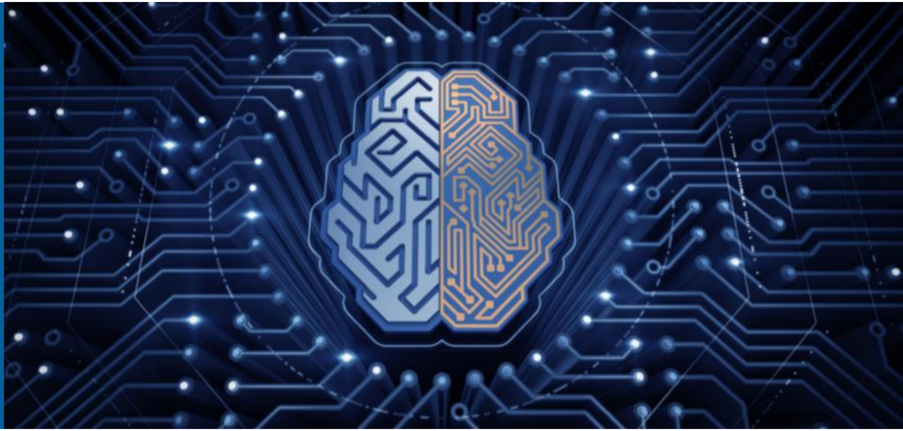
2023 DOE/NP AI-ML DATA SCIENCE PI EXCHANGE MEETING

USE OF AI-ML TO OPTIMIZE ACCELERATOR OPERATIONS & IMPROVE MACHINE PERFORMANCE



PRESENTER

Brahim Mustapha
Physics Division
Argonne National Laboratory

CONTRIBUTORS

Jose Martinez, Postdoc (ANL/PHY)
Anthony Tran, Student (MSU/FRIB)
Ian Sugrue, Student (NIU, ANL/HEP)

Yue Hao (MSU/FRIB)
John Power (ANL/HEP)
Philippe Piot (NIU, ANL/APS)
Nathan Krislock (NIU)

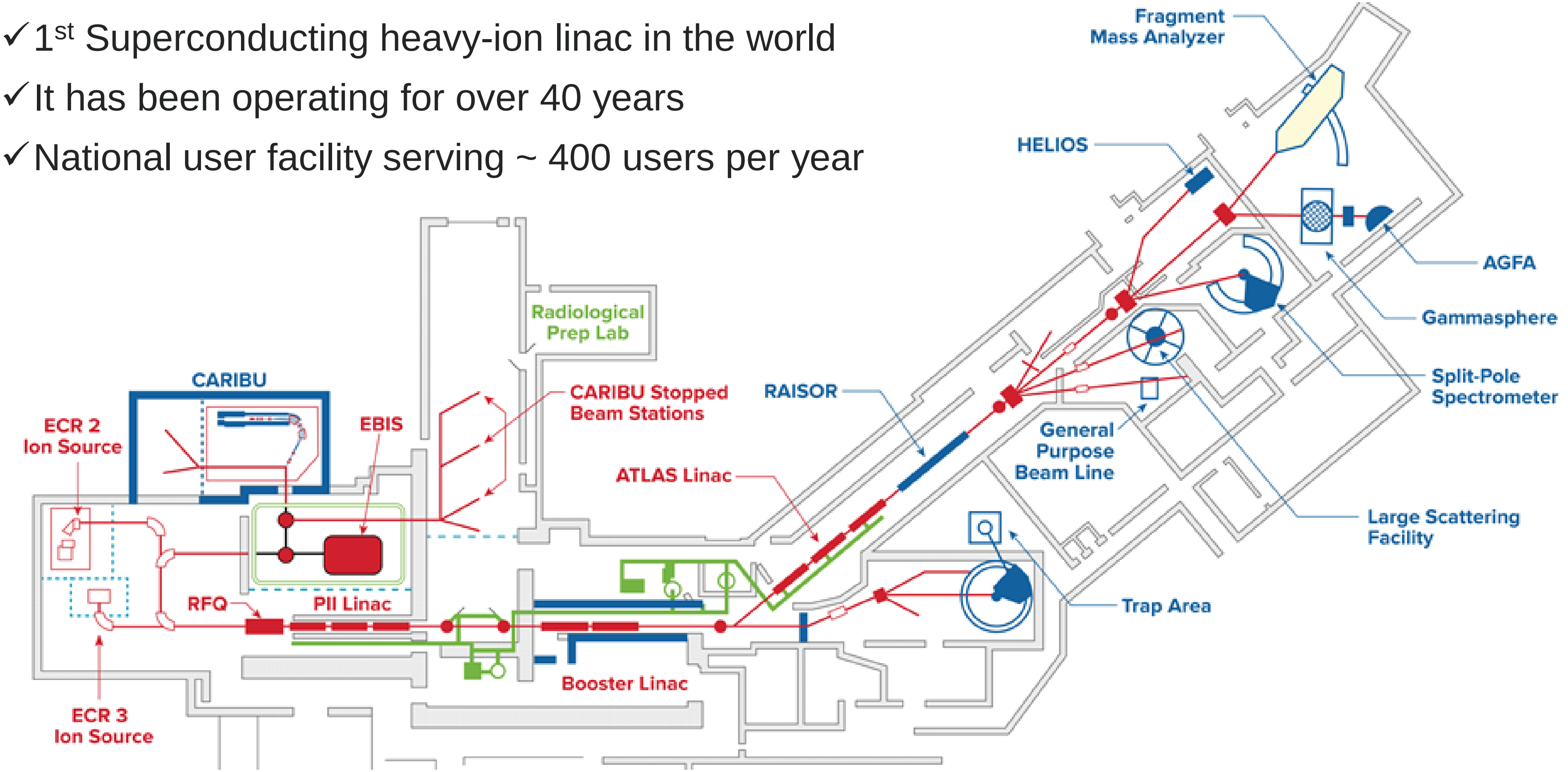
December 5th, 2023
Washington, DC

OUTLINE

- ❑Brief Overview of the ATLAS AI-ML Project and the Team
- ❑Project Status and Summary of Progress
- ❑Progress & Highlights at ATLAS
- ❑Progress Highlights at FRIB and AWA
- ❑Future Plans – Newly Approved Project

ATLAS: ARGONNE TANDEM LINEAR ACCELERATOR SYSTEM

- ✓1st Superconducting heavy-ion linac in the world
- ✓It has been operating for over 40 years
- ✓National user facility serving ~ 400 users per year



BRIEF OVERVIEW OF THE PROJECT

Use of artificial intelligence to optimize accelerator operations and improve machine performance

- At ATLAS, we switch ion beam species every 3-4 days … → Using AI could streamline beam tuning & help improve machine performance
- The main project goals are:
  - **Data collection, organization and classification, towards a fully automated and electronic data collection for both machine and beam data**… established
  - **Online tuning model to optimize operations and shorten beam tuning time in order to make more beam time available for the experimental program** … completed for several sections of the linac
  - **Virtual model to enhance understanding of machine behavior to improve performance and optimize particular/new operating modes** ... progress

THE TEAM / COLLABORATION

- ANL / PHY: B. Blomberg, D. Stanton, J. Martinez and K. Bunnell
  - J. Martinez, postdoc focused on ATLAS (finished recently, new hire …)

- MSU / FRIB: Y. Hao and A. Tran (PhD student started in May'21)
  - ATLAS and FRIB have a lot in common, any development for ATLAS will be useful for FRIB and vice versa (marry carry on to next project)

- ANL / AWA: J. Power, P. Piot and I. Sugrue (PhD student started in Jan'21)
  - AWA can serve as test bed for AI tools development and testing. Being a test facility, more beam time is available for testing tools useful for ATLAS (finished)

- ANL / DSL & ALCF: A. Ramanathan and V. Vishwanath
  - Consult & advise on AI/ML modeling, HP computing and data storage at ALCF

BUDGET SUMMARY & EXPENDITURE

	FY-21 (20)	FY-22 (21)	FY-23 (22)	Total/Actual
Funds allocated + c. o.	$280k	$440k	$425k	$840k
Actual costs to date	$120k	$295k	$335k	$750k
Uncosted commitments	$63k	$86k	$55k	$55k
Uncommitted funds	$97k	$59k	$35k	$35k

- ✓ Project officially started in January 2021 (FRIB started May 2021)
- ✓ Budget table above is as of the end of September 2023

