BERKELEY LAB

Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY

Office of Science

Machine Learning Optimization Upstream and Downstream of the Accelerator: The Cases of VENUS and GRETA

Funded under FY2023 Lab FOA

Heather Crawford, Damon Todd,

Chris Campbell, Mario Cromaz and Marco Salathe

NP AI/ML PI Exchange Meeting

November 20, 2025

Optimizing the Front-End and Experimental End-Station

The effective operation of any accelerator facility is not limited to the accelerator itself – fully optimized operation is realized by optimizing all parts of an experiment, and reducing down-time along the entire facility chain.

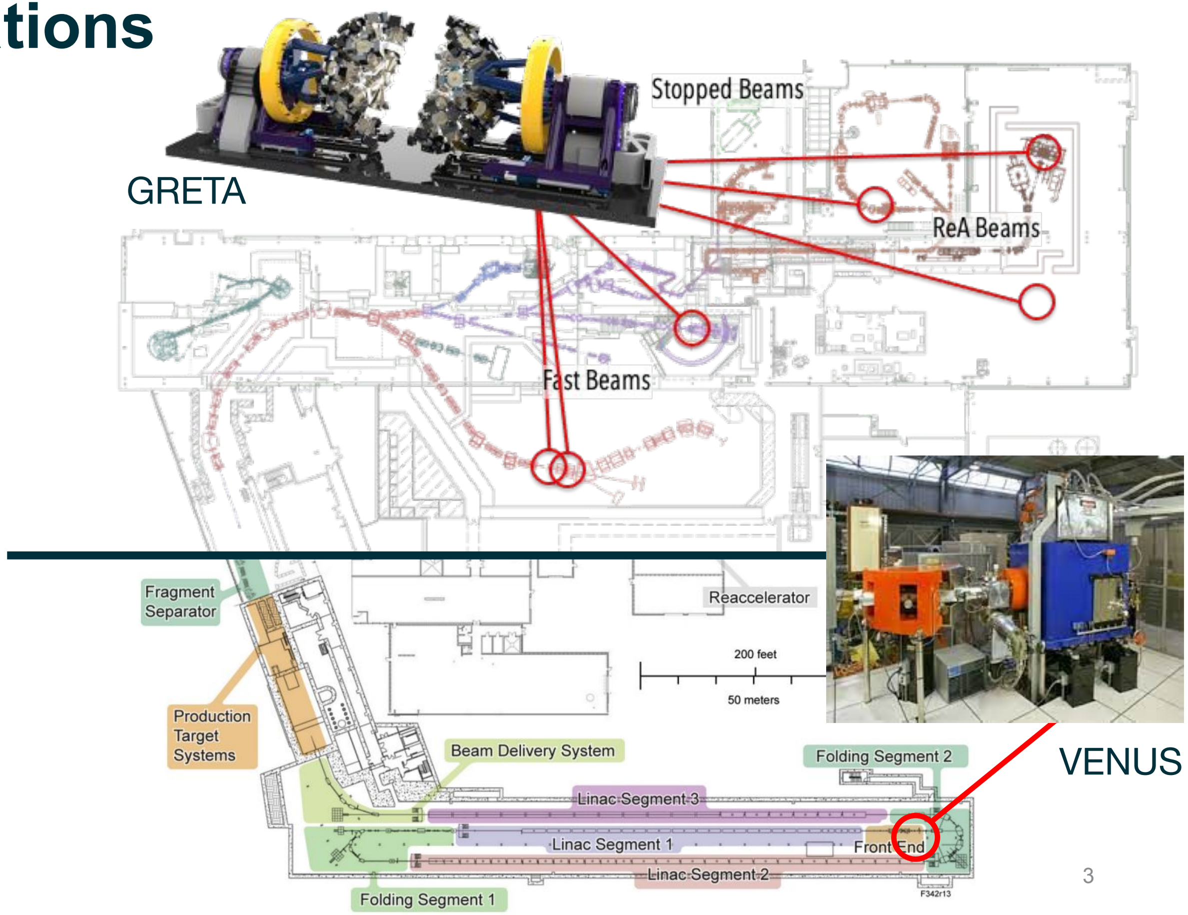
We focus on the front end and the end-point of a facility - the VENUS ion source and GRETA experiment.

Applying Machine Learning to LBNL Systems to Impact 88” and FRIB Operations

FRIB

88-Inch Cyclotron

BERKELEY LAB



Original FY21 ML/AI VENUS+GRETA Project

- The original effort was funded with a FY21 award, $1M split evenly across two years
- First effort focused on:
  - Readying VENUS for application of ML techniques – no data was recorded regularly, combination of EPICS and LabView interfaces needed to be made/re-written
  - Accumulating data from VENUS from human-driven tuning and source baking to provide a starting data set for ML applications
  - Automation of the frequent “baking” operation to reduce human time and improve efficiency
  - Initial demonstration of Bayesian optimized tuning within limited parameter space
  - Automating the optimization of the GRETA electronics signal chains for resolution, and providing complete calibration of a crystal including interface to GRETA EPICS systems and hardware
- Ended grant period with $360k of carryover (postdoc joined 10 months into the award period)

FY23 Award ML/AI VENUS+GRETA Project

- Award was a total of $1.098M split across two years
- Effort focused on:
  - Bayesian optimized tuning for VENUS across full parameter space
  - Computer-enhanced stability in running using data filtering techniques to detect and avoid source instabilities
  - Modest VENUS hardware upgrades to provide additional data (e.g. rapid charge state distributions)
  - Enhancement of the GRETA basis production pipeline for parameter sensitivity studies and ML-assisted optimizations

Budget (FY23 Award)

	Year 1	Year 2	Total ($k)
(a) Funds Allocated	228	870	1,098
(b) Actual Costs to Date	228	789	1,017

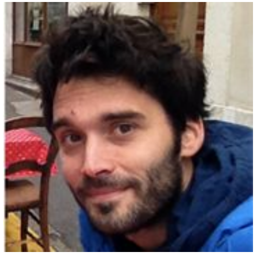
Research Team - Staff and Postdocs

Chris Campbell
Scientific Engineer
LE Program / GRETA

Mario Cromaz
Applied Physicist Staff Scientist
LE Program / GRETA

Marco Salathe
Applied Physicist Research Scientist
ANP Program

Jessica Rehak
Senior Scientific Engineering Associate
88 Ops Program



Nico Abgrall
Senior Scientific Engineering Associate
ANP Program

Yue Shi Lai
Applied Physicist Research Scientist
ANP Program

Victor Watson
Senior Scientific Engineering Associate
ML Project

Gabriel Garcia Jimenez
Postdoctoral Associate
ML Project

Research Team - Undergraduate Researchers

Ezra Apple
UCBerkeley
Electrical Engineering/Computer Science
Class of 2025

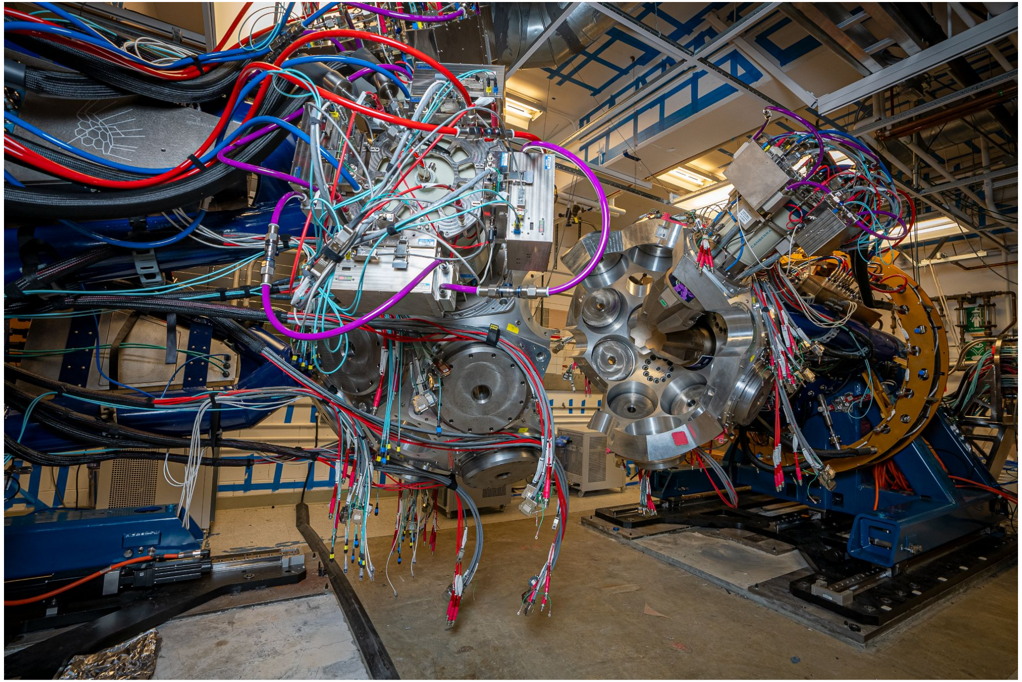
Julia Dreiling
University of Ohio
Data Analytics
Class of 2024
– now pursuing MSc at University of St. Andrews

Alex Kireeff
Carnegie Mellon University,
Electrical Computer Engineering
Class of 2024

Arin Manohar
UCBerkeley
Physics, Computer Science, Mathematics
Class of 2026

GRETA

Gamma-Ray Energy Tracking Array, GRETA



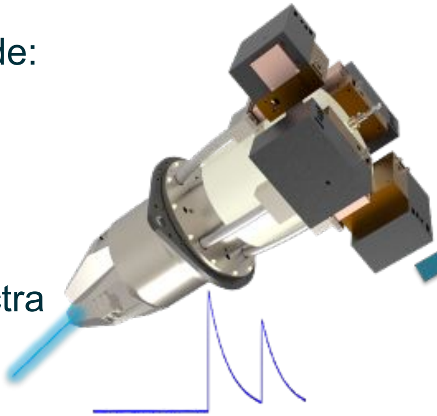
- U.S. implementation of a gamma-ray tracking array
- Complete 4π solid angle coverage of active high-purity germanium (HPGe), consisting of 120 individual detector crystals, each with 37 electrical signals
- Gamma-ray tracking and Compton suppression is enabled by signal decomposition algorithm which localized gamma-ray scatter events to within ~mm^3 volumes

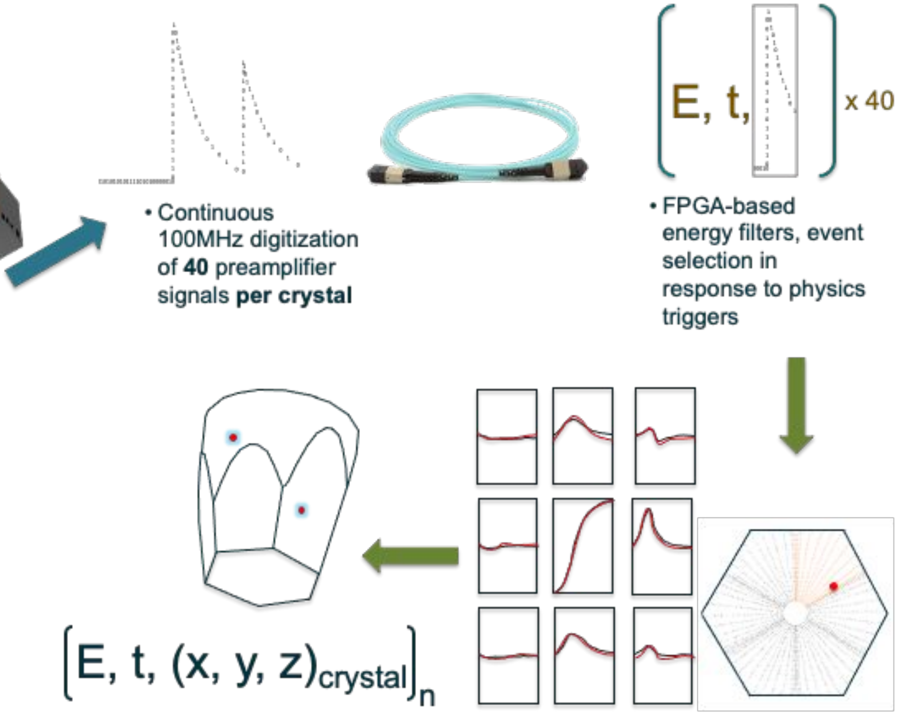
GRETA is currently being installed for the first time at FRIB.

GRETA Optimizations

Simple control parameters include:
- 4-6+ energy filter parameters per channel
- 2+ calibration parameters per channel

~ 30k knobs just for energy spectra





In addition to energy resolution optimization across the array, GRETA performance is tied to position resolution for reconstructing gamma-ray interaction points.

Position resolution depends on the **fidelity of the calculated response** of the HPGe crystals.

GRETA Goals and Status

- Improvement of the signal basis used for the process of signal decomposition, exploring improvement in the position resolution of interactions in GRETA by improving the calculated signals used in the fit through an ML-driven global optimization.

