Malachi Schram
Head of the Data Science Department
On behalf of the research from the JLab Data Science Department
Newport News, Virginia
June 13th, 2023

TJNAF is managed by Jefferson Science Associates for the US Department of Energy

Jefferson Lab
Thomas Jefferson National Accelerator Facility

Data Science at Jefferson Lab

Mission:

- Provide world-class data science solutions to advance research in nuclear physics by working with the subject matter experts at Jefferson Lab, partnering universities and Labs, and the Department of Energy.
- Provide world-class data science solutions to scientific applications relevant to the regional scientific community

Vision:

- Expand the capability and capacity of data science at JLab
- Create a collaborative data science research hub to:
  1. Work with regional partners on challenging scientific problems
  2. Champion education and research opportunities with regional universities and industry
  3. Reduce the carbon footprint by optimizing the data science workflow and algorithms

Current Portfolio

DOE Nuclear Physics:

- Quantom SciDAC (with ANL, VTech, ODU)
- Working with the experimental Halls (Tracking, etc.)
- Data Science contributing effort for AIEC (lead by EPSCI)

DOE Basic Energy Science:

- Machine Learning for Improving Accelerator and Target Performance (with ORNL)
- Collaborating with SLAC on application of ML-based controls for accelerators

DOE Advanced Scientific Computing Research:

- Data-Driven Decision Control for Complex Systems (with PNNL, ORNL, UC)

Non-DOE:

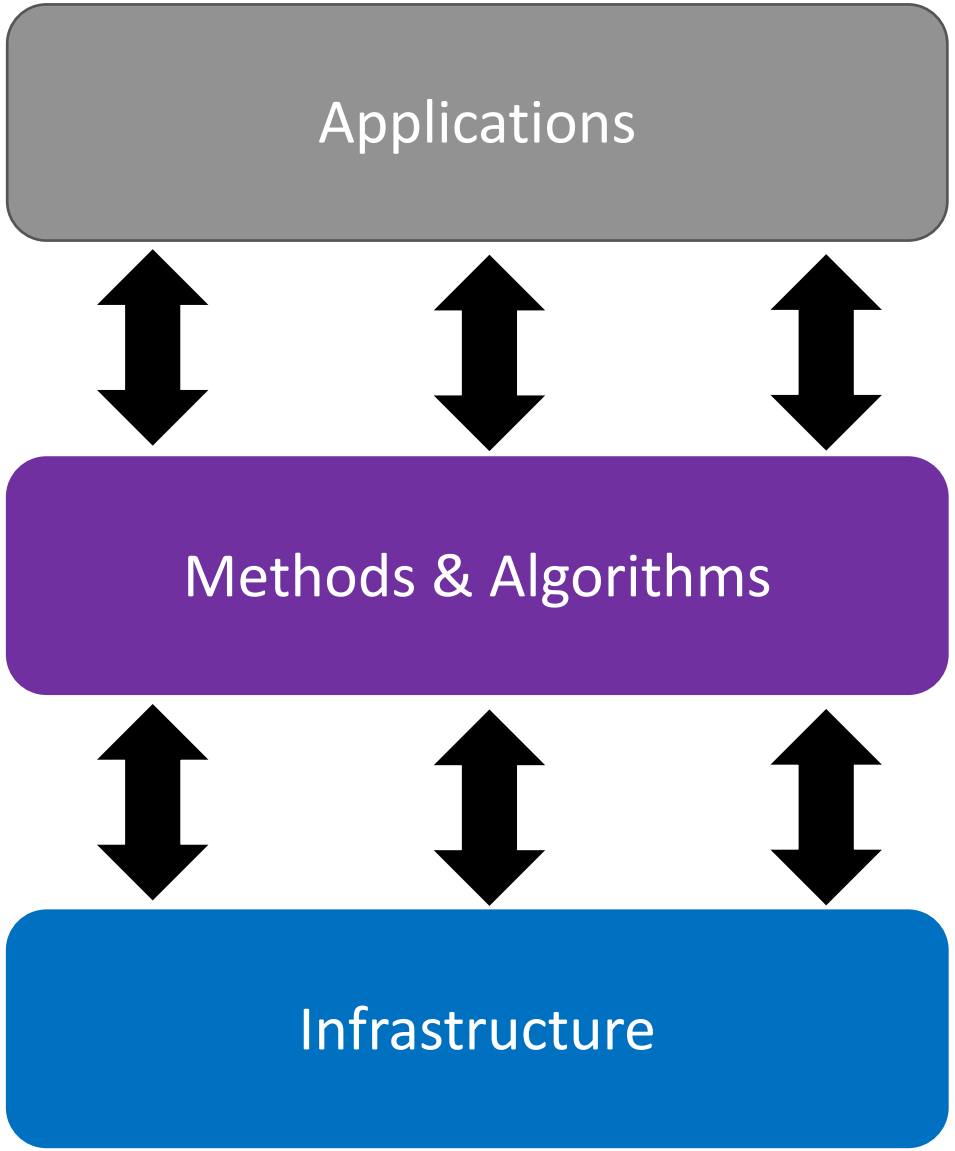
- Hampton Roads Digital Twin (with ODU)

Laboratory Directed Research & Development (LDRD):

- Multi-objective Optimization of Heat Load and Trip Rates in CEBAF (FY22)
- Adaptive Strategies for Optimal Computing Availability (FY23)

JLab Data Science Pillars

- **Applications**:
  - Nuclear Physics
  - Advanced Scientific Computing
  - Health & Climate
- **Focused Methods & Algorithms**:
  - Uncertainty Quantification
  - Interpretability and Explainability
  - Design & Control
- **Infrastructure:**
  - JLab ML & Data Hub
  - JLab Data Science software



DOE ASCR - BRN for SciML

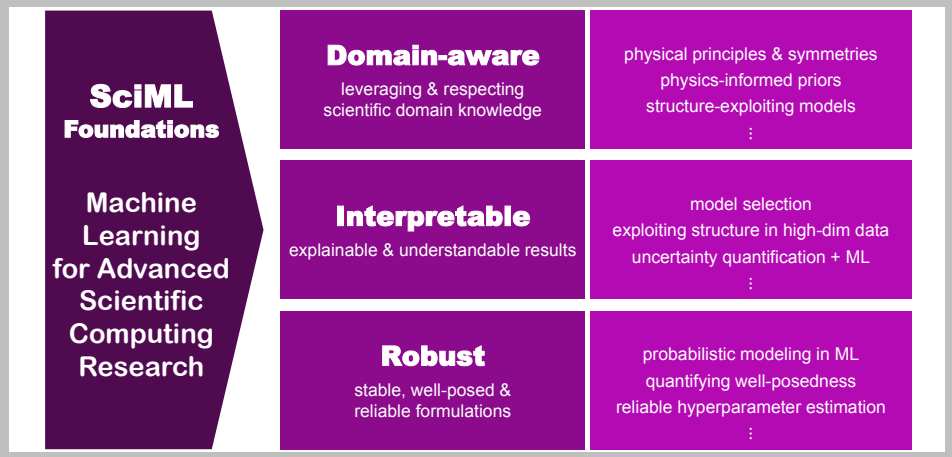


Figure 1: Foundational research themes of SciML must tackle the challenges of creating domain-aware, interpretable, and robust ML formulations, methods, and algorithms.

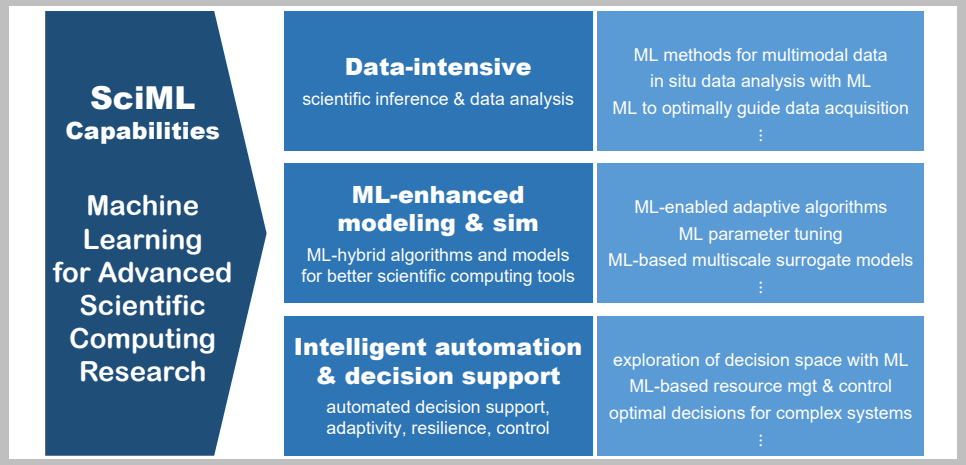


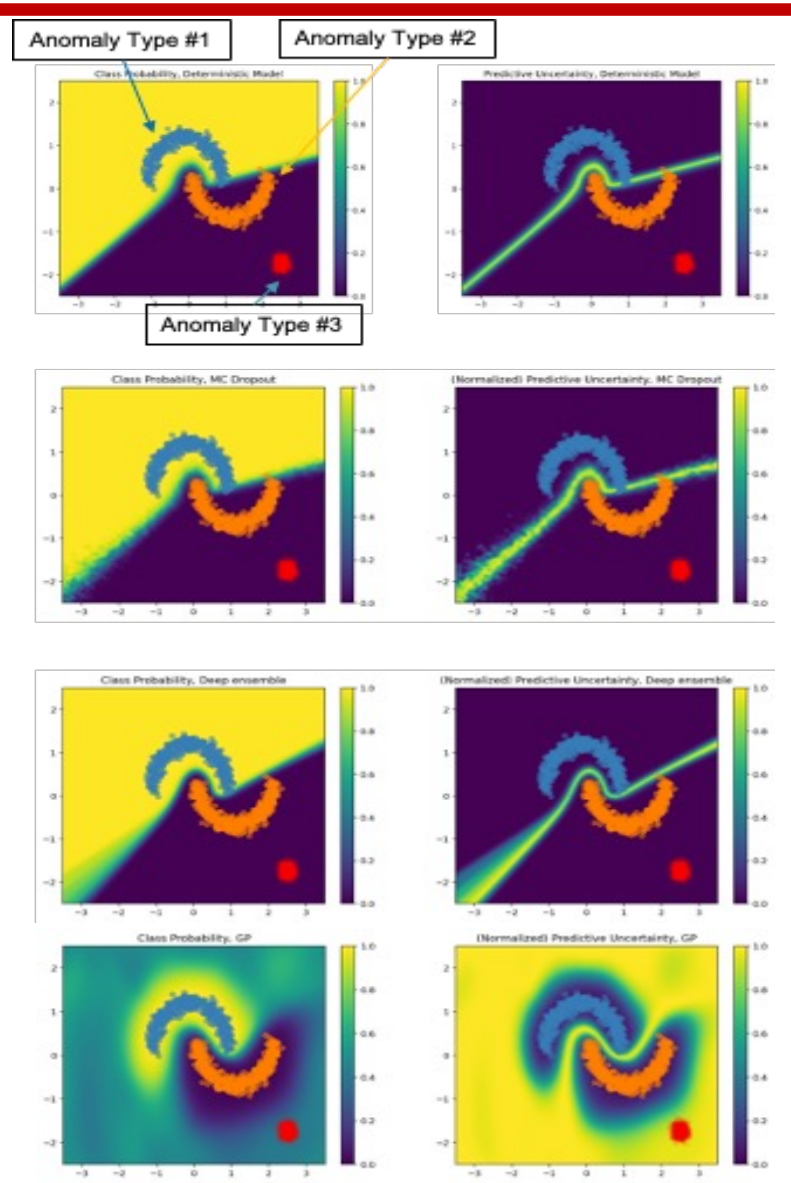
Figure 2: Opportunities for SciML impact arise in scientific inference and data analysis; in ML-enhanced modeling and simulation; in intelligent automation and decision support; and in related applications.

