Uncertainty Aware Machine Learning for Particle Accelerators

Malachi Schram, Kishansingh Rajput

Work in partnership with SNS, ORNL, FNAL, PNNL, University of California

Authored by Jefferson Science Associates (JSA) operating the Thomas Jefferson National Accelerator Facility for the U.S. Department of Energy under Contract No. DE-AC05-06OR23177. Part of this research is also supported by Office of Advanced Scientific Computing Research under Award Number DE-SC0021321. This research used resources at the Spallation Neutron Source, a DOE Office of Science User Facility at Oak Ridge National Laboratory operated by UT Battelle LLC under contract number DE-AC05-00OR22725.

OUTLINE

- Uncertainty Quantification in Deep Learning
- Errant Beam Prediction at SNS Accelerator
  - Uncertainty Aware Siamese Classifier
- Uncertainty Aware Booster Surrogate for FNAL
  - Uncertainty Aware Deep Regression with single inference

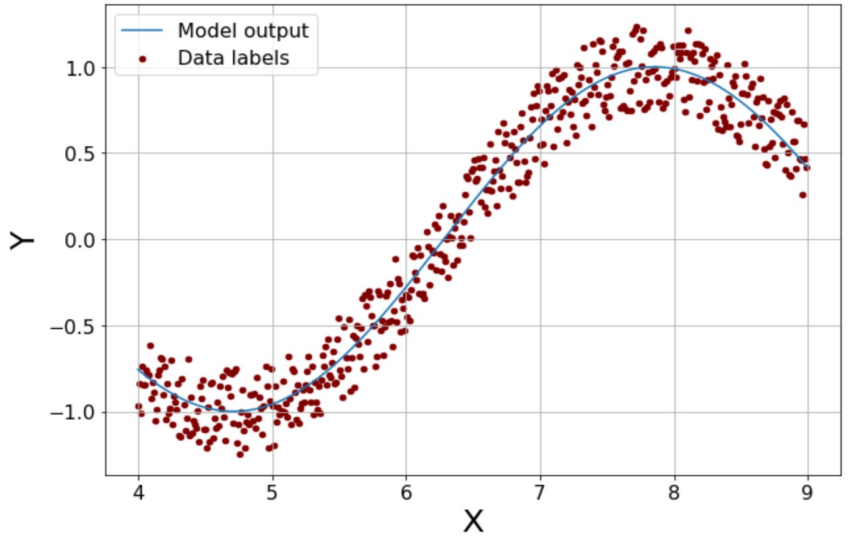
PROBLEM DEFINITION

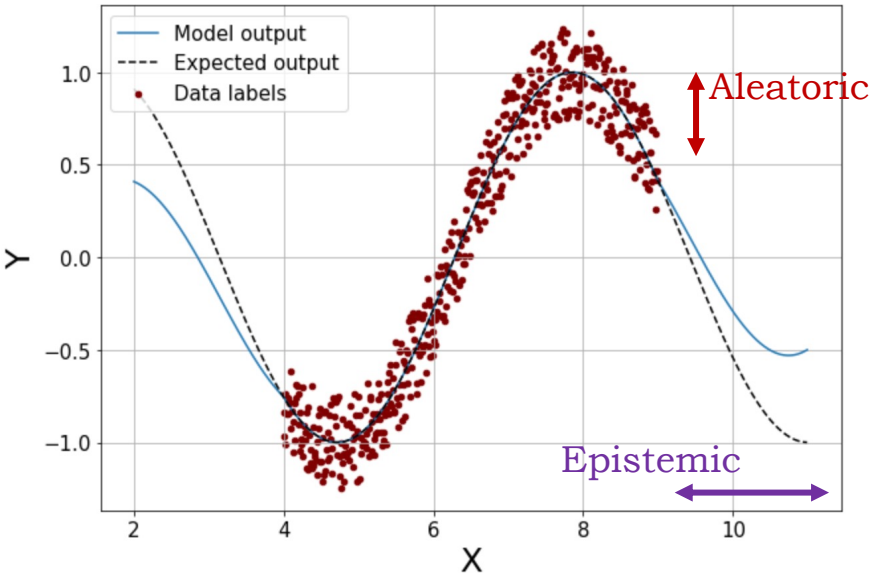
- We are focused on:
  1. Applications with **high-dimensional** continuous input features
  2. Focused on **large data sets** for DOE applications
  3. **Safety constraints** that should never or at least rarely be violated.
  4. **Inference** that must happen in **real-time** at the control frequency of the system.