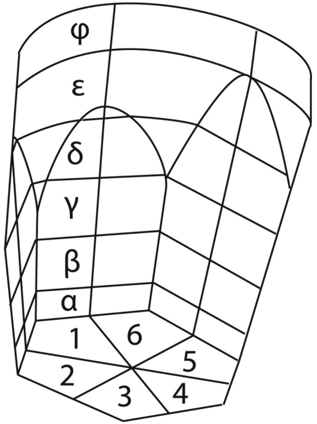
    - Signal basis parameters are now being fully explored, with automated and streamlined pipeline in place
    - Alternative detector simulation has also been implemented, enabling exploration of sensitivity of performance to additional parameters related to detector physics

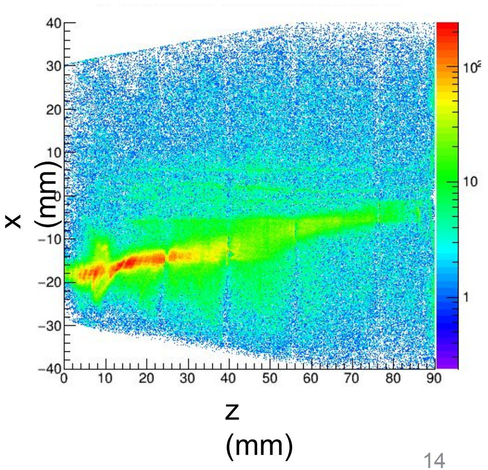
GRETA Basis Creation Optimization

GRETA Basis Creation Overview

The GRETA basis production has two distinct steps:

1. Pristine basis calculation and signal generation
   a. A calculation of the HPGe semiconductor is used to calculate the electric fields and weighting potentials within each crystal, and from this the shapes of signals on all crystal electrodes based on quantities e.g. **material impurity (profile), temperature, bias voltage, dead layers** (dozens of parameters)
2. Electronics response correction
   a. The real data folds the innate crystal response with the response of the signal processing electronics – includes **shaping times, cross-talk (integral + differential), rise times** (includes several hundred parameters)





GRETA Basis Generation Pipeline

Generating a basis for GRETA requires multiple calculation steps:

1. Calculation of electric field and weighing potentials
2. Gridding of crystals and modeling of signal shapes at each point
3. Combination of all signals into the pristine basis
4. Correcting the crystal response with the electronics response.

Completing the full pipeline required multiple points of manual intervention:

- Compilation of the binaries to run each step, which may involve editing the source C files
- Editing output file contents to prepare it for the next step
- Moving or renaming files to the expected input of the next binary
- Combining files to create a “full output”

Automated GRETA Basis Generation

An automated workflow was implemented in `python3` to simplify use of the pipeline and eliminate manual intervention. The automation:

- Appropriately runs the binaries for each step
- Modifies files as needed to act as input for the next step
- Allows for running **some** or **all** of the steps
- Implements configuration in a single yaml configuration file
- Can be run in a Docker container eliminating the need to build binaries locally

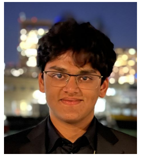
```yaml
static_inputs:
  static_input_root: static_inputs # resolved at runtime
  crystal_data: crystal_data.csv
  detector_geometry_file: geometry_setup.dat
  signal_calculation_configuration: calc_signal_setup.dat
  drift_velocity_correction: drift_vel_tcorr.dat
  cross_talk_parameters: pars669.txt
  geant_simulation_points:
    A: sel_sim_Atype_125.txt
    B: sel_sim_Btype_124.txt
  superpulse_measurements: superpulse/

results_root: output

pipeline:
  - name: crystal_impurity_scaling
    routine: impurity_scaling
    run: true
    results_path: impurity_scaling
    script_path: find_impurity_scaling.pl
    binary: f3d_gretina

  - routine: crystal_scaling_update
    # Processes final impurity scaling, updates crystals and writes final
    # field calc setup files in results_path
    run: true
    input_step: crystal_impurity_scaling
    results_path: field_calc_setup

  - routine: calculate_fields_and_potentials
    # Runs the field and weighing potential routine one final time on the
    # field calc setup files from the last step and places the results in
```

compose.yaml

```yaml
services:
  greta-pipeline:
    image: ghcr.io/lbnl-ai88/greta-pipeline-docker:latest
    user: root
    volumes:
      - ./config.greta.yaml:/greta-pipeline/config.greta.yaml:ro
      - ./static_inputs:/greta-pipeline/static_inputs:ro
      - type: bind
        source: ./output
        target: /greta-pipeline/output
        bind:
          create_host_path: true
```

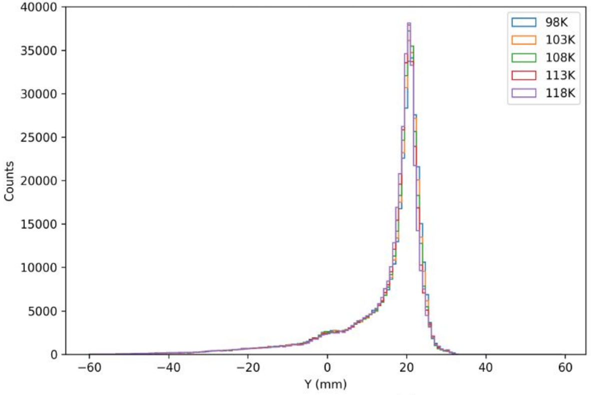
Superpulse Fitting and Detector Response

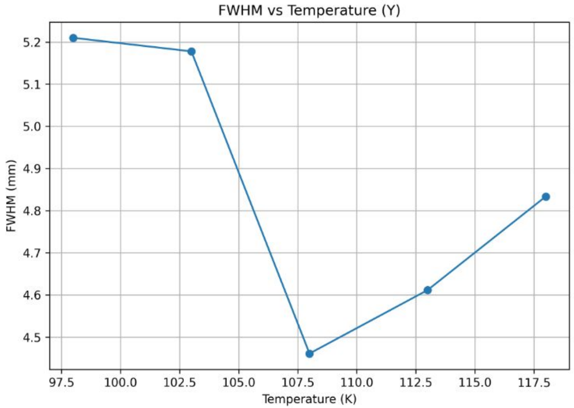
Work performed by Arin Manohar (undergraduate student researcher)



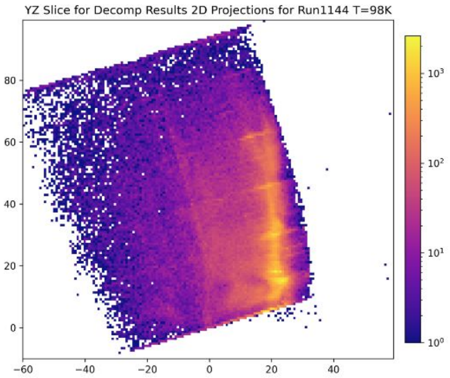
- Goal was to streamline the procedure from Geant4 and basis points data to fitting the superpulse, to automate procedure and improve the current fit model
  - Entire process has been implemented in Python, fitting using a trust-region reflective algorithm for convergence stability and computational efficiency
- With the full GRETA/GRETINA electronics response function implemented, the parameter sensitivity was investigated
- Insensitive per-segment delay parameters were removed, and the quality of the GRETA basis was confirmed to be fully maintained

Exploring Crystal Parameters - e.g. Temperature





- Combining the containerized basis production pipeline and the Python-based superpulse fitting, now exploring sensitivity to crystal parameters
- First to be explored is crystal temperature
- Optimizing position resolution of a pencil beam measured for a given GRETA crystal

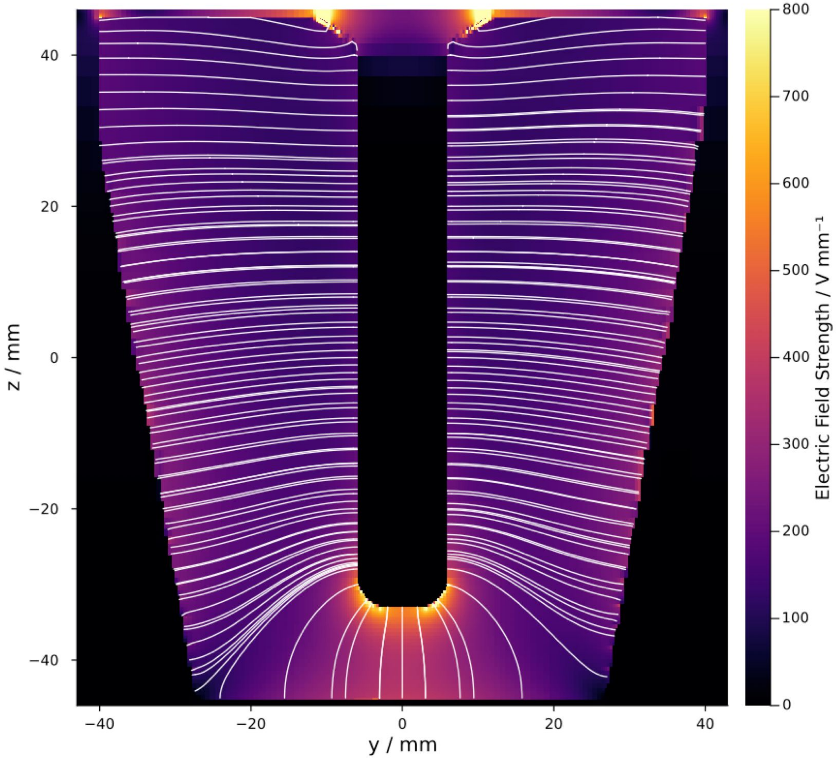


- Demonstrated that optimal temperature is above what was previously assumed; now extending optimization to multiple crystals to understand systematic behaviour of detectors

Manohar *et al.*, to be submitted December 2025.

Work performed by Gabriel Garcia Jimenez (PD)

Alternative Basis Production - Additional Physics

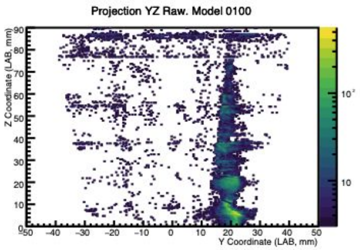


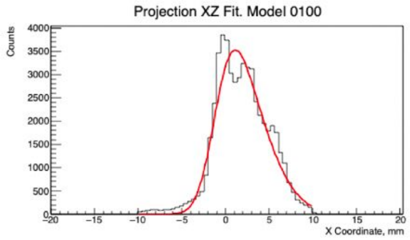
- An alternate pipeline was also established using the Julia-implemented SSD (Solid State Detectors) framework
- In addition to having an active community and documentation, this simulation enables additional physics
  - Charge cloud simulation
  - Non-linear impurity profiles
  - Alternate charge drift models

Impact of Charge Drift Models

- Different charge drift model parameterizations were investigated including the GRETINA-adopted model and the AGATA Detector Library model, and applied across several data sets to explore the impact (in combination with several other parameters)
- Optimization was across a discrete set of options, but shows a clear preference for the GRETINA-adopted drift velocity parameterization

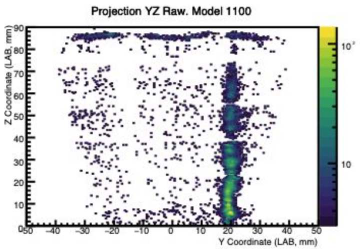
ADL charge mobilities

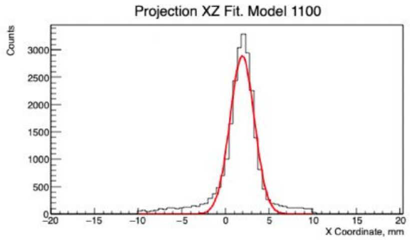




IQR: 4.2

GRETINA-adopted charge mobilities



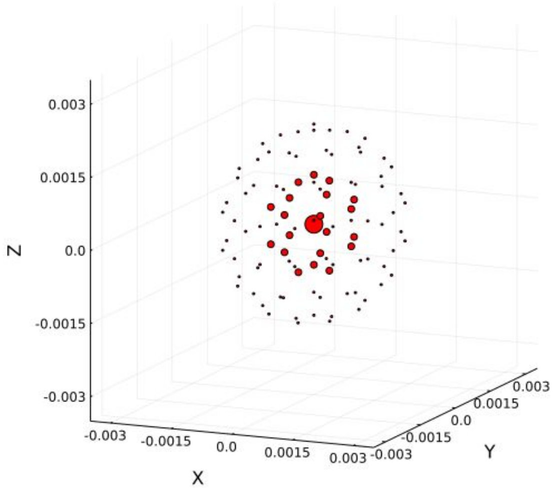


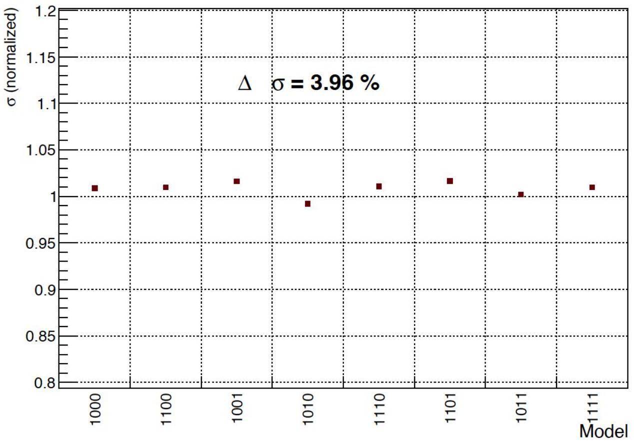
IQR: 2.0

Impact of Charge Clouds

- Prior GRETA basis assumed a point charge moving in the crystal to generate the calculated pulses
- SSD pipeline enables exploration of the impact of charge cloud diffusion and self-repulsion

- Optimization was performed across discrete model options including electron diffusion, self-repulsion
- Considered different sized charge clouds
- **No sensitivity** in position resolution with respect to charge clouds





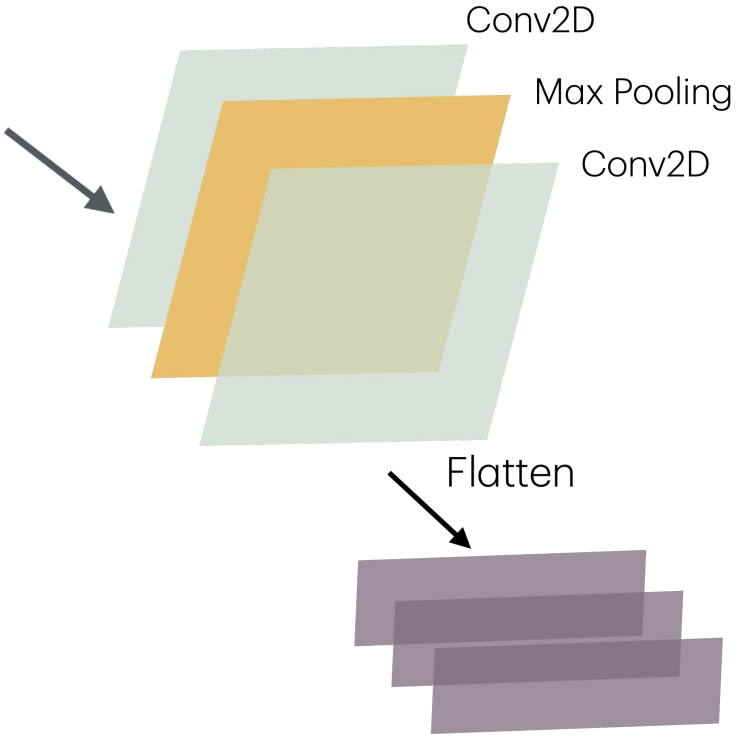
Garcia Jimenez *et al.*, in preparation.