PROGRESS & HIGHLIGHTS - ATLAS

SUMMARY OF PROGRESS & HIGHLIGHTS

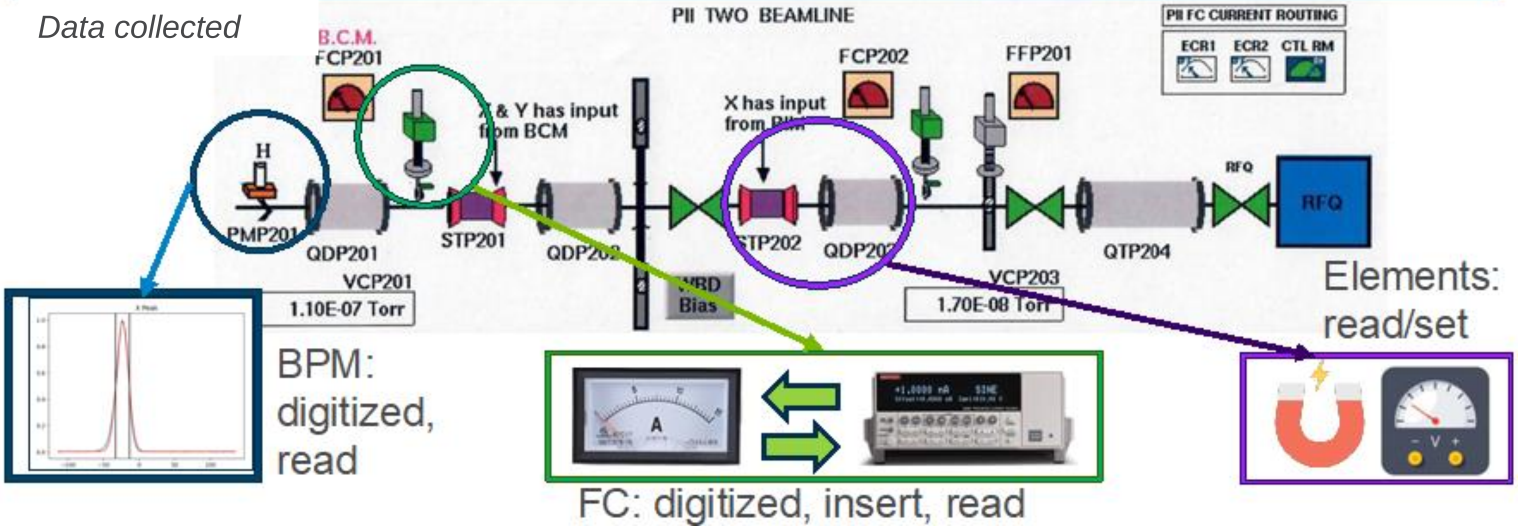
- Automated data collection and two-way communication established
- Bayesian Optimization (BO) successfully used for online beam tuning
- Multi-Objective BO (MOBO) to optimize transmission and beam size
- AI-ML supporting the commissioning of a new beamline (AMIS)
- Transfer learning from one ion beam to another (BO)
- Transfer learning from simulation to online model (BO with DKL)
- Reinforcement Learning for online beam tuning – First expr. Success
- Some progress on the virtual machine model / physics model

AUTOMATED DATA COLLECTION ESTABLISHED

- ✓ Beam currents and beam profiles digitized
- ✓ A python interface developed to collect the data automatically

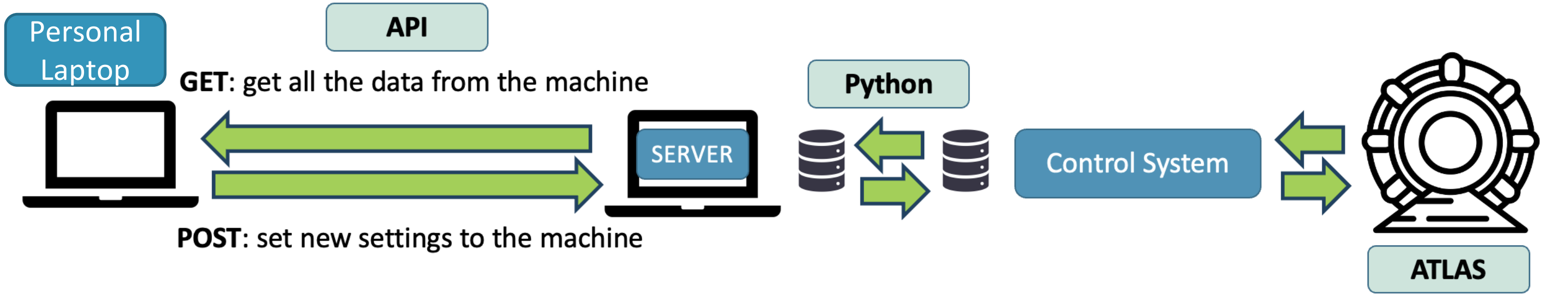
Schematic of data collection interface

Data collected



Now working on reducing acquisition time …

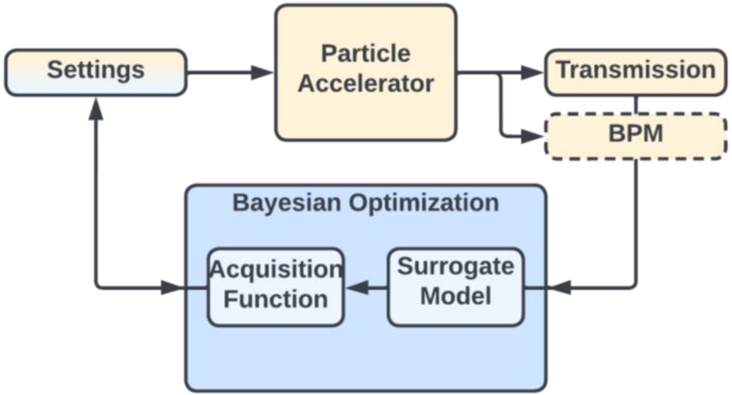
ONLINE – INTERFACE WITH CONTROL SYSTEM



OFFLINE – INTERFACE WITH BEAM SIMULATIONS

- ✓ Python wrapper for TRACK (Simulation Code)
- ✓ Generation of simulation data
- ✓ Different conditions and inputs
- ✓ Integration with AI/ML modeling

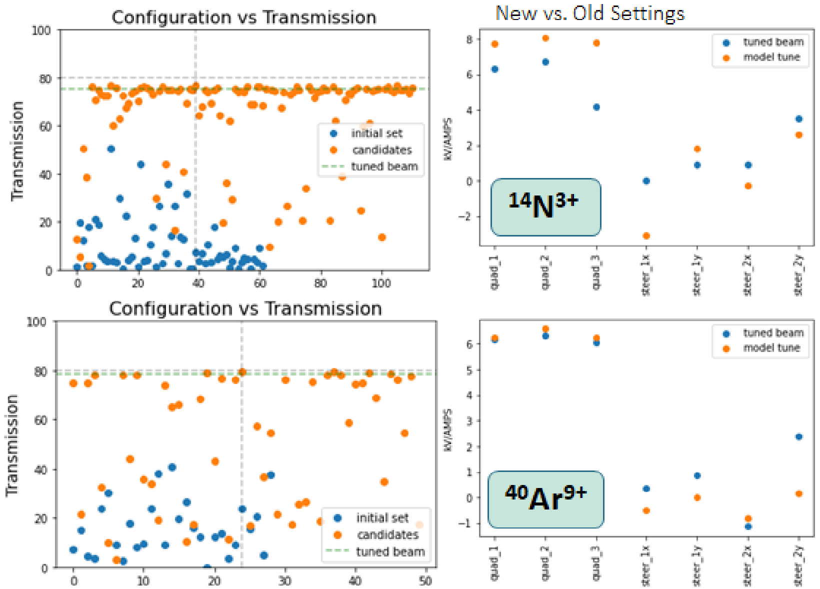
BAYESIAN OPTIMIZATION USED FOR BEAM TUNING



- **<u>Surrogate Model</u>: A probabilistic model approximating the objective function** [Gaussian Process with RBF Kernel and Gaussian likelihood]

- **<u>Acquisition Function</u> tells the model where to query the system next for more likely improvement** [EI]

