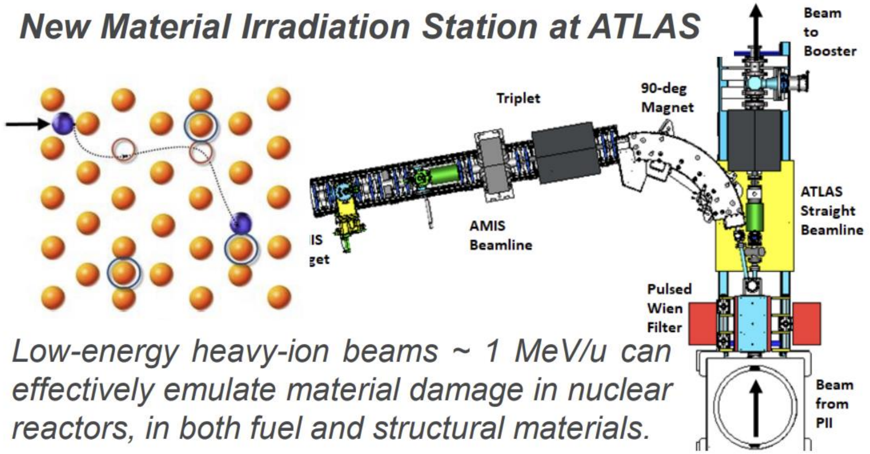


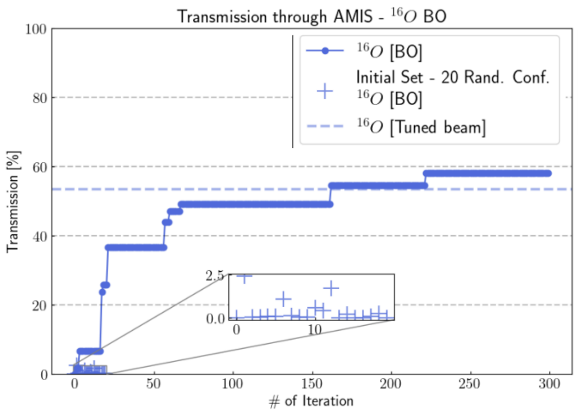

@ B. Mustapha

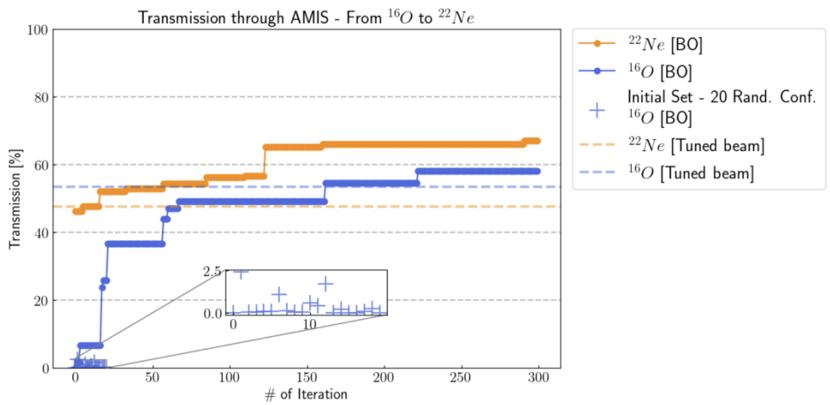
Pareto front of two objects.