Exploring a Neural Network to Extract Positions in GRETA

$$\begin{pmatrix} a^i_{0,0} & \cdots & a^i_{0,49} \\ \vdots & & \vdots \\ a^i_{36,0} & \cdots & a^i_{36,49} \end{pmatrix}$$

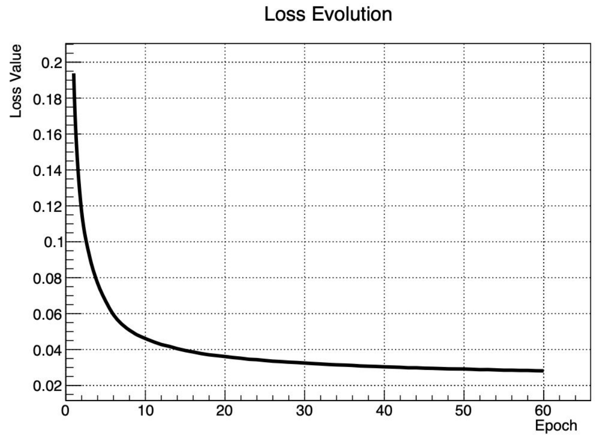
$$+$$

$$(\hat{r}^i, \hat{\phi}^i, \hat{z}^i)$$

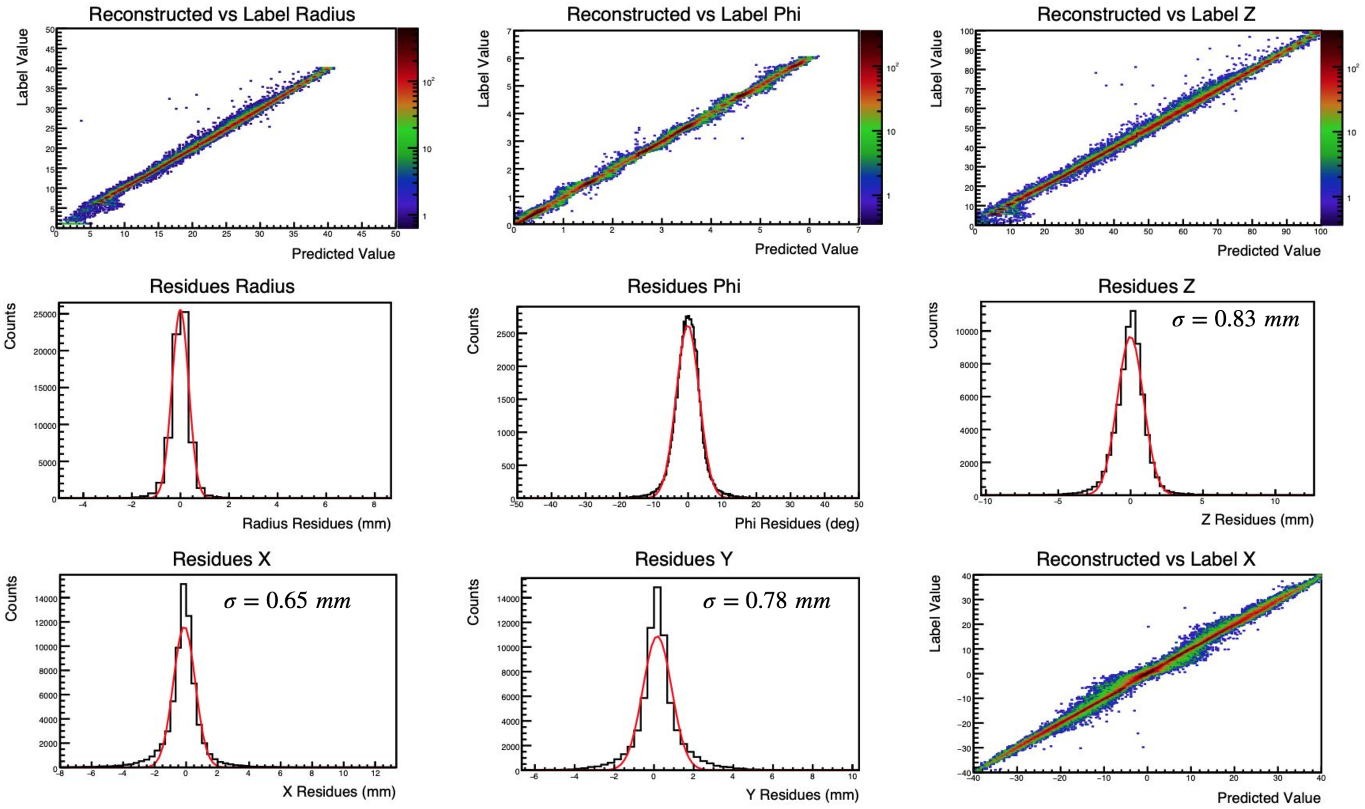


$$(r^i, \phi^i, z^i)$$

- 9990 Parameters
- ~ 229.000 events (the basis)
- 37 x 50 matrices
- 60 epochs
- 6/10 training, 3/10 test, 1/10 validation
- ~ 30 min training



Results - Calculated Signal Basis



- Promising!
- Next step is to test with calculated signals including added noise (white + correlated)
- Single-interaction data will follow

GRETA Project Goals and Status

WBS	Milestone	Description	
2	GRETA Staffing Requirements Met	Advertise and hire an undergraduate student and postdoc to work on GRETA scope.	✅
2.2.1	Develop Python utilities for signal basis representation	Develop a library of Python tools for signal basis representation and visualization, including pulses at individual interaction points.	✅
2.2.1	Define electronics response function	Define a parameterization for the electronics response function for basis generation.	✅
2.2.1	Explore sensitivity of superpulse types to parameters	Characterize the sensitivity of different measurement types (superpulse types) to parameters in the electronics response function.	✅
2.2.2	Evaluate hyperparameter search tools for use in GRETA case	Explore the available hyperparameter search tools that we can consider for use in optimizing the electronics response and crystal parameters.	
2.1	Demonstrate (up to) 120 crystal simultaneous optimization	Extend the optimization and calibration code to tackle 120 crystals at once.	✅
2.2.1	Implement updated signal basis generation tool chain	Implement and configure complete signal basis generation tool chain with updated utilities for automated basis generation.	✅
2.2.2	Develop parameterization for crystal description	Define a parameterization of the crystal properties such as impurity profile etc.	✅
2.3	Evaluate opportunities for direct ML inference of basis signals	Look into techniques that can generate a signal basis without the crystal properties calculation based on data only.	✅
2.2.2	Complete final code base for open-source distribution.	Assuming success for previous steps, clean up code and package for open-source distribution following LBNL policies.	✅

VENUS

The Electron Cyclotron Resonance (ECR) Ion Source VENUS

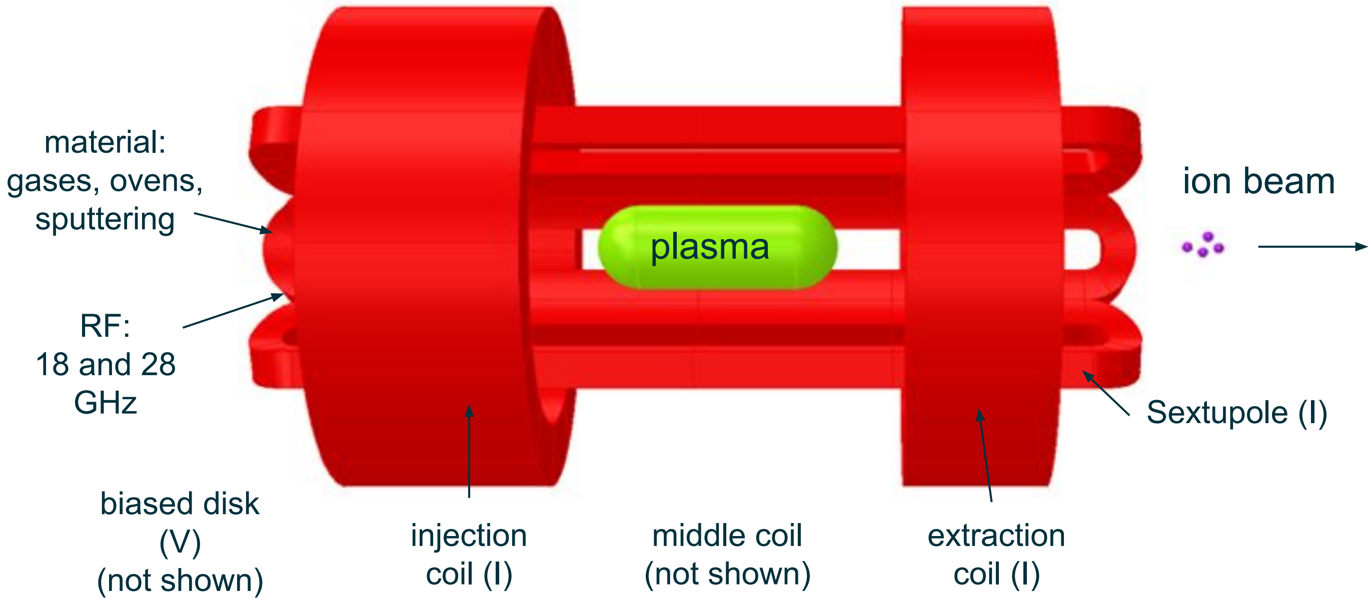
VENUS @ LBNL

- World's first fully-superconducting ECR ion source designed for 28 GHz operation
- One of the world's two highest-performing ECR ion sources
- Injector for LBNL's 88" Cyclotron
- Prototype ECR ion source for FRIB, where a near-identical copy has been installed, and two more are under construction

<u>Example beams</u>

high currents	high charge states
> 4.7 mA O^{6+}	$^{197}Au^{61+}$ out of cyclotron
> 20 mA He^{+}	> 2.3 GeV!

VENUS Primary Control and Diagnostic Parameters

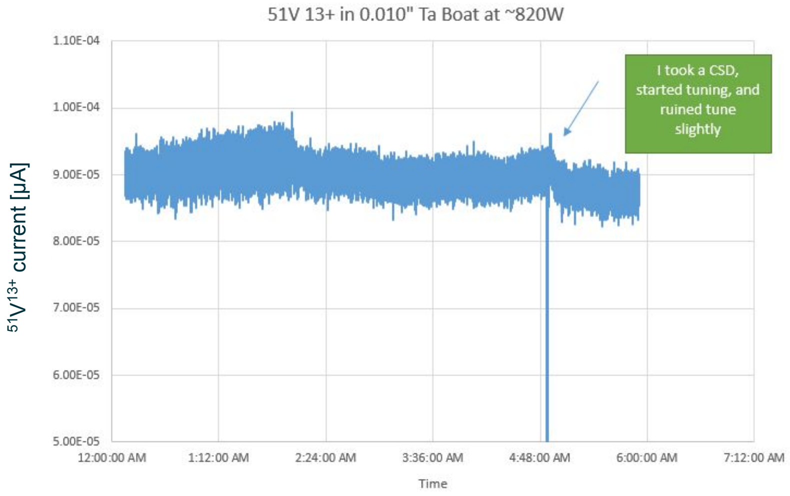


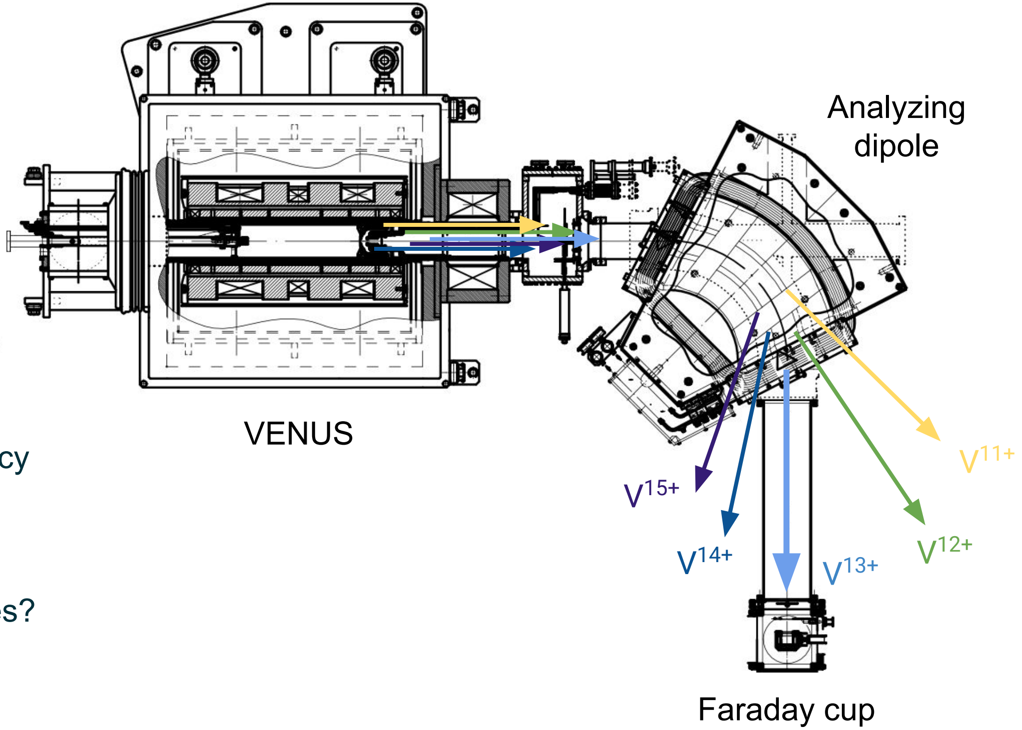
Diagnostics:
- Faraday cup
- Charge state distributions
- Emittance scanner
- Drain, bias currents
- Cryostat x-ray load

10-20 control parameters

Important: no reliable model exists!