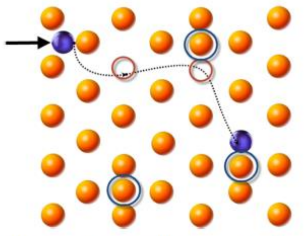
- **Bayesian Optimization with Gaussian Processes gives a reliable estimate of uncertainty and guides the model**

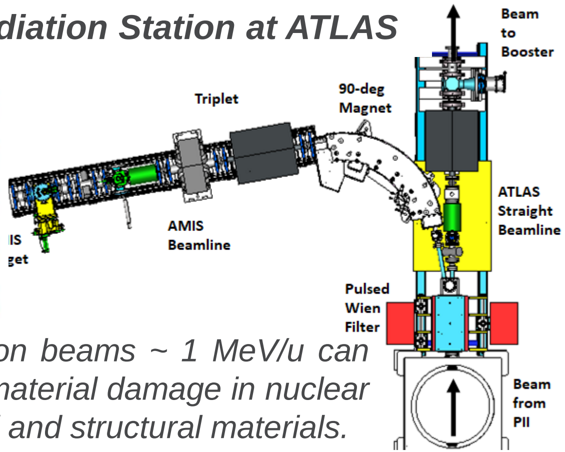


- **7 varied parameters (3 quads + 2 steerers)**
- **Optimization of beam transmission**
- **Case of $^{14}N^{3+}$: 29 historical + 33 random tunes**
- **Case of $^{40}Ar^{9+}$: 29 historical tunes**

AI/ML SUPPORTING AMIS LINE COMMISSIONING

New Material Irradiation Station at ATLAS

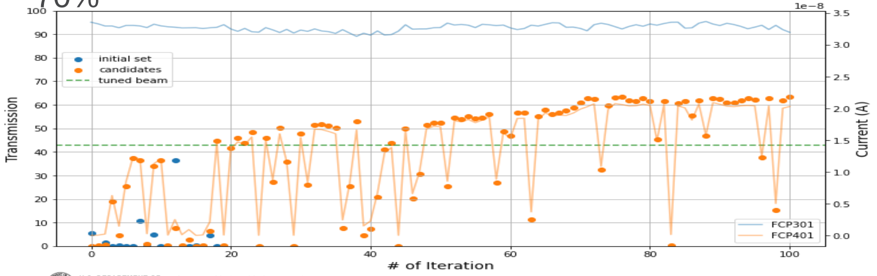




Low-energy heavy-ion beams ~ 1 MeV/u can effectively emulate material damage in nuclear reactors, in both fuel and structural materials.

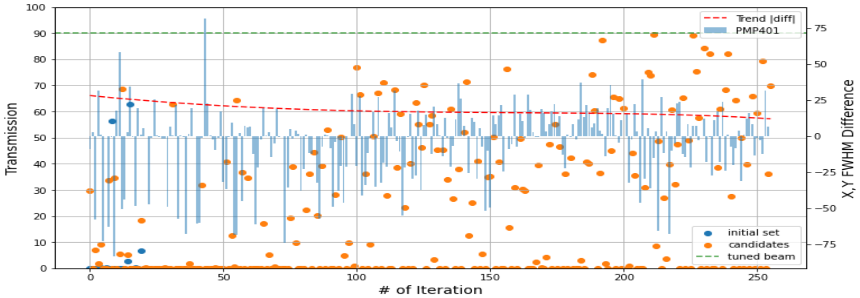
Improving Beam Transmission

Problem: Maximize beam transmission by varying a triplet, two dipoles and two steerers [BO]; **Results**: 40 → 70%

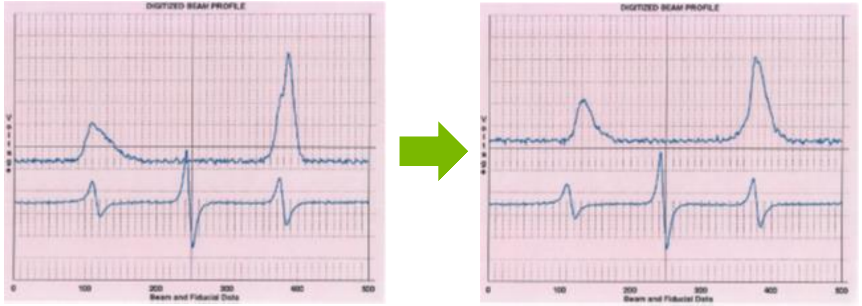


Improving Beam Profiles

Problem: Produce symmetric beam profiles by varying a triplet and a steerer [BO]



Training online, slow convergence but steady progress. Competition between nice profiles and beam transmission!



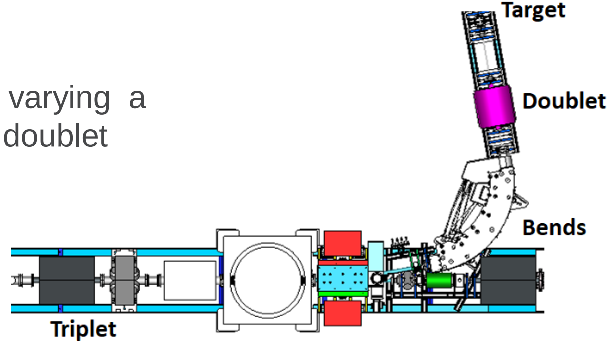
Very encouraging first results!

MULTI-OBJECTIVE BAYESIAN OPTIMIZATION

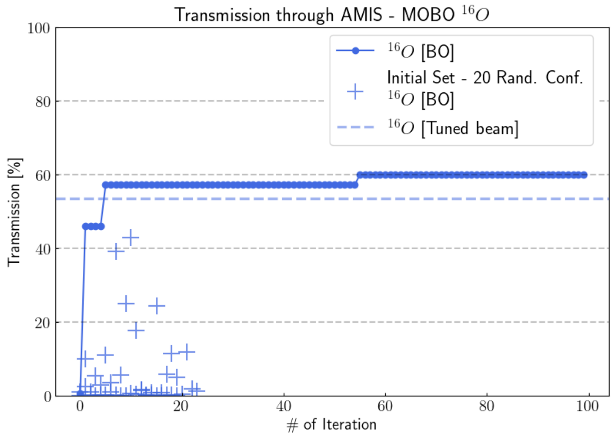
Multi-Objective Problem: Optimize transmission and beam profiles on target - Not easy for an operator!

Improving Beam Transmission & Improving Beam Profiles

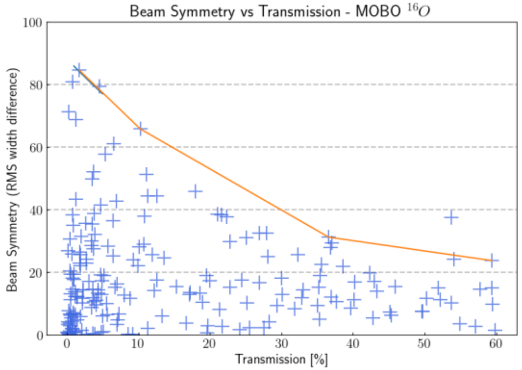
AMIS line: varying a triplet and a doublet



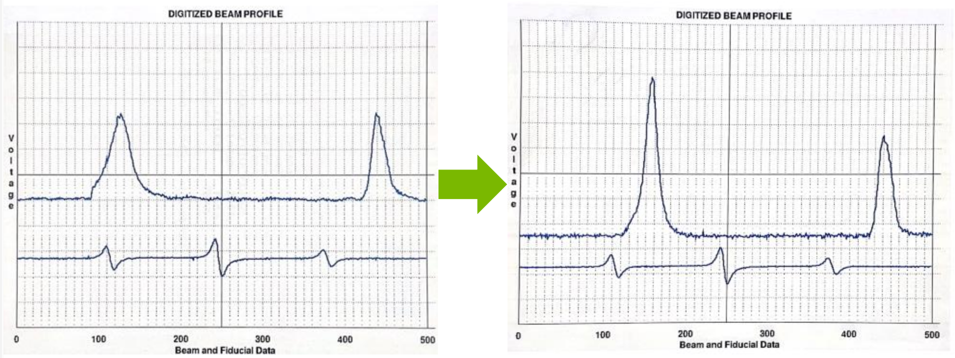
MOBO Results: 53→ 60% Beam transmis.