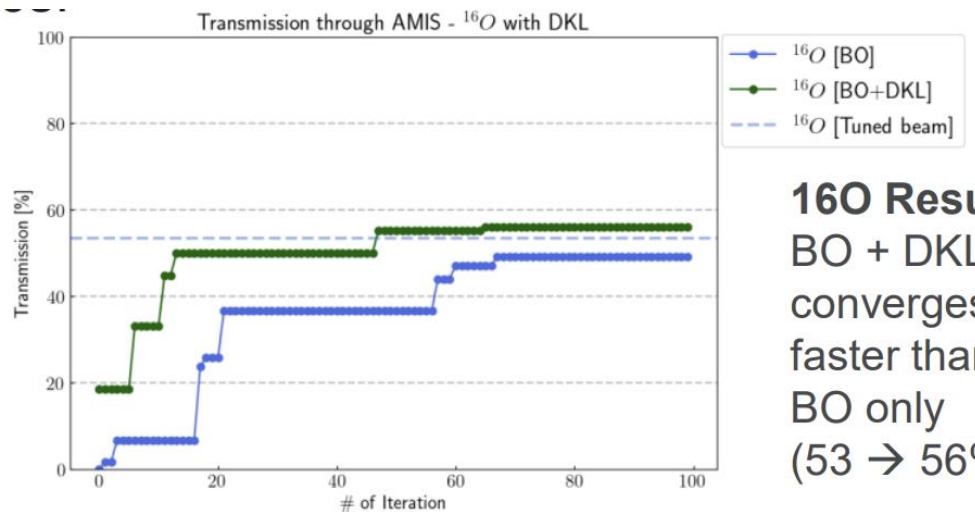
# ATLAS Tuning: Transfer learning




Establish model using $^{16}$O then use on $^{22}$Ne, with scaling

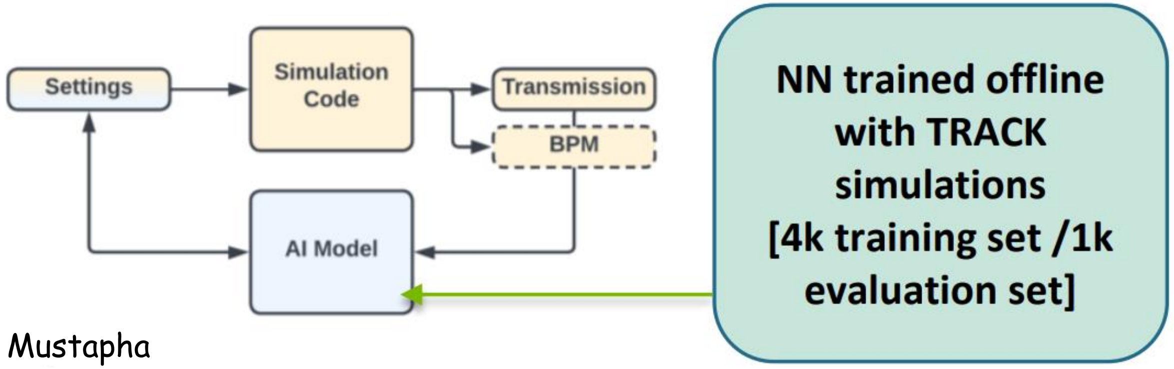


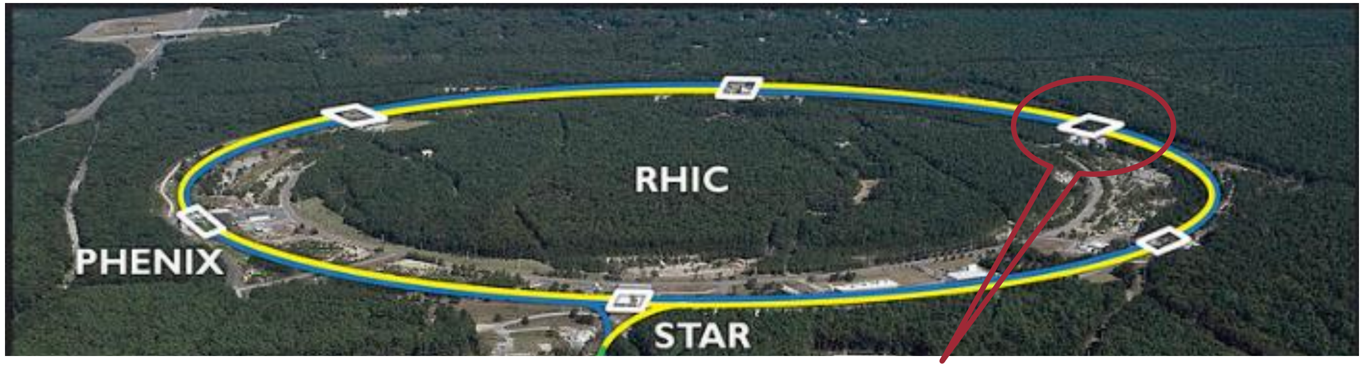

**16O Results**:
BO + DKL converges faster than BO only
(53 → 56%)

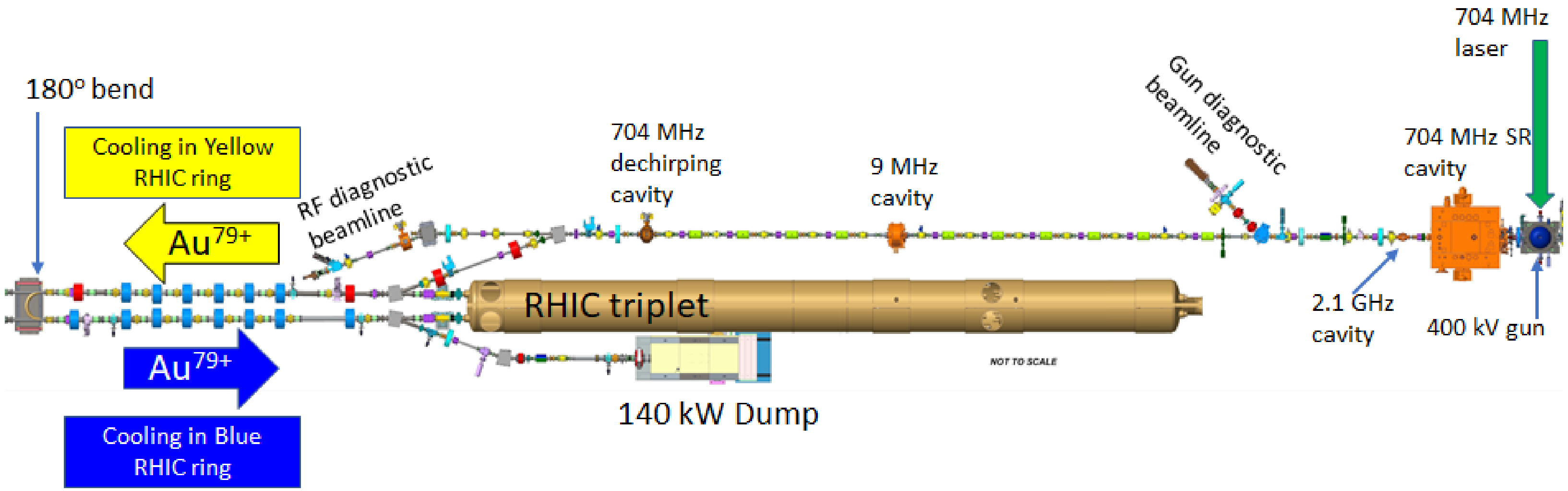
Transfer simulation knowledge using Deep Kernel Learning