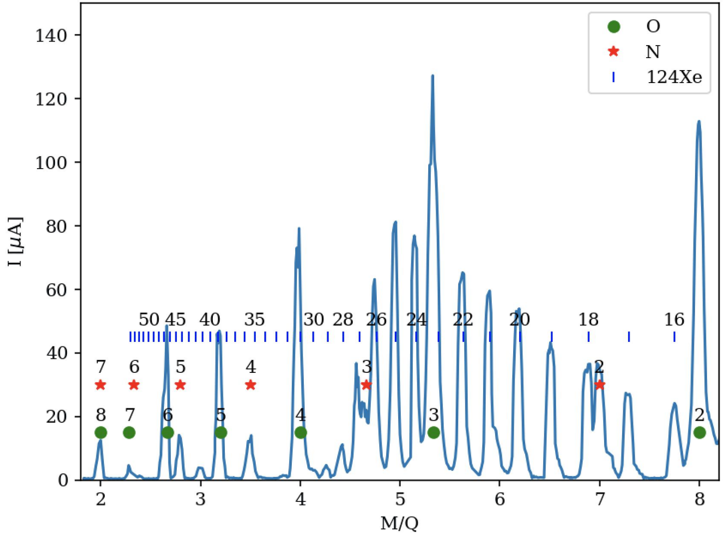
Beam current as primary diagnostic

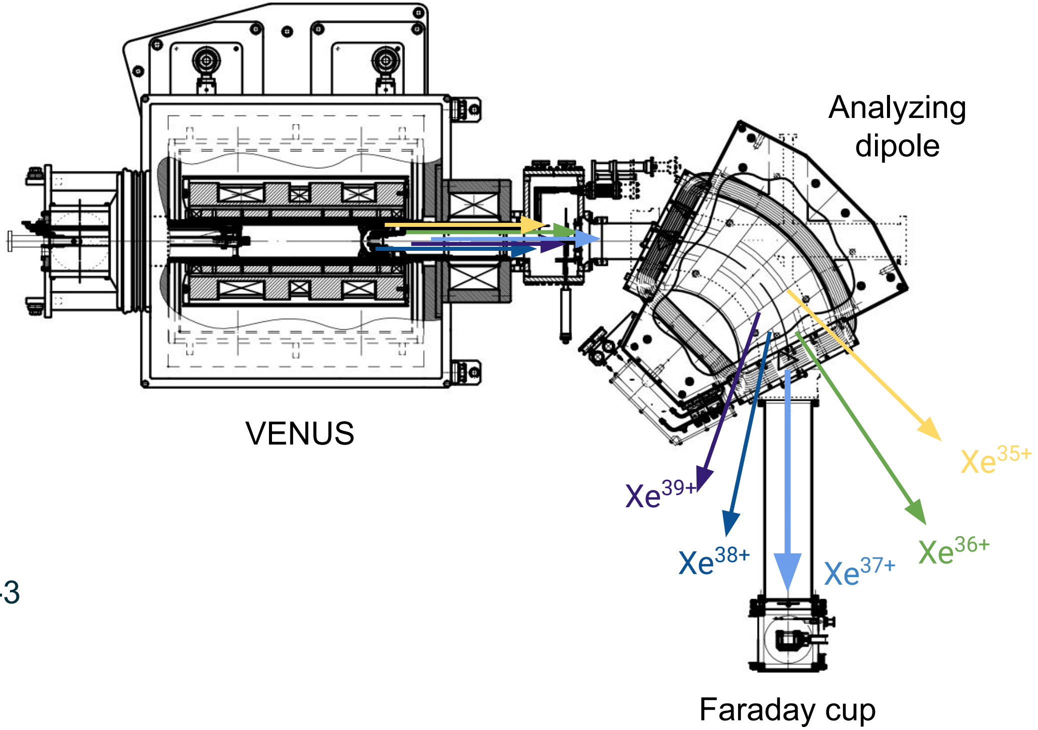




- We have collected source settings and species current over 1200 days of data at ~ 1 Hz frequency
- What's missing:
  - intent?
  - was the dipole adjusted to peak species?

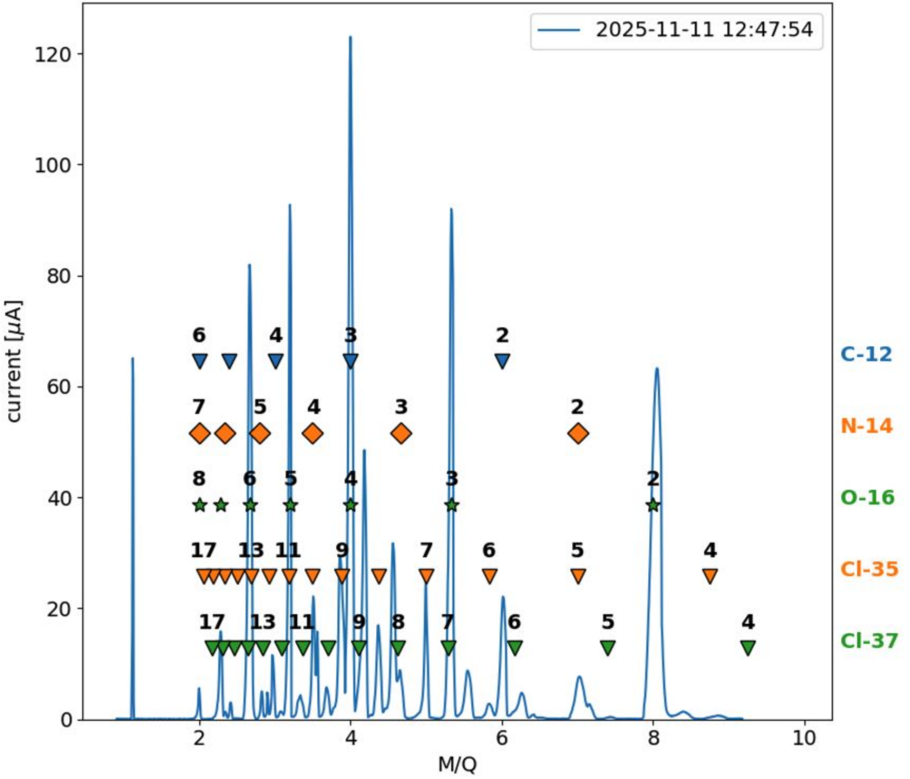
Beam current as primary diagnostic





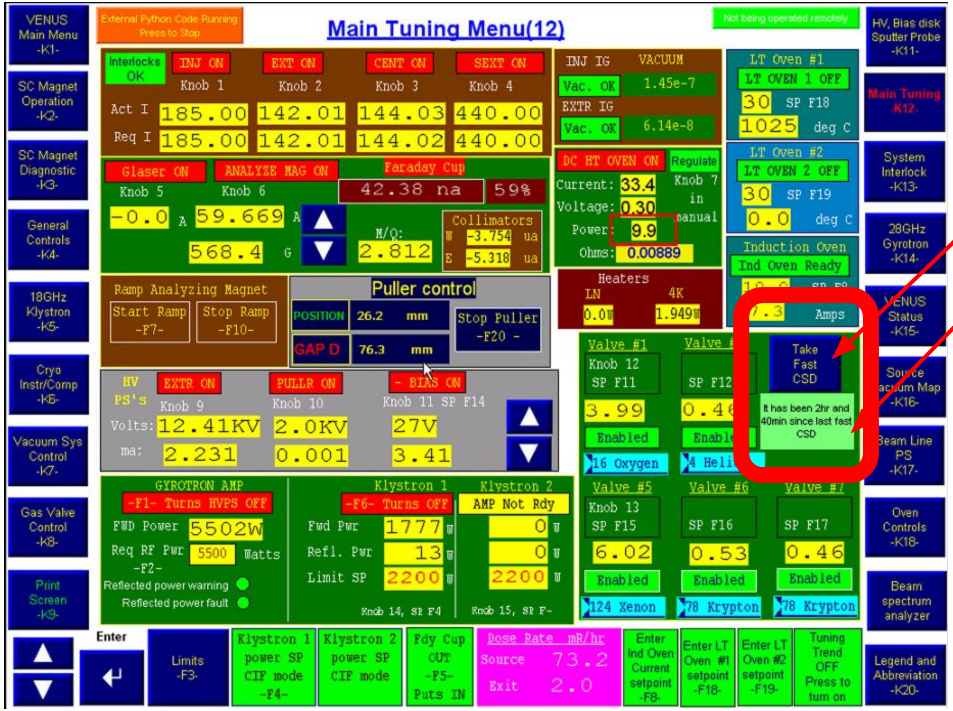
- A better diagnostic: the charge state distribution
- Historically done infrequently as they are slow (2-3 minutes)

Faster charge state distributions (CSDs)



- Developed ability to take a charge state distribution in as little as 6 seconds

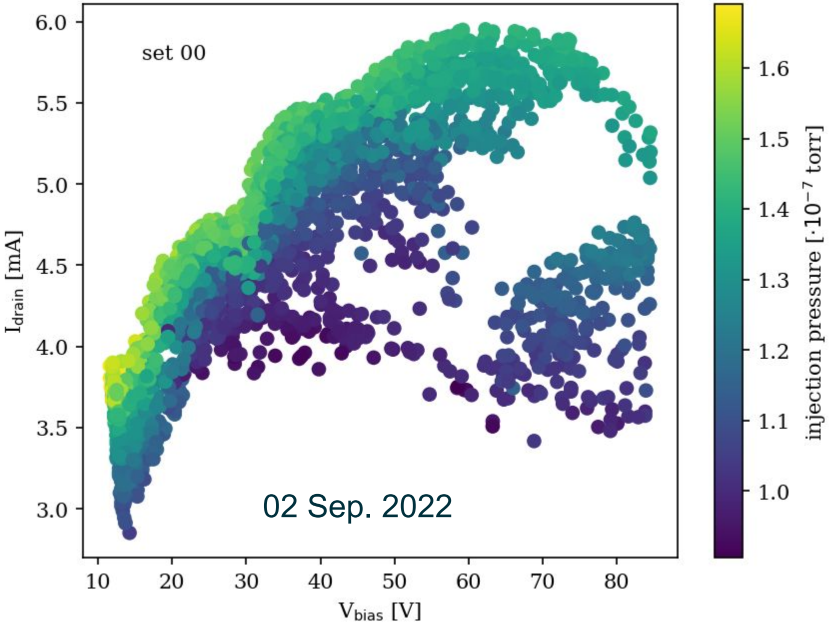
Charge state distributions (CSDs) part of normal operation

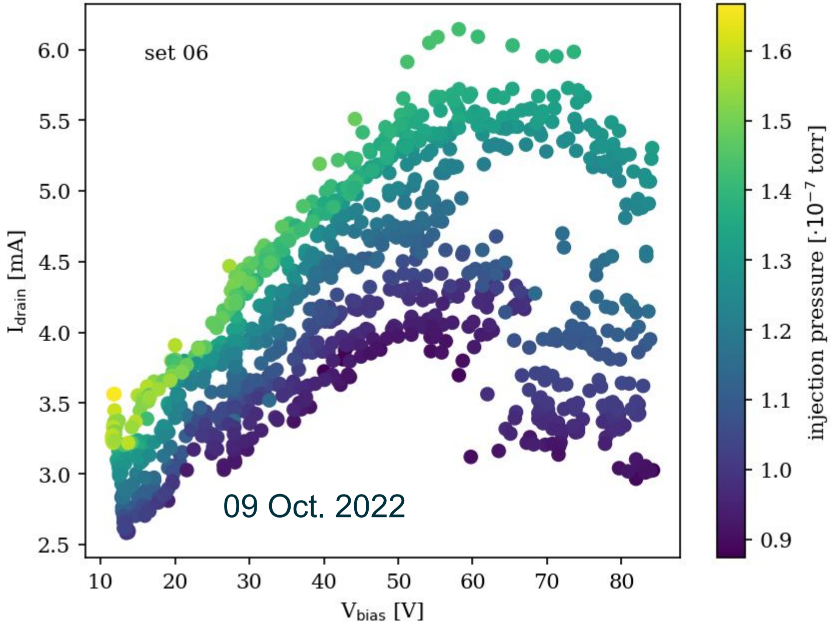


- Added single button operation
- Reminders alert cyclotron operators to perform CSD one hour after previous
- Faraday cup inserted, CSD completed, Faraday cup removed, and beam back on target within 30 seconds
- Good compliance by cyclotron operators: over 2100 CSDs recorded since March 2025
- Better data for machine learning codes!