Jefferson Lab

Uncertainty Quantification for ML

Develop methods that include uncertainty estimates in machine learning models

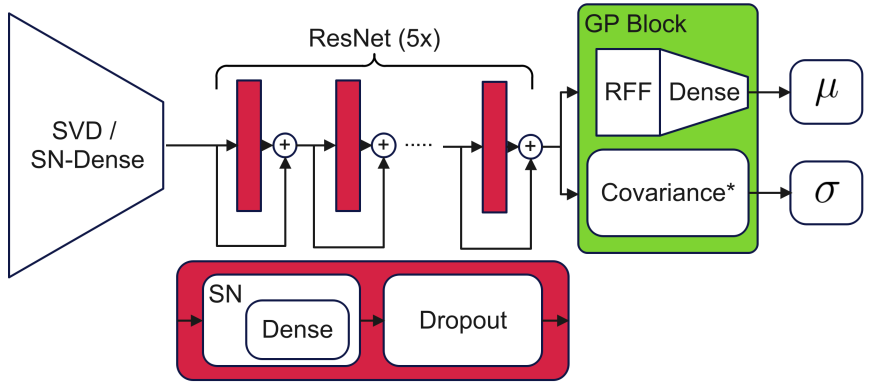
- Applications:
  - Data driven ML-based surrogate models
  - Real time controller
  - Anomaly detections
- Requirements:
  - Out-of-distribution uncertainties
  - Auto-calibration
  - Single inference
- Hardware considerations:
  - Memory
  - Inference time
  - Performance trade-off due to approximations

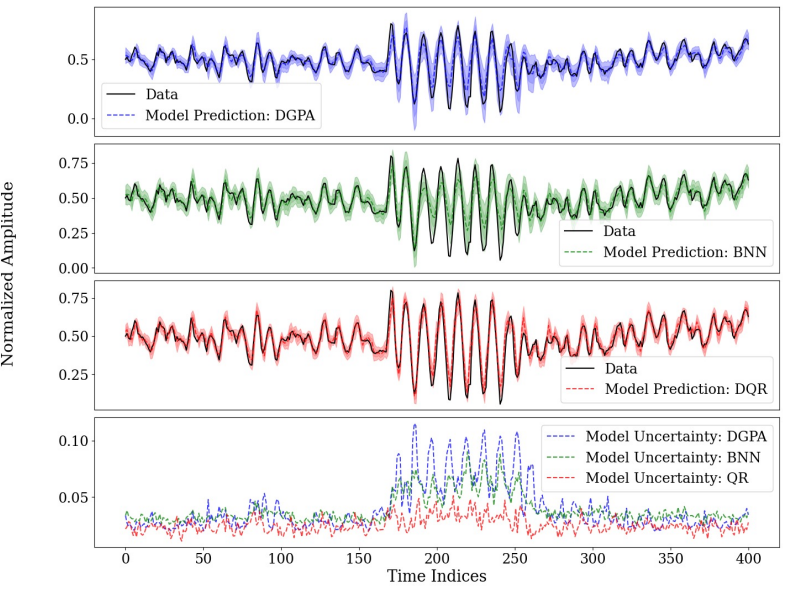


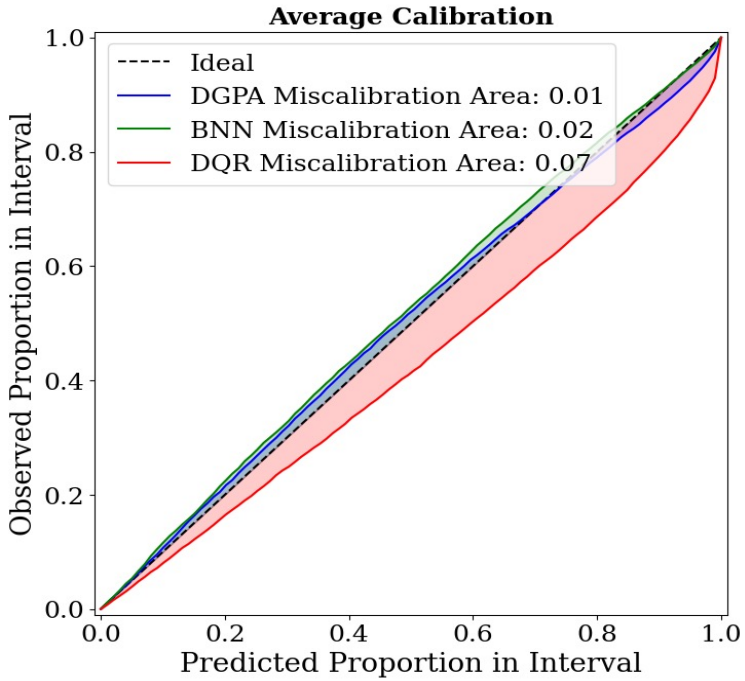
Jefferson Lab

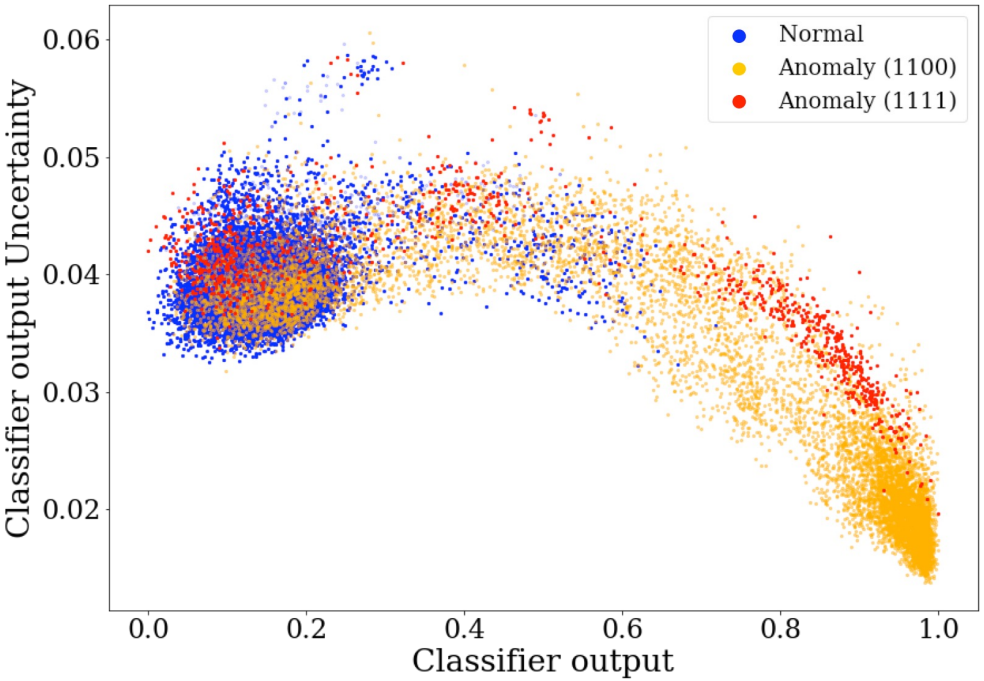
Uncertainty Quantification For Reliable AI/ML

- **Goal:** Develop and integrate new UQ methods for reliable AI/ML
- **Problem space:**
  - Applications with high-dimensional continuous input features
  - Focused on large data sets for DOE applications
  - Safety constraints that should never or at least rarely be violated.
  - Inference that must happen in real-time at the control frequency of the system.
- **Applications:**
  - Anomaly and prognostication
  - ML-based controls



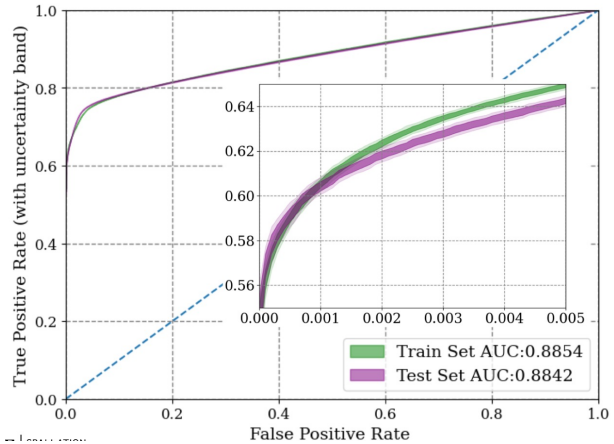


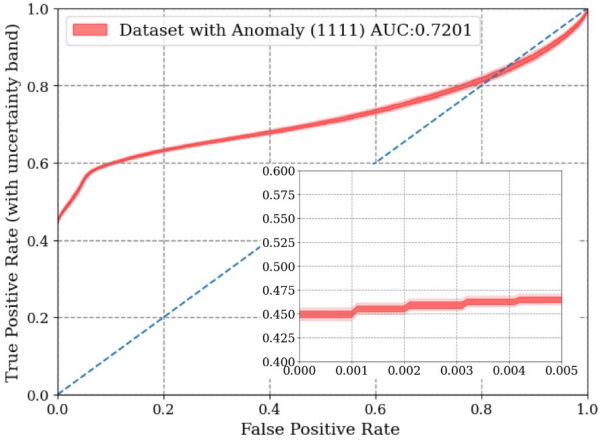


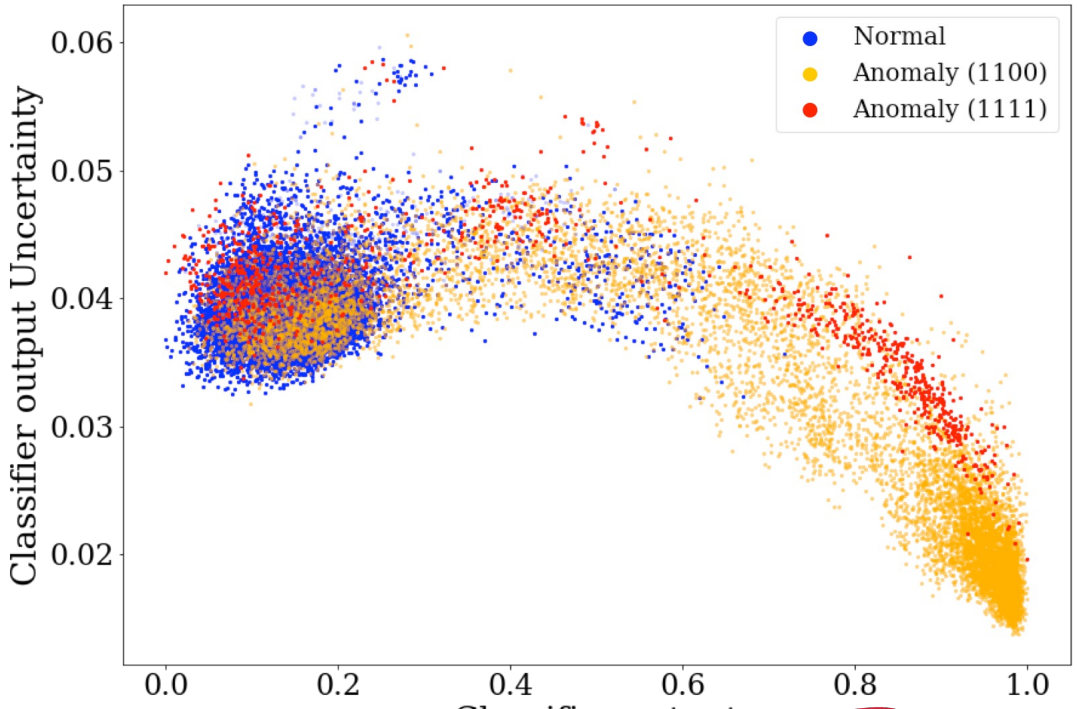


Uncertainty Aware Siamese Model ("Classification")

- We enhanced our models by adding GP approximation layer which provides the uncertainty estimate
- Results from similarity model showed a ~4x improvement in performance over previously published results, it is also much better than a vanilla Auto-encoder
- The ROC curves show true fault detection rate above 60% while keeping the false alarms below 0.5% (not optimized)
- We introduced an out-of-domain anomaly, labelled 1111 (red), the UQ-based model performed similar in classifying the anomalies and indicated high uncertainty (as expected)