MOBO Results: Pareto Front



MOBO Results: More symmetric beam profiles

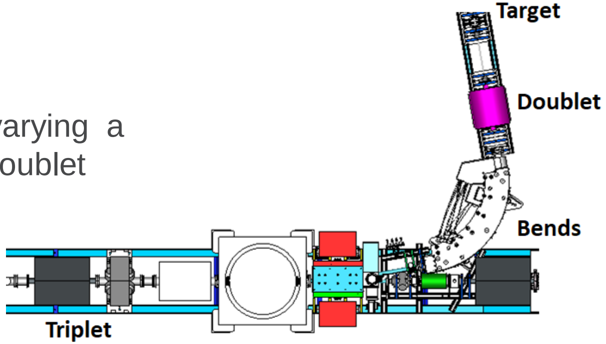


TRANSFER LEARNING FROM ^{16}O TO ^{22}NE - BO

Goal: Train a model using one beam then transfer it to tune another beam → Faster switching and tuning

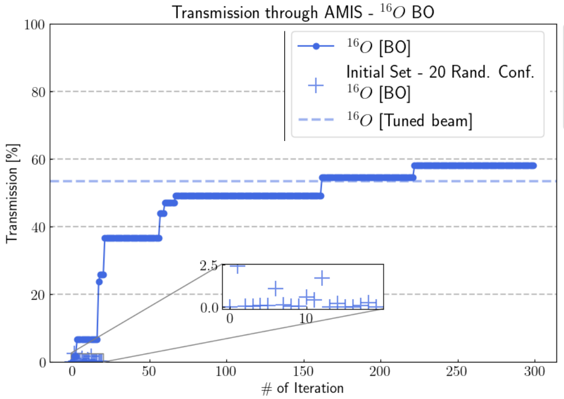
Training model on ^{16}O → Applying same model to ^{22}Ne

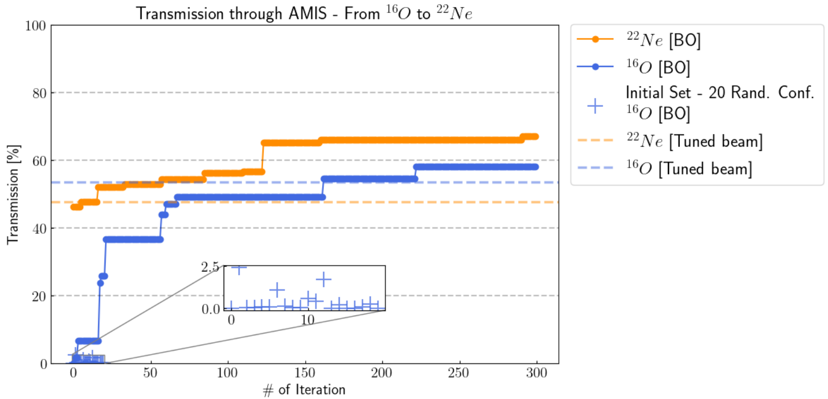
AMIS line: varying a triplet and a doublet



BO Training:
Over 300 iterations
53 → ~ 60%
Beam transmis.

Model saved & exported





16O Model loaded for 22Ne: Initial transmission improved in 7 iterations: 48 → 55 %

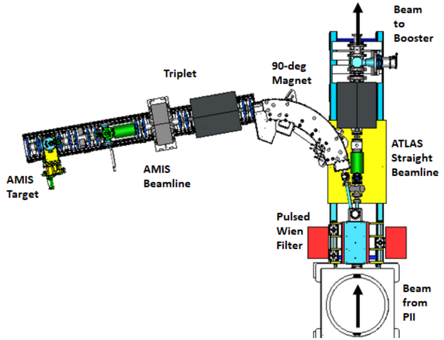
With more training for 22Ne: 48 → 67%

Scaling was applied from 16O to 22Ne, re-tuning is often needed because of different initial beam distributions

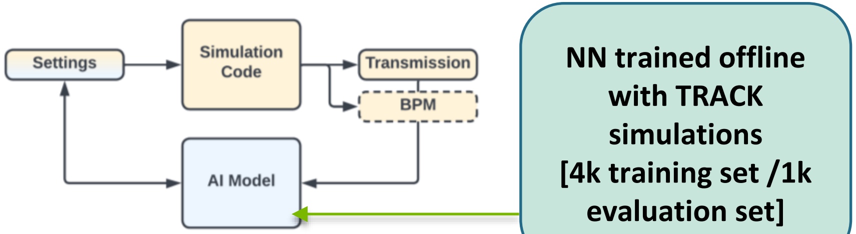
TRANSFER LEARNING FROM SIMULATION TO ONLINE

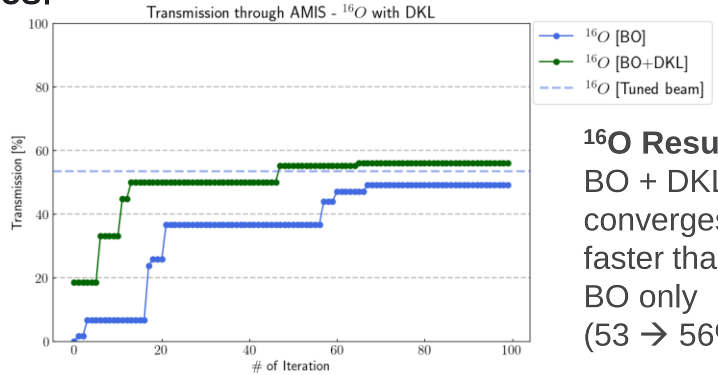
Goal: Train a model using simulations then use it for online tuning → Less training & fast convergence online

Method: Deep kernel learning (DKL) to **combine the representational power of neural networks with the reliable uncertainty estimates of Gaussian processes**.

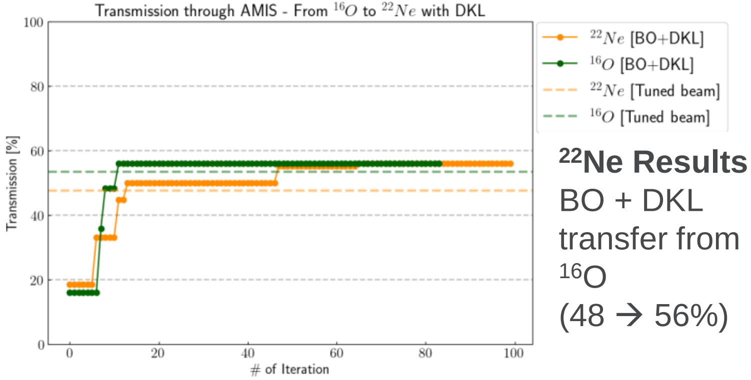


AMIS Line: Maximize beam transmission by varying a triplet [BO+DKL]



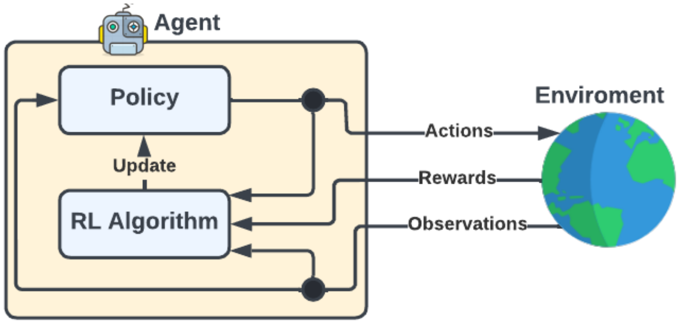


^{16}O Results: BO + DKL converges faster than BO only (53 → 56%)