Modeling VENUS Across Data Sets with Random Forests and Neural Networks

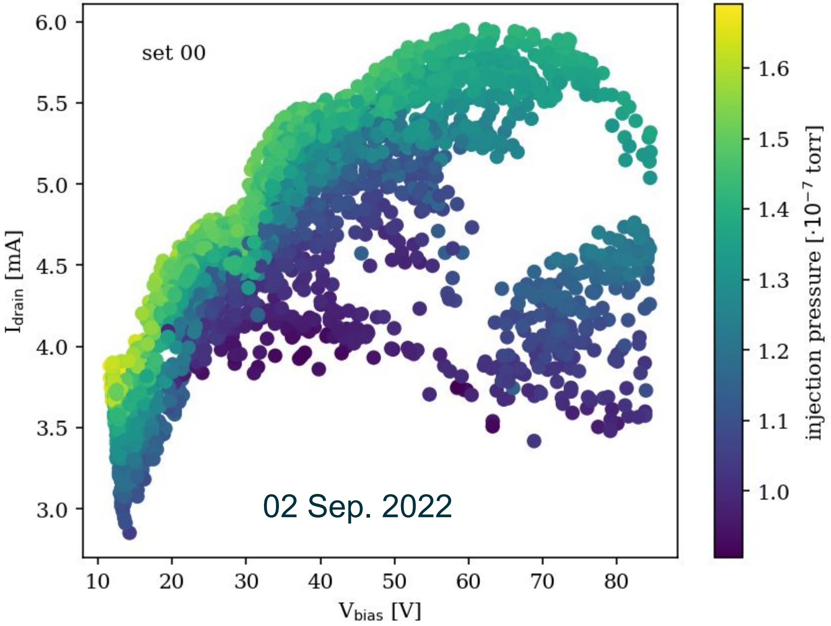
Data collection for Random Forest and Neural Networks



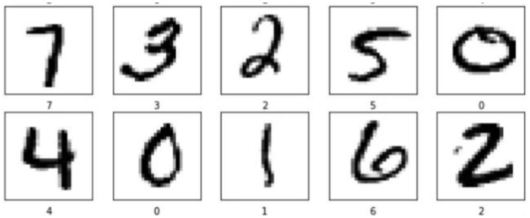


- Encouraging week-to-week similarities!
- Use learning on past to predict future?

Data collection for Random Forest and Neural Networks



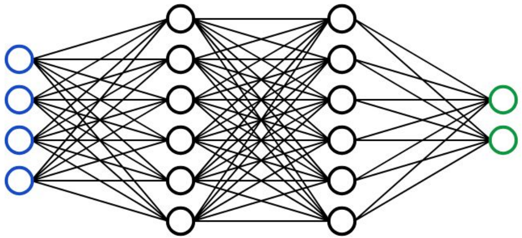
- Visual similarity of data sets encouraging for typical machine learning problems

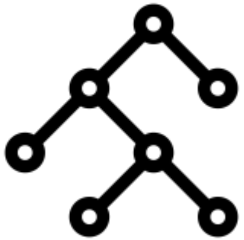


- Solution: apply typical tools (e.g. neural networks and random forest modeling) to new problems

Two approaches: neural networks and random forests

Work performed by Ezra Apple (undergraduate student researcher)

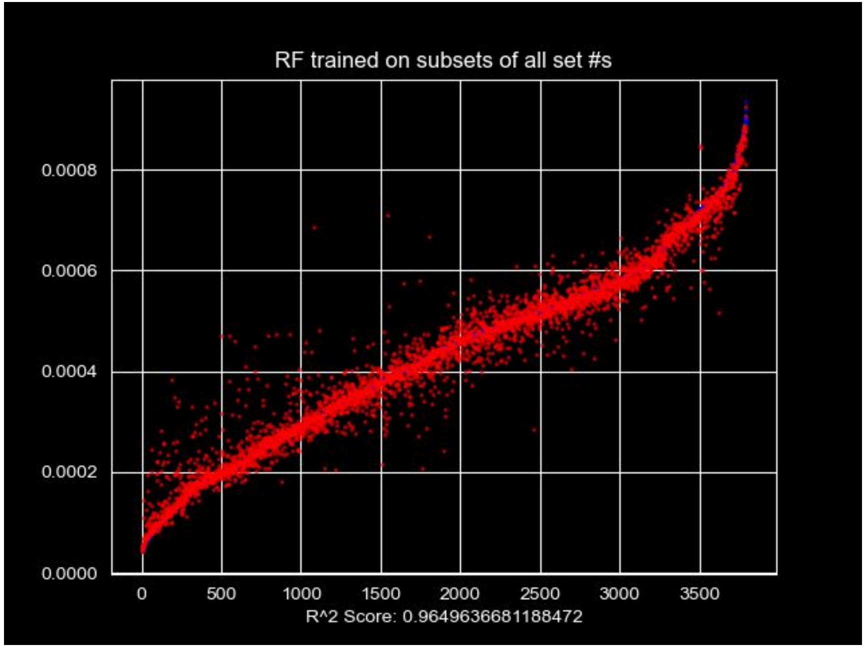




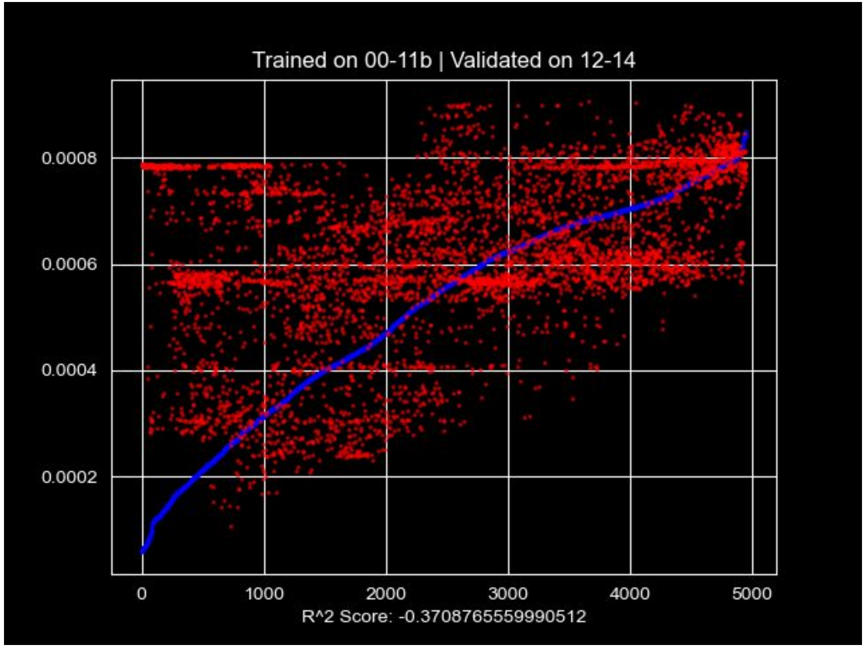
- Apply neural networks: use source settings as the input layer and output beam current measurements as output layer to train hidden layers
- Results:
  - Network trained on a given subset of a weekend's data performed well on that weekend
  - Same network performed poorly on some separated weekends

- Random forest for regression: apply an ensemble method with multiple decision trees, and make predictions using the average output of all trees
- Results:
  - Results similar to neural network (good on one weekend, poor weekend-to-weekend)

Example data from Random Forest



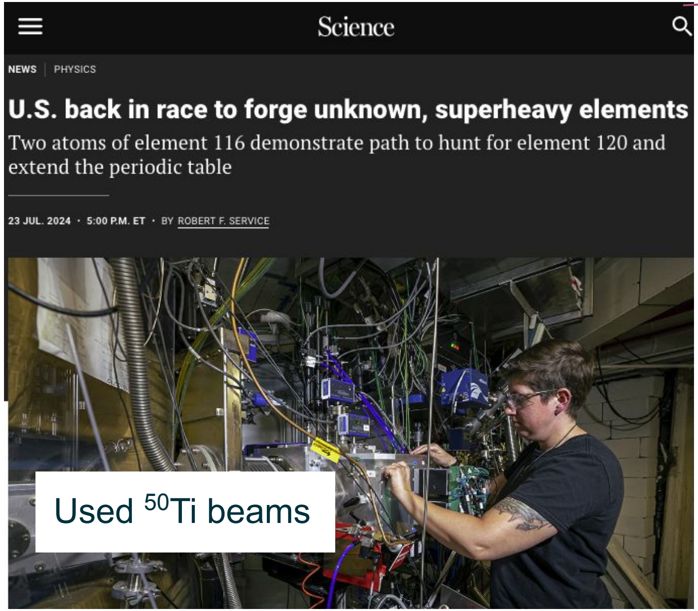
- If the random forest has seen some of a new data set in its training, it performs very well



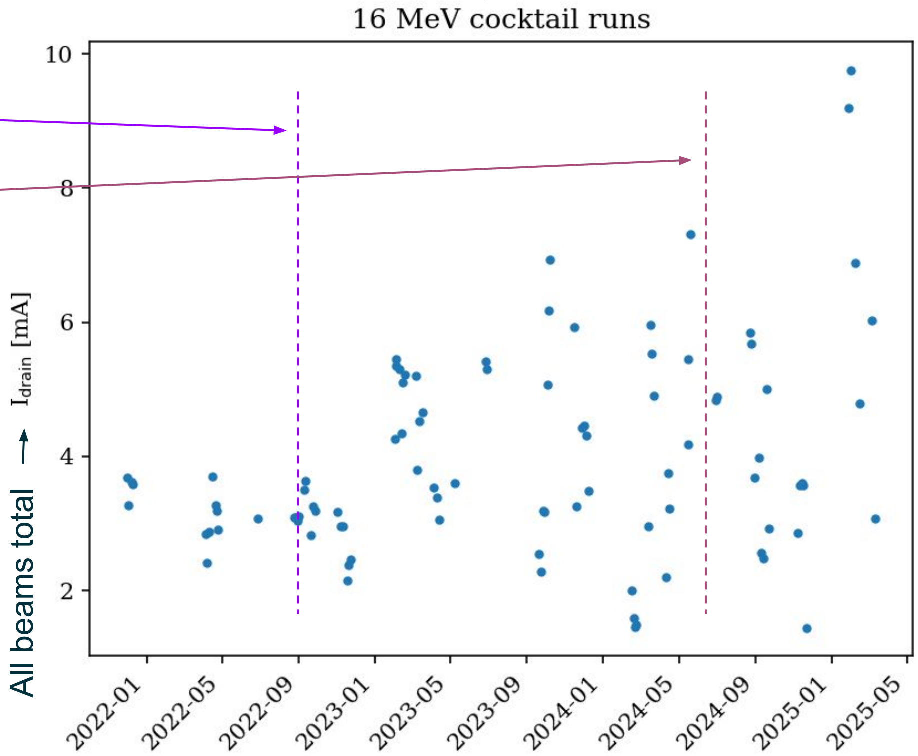
- If the data set is brand new, it performs horribly

Set-to-set reliability

- January 2022: machine-learning-related VENUS data collection
- September 2022: machine learning neural network work begins



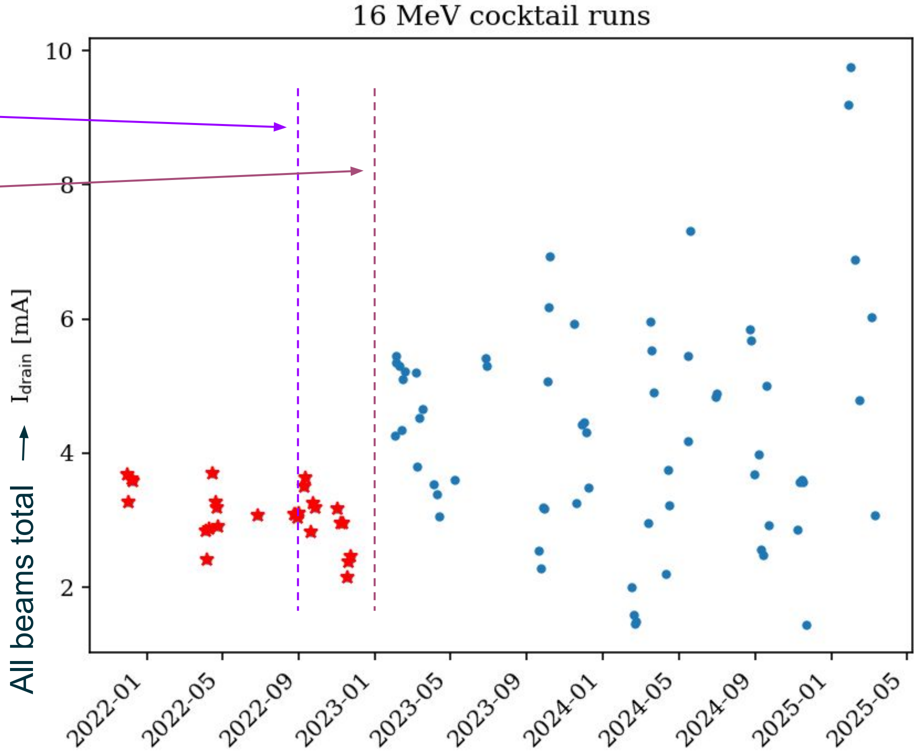
Frequently run beam for cyclotron



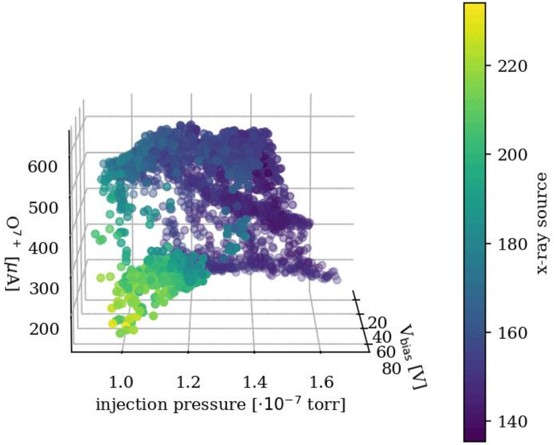
Set-to-set reliability

- January 2022: machine-learning-related VENUS data collection
- September 2022: machine learning neural network work begins
- January 2023: long ^{50}Ti campaigns begin for superheavy research
  - Note: Little-to-no ^{50}Ti in spectrum!!