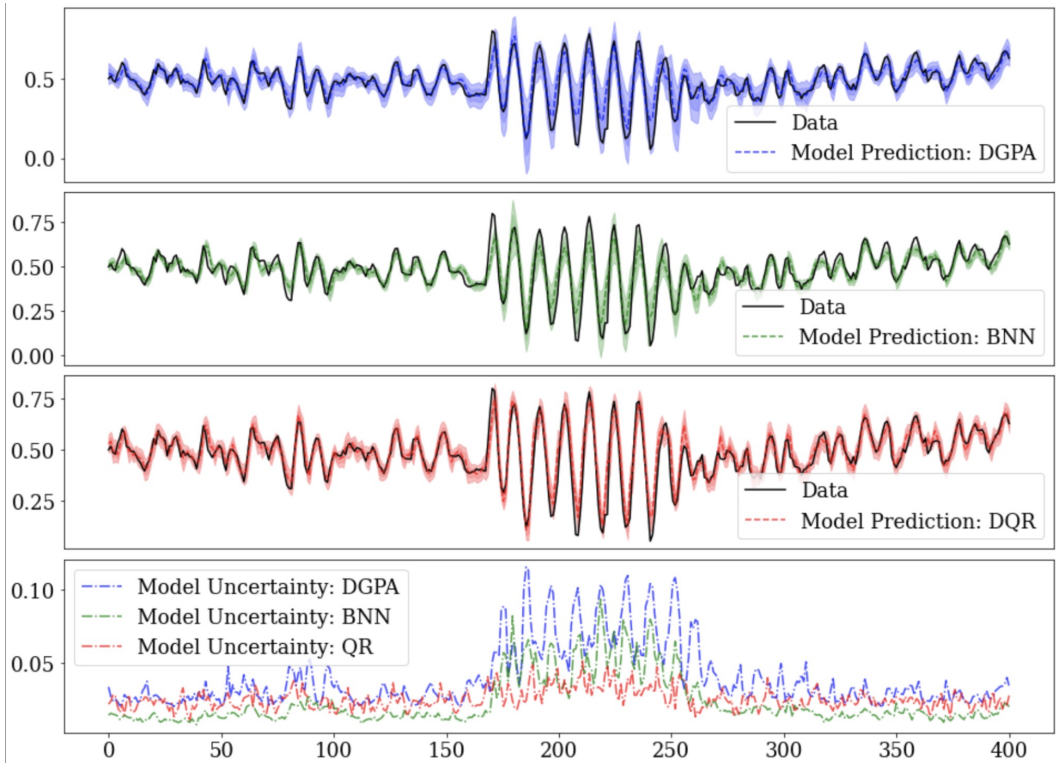


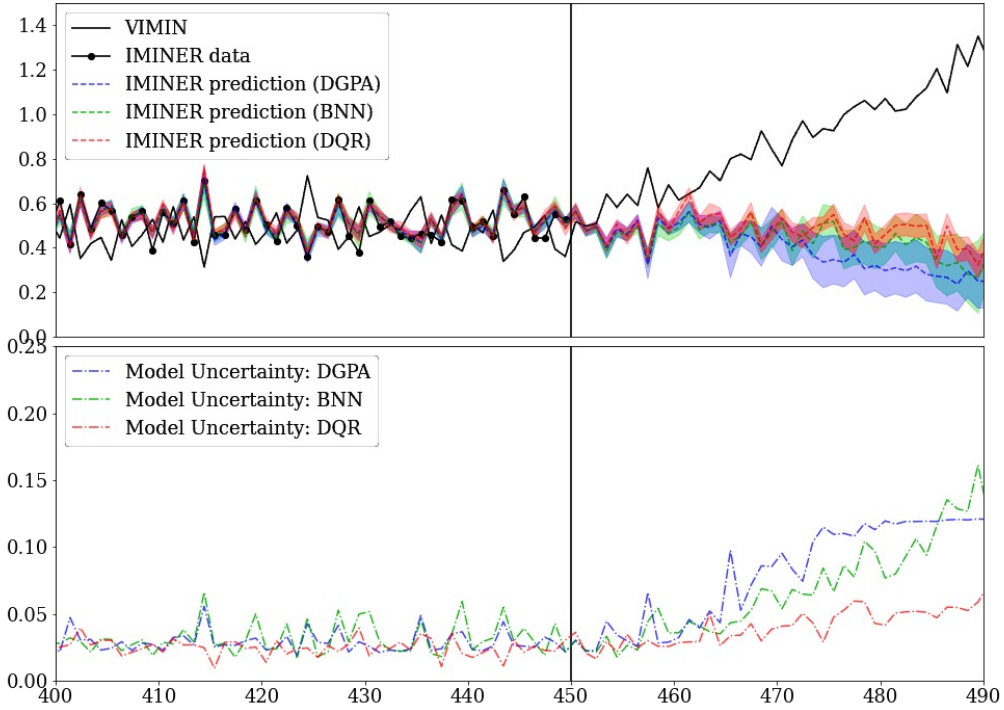
W. Blokland, K. Rajput, M. Schram, T. Jeske, *et al* 2022 Phys. Rev. Accel. Beams 25, 122802

Jefferson Lab

Data Driven UQ ML-based Surrogate Models (Regression)

- Compare different techniques: DQR, BNN, DNGPA
- DQR models have great performance for training distribution but not for OOD
- BNN models do a better job to estimate OOD
- DGPA models are distance aware by design resulting in better OOD estimation





M. Schram, K. Rajput, *et al* 2023 Phys. Rev. Accel. Beams 26, 044602

Jefferson Lab

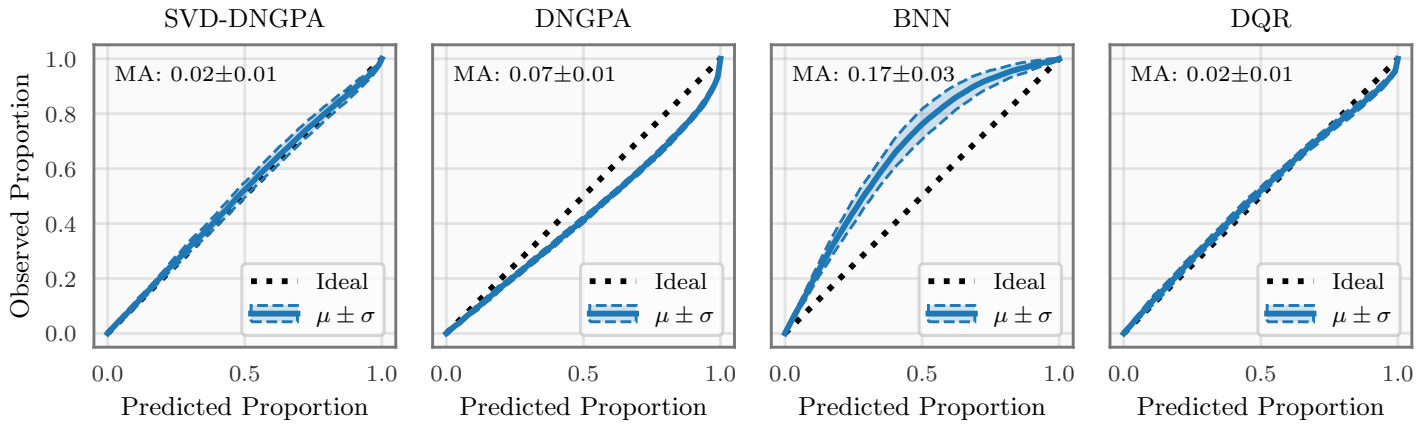
Predicting HVCM Capacitance with Uncertainty Quantified ML

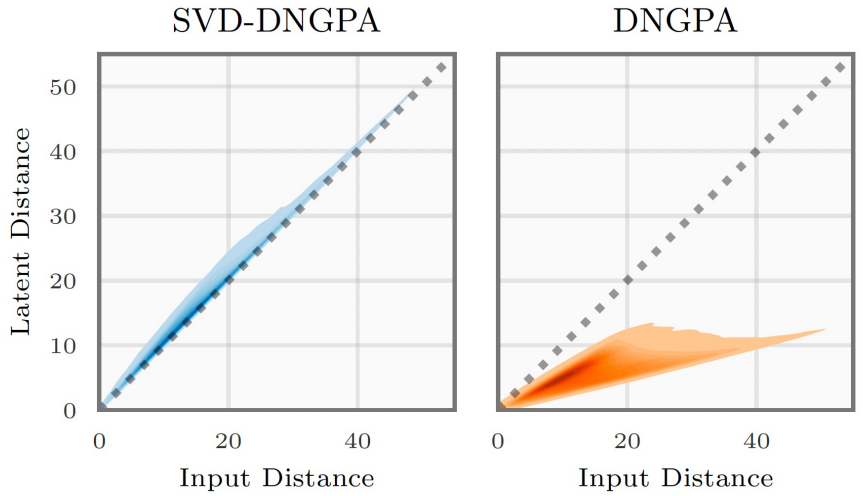
- **Goal:**
  - Develop a UQ model to predict the capacitance of 3 capacitors in the SNS HVCM system using currently available sensor data
- **Techniques:**
  - Singular Value Decomposition, Residual Networks, Distance Preservation and bi-Lipschitz Constraints, Gaussian Process Approximation
- **Results:**
  - Achieves <1% error with accurate UQ for in-distribution data
  - Provides significantly better OOD performance than other methods
  - Submitted to Machine Learning with Applications

Distance Preserving Machine Learning for Uncertainty Aware Accelerator Capacitance Predictions

Steven Goldenberg[a,*], Malachi Schram[a], Kishansingh Rajput[a], Thomas Britton[a], Chris Pappas[b], Dan Lu[b], Jared Walden[b], Majdi I. Radaideh[c], Sarah Cousineau[b], Sudarshan Harave[b]

[a] *Thomas Jefferson National Accelerator Facility Newport News VA 23606 USA*
[b] *Oak Ridge National Laboratory Oak Ridge TN 37830 USA*
[c] *Department of Nuclear Engineering and Radiological Sciences, The University of Michigan, Ann Arbor, Michigan 48109, USA*





Jefferson Lab

MULTI-MODULE CVAE TO PREDICT HVCM FAULTS IN THE SNS

- **Goal**: Predict an upcoming machine failure before it occurs to improve the reliability of the HVCMs and reduce the down time for the SNS facility
- **How:** We use pulses leading to failure to allow for future forecasting
- **Method:** Condition the Variational Autoencoder (VAE) model on the module unique identifier to learn the association between waveforms and their modules
  - By using all 15 modules, we eliminate the need to train a Single-module and increase the number of statistics
- **Paper**: Accepted for publications at Machine Learning with Applications journal: https://arxiv.org/pdf/2304.10639.pdf

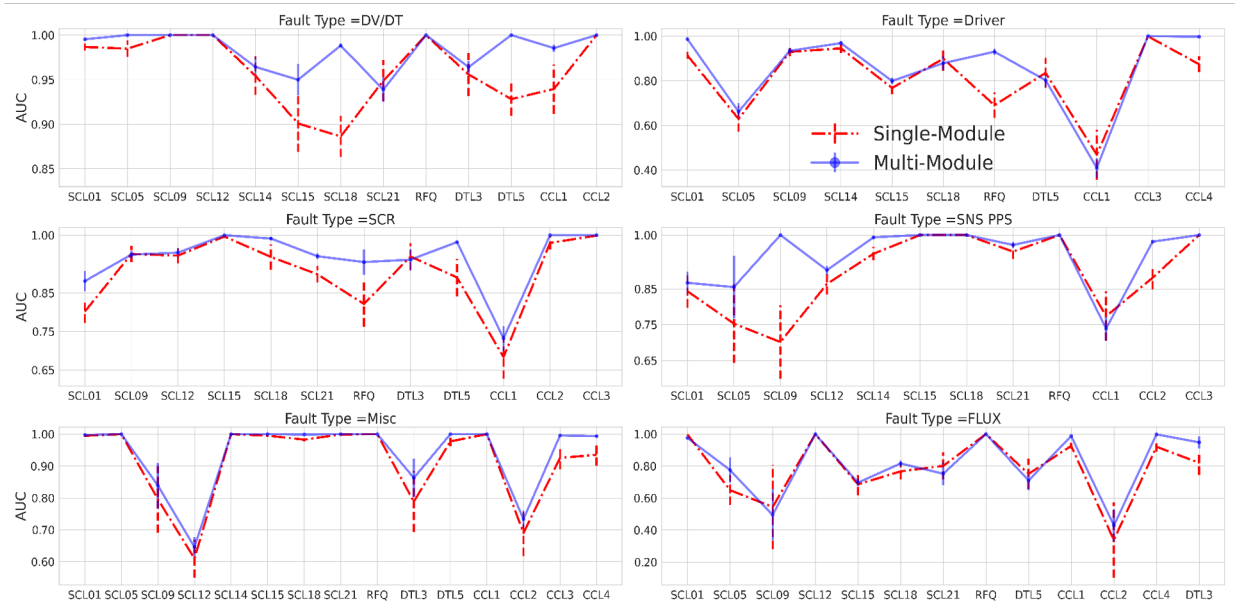


Figure 2: Compare the AUC values between single-module and multi-module using six types of faults across several modules. The error bar is ± 1 Standard Deviation (SD) error generated by sampling the latent Z of each Method.

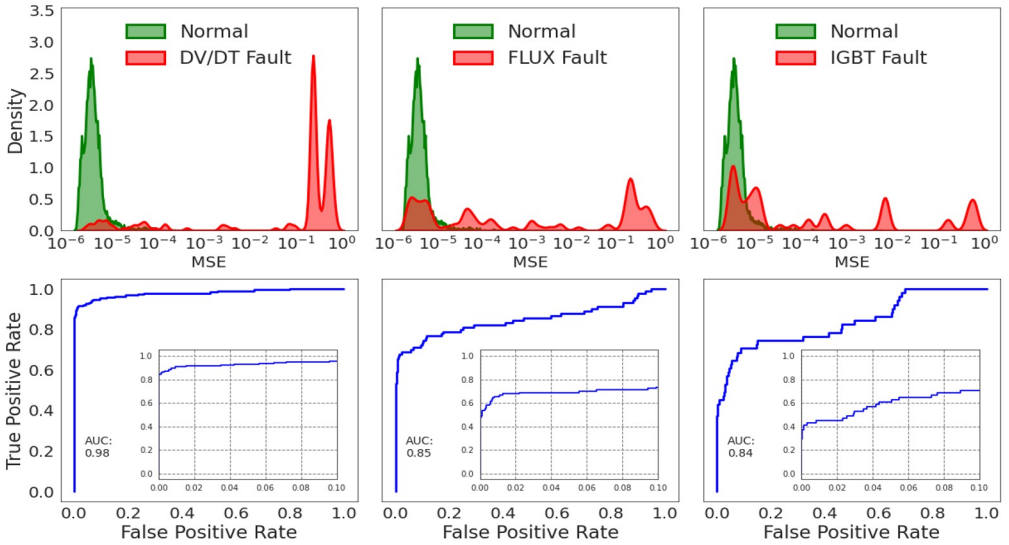


Figure 1: KDE distributions of the MSE from reconstructing normal (green) and faulty waveforms (red) for three fault types, with the corresponding ROC curve for each fault.