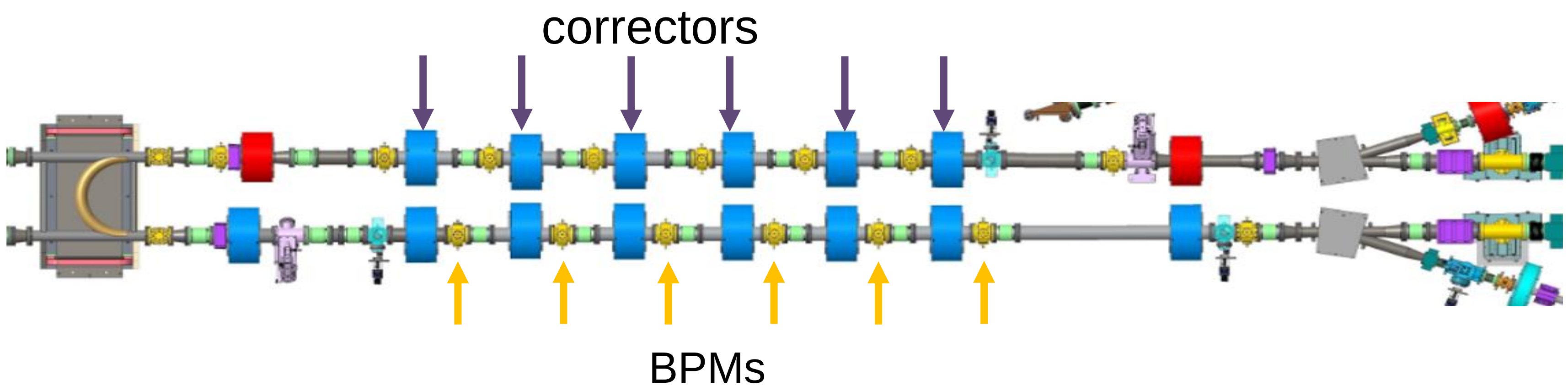

@ B. Mustapha

# AI/ML application in Cooling Exp at RHIC



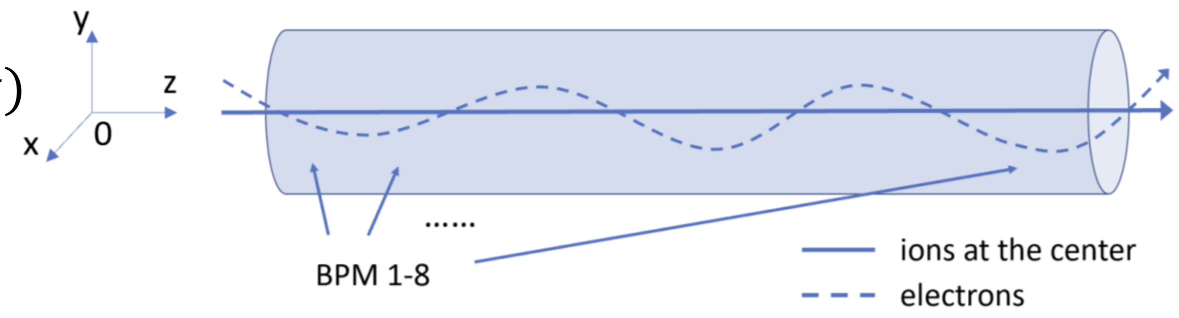
# RHIC: BO to optimize cooling rate




- Only the first 4 BPMs are considered;
- Cooling performance is measured by the cooling rate;
- Decreasing speed of transverse ion beam size:

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

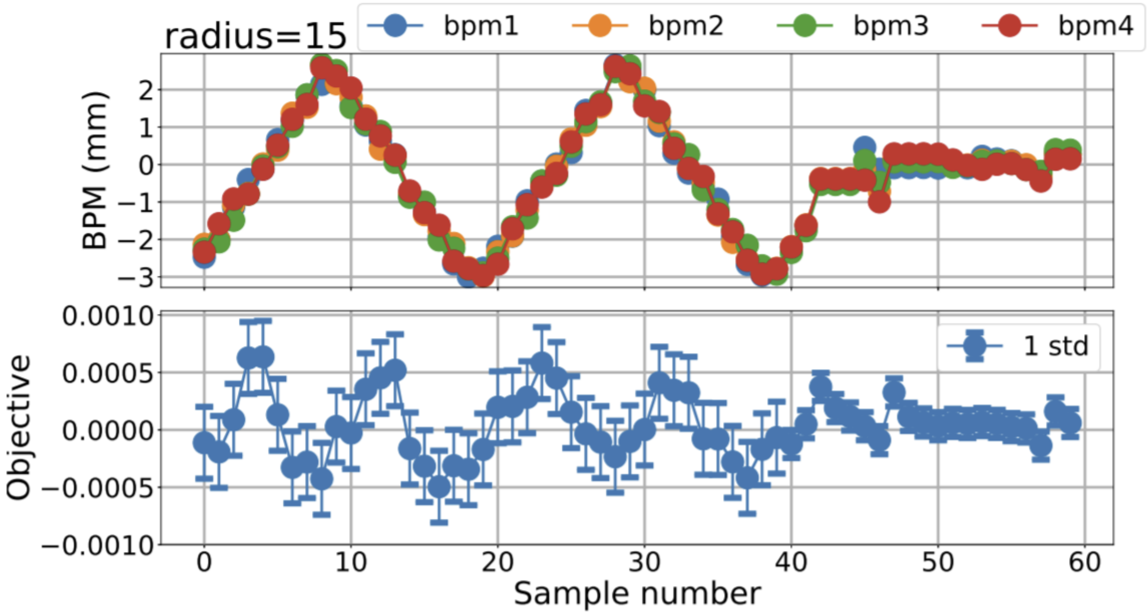
- A more negative $\lambda$ means a faster cooling rate;
- Ions are assumed in the center position (x=0, y=0).