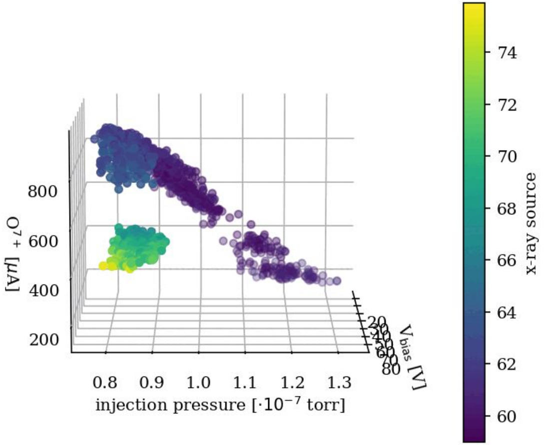
Frequently run beam for cyclotron



Data collection for Random Forest and Neural Networks



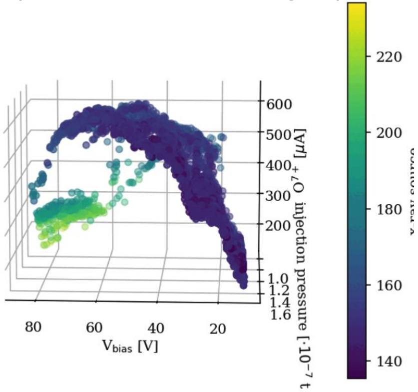
02-06 Sep. 2022
(before ^{50}Ti campaigns)



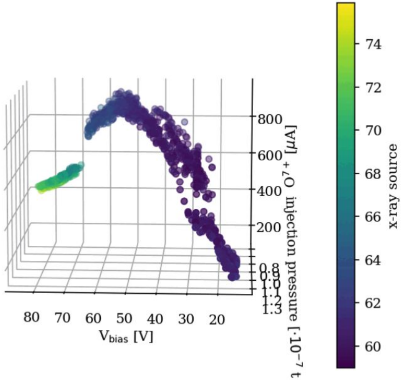
23-25 July 2023
(between ^{50}Ti campaigns)

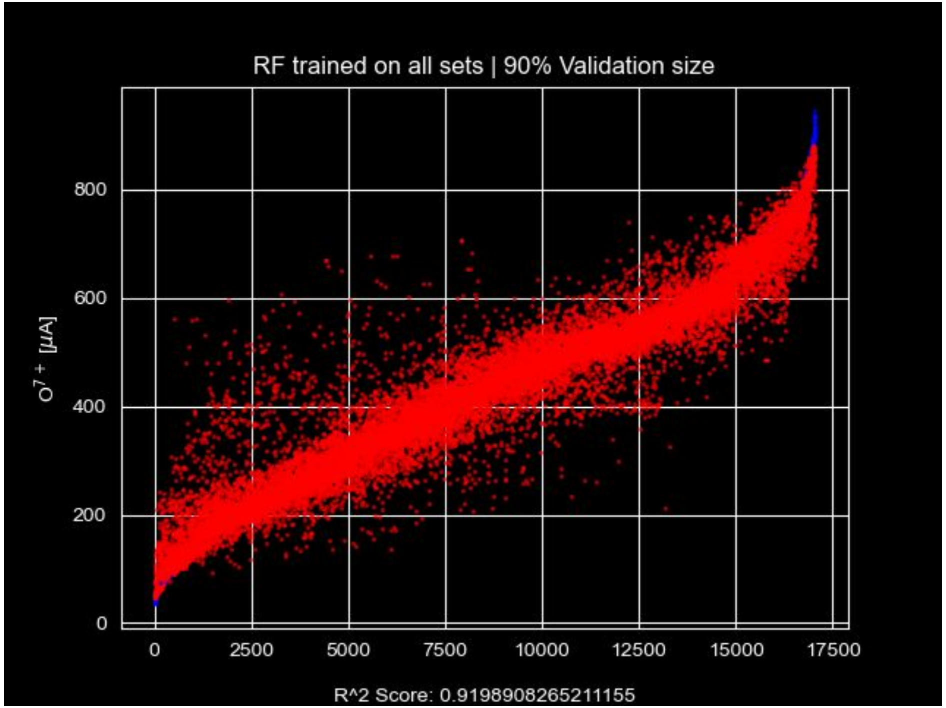
Introducing new data set to random forest

02-06 Sep. 2022
(before ^{50}Ti campaigns)



23-25 July 2023
(between ^{50}Ti campaigns)

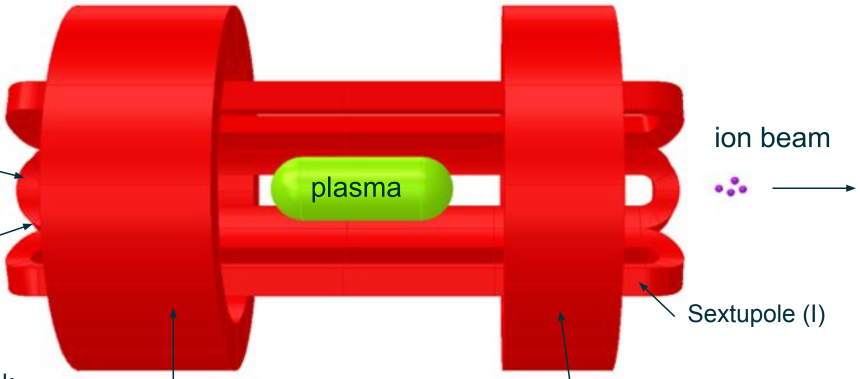




- Hidden variable (^{50}Ti), but data is still "similar"
- Training random forest on small subset (10s to 100s of measurements) of new data space produces good predictions

VENUS Bayesian Optimization

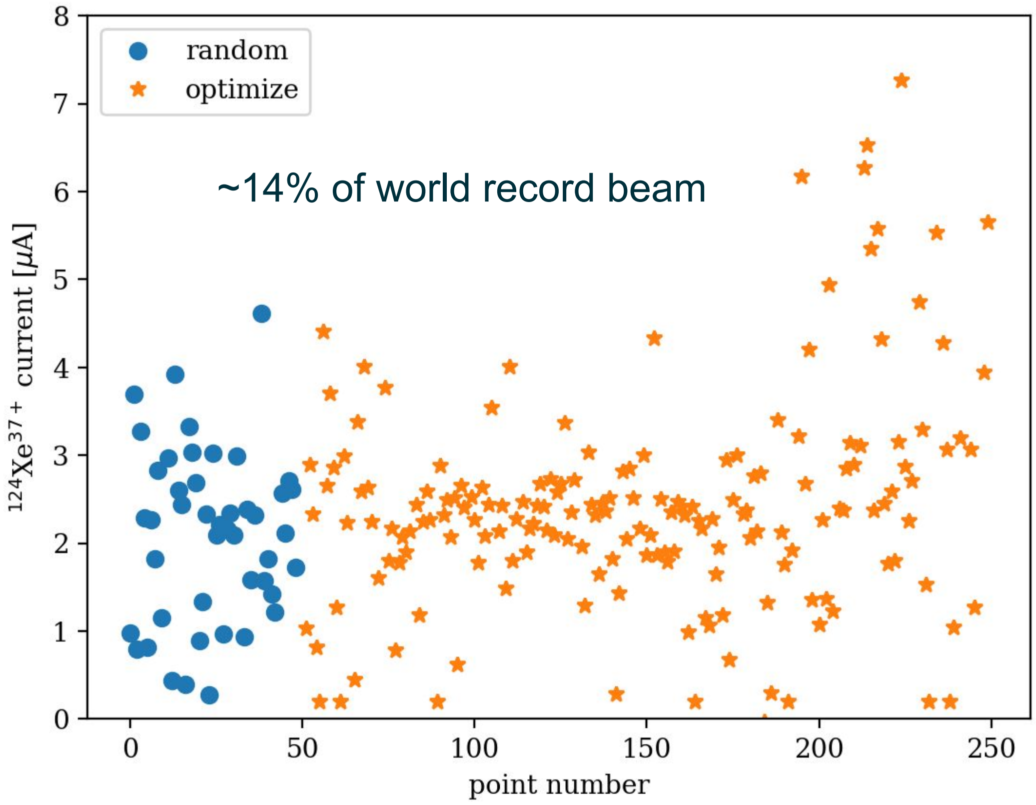
Machine Learning: Full Bayesian Optimization of $^{124}Xe^{37+}$



Parameter	Min	Max
Bias voltage [V]	40	105
Oxygen valve	11.6	12.5
Xenon valve	8.0	13.0
Inj coil [A]	185.6	186.0
Ext coil [A]	136.6	136.8
Mid coil [A]	152.0	152.3
Sext coil [A]	430.3	430.5
18 GHz [kW]	1.4	1.8
28 GHz [kW]	5.2	6.0

- VENUS completely under computer control
- Computer “knows” nothing about VENUS

Machine Learning: Full Bayesian Optimization of $^{124}Xe^{37+}$



(~5 min/point)

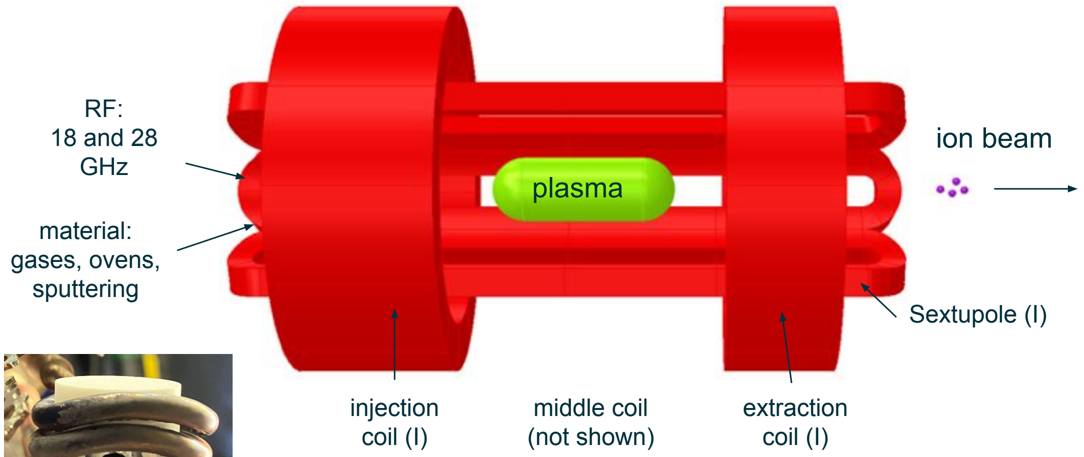
Parameter	Min	Max
Bias voltage [V]	40	105
Oxygen valve	11.6	12.5
Xenon valve	8.0	13.0
Inj coil [A]	185.6	186.0
Ext coil [A]	136.6	136.8
Mid coil [A]	152.0	152.3
Sext coil [A]	430.3	430.5
18 GHz [kW]	1.4	1.8
28 GHz [kW]	5.2	6.0

- VENUS completely under computer control
- Computer "knows" nothing about VENUS

Machine Learning: will attempt world record optimization with new plasma chamber

- New plasma chamber with microchannel cooling constructed 2025
- Once installed (2026), will attempt world record beam attempt using Bayesian optimization