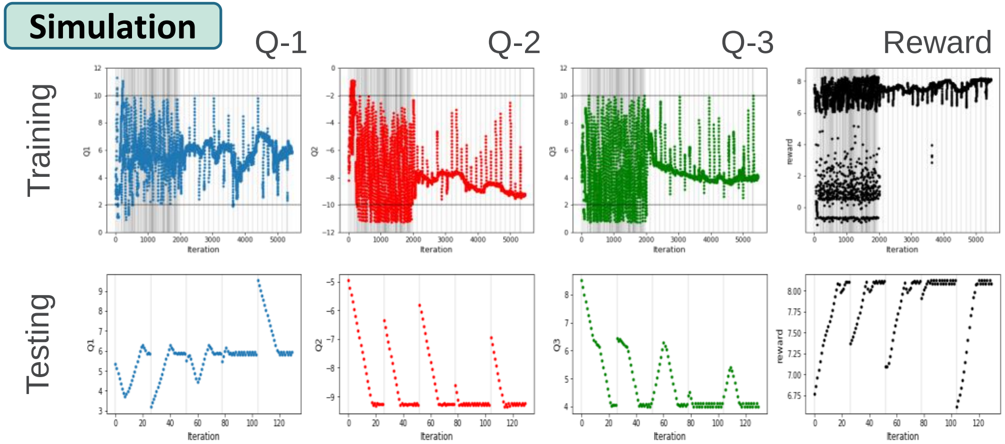


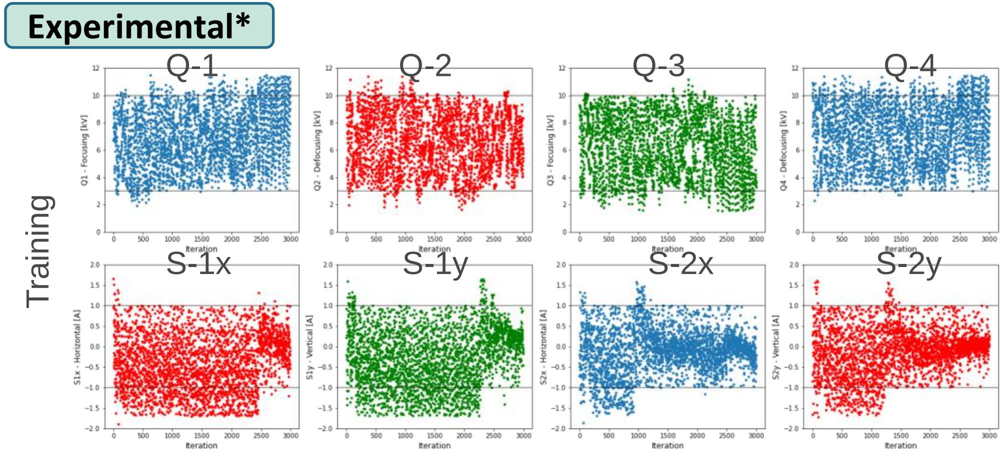
^{22}Ne Results: BO + DKL transfer from ^{16}O (48 → 56%)

REINFORCEMENT LEARNING FOR FINE TUNING



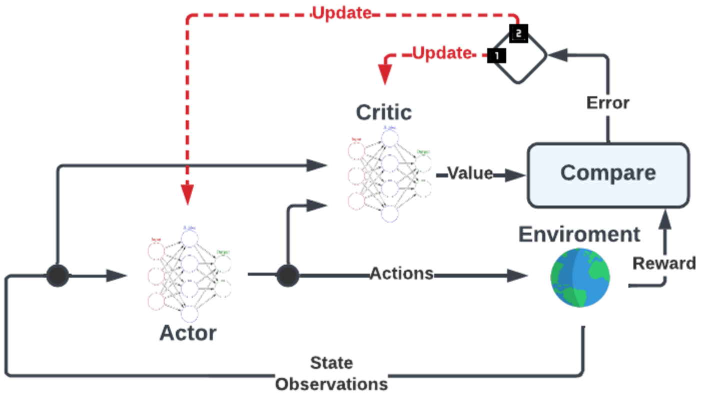
- ✓ **Method**: Deep Deterministic Policy Gradient (DDPG); Actor-Critic Approach
- ✓ **Simulation Case**: Focusing beam on target using a triplet (3 Quadrupoles)
- ✓ **Experimental Case**: Maximizing beam transmission using 4 quads and 2 steerers
- ✓ Electrostatic Quadrupoles :
  - 2 kV to 10 kV
  - Max action +/- 0.25 kV
- ✓ Steering Magnets:
  - -1 A to 1 A
  - Max action +/- 0.25 A





REINFORCEMENT LEARNING: FIRST EXP. SUCCESS

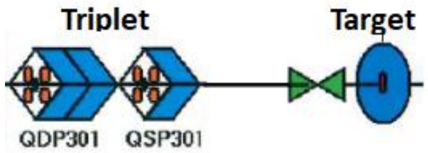
Goal: Demonstrate Reinforcement Learning experimentally and compare with Bayesian Optimization



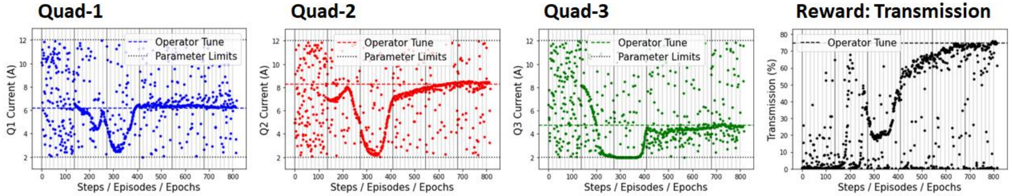
- ✓ **Essence:** Learning from experience based on interaction with the environment
- ✓ **Action:** Varies the parameters/variables of the problem
- ✓ **Reward:** Measures the goal function to maximize/optimize
- ✓ **Policy:** How the process evolves/learns
- ✓ **Algorithm used:** Deep Deterministic Policy Gradient (DDPG); Actor-Critic Approach

Experimental Case: Maximize beam transmission to target

- Varying 3 magnetic quads
- Current limits: 2 – 12 Amps
- Max. Action: Full range

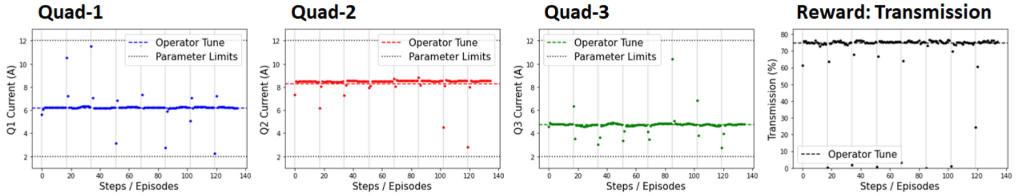


Training - Online



➢ Training done in 816 total steps/evaluations (48 episodes)

Testing - Online

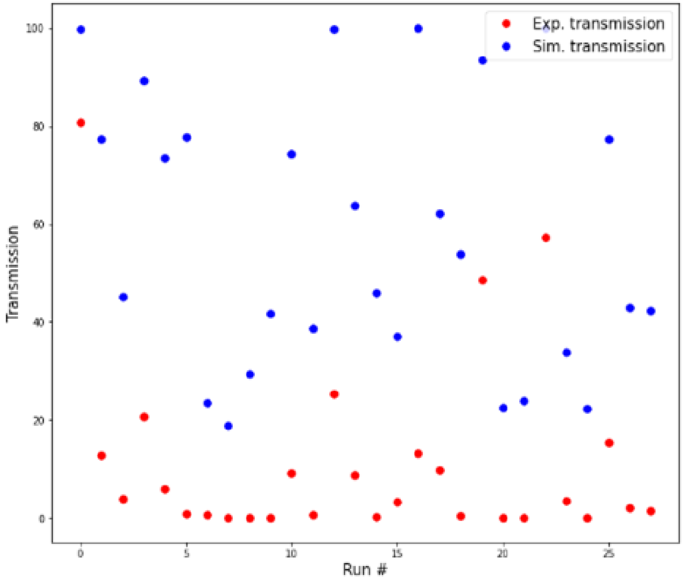


➢ Testing done for 8 episodes (16 steps/episode)

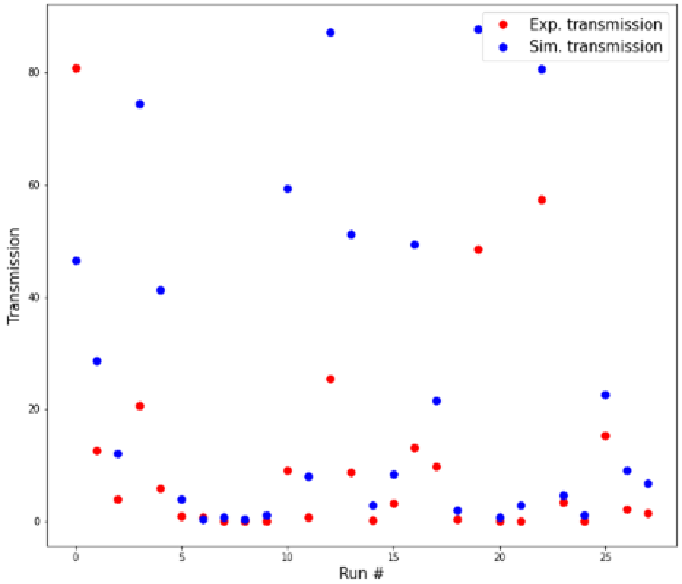
➢ Model converges in 2-3 steps, starting from random conf.