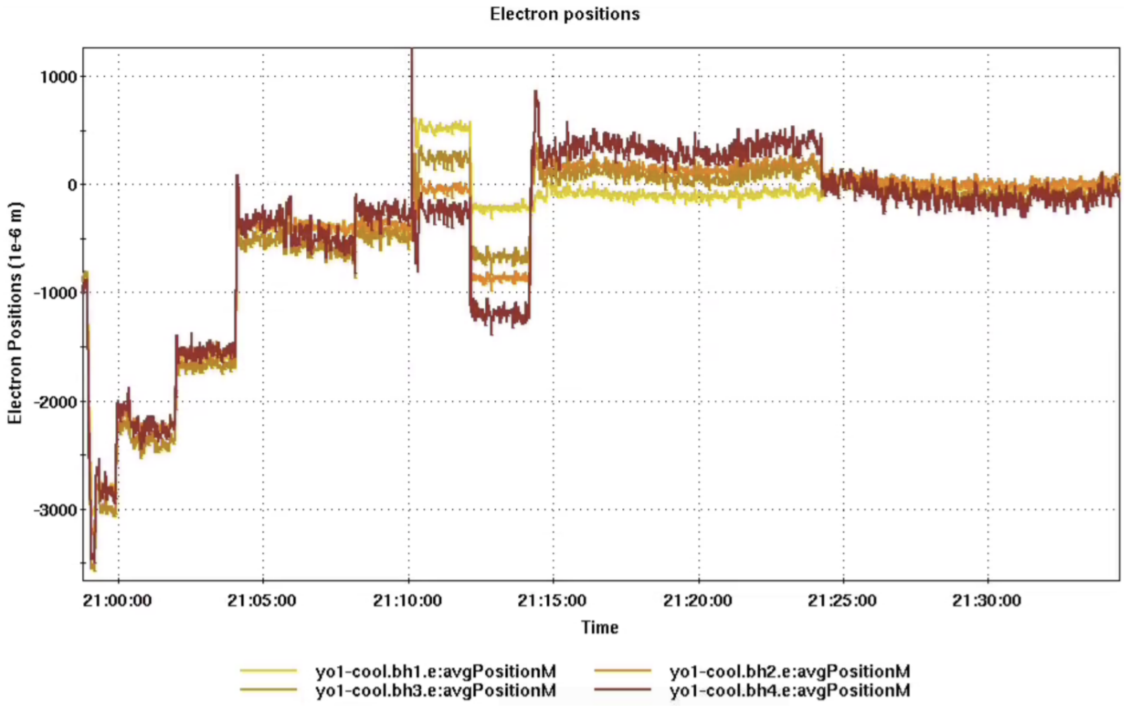



@ Y. Gao and W. Lin

# RHIC: BO to optimize cooling rate

- 4 BPMs are used to optimize the cooling rate.  Cooling rate is observed to be optimized with zero BPM offsets.



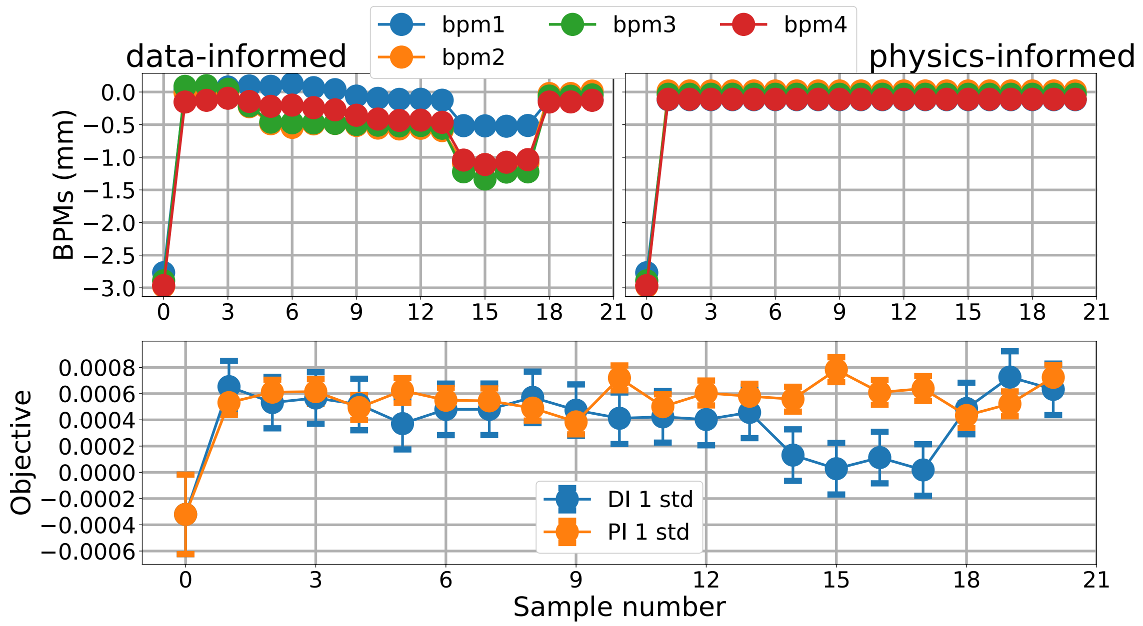




@ Y. Gao and W. Lin

# RHIC: Physics informed and Contextual GP

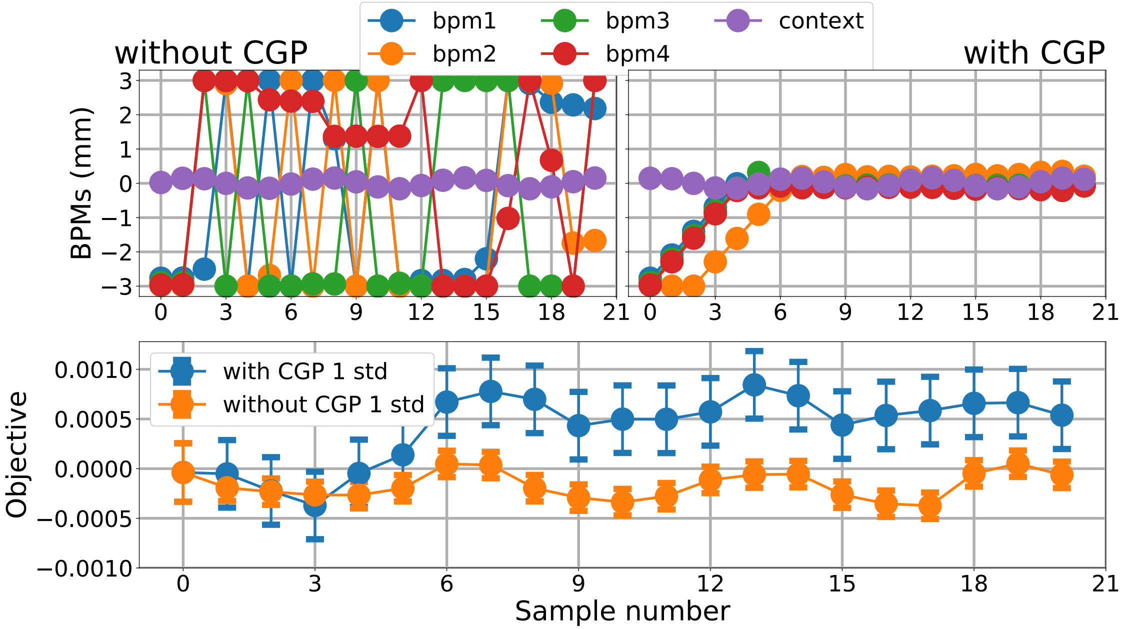
## Physics-Informed GP

Replace the kernel (RBF) with the Hessian from the simulation data near the optimum point