- To tackle some of these points would need:
  - Integration of uncertainty quantification (UQ) to provide safety
    - Including **out-of-distribution** uncertainty
  - **Single inferences** model estimation with UQ

UNCERTAINTY QUANTIFICATION

- Deep Learning (DL) models are deterministic transformation functions from an input to the output
- DL models are very powerful and expressive
- It is important to know the confidence associated with each prediction from a DL models for decision making

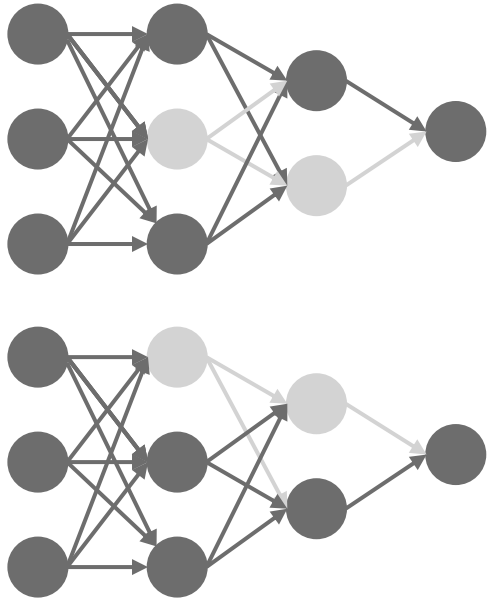




Uncertainty Types: Aleatoric vs Epistemic uncertainties

- Aleatoric → Data uncertainties
- Epistemic → Out of training distribution uncertainty (OOD)

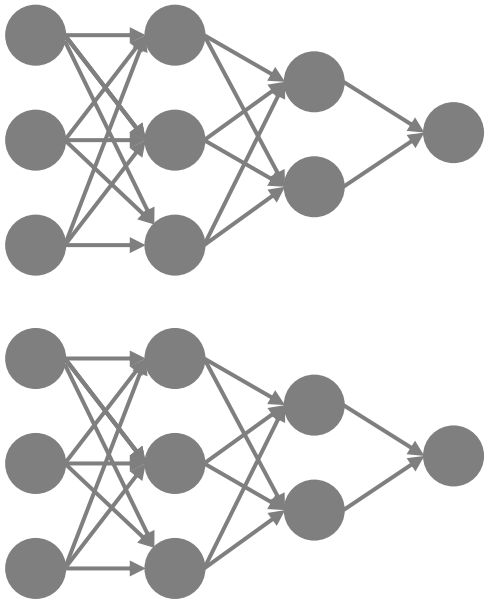
COMMON UNCERTAINTY ESTIMATION METHODS IN DEEP LEARNING



(a) MC Dropout

Use MC dropout during inference with dropout layers on can provide uncertainty prediction.

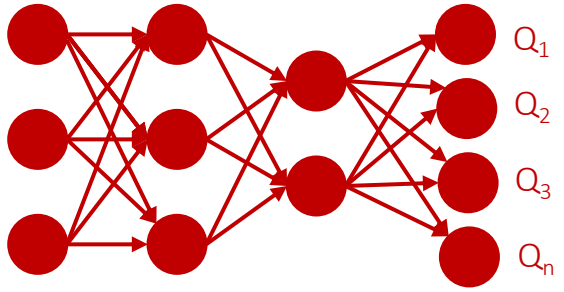
However, it slow and requires offline calibration.



(b) Ensemble

Create multiple copies of the same model architecture trained with different parameters initialization.

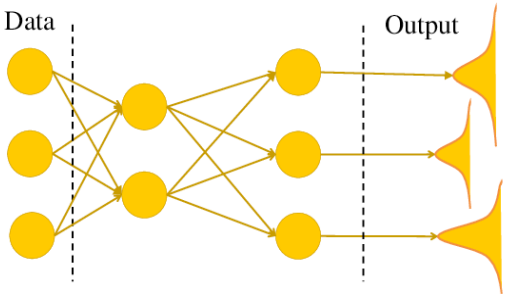
However, it's requires a lot more memory, it's slower (aggregate results) and requires calibration after training.



(c) Quantile Regression

Model is trained to predict quantiles for the regression problem.