PROGRESS ON THE VIRTUAL / PHYSICS MODEL

No Steering (<Diff> ~ 46%)



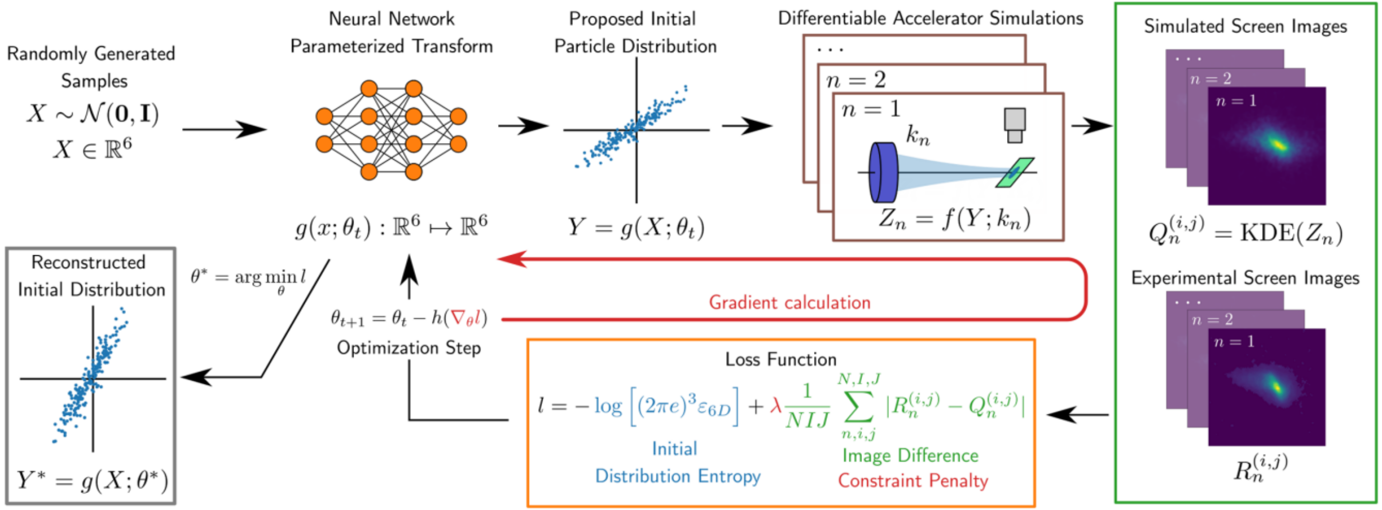
With Steering (<Diff> ~ 16%)



- ✓ In order to develop a realistic virtual machine mode, we need first to improve the predictability of the physics model based on TRACK simulations.
- ✓ Significant improvement was realized by adding the steering effects, adding information on misalignments and initial beam distribution should close the gap further.
- ✓ Once the agreement is ~ 1%, a surrogate model will be developed based on the simulations.

PROGRESS HIGHLIGHTS - FRIB

4D BEAM TOMOGRAPHY - ML BASED



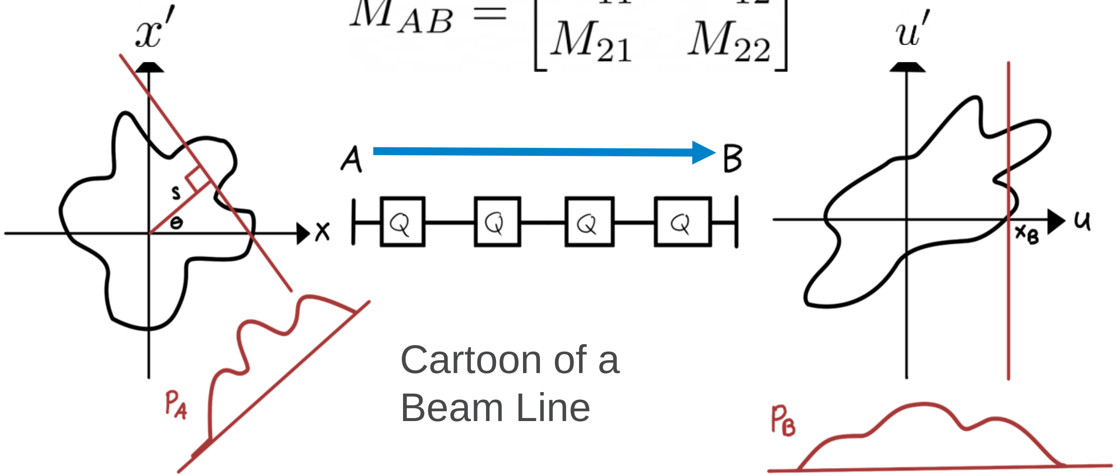
Work done at Argonne Wakefield Accelerator (AWA)

R. Roussel et al., "Phase Space Reconstruction from Accelerator Beam Measurements Using Neural Networks and Differentiable Simulations" in Phys. Rev. Lett., 130, p. 145001 (2023)

1. We start with any distribution of particles. Here, a simple 4D gaussian is used.
2. The method is to move the particles iteratively until they form our initial distribution in 4D.
3. Particles are propagated through the beamline and compared to the measured 2D images.

2D BEAM TOMOGRAPHY: RECONSTRUCTING 2D DISTRIBUTION FROM 1D PROFILES

$$M_{AB} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$



Cartoon of a Beam Line

$$\tan\,\theta = \frac{M_{12}}{M_{11}}$$

$$a = \sqrt{M_{11}^2 + M_{12}^2}$$

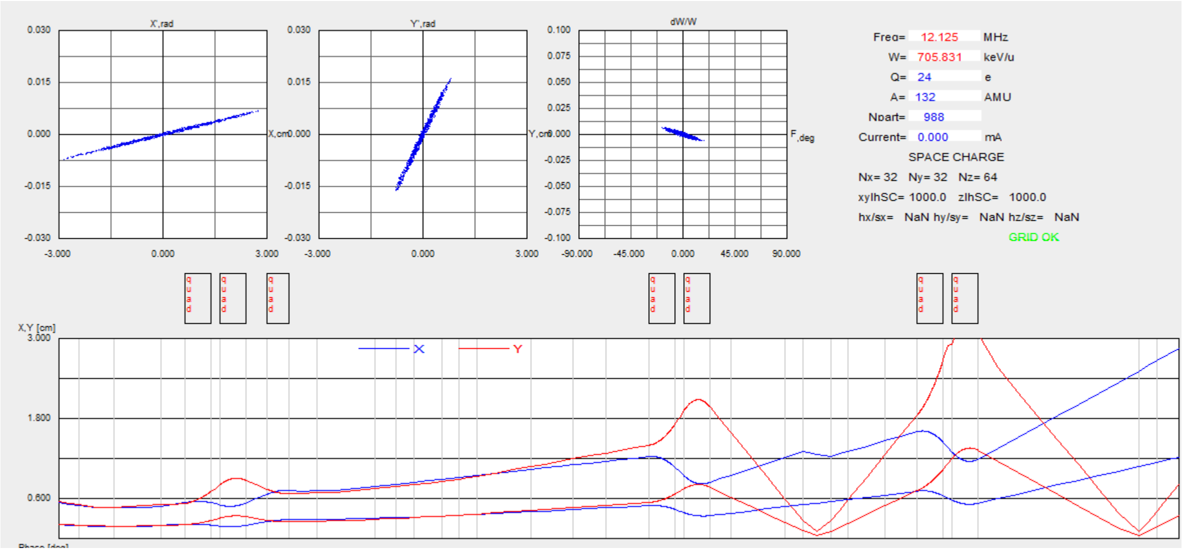
$$s = \frac{x_B}{a}$$

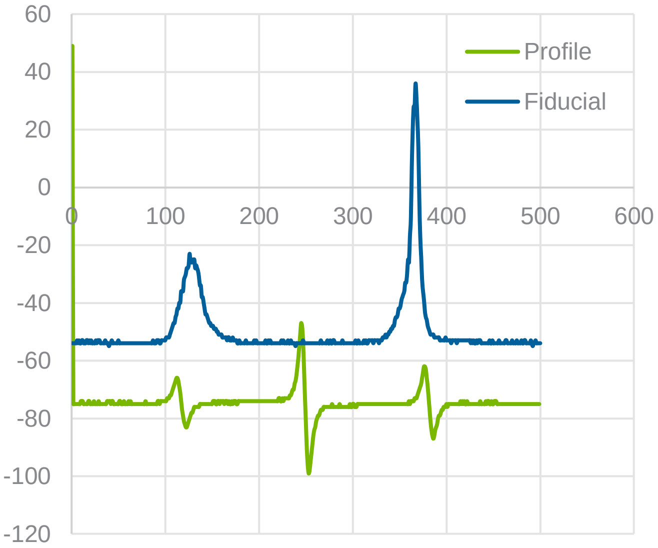
$$p_A = a p_B.$$

- If beams are decoupled, dynamics reduces to 2D matrix transformations
- There exist a simple method to transform projections at B to projections at A
- We get different projections depending on the beam rotation angle
- Requires at least 180-deg total rotation sampling the profile at diff. angles