## Contextual GP

Handel the environmental change, such as the intensity change during one ‘store’, using Contextual-UCB

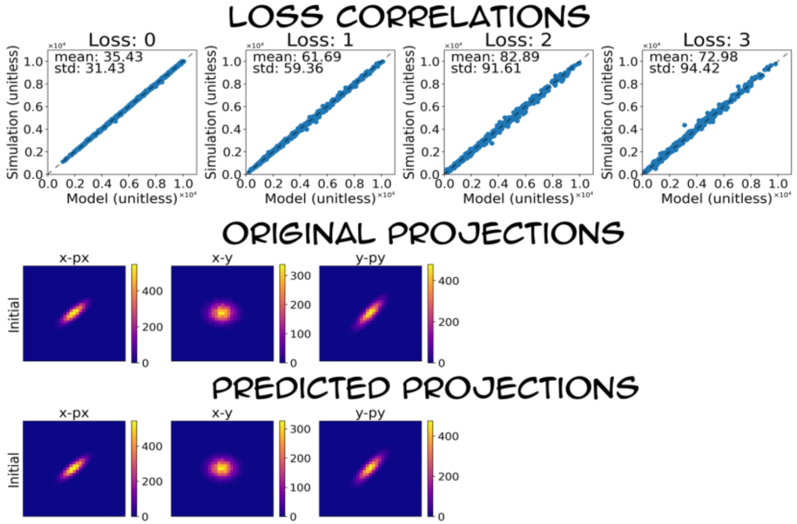


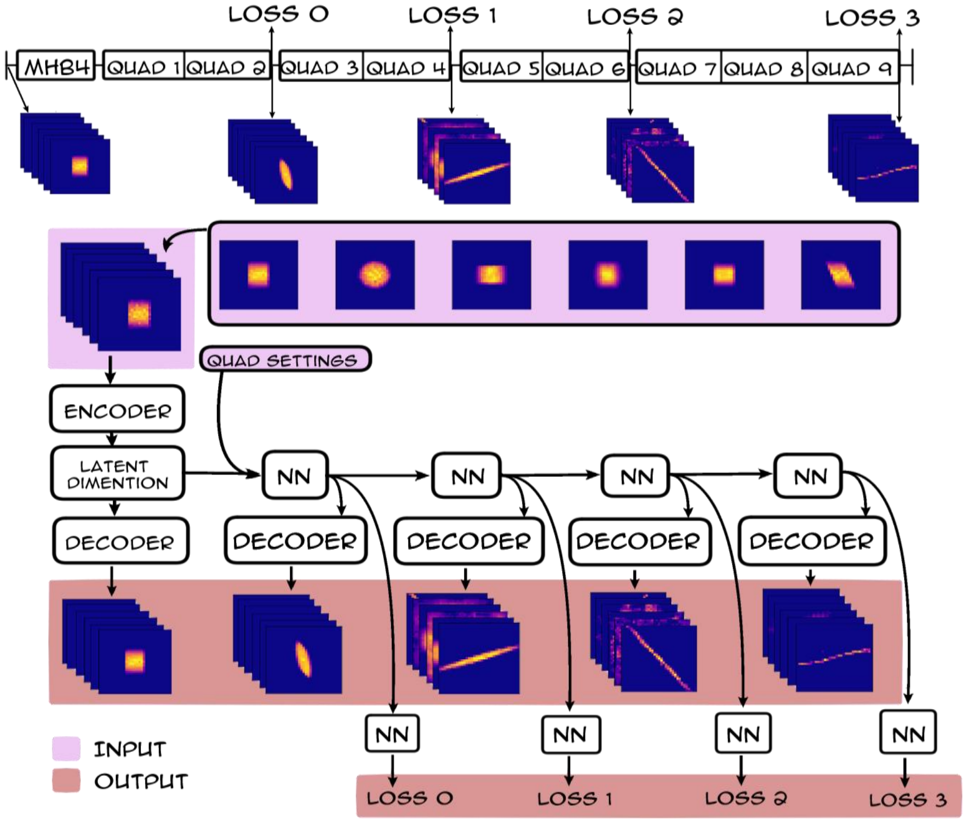

@ Y. Gao and W. Lin

# Beam Loss, Beam Profile, and Tomography

# Beam distribution in NP accelerators

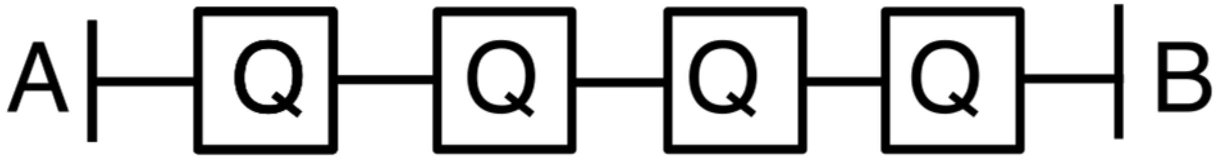
- Unlike light sources, for NP accelerators, the knowledge of beam distribution (other than the 2nd order moments) are used to provide **better beam matching** and **minimize the uncontrolled losses**.
- Beam distribution inferred from a series of 2-D profiles in latent space.
- Associate latent space with beam loss.
- No accelerator physics is used.