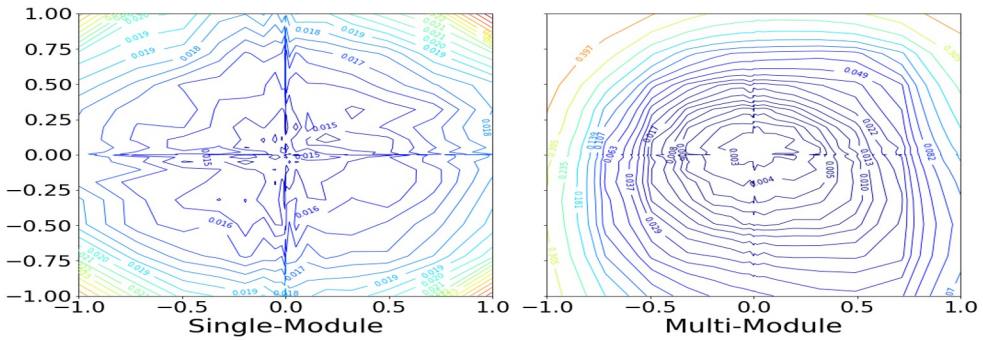


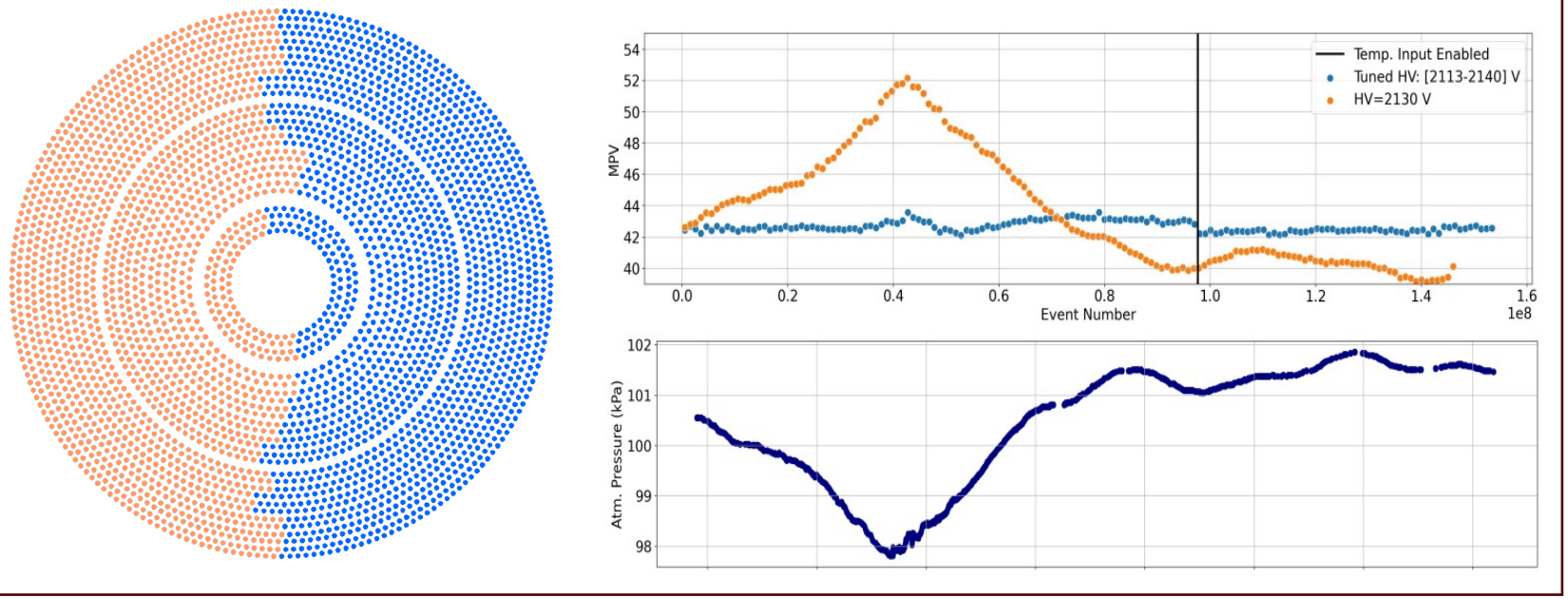
Figure 3: Loss surface, where x- and y-axis are two random directions in weights space generated using filter normalization method. Multi-Module shows convex-like surface, while Single-module hs chaotic behaviour.

Jefferson Lab

Improving System Controls Through AI/ML

- **Goal:** Apply UQ based AI/ML methods to improve system controls
- **Applications:** Detector and Accelerator controls
- **Techniques:** Bayesian Optimization, Genetic Algorithms, Model Predictive Control, Reinforcement Learning

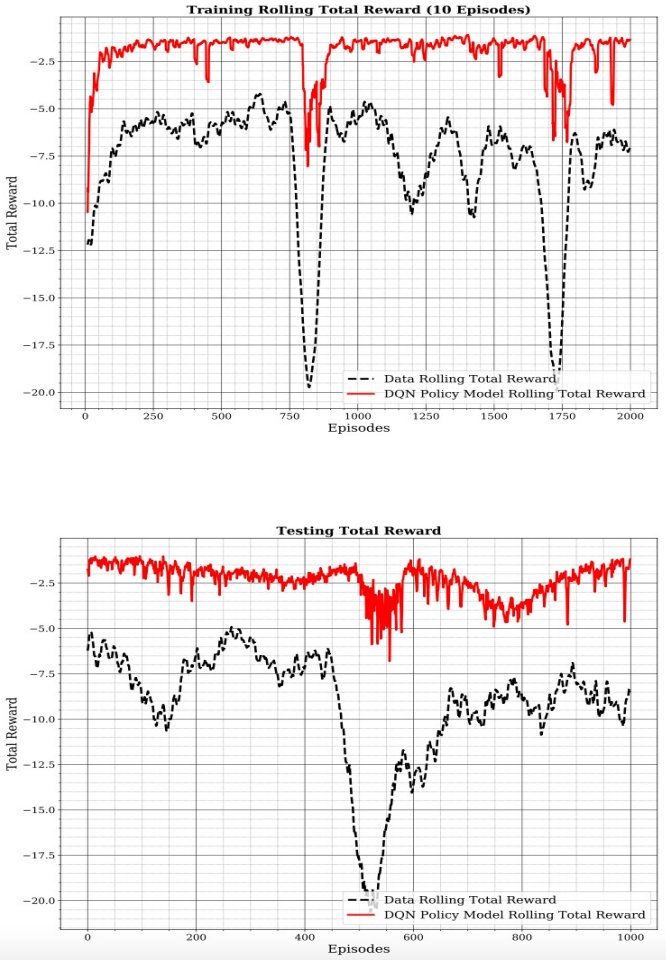
JLab Hall D CDC Calibration (lead by EPSCI)



T. Jeske *et al* 2022 *JINST* **17** C03043

T. Jeske *et al* 2023 *J. Phys.: Conf. Ser.* **2438** 012132

Applied RL for FNAL Booster

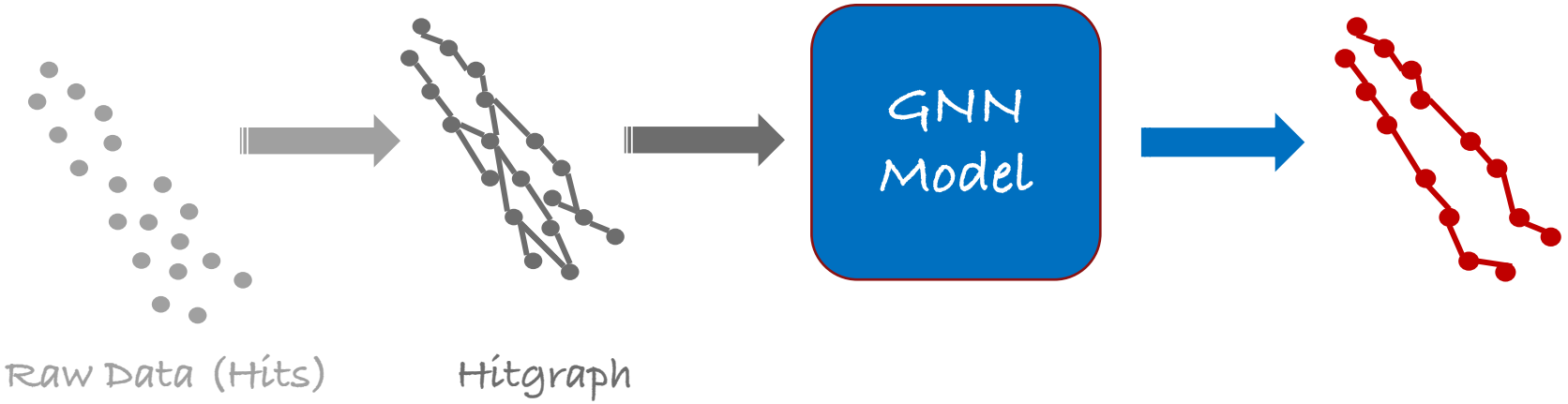


M. Schram *et al* 2022 Phys. Rev. Accel. Beams 24, 104601

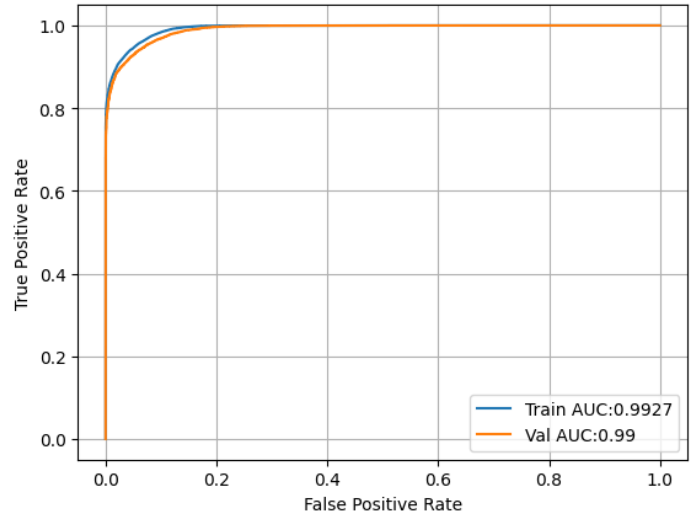
GRAPH NEURAL NETWORKS FOR TRACKING

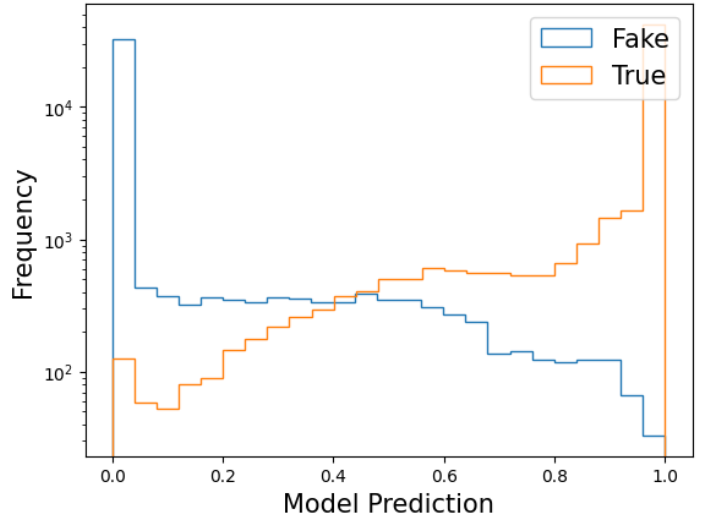
Initial Results

- **Goal:** Improve charged particle tracking with the help of Graph Neural Networks (GNNs) and potentially run in near real time on edge computing
- **Application:** Charged Particle Tracking at GlueX, CLAS, EIC, and many more…
- **Techniques:** Graph Analytics, GNNs, GNNs on FPGA, Physics Informed ML