Machine Learning: Bayesian Optimization of $^{48}Ti^{13+}$

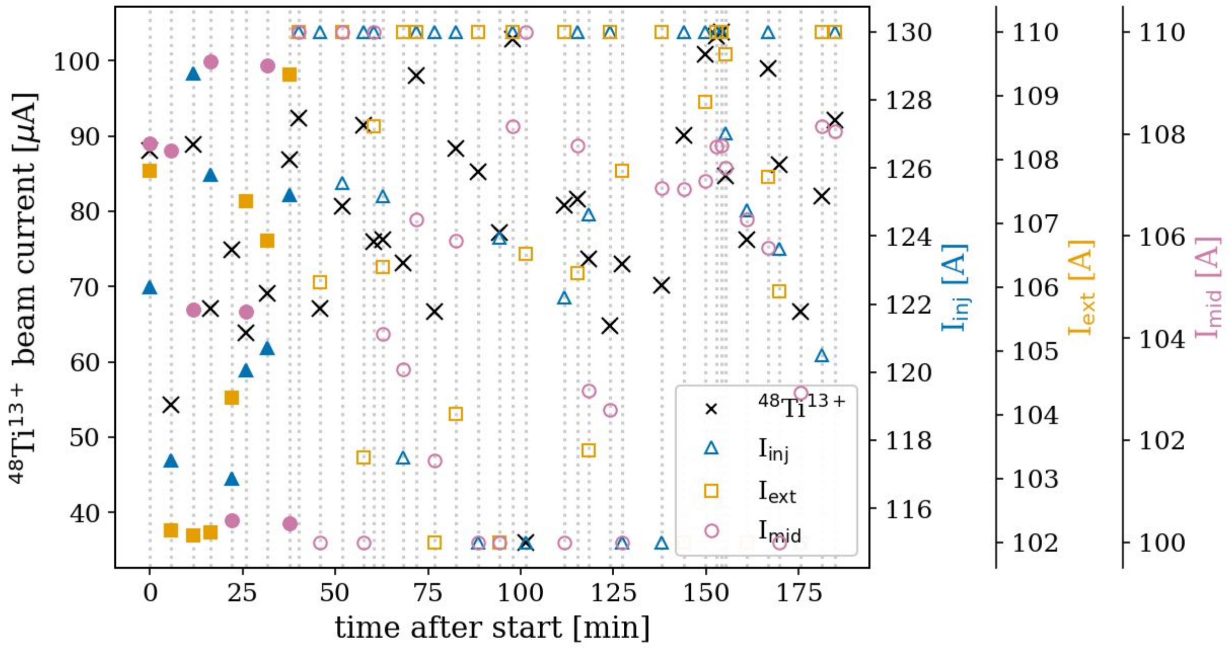


inductive
oven

- **Don't** use oven as an optimization parameter
- Instead get as much beam as possible for a given material input

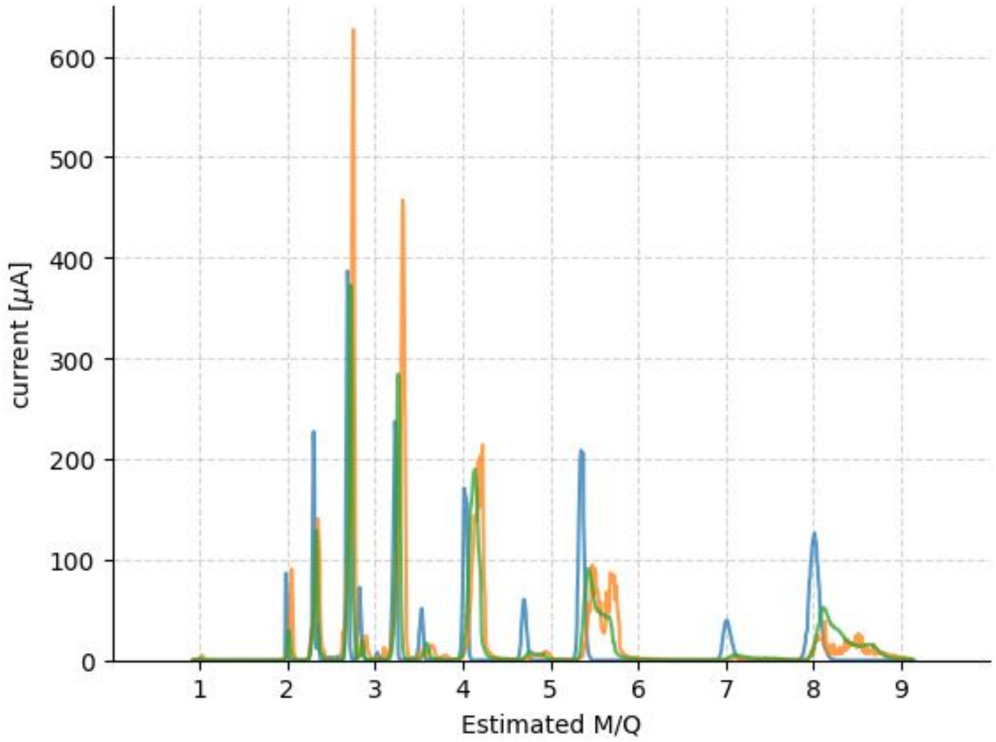
- Bayesian optimization performed well for Ti optimization
- Greater difficulty was pressure control: Ti is a getter metal and large pressure changes were hard to recover from when wildly varying the plasma
- Will employ PID (proportional-integral-derivative) control for pressure and retry

Machine Learning: Bayesian optimization of superconducting coils for $^{48}Ti^{13+}$ production



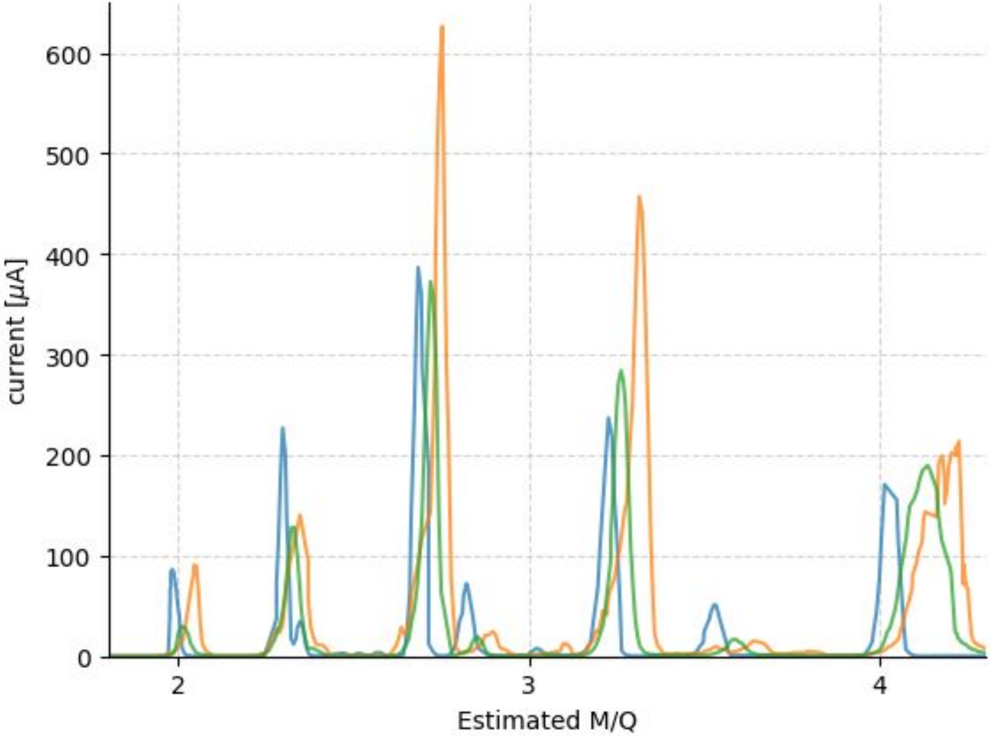
- Superconducting coils are often avoided by human operators as they can take ~5 minutes to settle
- Typical operating fields produced 82 μA, while Bayesian optimizer was able to find solutions as high as 104 μA
- Considerable (tedious) human tuning time saved.

Applying Bayesian optimization to CSDs

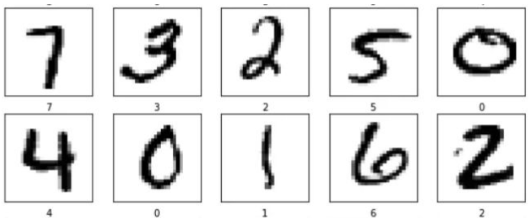


- Peak identification is a tedious human task
- Dipole hysteresis, changes to the plasma potential, and CSD shifts to lower or higher charge state can move the peaks

Applying Bayesian optimization to CSDs



- Peak identification is a tedious human task
- Dipole hysteresis, changes to the plasma potential, and CSD shifts to lower or higher charge state can move the peaks
- Clear application point for machine learning to identify peaks for the 2000+ current and all future CSDs



CSD Oxygen Peak Identification Results

Neural network trained: Multi-Layer Perceptron (MLP) Classifier

Training set: 15100 sets of peaks, 302 identified by hand as oxygen peaks

Testing: held out 3020 sets (20%) for testing,

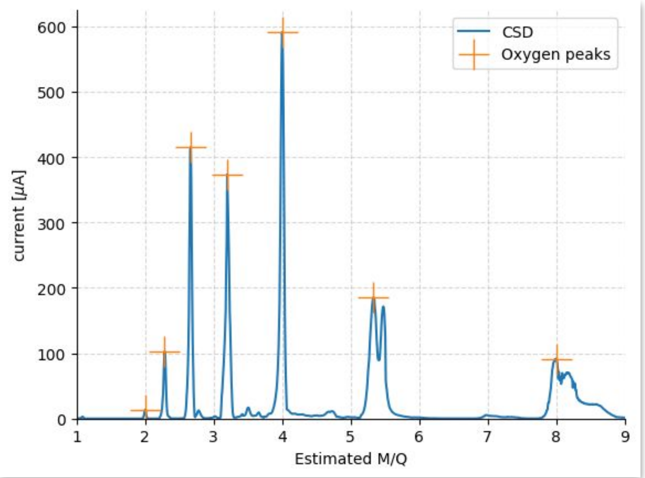
Results:

- Model correctly identified all 60 oxygen peaks in the test set (no false negatives)
- 11 false positive

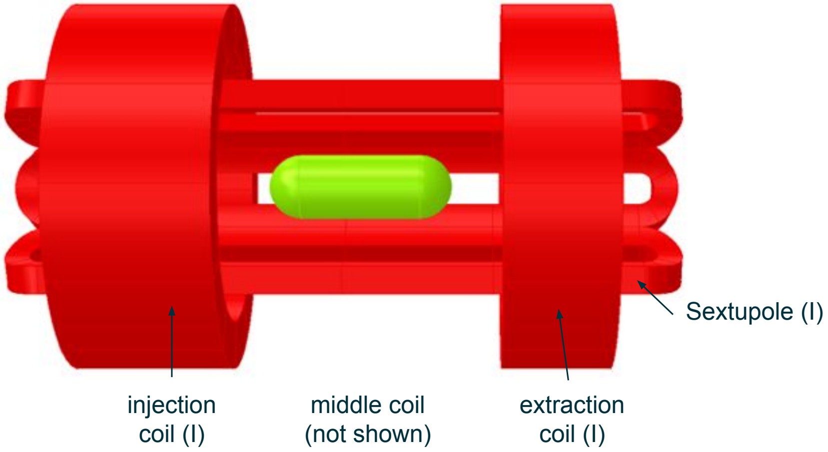
Model parameters:

- ReLU activation function
- 3 hidden layers (30, 15, 7)

```
ROC-AUC   : 0.9991
Classification report
              precision    recall  f1-score   support

           0     1.0000    0.9963    0.9981      2960
           1     0.8451    1.0000    0.9160        60

    accuracy                         0.9964      3020
   macro avg     0.9225    0.9981    0.9571      3020
weighted avg     0.9969    0.9964    0.9965      3020

Confusion matrix
[[2949   11]
 [   0   60]]
```

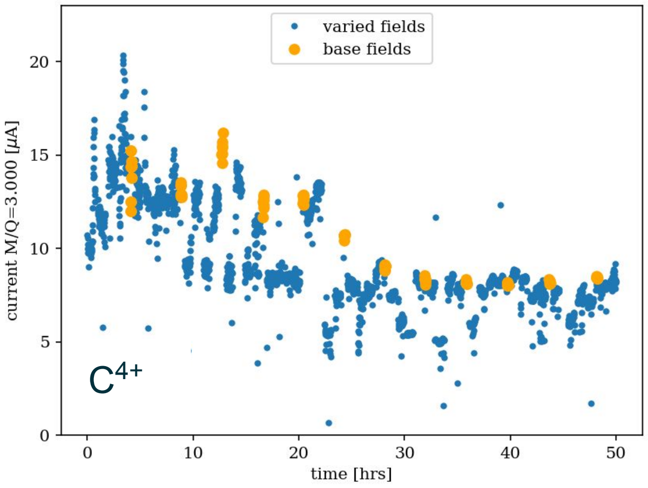


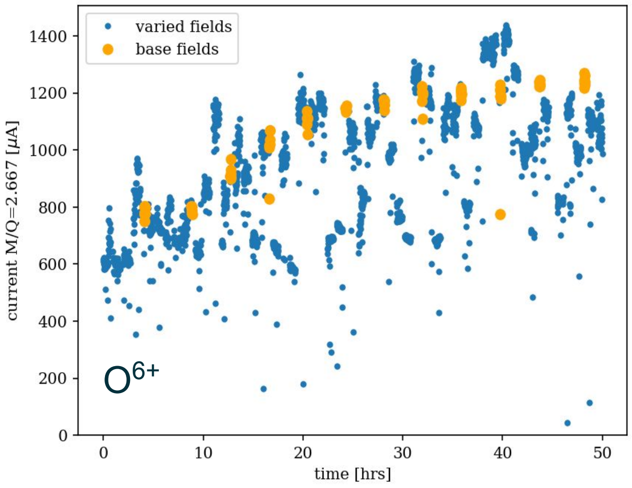
Baking optimization



- Baking, the removal of unwanted particles from the plasma chamber and source to improve performance, can be a days-long process
- This process can be sped up by varying the plasma-confining magnetic fields to have the plasma “clean” — humans typically vary the fields occasionally (if at all)