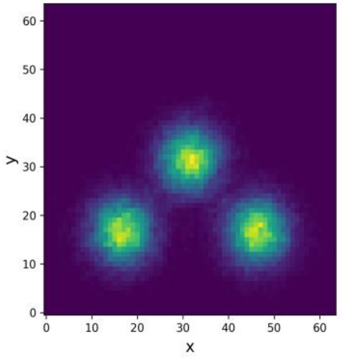




~1% accuracy achieved in this simulation demonstration, with aggressive data needs

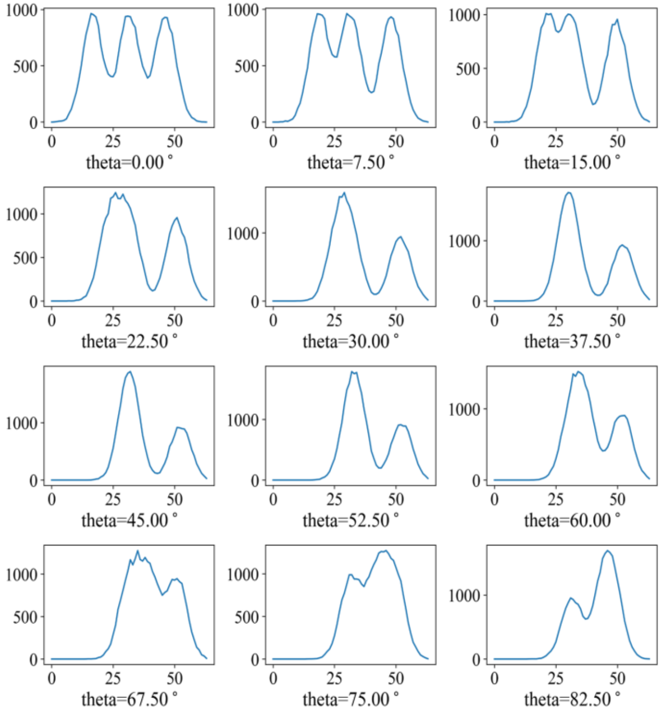
# Tomography




Ground Truth

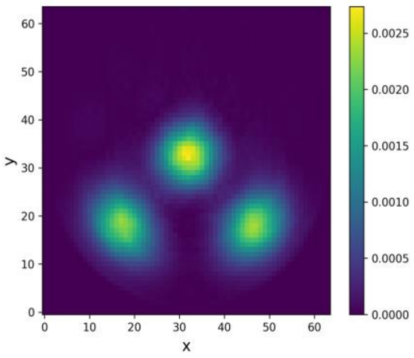


$M(A \rightarrow B)$, Projection

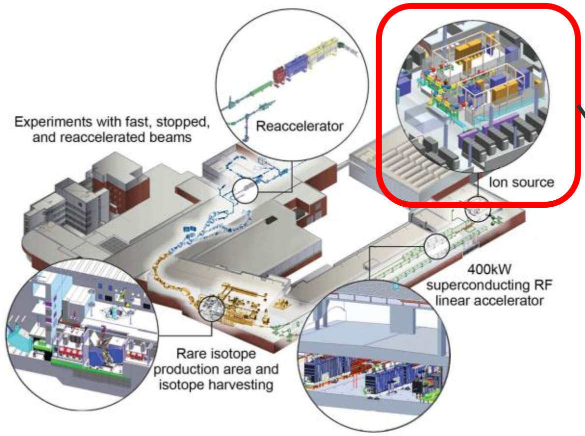



$M^{-1}(A \rightarrow B)$, Infer

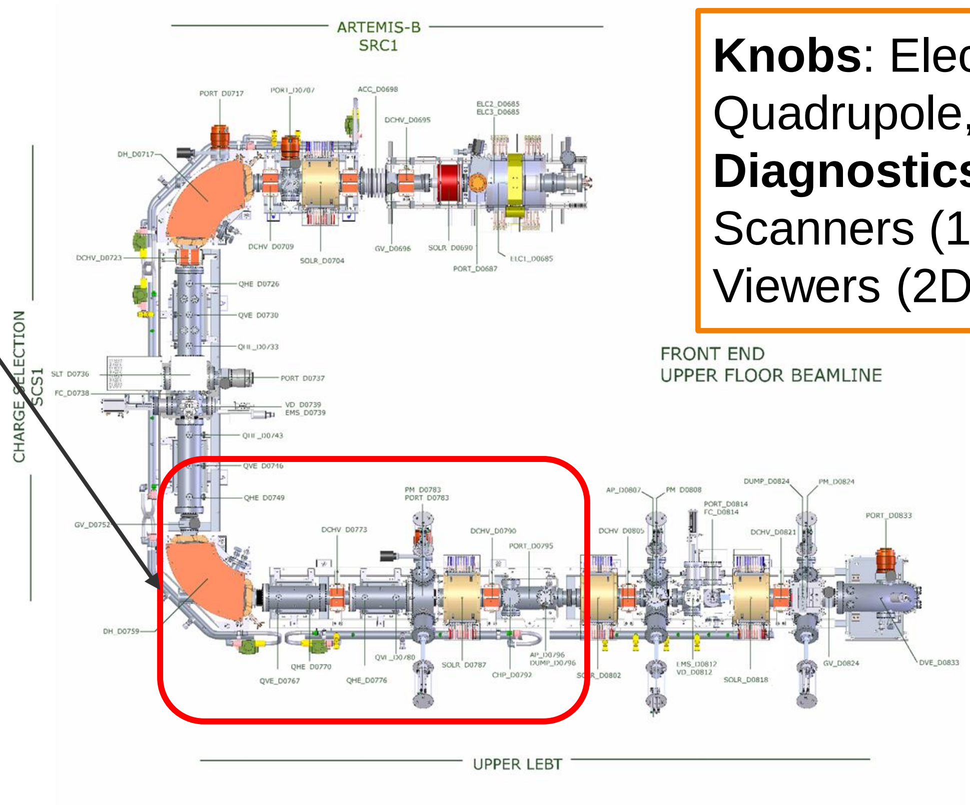
Reconstruction



# Tomography at FRIB - Simulation




- Linear accelerator
- Rare isotope production with primary beams up to 400 kW, 200 MeV/u uranium
- Understanding the beam will help control beam loss at high intensity