Predicting True edges belonging to tracks with GNNs

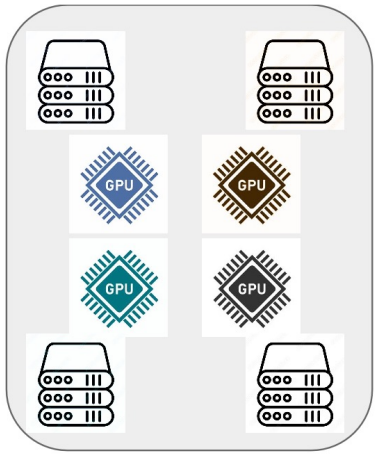


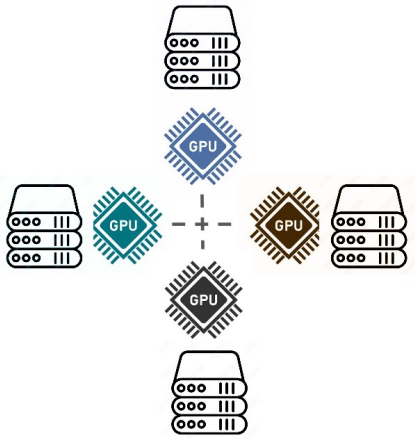


Jefferson Lab

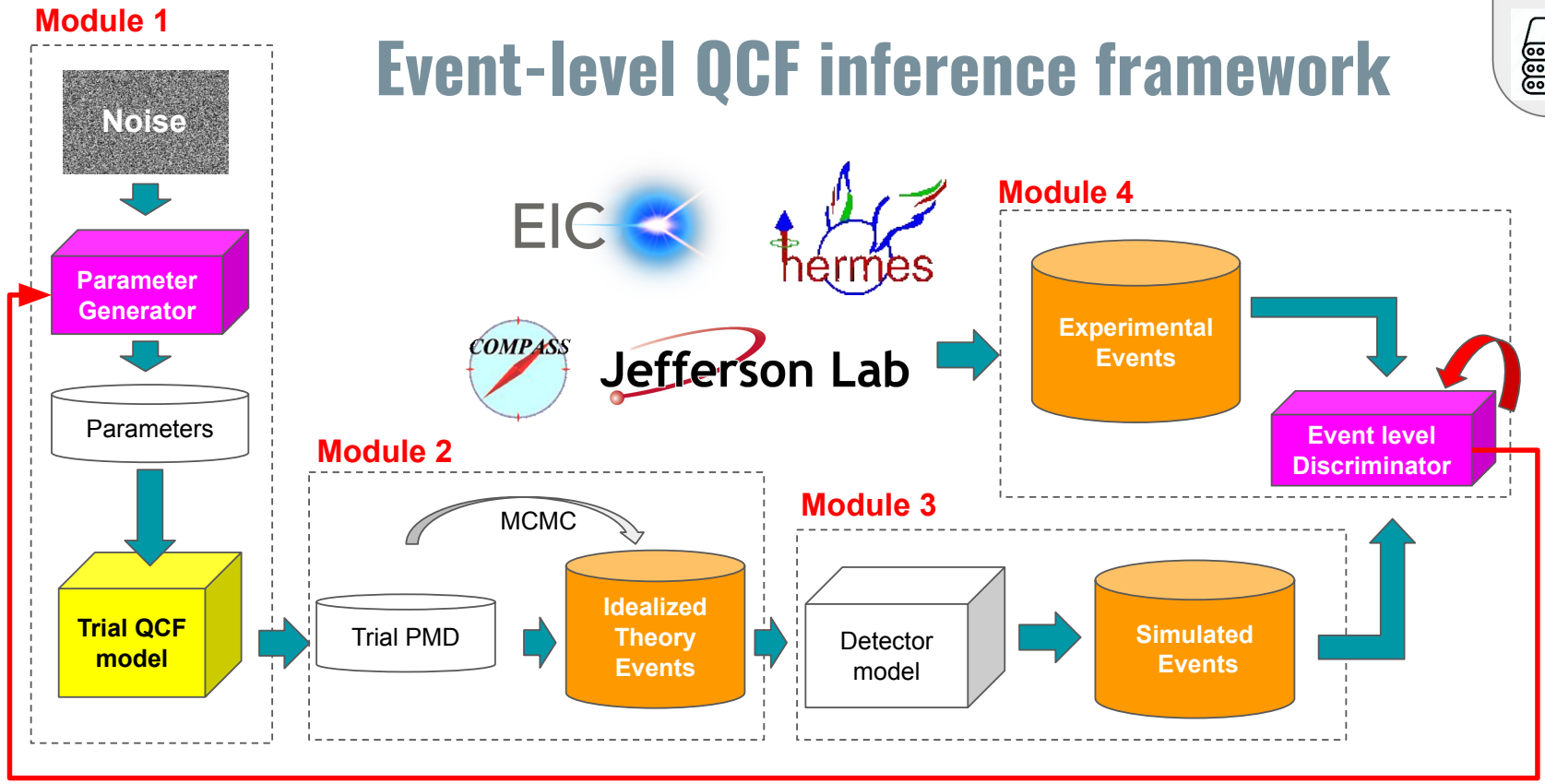
FEMTOSCALE IMAGING OF NUCLEI USING EXASCALE PLATFORMS

- The goal is to extract a quark and gluon tomography of nuclei and answer important questions on the nature of visible matter at the femtoscale
- Develop modular components to dynamically compose workflows
- Multiple AI/ML components that need to scale to LCFs





Event-level QCF inference framework



TensorFlow

PyTorch

HOROVOD

Jefferson Lab

FRAMEWORK

- Developed a common and modular framework
- Includes:
  - Core base classes
  - ProxyApp, GPDs Theory
  - Experimental (filter, detector, etc.)
  - GAN workflows

cfg	Updated visualization tools
core	Implemented visulation tools and helper scripts
demos	Implement requested changes and information capture
discriminator_module	fix useBias bug in disc
eventselection_module	Implemented visulation tools and helper scripts
expdata_module	Update with code from pub_demo that appeared to be missing.
experimental_module	Add first readme version for the experimental module
generator_module	configurable-N events per parameter set
sample_data	Add sample data for test use cases.
theory_module	Updated visualization tools
utils	Updated visualization tools
workflow	Updated visualization tools

master · 36 branches · 0 tags · Go to file · Add file · Code

schr476 Include MacOS env builder (testing). 8e78e1f 2 weeks ago 111 commits

tomography_toolkit_dev	Updated visualization tools	last month
utests	Implement requested changes and information capture	last month
.gitignore	add scripts dir to ignore	last month
HOWTO_CONTRIBUTE.md	Write documentation file: HOWTO_CONTRIBUTE.md	last month
README.md	Merge branch 'master' into 80-update-conda-requirements-file	last month
env-metal-arm64.yaml	Include MacOS env builder (testing).	2 weeks ago
env.yaml	Update the conda setup file and the requirements.txt	last month
requirements.txt	Update the conda setup file and the requirements.txt	last month
setup.py	Implemented visulation tools and helper scripts	last month

README.md

SciDAC Quantom Collab Dev [version 0.1]

Dev repository for SciDAC Quantom project

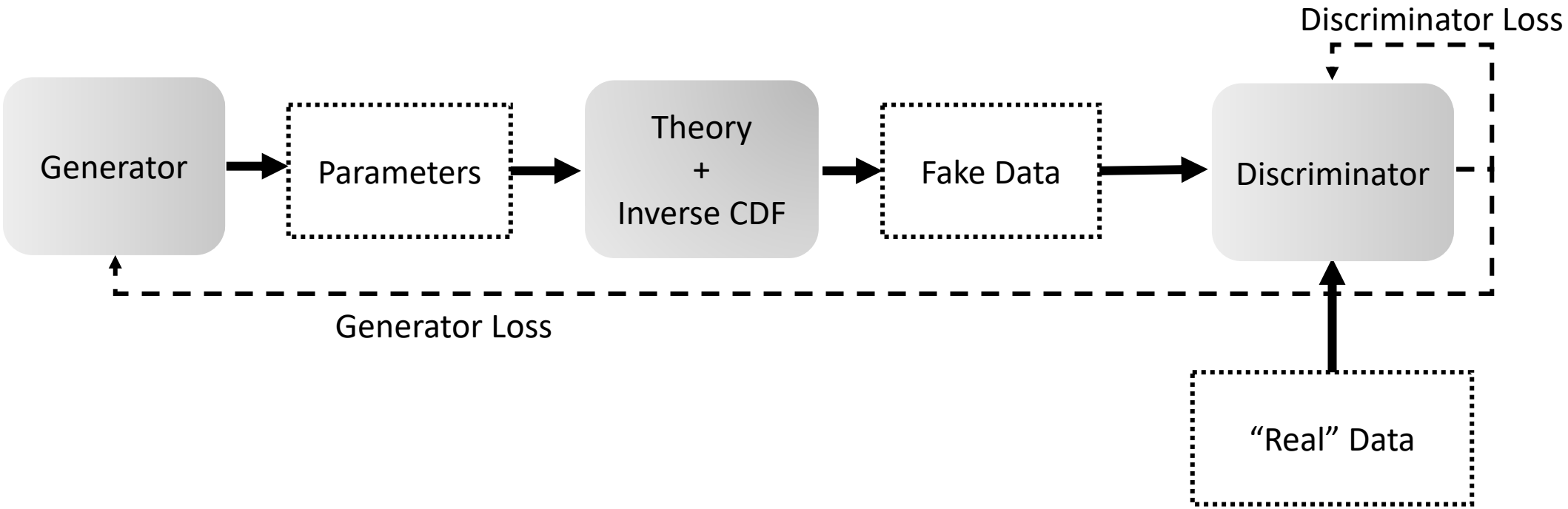
Code contribution proceedure

1. Create an issue or access assigned issue
2. Create a new branch using the issue id
3. When your code is complete and updated in the branch then make a pull request

More details are found here: HOWTO_CONTRIBUTE.md

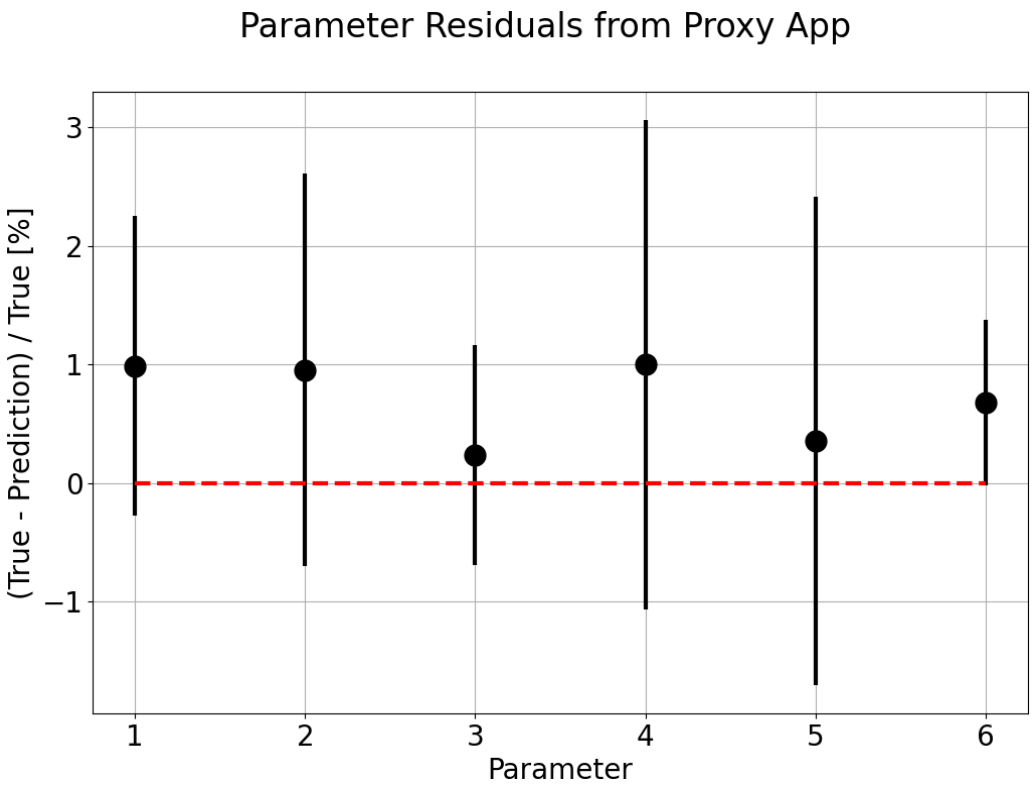
PROXY APP (NOT REAL PHYSICS)

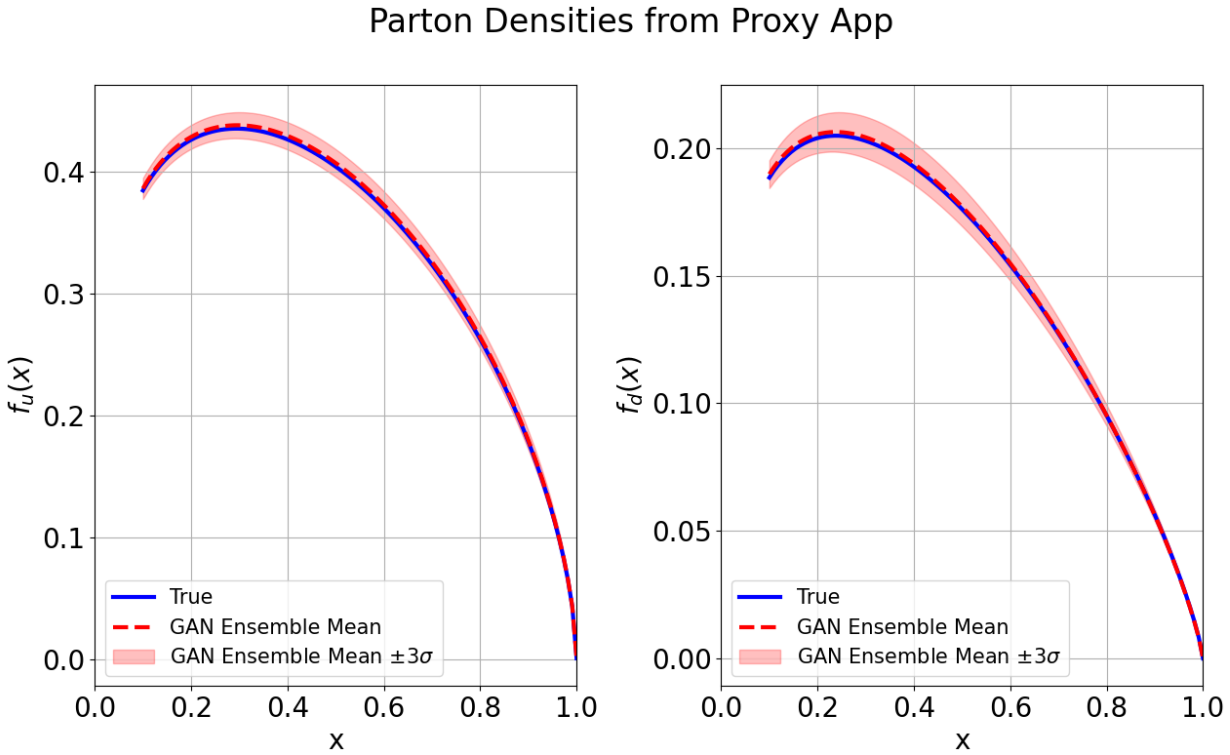
- Only considering u and d contribution and has not physical equivalent