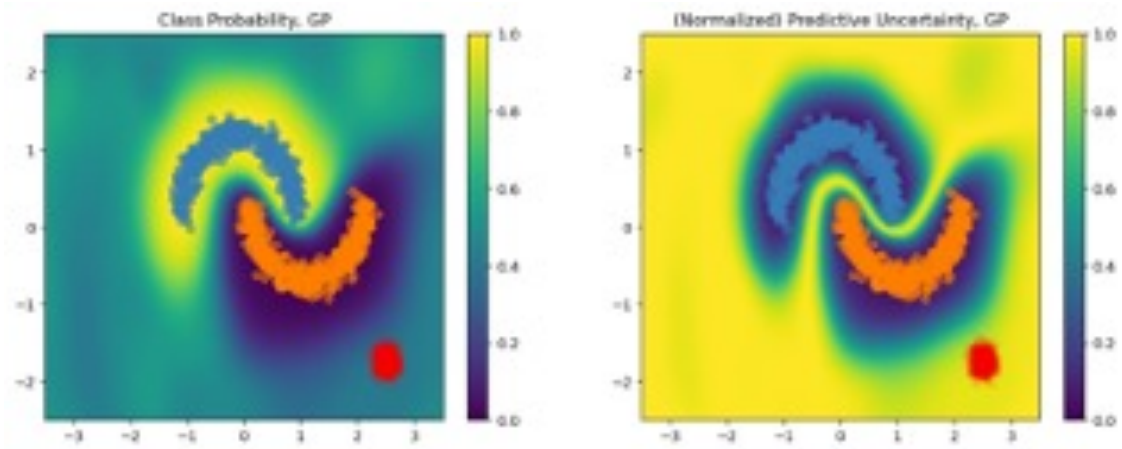
However, we'll see it doesn't account for out-of-distribution uncertainty.