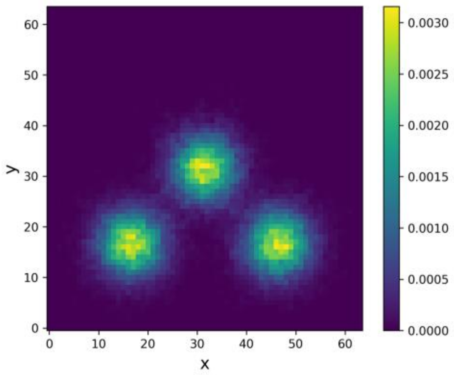



**Knobs**: Electric Quadrupole, Solenoids
**Diagnostics**: Wire Scanners (1D), Viewers (2D)

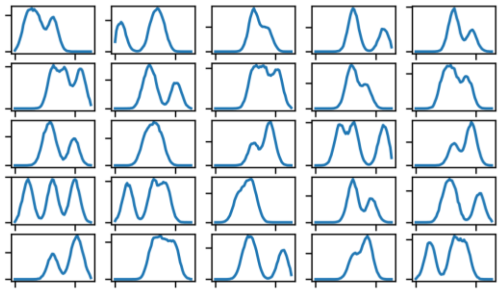
Fox, https://www.fox47news.com/neighborhoods/msu-campus/after-13-years-and-close-to-1-billion-msus-facility-for-rare-isotope-beams-is-up-and-running

# Maximum Entropy (MENT) method

Distributions:




Measurements Projections



The entropy of a beam distribution:

$$H(f) = -\int\int dxdy f(x,y)\ln f(x,y)$$

Constrains are the profile measurement (projections) down stream, denoted p(s). Using Lagrange multiplier:

$$\psi(f,\lambda) = H(f) + \sum_{n=1}^{N}\int ds\lambda_n(s)\left[\int dt f(x_n,y_n) - p_n(s)\right]$$

$$\frac{\partial\psi}{\partial f} = 0, \qquad \frac{\partial\psi}{\partial\lambda} = 0$$

It can be solved, albeit with computational difficulties

$$f(x,y) = \prod_{n=1}^{N} h_n[s_n(x,y)] \qquad h_n(s) = \frac{p_n(s)}{\int dt \prod_{k\neq n} h_k[s_k(x_n,y_n)]}$$

# MENT example for 4-D Distribution

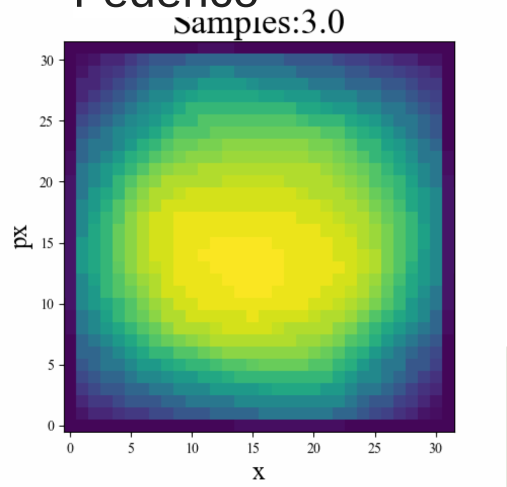
$$\vec{x}_s = R_{4\times 4}\vec{x_{s0}}$$

$$R_{4\times 4} = \prod_{i\neq j} R_{ij},$$

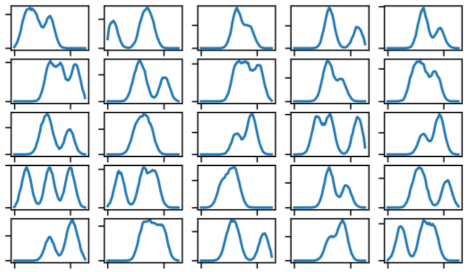
$$i, j \in (x, x', y, y')$$

$$R_{4\times 4} \to \begin{bmatrix} R(\alpha_1) & 0 \\ 0 & R(\alpha_2) \end{bmatrix}$$

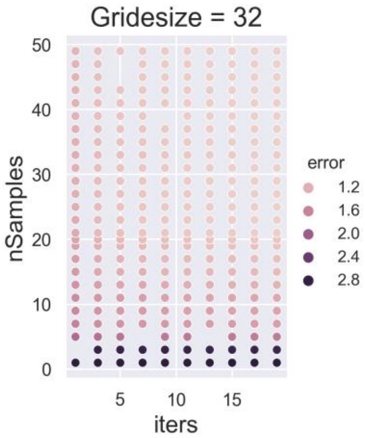
Change in reference frame [6] T. Federico