CLOSURE TEST

- Use the Proxy App for closure tests and scaling
- We generate toy data (1M events) using fix parameters in the theory module
- We train an ensemble of 15 GAN workflows
- Results are from ideal setup

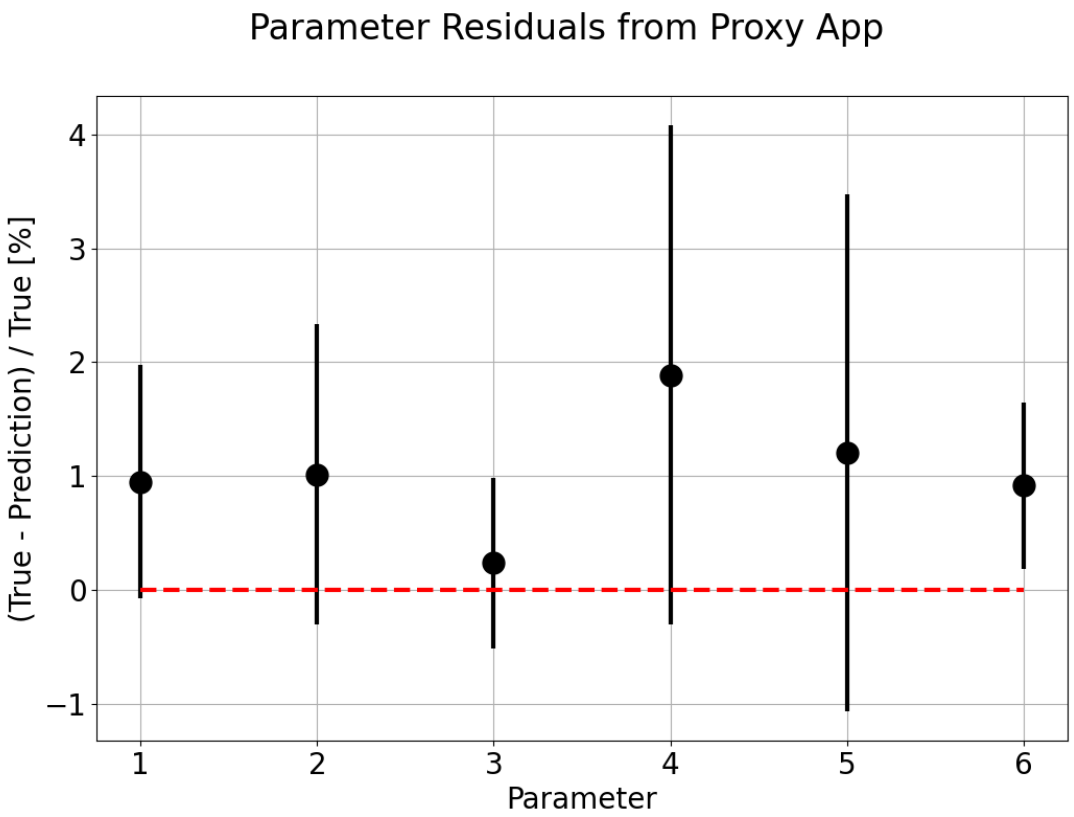


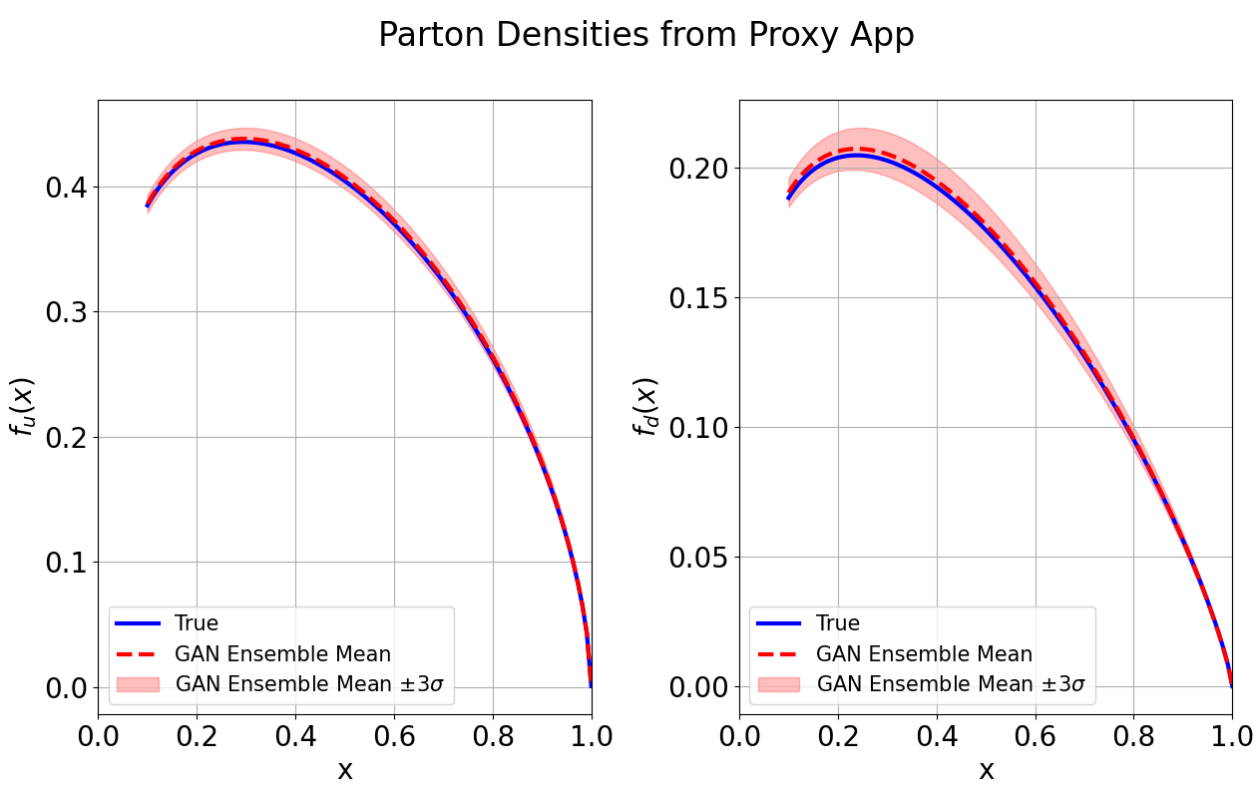


Jefferson Lab

ADDITIONAL PROXY APP RESULTS

- Including some detector effects:
  - Detector with 5% / 12% resolution effect on sigma1 / sigma2
  - Added correlations between sigma1 / sigma2

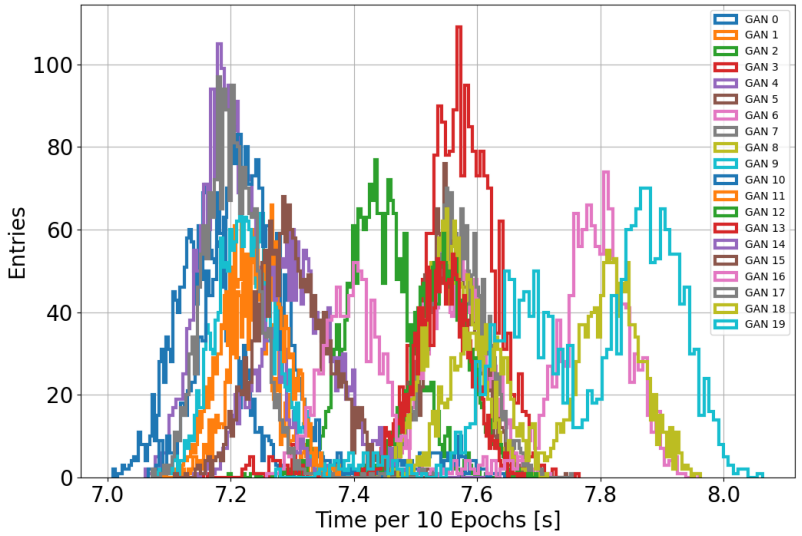


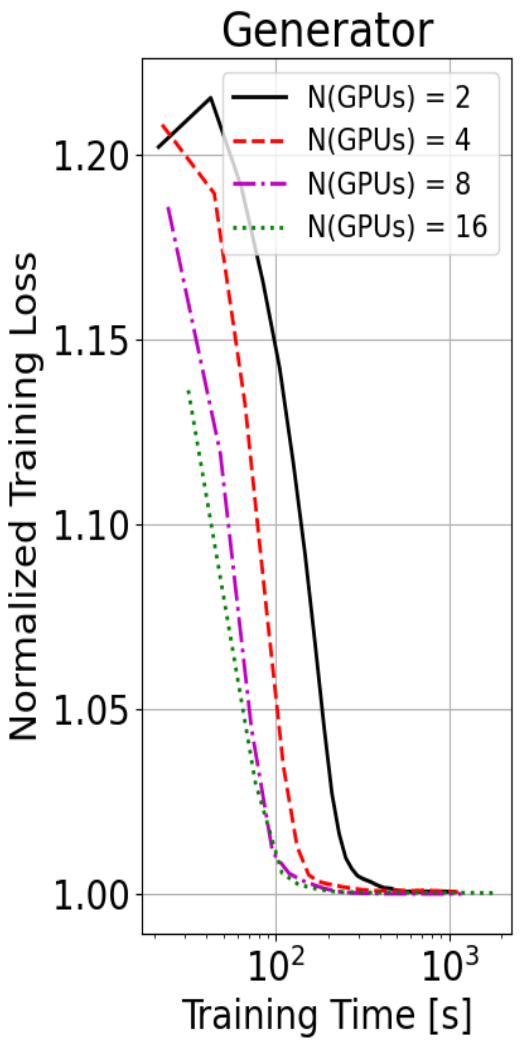


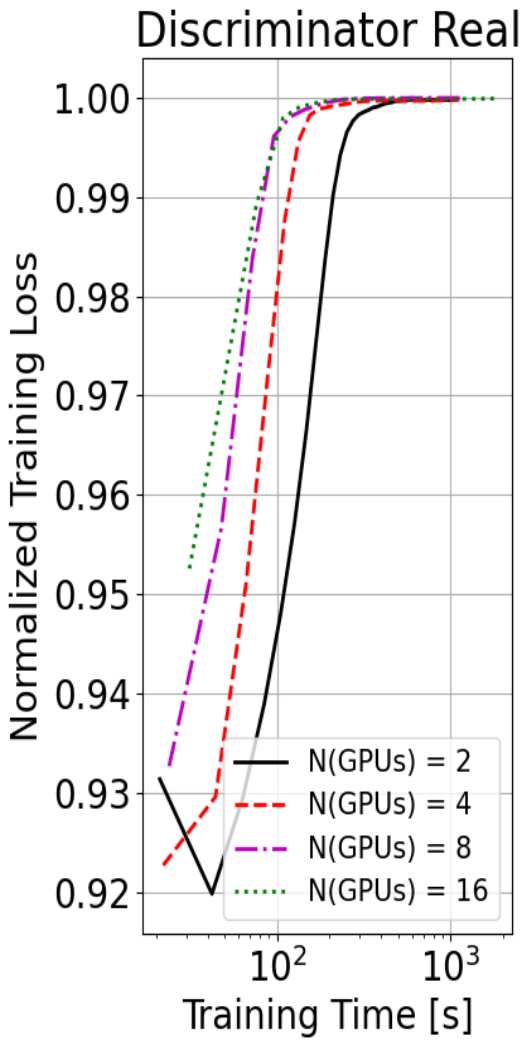
Jefferson Lab

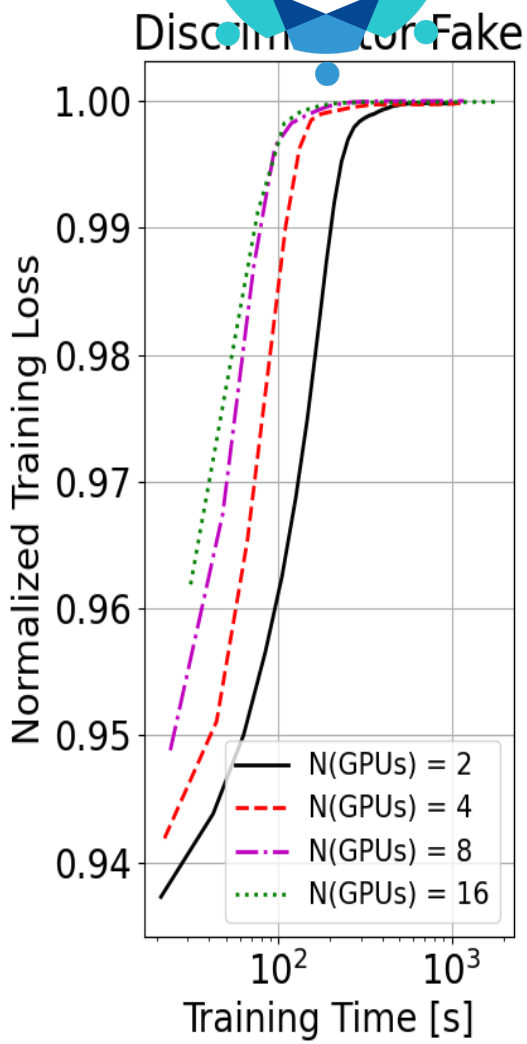
SCALING USING COMMON TOOLS

- Initial scaling studies where based on Horovod and Fairscale
- Both didn't scale very well
- The stochastic nature of the workflow doesn't work well with allreduce techniques