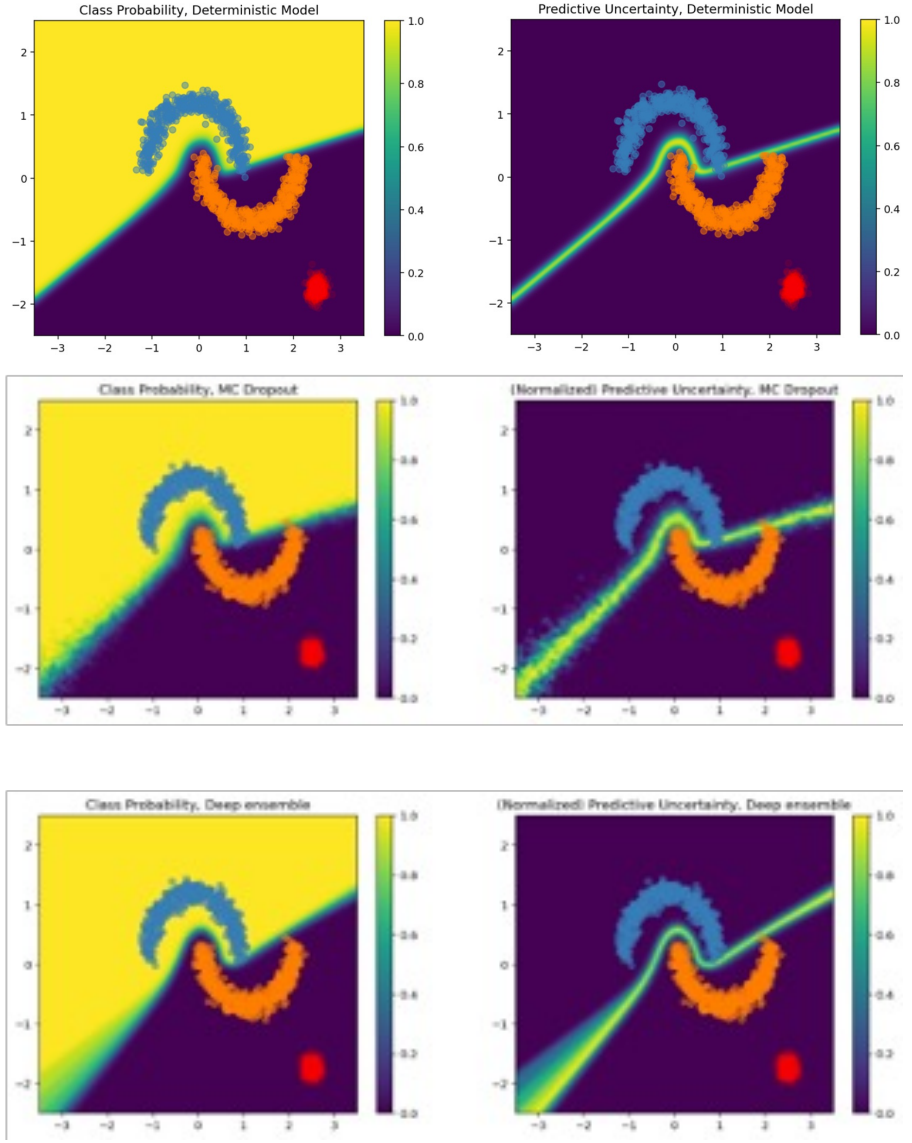


Jefferson Lab

POPULAR METHODS FOR UNCERTAINTY QUANTIFICATION IN DEEP LEARNING

- Unfortunately, majority UQ methods for DL do not account for OOD uncertainty
- This is critical in optimization or control problems
- For example, different methods yield vastly different uncertainty estimation
  - Deterministic (Prediction value)
  - MC Dropout
  - Deep Ensemble
  - Gaussian Processes (GP)





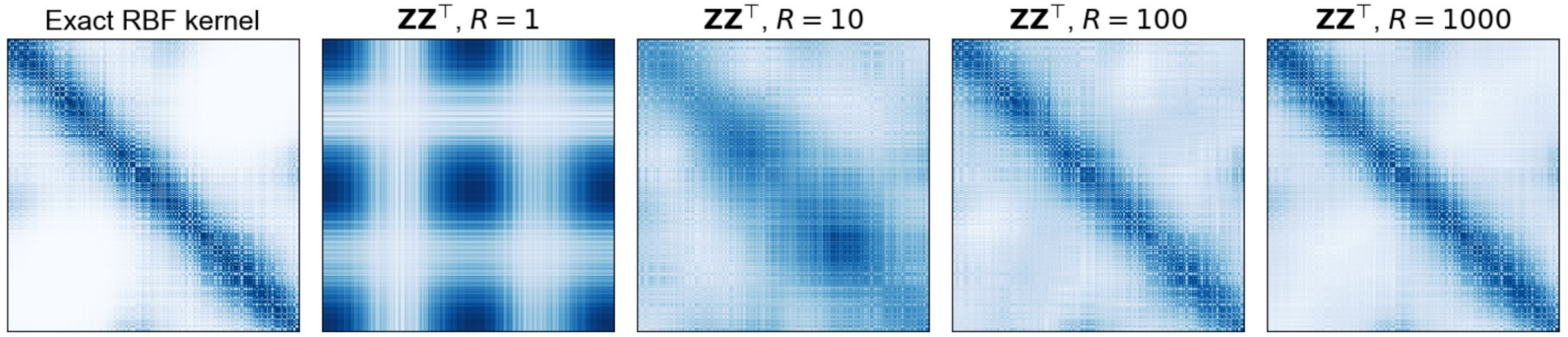
Simple and Principled Uncertainty Estimation with Deterministic Deep Learning via Distance Awareness (https://arxiv.org/abs/2006.10108)

Jefferson Lab

GAUSSIAN PROCESSES AND RANDOM FEATURES

- Gaussian processes scales very poorly with high dimensions and large datasets
- Random Fourier Features have been used to approximate the kernel (for specific conditions) to significantly reduce the computational cost for large dimension and big data problem