Baking optimization



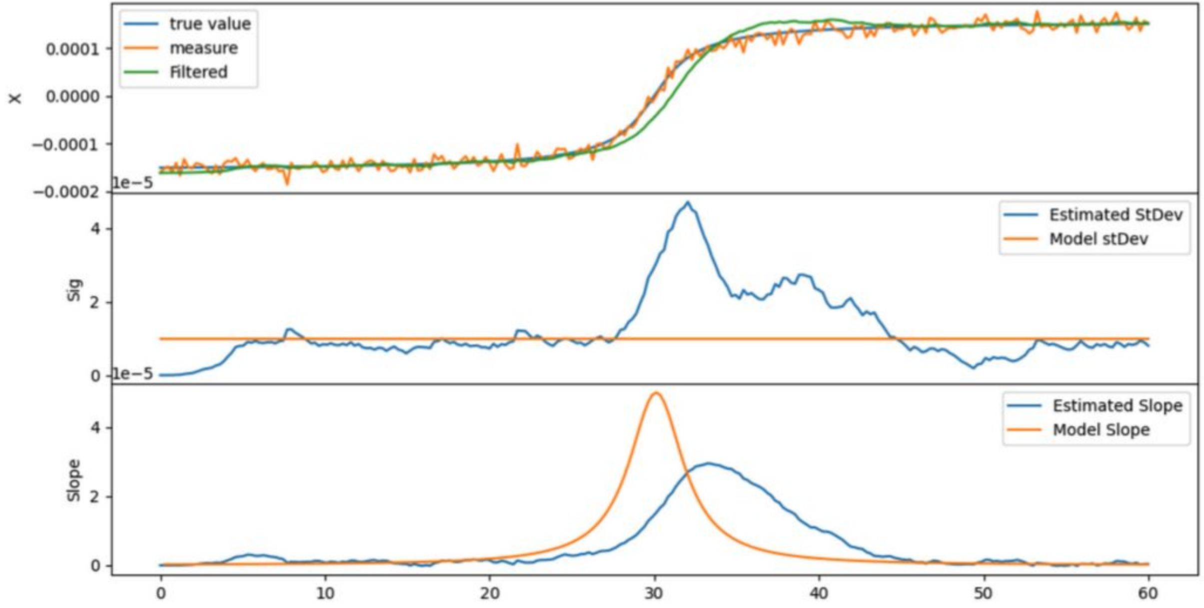


- Currents of unwanted material at "base" coil settings reduce with time
- System baked after ~30 hours

- Currents of desired material asymptotically approach final values until ~30 hours

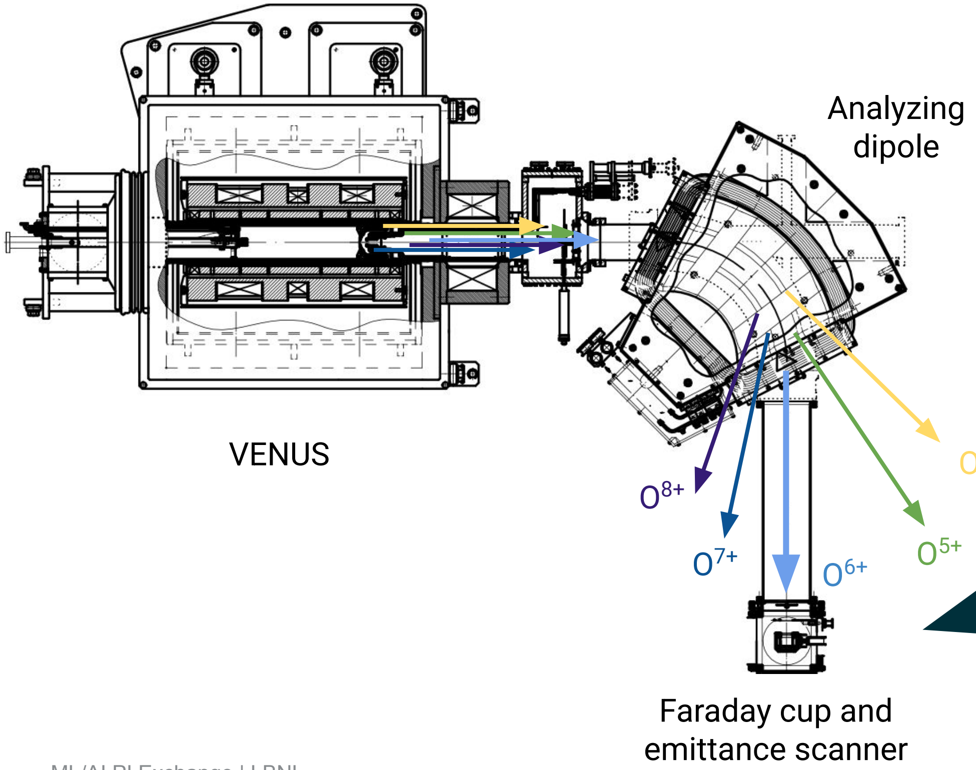
Pro: A faster and better baked source with no human effort!
Con: Too small of data set for true machine learning (yet!)

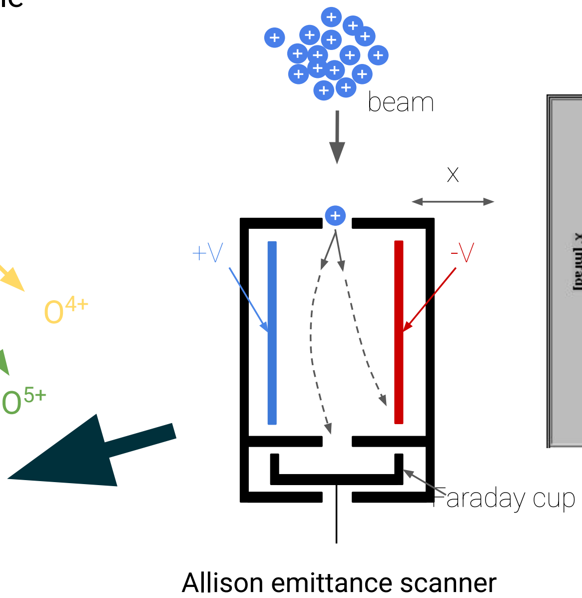
Kalman filtering for stability and settling

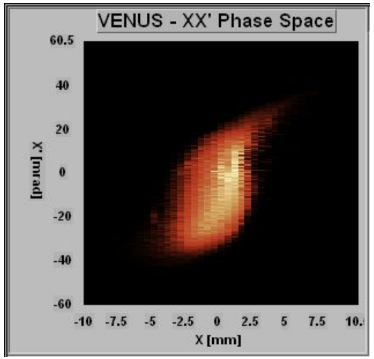


- First Bayesian optimizations used constant wait times after each system change. New values were accepted if determined stable and settled
- 2025: incorporated a multi-parameter Kalman filter applicable to all measured values.
- Slopes (e.g. beam current, vacuum pressure, etc.) can be used to determine if still changing
- Standard deviation (typically 5% relative) is also used as a settling and acceptance parameter

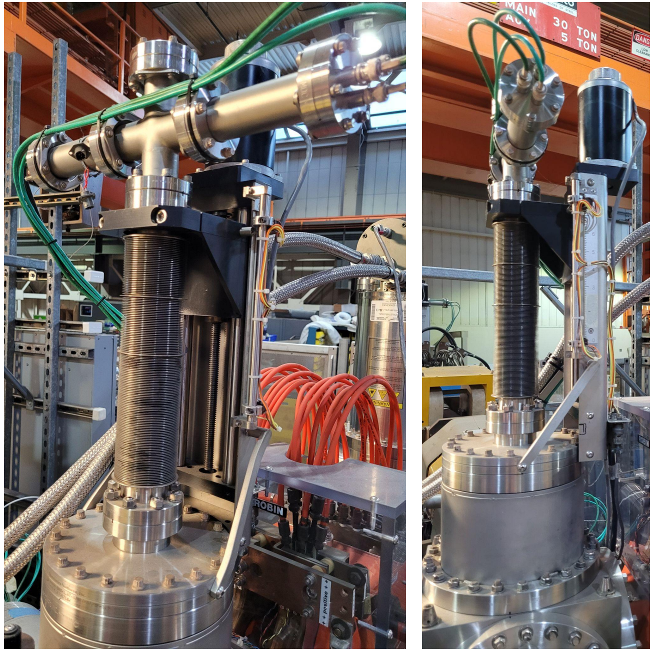
Emittance scanning with VENUS



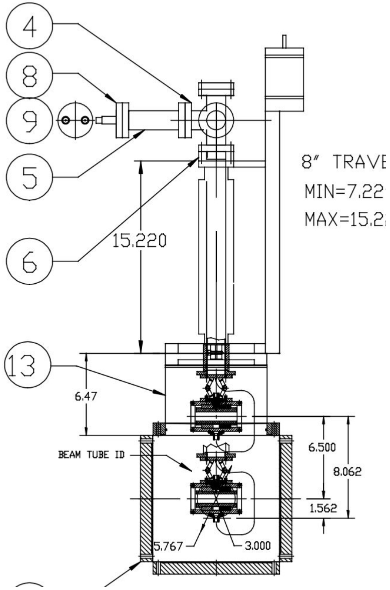




VENUS optimization using emittance scans



VENUS vertical emittance scanner



- Emittance scans rarely performed previously:
  - ~3 minutes to perform
  - finicky Labview program only runs on Windows

Successes:
- Control and analysis system ported to Python
- Motor control motion optimized: scans reduced to less than one minute
- Scanner easily controlled and output easily read by optimizer

Still to come:
- Dedicated optimization run

Reinforcement Learning for VENUS

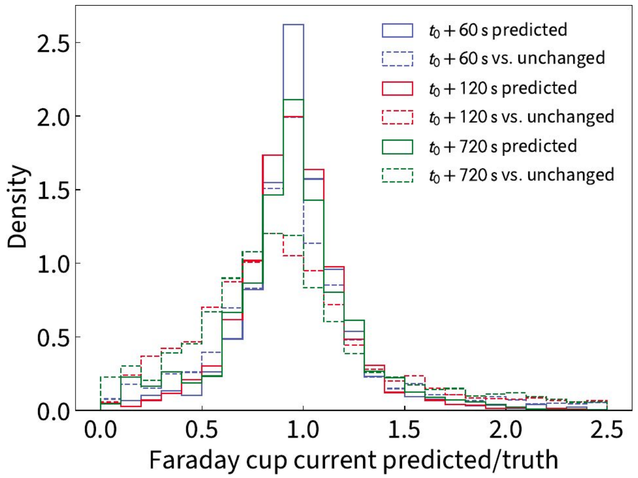
VENUS reinforcement learning updates

Reinforcement learning (RL), using rewards and/or penalties to train decision making, has now been used offline and online

- FY2024: offline RL (computer learning from previous data and then making VENUS tuning decisions) demonstrated
- FY2025: online RL demonstrated using a surrogate model of VENUS

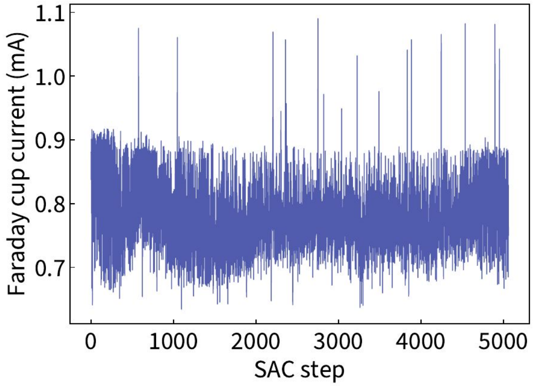
VENUS surrogate model:

- built using a recurrent neural network trained on previous VENUS optimization runs.
- based on current source state can predict the source state up to 10 minutes in the future



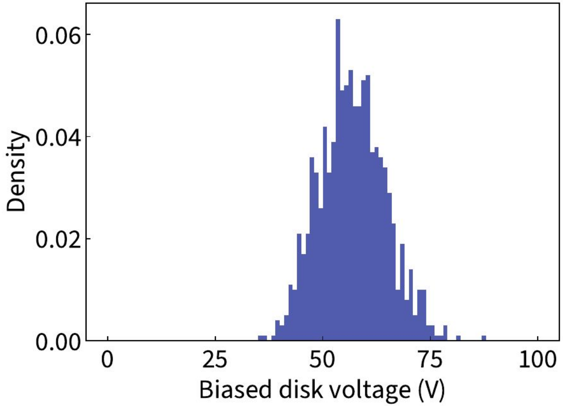
- Probability distribution of the ratio of predicted beam current to measured for 60, 120, and 720 seconds after initialization
- Solid curves are predictions, dashed curves assume no control system changes over time period

VENUS reinforcement learning updates



Difficulties of online reinforcement learning with surrogate model:

- system prone to over-estimating rewards when extrapolating into operation space outside model training range. This is indicated by the spikes in current



Offline reinforcement learning model of biased disk operation was better at setting voltage to a stable range operating live on VENUS.

Offline reinforcement learning, especially using CSDs, likely holds the most promise for VENUS optimization

VENUS Project Goals and Status

VENUS Project Goals and Status

WBS	Milestone	Description
1.1	Implement a monitoring code to predict/warn of instabilities	Based on training with recorded data, implement an online stability monitoring program for VENUS. ✅
1.2	Incorporate emittance scanning into VENUS optimization	Following upgrade of emittance scanner hardware incorporated into optimization as a separate parameter to optimize.
1.3	Implement ML-driven baking for VENUS	Implement an ML-based program for baking VENUS and benchmark performance against human and automated script.

BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY
Office of Science

Thank you

Work is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Contract No. DE-AC02-05CH11231.