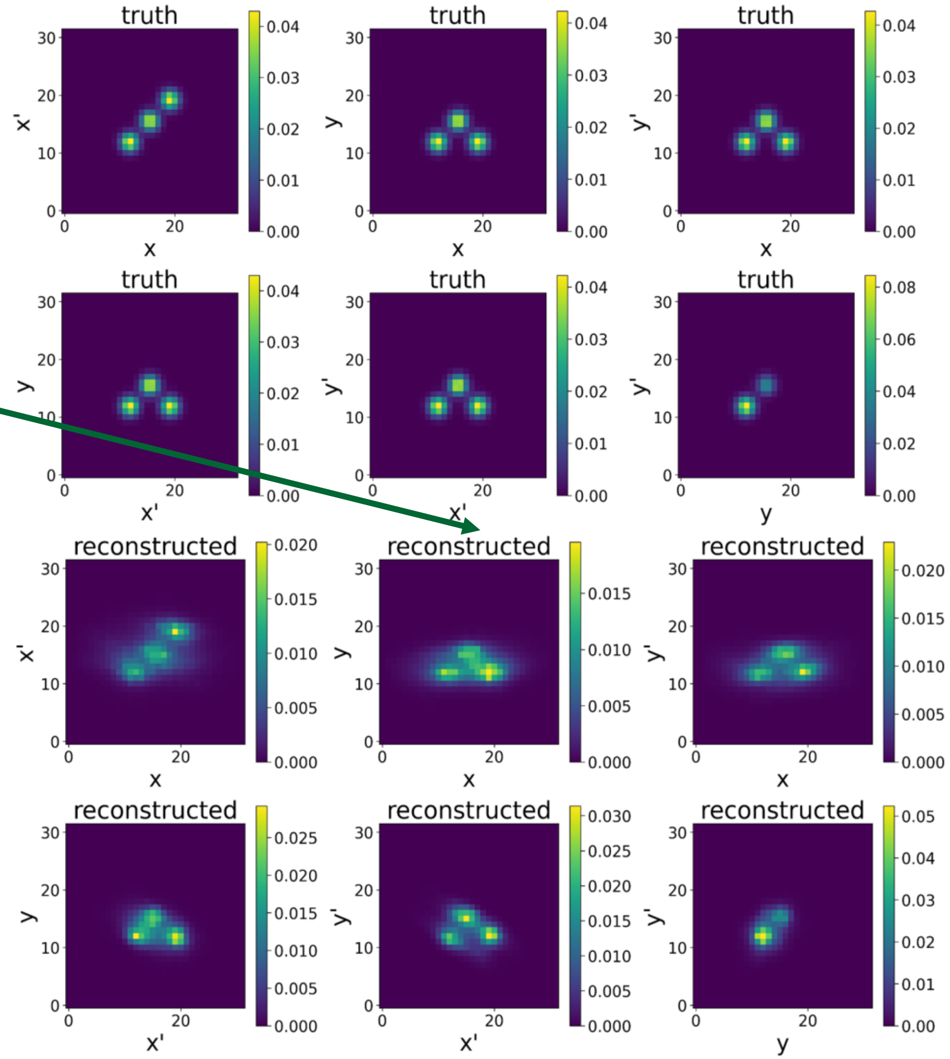

Conditions for “trick 2” to work?



Ground Truth



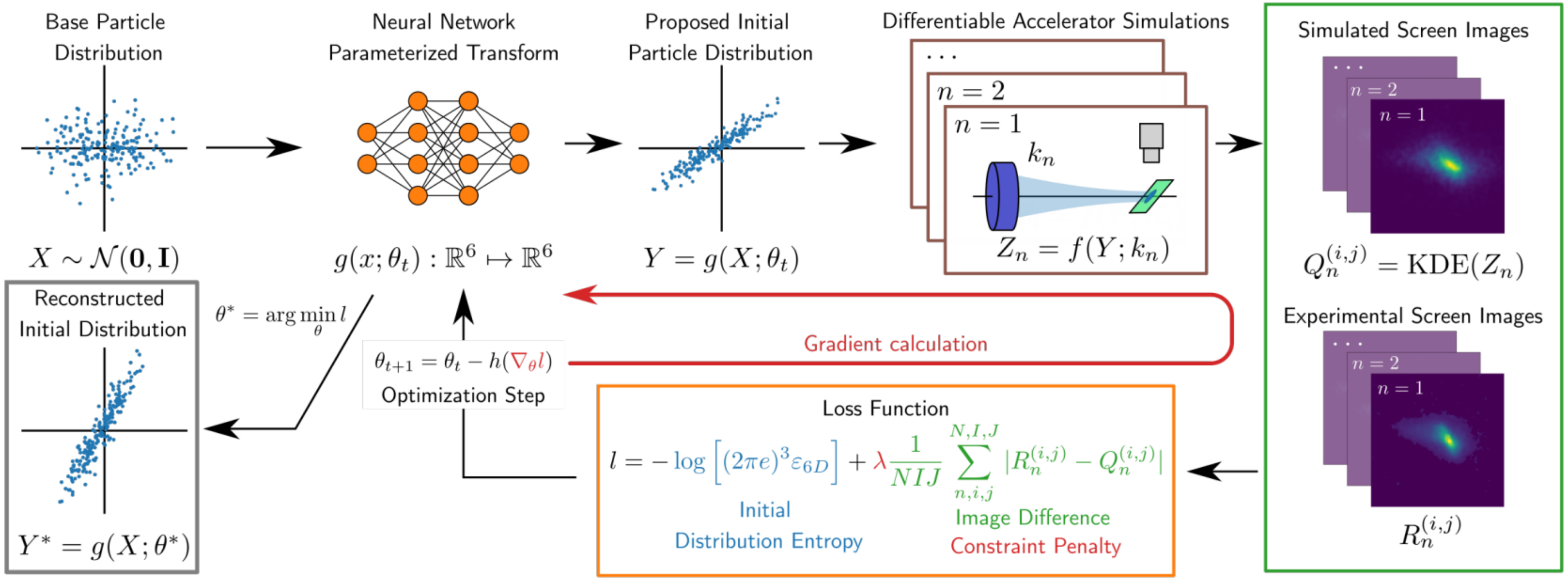

Reconstruct - 45 1D samples



# ML based MENT

Recent paper has demonstrated that MENT is ML friendly:

@ R. Roussel, et.al. PRL 130 145001, 2023




- ML approach bypasses the challenging iteration process.
- Predict the core of the beam well.
- ➢ Is finite number of measurement good enough for the beam tail 's information?

# Summary

- Machine Learning techniques are powerful tools to boost accelerator performance.
- Currently the demonstrated ML application is limited to 10 knob/feature, order of magnitudes less than number of control knobs in accelerator complex (~10K knobs)
- Problem Isolation and dimension reduction are the key.
- Beam matching and loss/background control is a challenging topic for NP accelerators.
- The link of beam distribution and losses is a challenging problem even for ML methods.
- A combination of data-driven and physics-based approach seems the promising way to proceed.

# Acknowledgement

- The materials are provided by colleagues working on ATLAS, RHIC and FRIB
  - Brahim Mustapha, Jose Martinez (ATLAS, ANL)
  - Yuan Gao, Kevin Brown (RHIC, BNL); Weijian (Lucy) Lin, Georg Hoffstaetter (Cornell U)
  - Kilean Hwang, Anthony Tran (FRIB, MSU)

*Thank you.*