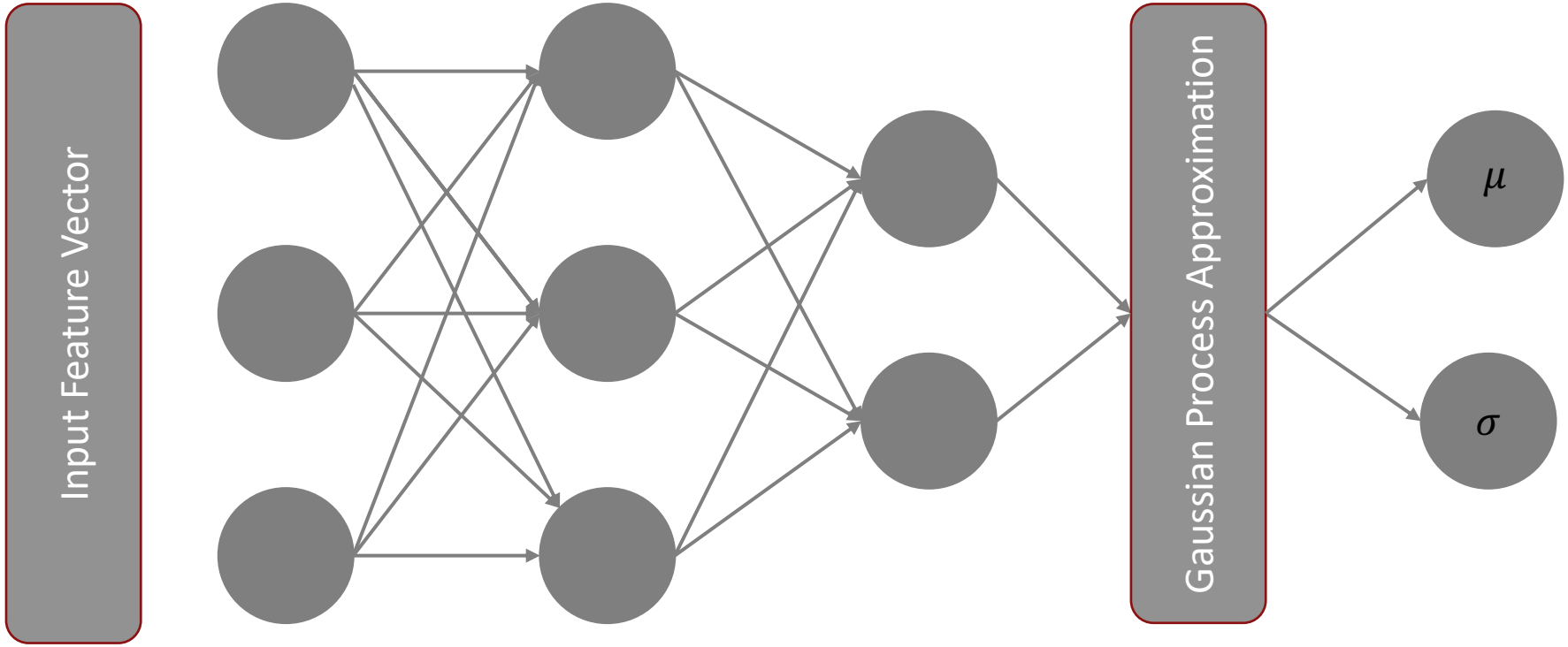
$$k(x, y) \approx z^T(x)z(y)$$



- Select research on reducing the high dimension using deep model:
  - Random Features for Large-Scale Kernel Machines (https://proceedings.neurips.cc/paper/2007/file/013a006f03dbc5392effeb8f18fda755-Paper.pdf)
  - Deep Kernel Learning (https://arxiv.org/abs/1511.02222)
  - Simple and Principled Uncertainty Estimation with Deterministic Deep Learning via Distance Awareness (https://arxiv.org/abs/2006.10108)
  - On Feature Collapse and Deep Kernel Learning for Single Forward Pass Uncertainty (https://arxiv.org/abs/2102.11409)

DEEP GAUSSIAN PROCESS APPROXIMATION

1. Reduce the high dimensional input feature vector using a neural network



2. Take the reduced latent space as input to the Gaussian Process approximation

$$k(h, h') \approx z^T(h)z(h')$$

BI-LIPSCHITZ CONSTRAINT AND FEATURE COLLAPSE

- A problem when introducing a deep model to reduce latent space is it doesn't guarantee that the distance between the input features is preserved in the latent space

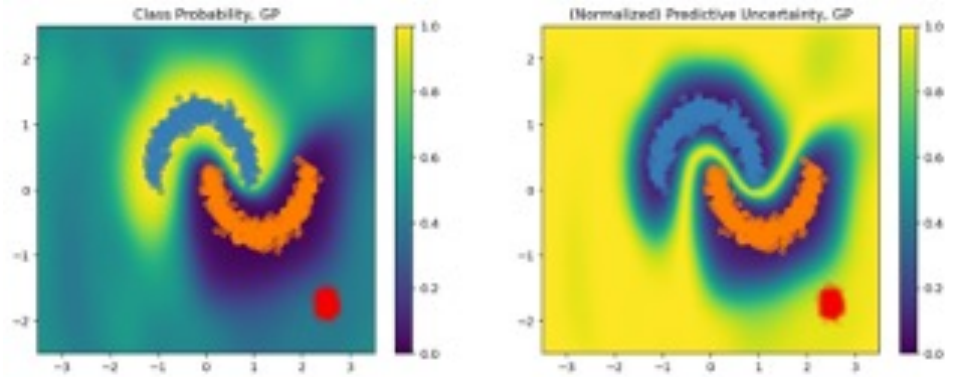
- This is typically handled using the bi-Lipschitz constraint:

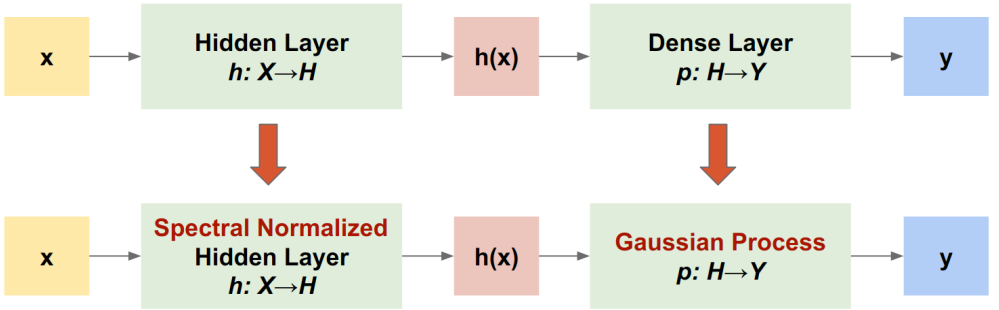
$$L_1 * ||x_1 - x_2||_X \leq ||h_1 - h_2||_H \leq L_2 * ||x_1 - x_2||_X$$

- The lower bound avoid feature collapse
- The upper bound ensure feature similarity
- We enforce this constraint using a loss penalty but will revisit other techniques

GAUSSIAN PROCESSES FOR UNCERTAINTY QUANTIFICATION IN DEEP LEARNING MODELS

- GP transforms the input space into a higher dimensional space with the help of a kernel
- The inferences are based on the distance measure between different input samples
- This allows GP to intrinsically provide uncertainty estimates including OOD
- GP is limited in terms of Scaling and data reduction techniques are usually required for large data sets
- Recent study presented a way to introduce Gaussian Process approximation within a neural network
- This allows to use highly expressive deep networks and provide uncertainty estimation