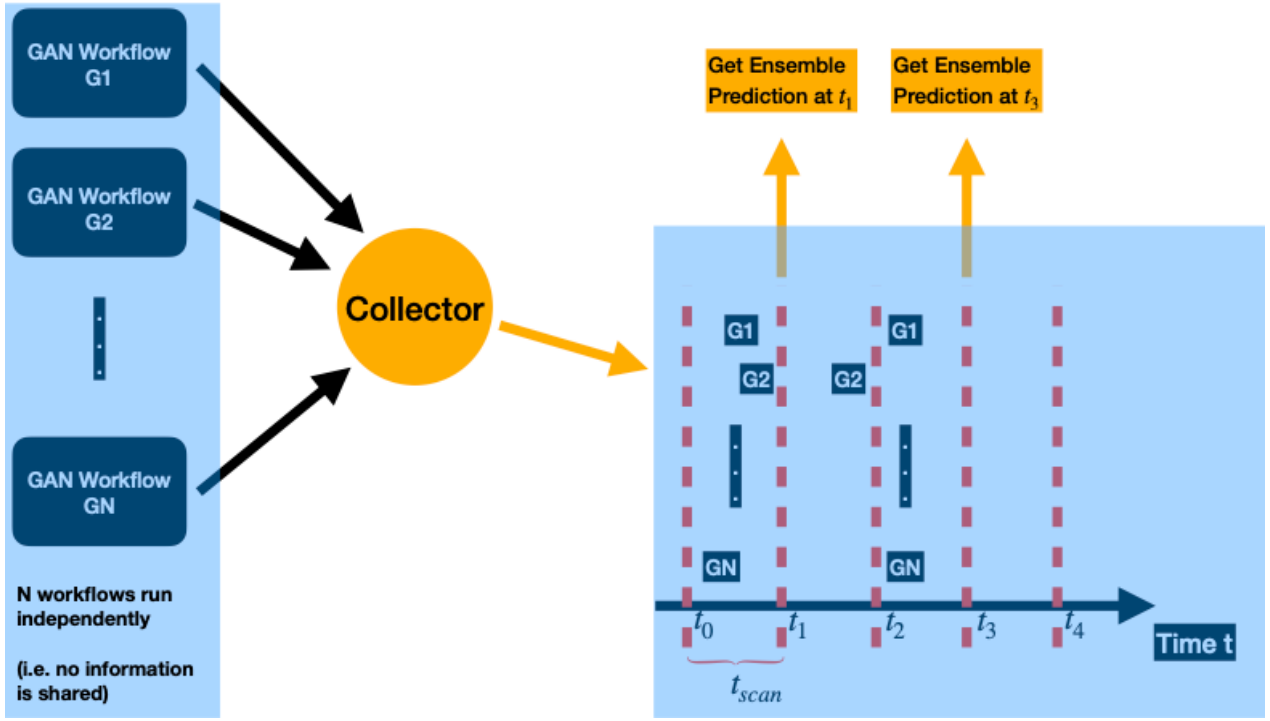


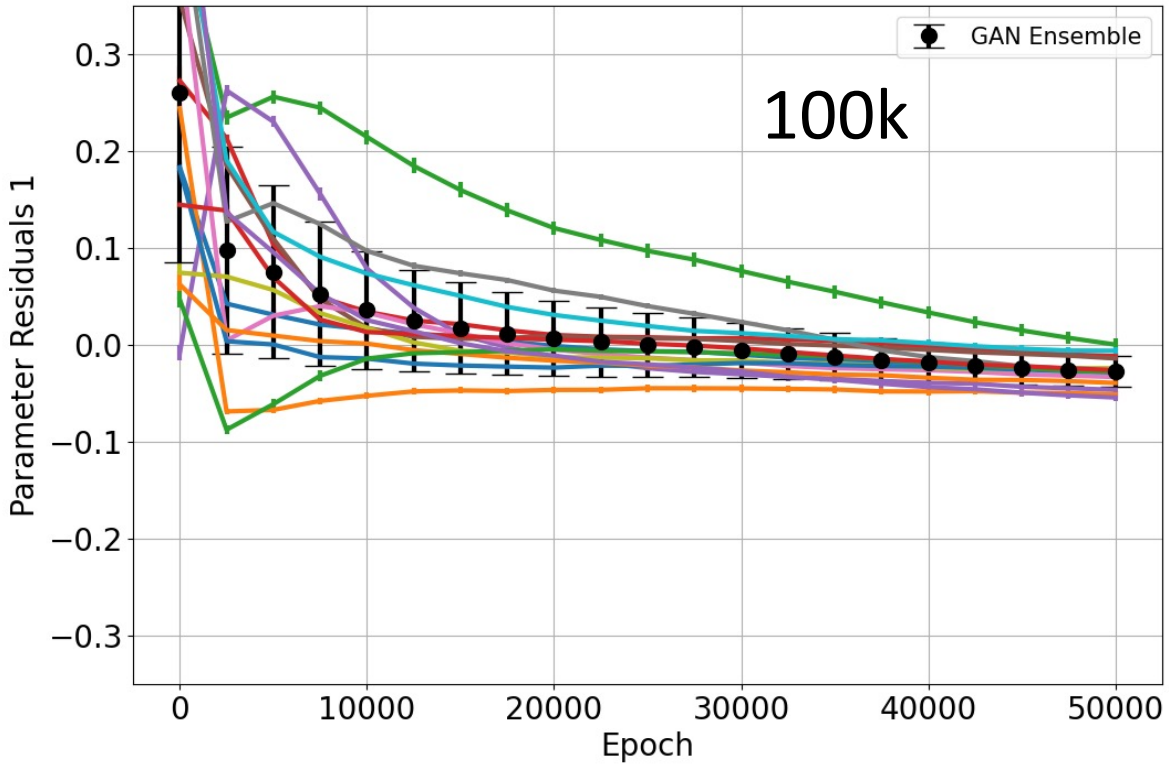




EXPLORING SCALING USING ENSEMBLES

- Simplest approach is to have completely decoupled learning and take aggregate predictions at fixed time intervals
- This allows us to study the model stability and accelerate convergence