EXPERIMENT DONE AT ATLAS – PII TO BOOSTER LINE

PII to Booster line: ~ 13 m long, 7 quads, 2 bunchers





- A good amount of data was taken; beam profiles and beam transmission
- The challenge: We didn't get the full 180-deg rotation without beam loss
- Analysis is in progress …

PROGRESS HIGHLIGHTS - AWA

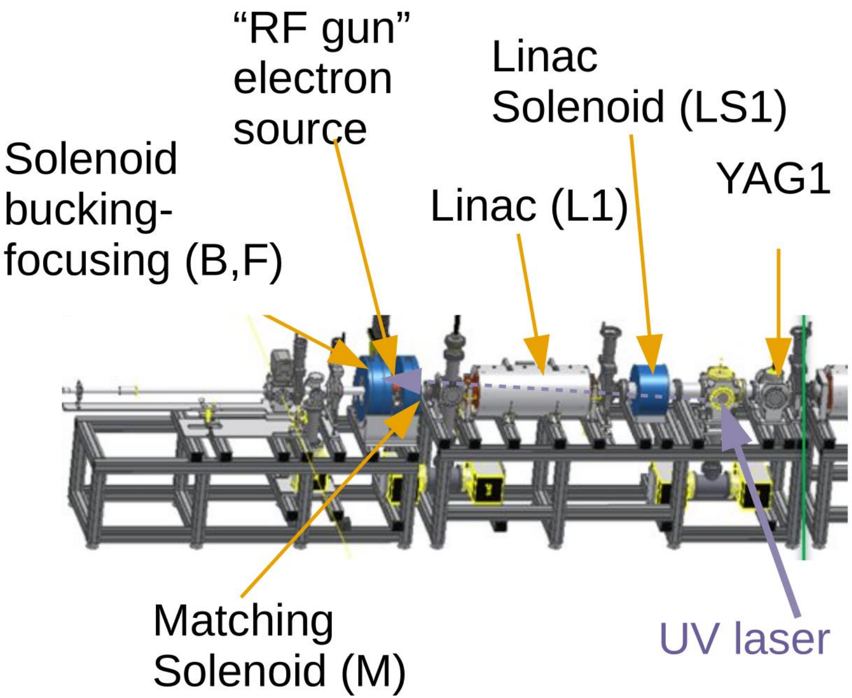
PROGRESS SUMMARY - AWA WORK

Idea: AWA can be used as a testbed for ML-based machine tuning and virtual diagnostics development

Progress made so far

- ❑ Improved surrogate model for beam image prediction: Improved simulation data and PCA decomposition
- ❑ Least squares minimalization applied to retrieve the actual beamline elements settings for a given beam image
- ❑ Method tested first on simulation data with known settings and added noise to image – controlled or supervised fitting
- ❑ When tested on experimental data, some parameters are predicted very well but not the rest – work in progress

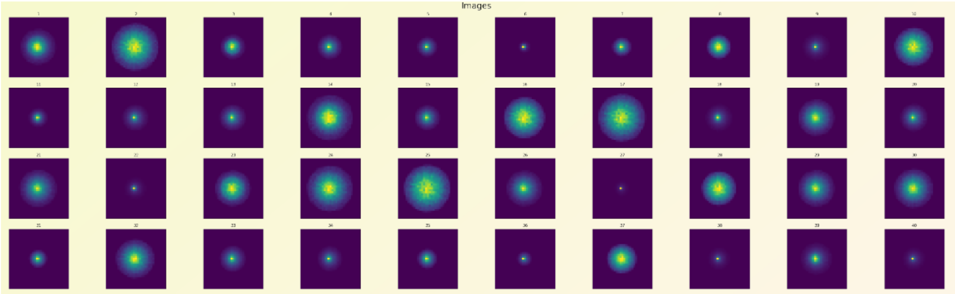
Lattice & beamline parameters

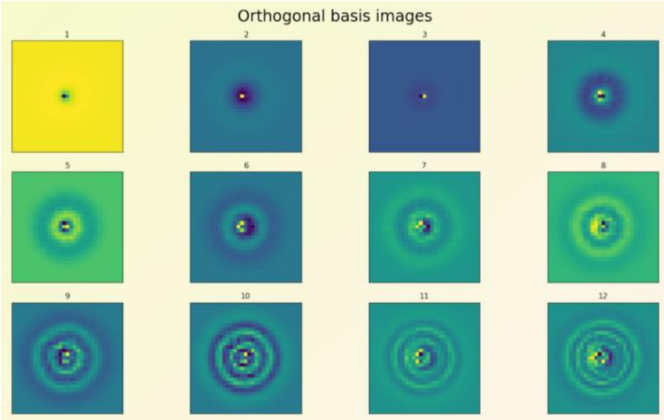


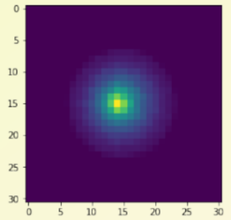
IMPROVED SURROGATE MODEL FOR BEAM IMAGE PREDICTION

Goal: Associate a given image to given input lattice parameters

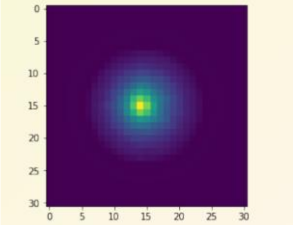
Improved simulation data & PCA decomposition





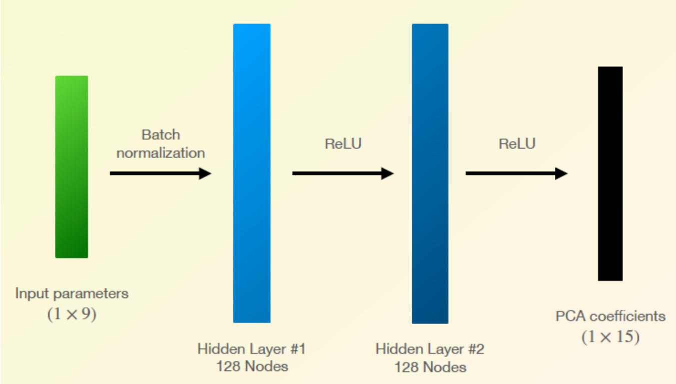


- Before PCA



- After PCA (15 Components)

Surrogate model: NN architecture



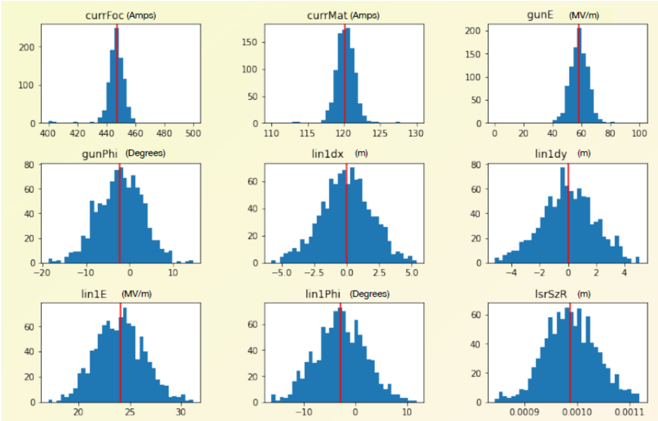
- ✓ 9 Input lattice parameters
- ✓ Images reduced to 15 PCA components
- ✓ Two hidden layers of 128 nodes each
- ✓ ~ 500 epochs, default batch size (32), MSE loss function

LEAST SQUARES MINIMIZATION: TEST ON SIMULATION DATA

Problem: What are the real lattice parameters for given beam image?