Spectral Neural Gaussian Process*

https://arxiv.org/abs/2006.10108

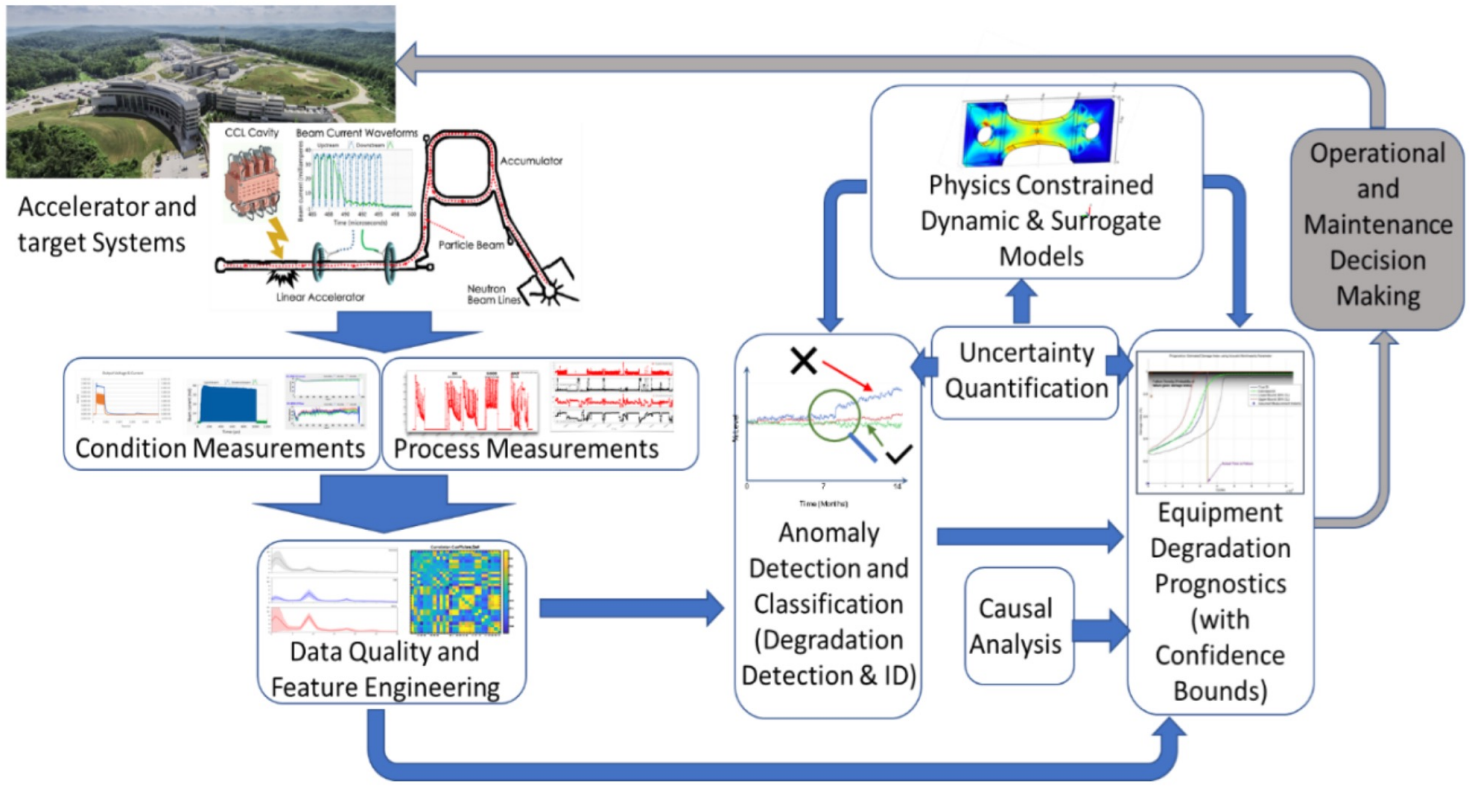
Jefferson Lab

OUTLINE

- Uncertainty Quantification in Deep Learning
- Errant Beam Prediction at SNS Accelerator
  - Uncertainty Aware Siamese Classifier
- Uncertainty Aware Booster Surrogate for FNAL
  - Uncertainty Aware Deep Regression with single inference

Errant Beam Prediction for SNS Accelerator

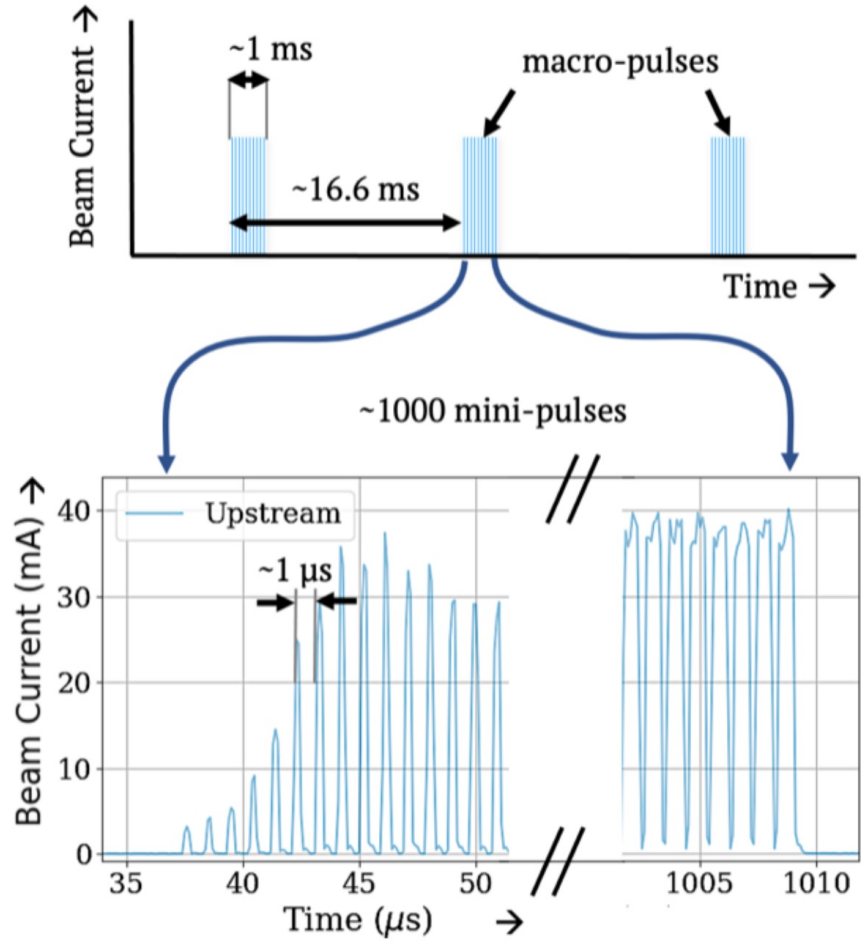
- Spallation Neutron Source (SNS) accelerator at ORNL delivers 1.4 MW of a 1 GeV pulsed beam at 60 Hz
- Ongoing work to **predict errant beam pulses** as well as equipment degradation and prognostics
- Continuous data collection is done by **Differential Current Monitor (DCM)**, Beam Position Monitor (BPM) etc.
- Errant beam prediction on one pulse before it occurs to potentially avoid it