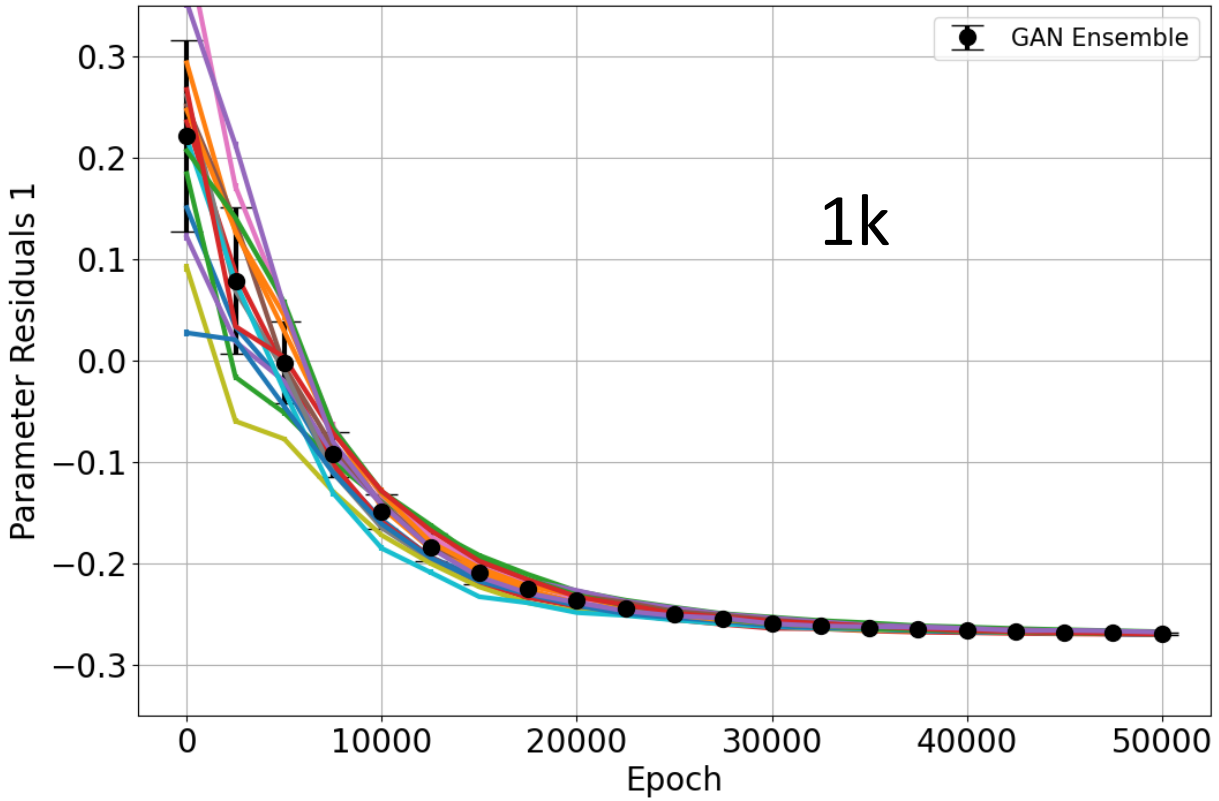


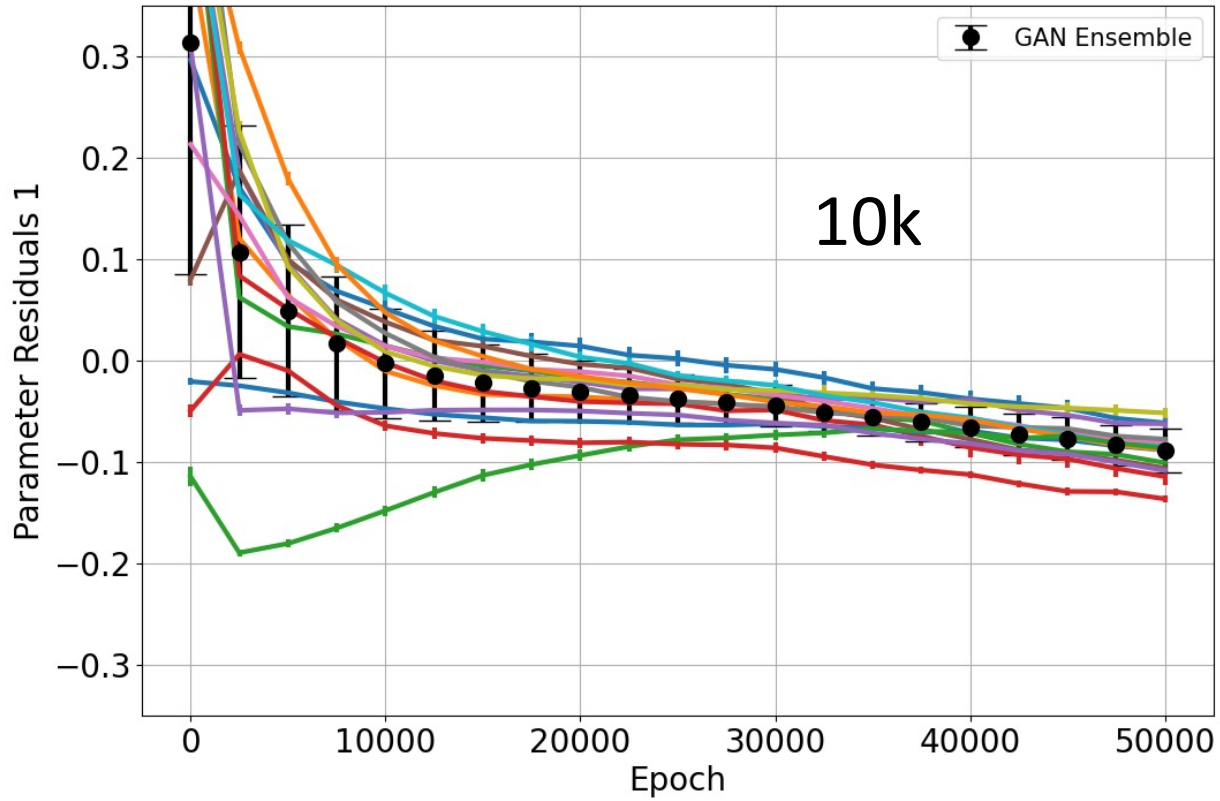


Jefferson Lab

IMPACT FROM SAMPLE STATISTICS

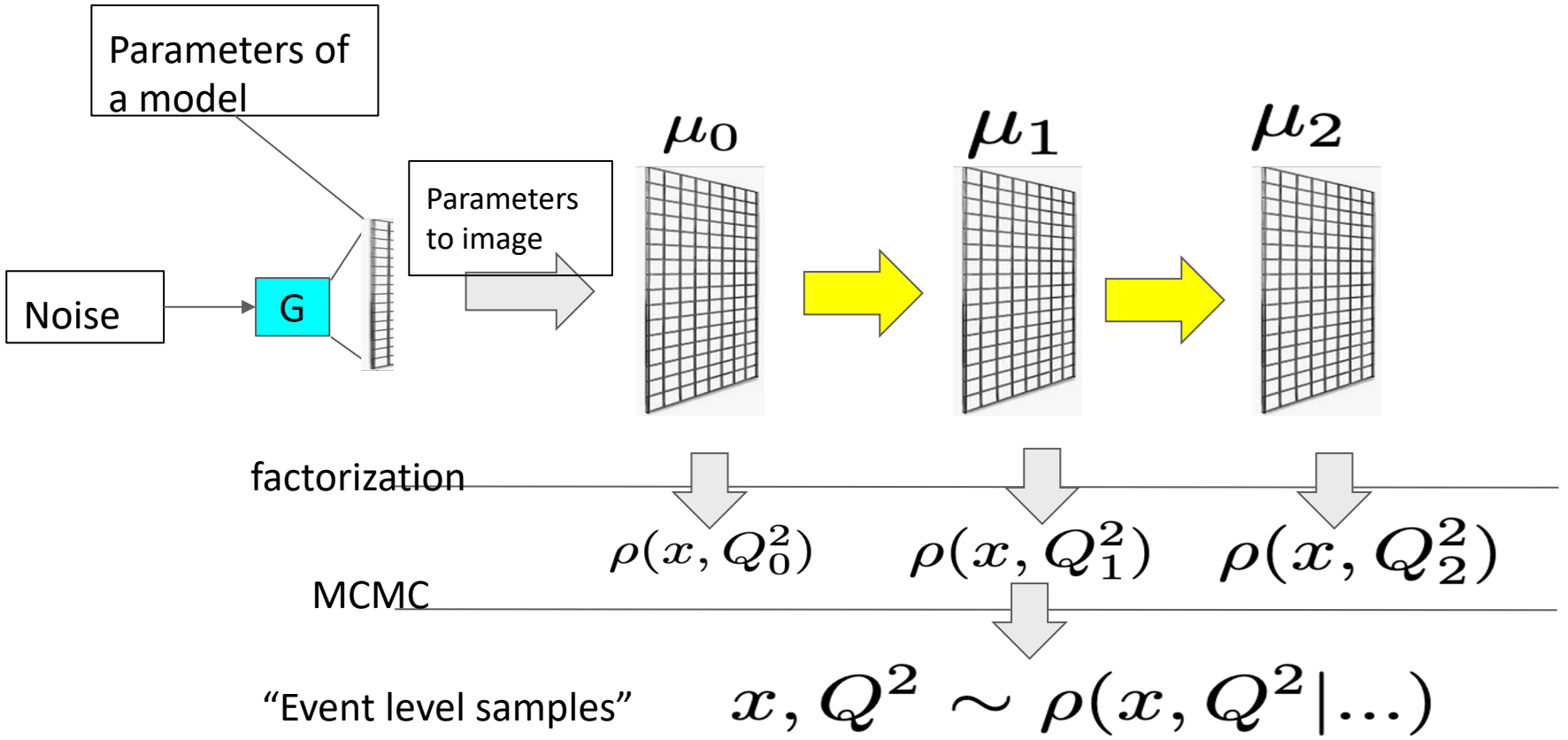
- Statistical impact in results: 1k samples (left) and 10k samples (right)
- Clear bias that need to be understood (difference between parent and sampled distributions)





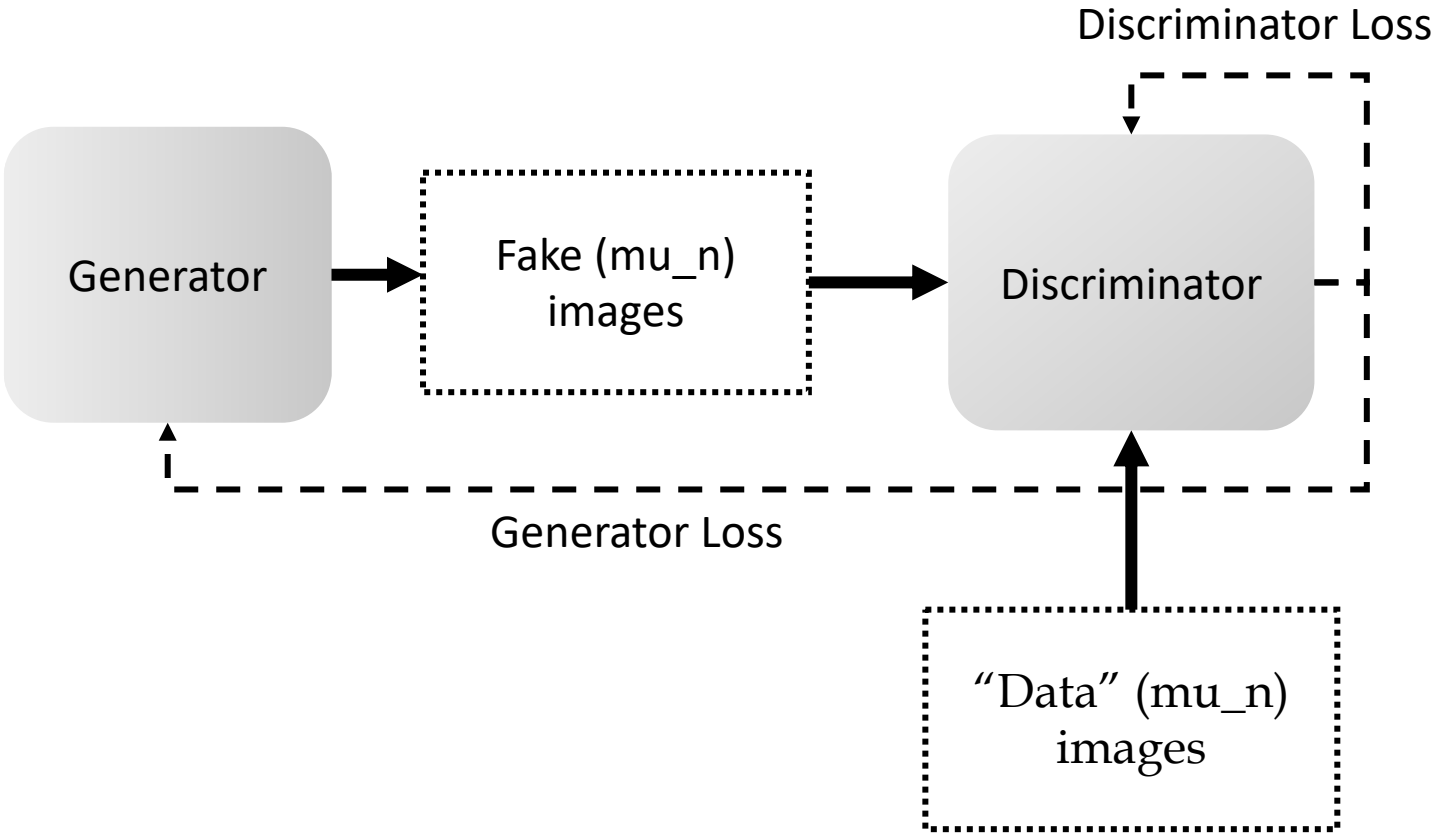
BEYOND THE PROXY APP

- The proxy app allows us to study the framework and scaling
- We need to also study potential real use case (inclusive DIS, etc.)

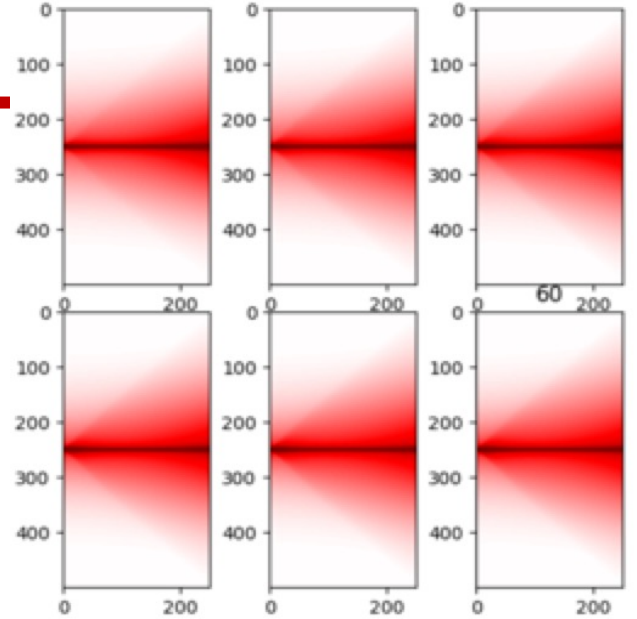


LEARNING FROM GPD IMAGES

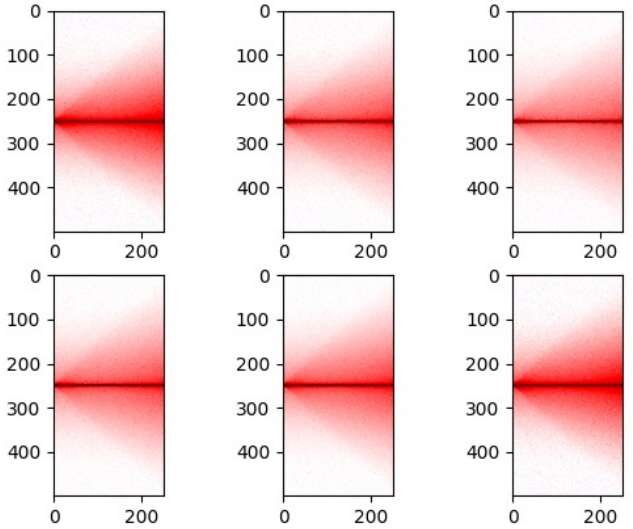
- The initial goal is to ensure that we can generate images that look like the designed “data” images
- This is the MNIST example for NP ☺



Real



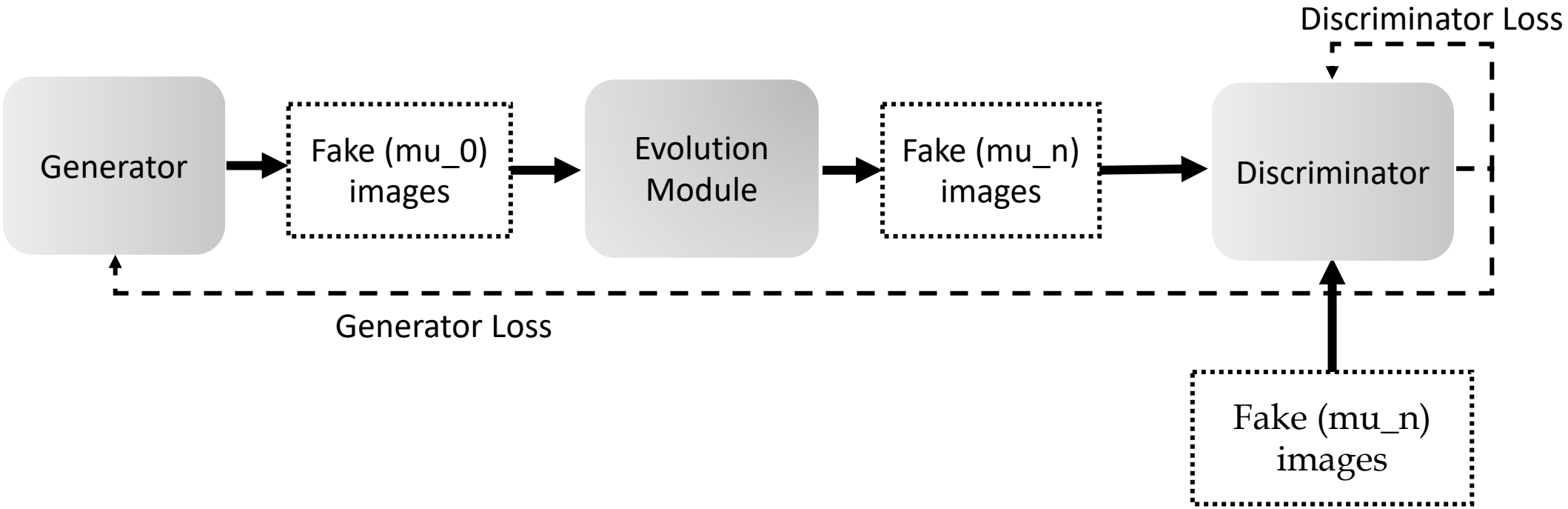
Training



Jefferson Lab

LEARNING FROM GPD IMAGES

- We now add the evolution code in the mix
- The evolution code is slow and require a lot of memory
- In fact, we cannot run on GPU due to memory requirement (on A100)



CHALLENGES

- Gradient-based approach require all components to be differentiable for backpropagation
  - This is a challenge for traditional sampling techniques

- Gradient-based optimizers store large amount of information
  - This is a problem when backpropagating through evolution code, etc.

- Elements of the current workflow have stochastic execution times which doesn't scale using allreduce methods


Thank you