- **Method**: Minimize $\|f(x) - y\|_2^2$, where f(x) is the surrogate model output with input parameters x, and y is the PCA coefficients of the image.
- The initial input is the vector of experimental parameters x_0, and the result of the least squares optimization is an approximate solution of the true parameters.
- Test the optimization by pretending we have experimental data (noisy input) and that we know the true parameters (true input).
- Let v be a vector of random noise in R^9, and let x_t be the true input parameters. Minimize $\|f(x_t) - f(x)\|^2$ when $x0 = x_t(1+ v)$.
- We run this 1000 times, each with a different noise vector. The results are shown next …

Results of Least squares Minimization



Blue is the least squares estimate.

Red is the true parameter value

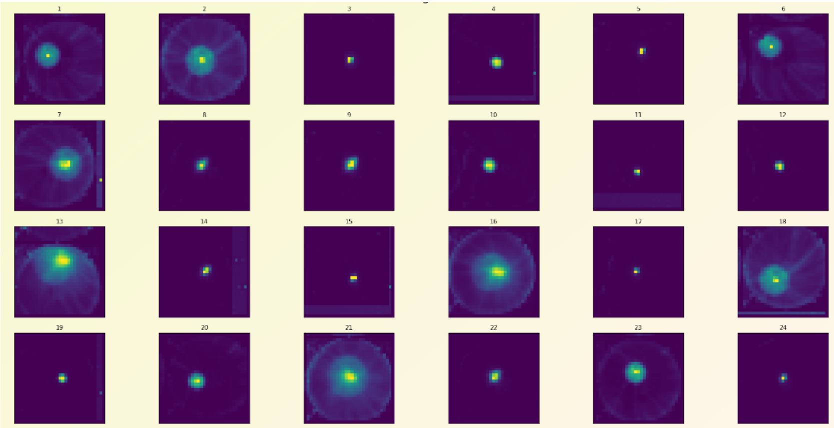
Results seems to be much closer to reality for the first 3 parameters than for the rest of them!

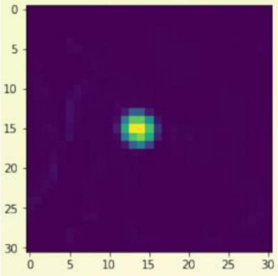
There seem to be large uncertainties on the misalignment parameters of the first linac cavity; lin1dx and 1lin1dy

LEAST SQUARE MINIMIZATION: TESTS ON EXPERIMENTAL DATA

Problem: What are the real lattice parameters for given beam image?

Experimental beam images & Related PCA





Original Image

100 experimental images of size 31×31, flattened to a single array of size 100×961

Simulated Data = X

$\hat{X} = X - \mu$ (mean of each row of X)

$\hat{X} = U\Sigma V^{\top}$

Array of coefficients is $C = \hat{X}V_k$ (first k rows of V)

Experimental Data = Y

PCA coefficients using existing basis images is $C_{\text{exp}} = YV_k$

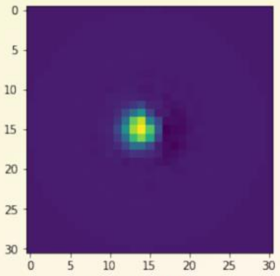
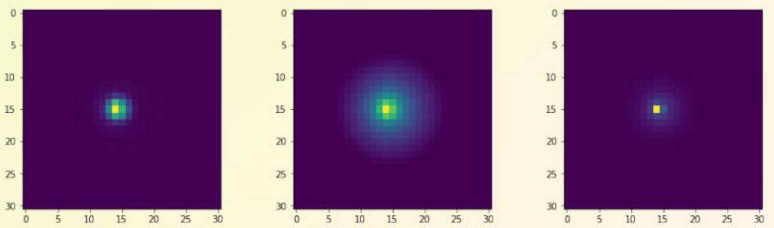


Image from PCA reconstruction

$V_k^{\top} C_{\text{exp}}^{\mathrm{l}} + \mu$

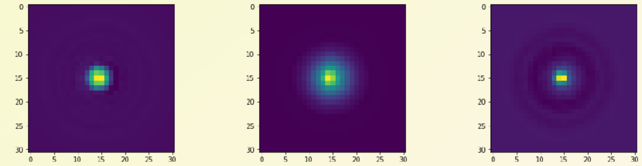
Experimental Test #1



- Experimental Image #7 (after PCA)
- Surrogate Model Prediction
- Surrogate Model after Least Squares

Experimental Test #2

Experimental Image #47 (after PCA) | Surrogate Model Prediction | Surrogate Model after Least Squares



	Experimental Params	Least Squares Result	Difference
currFoc	4.981189243296668110e+02	4.981189243296668110e+02	0.000000000000000000e+00
currMat	1.416873704809135006e+02	1.416873704809135006e+02	0.000000000000000000e+00
gunPhi	6.000000000000000000e+01	6.000000000000000000e+01	0.000000000000000000e+00
gunE	0.000000000000000000e+00	0.000000000000000000e+00	0.000000000000000000e+00
Lin1dx	0.000000000000000000e+00	4.705160544383231809e+02	4.705160544383231809e+02
Lin1dy	0.000000000000000000e+00	-2.253368202526503410e+02	-2.253368202526503410e+02
Lin1Phi	2.200000000000000000e+01	2.200000000000000000e+01	0.000000000000000000e+00
Lin1E	0.000000000000000000e+00	-1.879956681709598598e+01	-1.879956681709598598e+01
lsrSzR	1.000000000000000021e-03	1.000000000000000021e-03	0.000000000000000000e+00

FUTURE PLANS – NEW PROJECT

NEW AI-ML PROJECT: BRIEF OVERVIEW

Same title: Use of artificial intelligence to optimize accelerator operations and improve machine performance

❑ The main objectives of the new project are:

- Deploy the autonomous beam tuning tools developed during our previous project, evaluate their impact on both automating the tuning process and saving on tuning time.

- Develop tools for new operating modes such as multi-user operation of the ATLAS linac and high-intensity beams, as well as developing virtual diagnostics to supplement existing ones.

THREE COMPONENTS OF THE NEW PROJECT

- ❑ Stable beams in ATLAS – Brahim Mustapha
- ❑ Inflight radioactive beams from RAISOR – Calem Hoffman
- ❑ Radioactive beams from CARIBU – Daniel Santiago

- ❑ Close collaboration, exchange of ideas and codes and effort if needed
- ❑ Two new postdocs will join the ATLAS and CARIBU projects soon

THANK YOU

MANY THANKS TO

- ATLAS Controls Team:

  D. Stanton, K. Bunnell and C. Dickerson

- ATLAS Operations Team:

  B. Blomberg, E. Letcher, G. Dunn and M. Hendriks

- ATLAS Liaison and beam time schedular:

  D. Santiago

- …

RECENT TALKS AND PUBLICATIONS

- “Reinforcement Learning and Bayesian Optimization for Ion Linac Operations”, J. Martinez, B. Mustapha et al, Invited talk at the Heavy Ion Accelerator Technology (HIAT) Conference, Darmstadt, Germany, June 27 - July 1 2022

- “Machine Learning to support the ATLAS Linac Operations at Argonne”, B. Mustapha et al, Poster & Paper at NAPAC’22, August 7-12th, 2022, Albuquerque, New Mexico & ICFA Workshop on Machine Learning for Accelerators, Nov. 1-4, Chicago, Illinois

- “Machine Learning Tools to support the ATLAS Ion Linac Operations at Argonne”, J. Martinez, B. Mustapha et al, Talk at the ICFA Workshop on Machine Learning for Accelerators, Nov. 1-4, Chicago, Illinois

- “Model-based Calibration of Control Parameters at the Argonne Wakefield Accelerator”, I. Sugrue et al, NAPAC’22, August 7-12th, 2022, Albuquerque, New Mexico

- “Predicting beam transmission using 2-dimensional phase space projections of hadron Accelerators”, A. Tran et al, Front. Phys. 10:955555. doi: 10.3389/fphy.2022.955555

MORE PUBLICATIONS …

- ❑…