Data Collection and Preparation

How was the data collected and labeled?

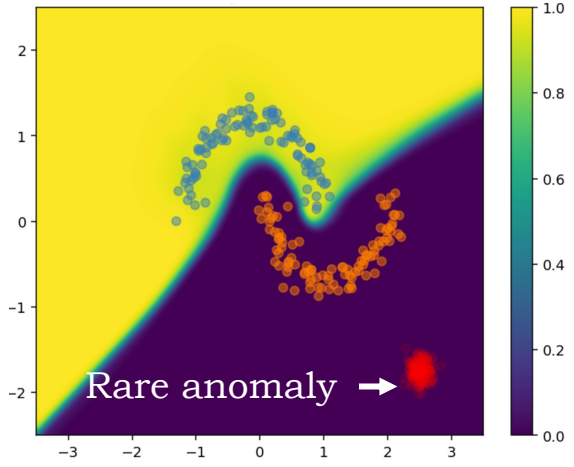
- DCM creates a series of pulses ("macro-pulses") with each macro-pulse composed of ~1k mini-pulses
  - An errant-beam data file is composed of 25 "good" macro-pulses followed by the errant beam pulse
  - A "normal" data file has no errant beam pulse
- We used the macro-pulse before the errant beam pulse (and labeled it as anomaly) and macro-pulses from the normal file (and labeled them as normal) for our studies
- Our hypothesis: there is a sign about upcoming anomaly in macro-pulses even before it happens
- We also need to forecast the fault within a short time window to be actionable
- Samples were divided into 3 orthogonal dataset:
  - Train (64%)/Test(20%)/Validation(16%)

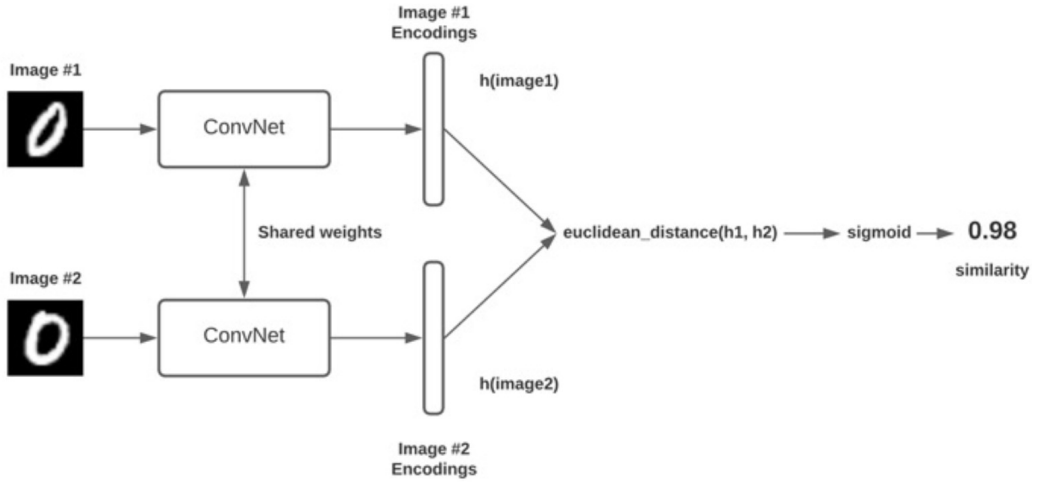


Siamese Neural Network (SNN) Model

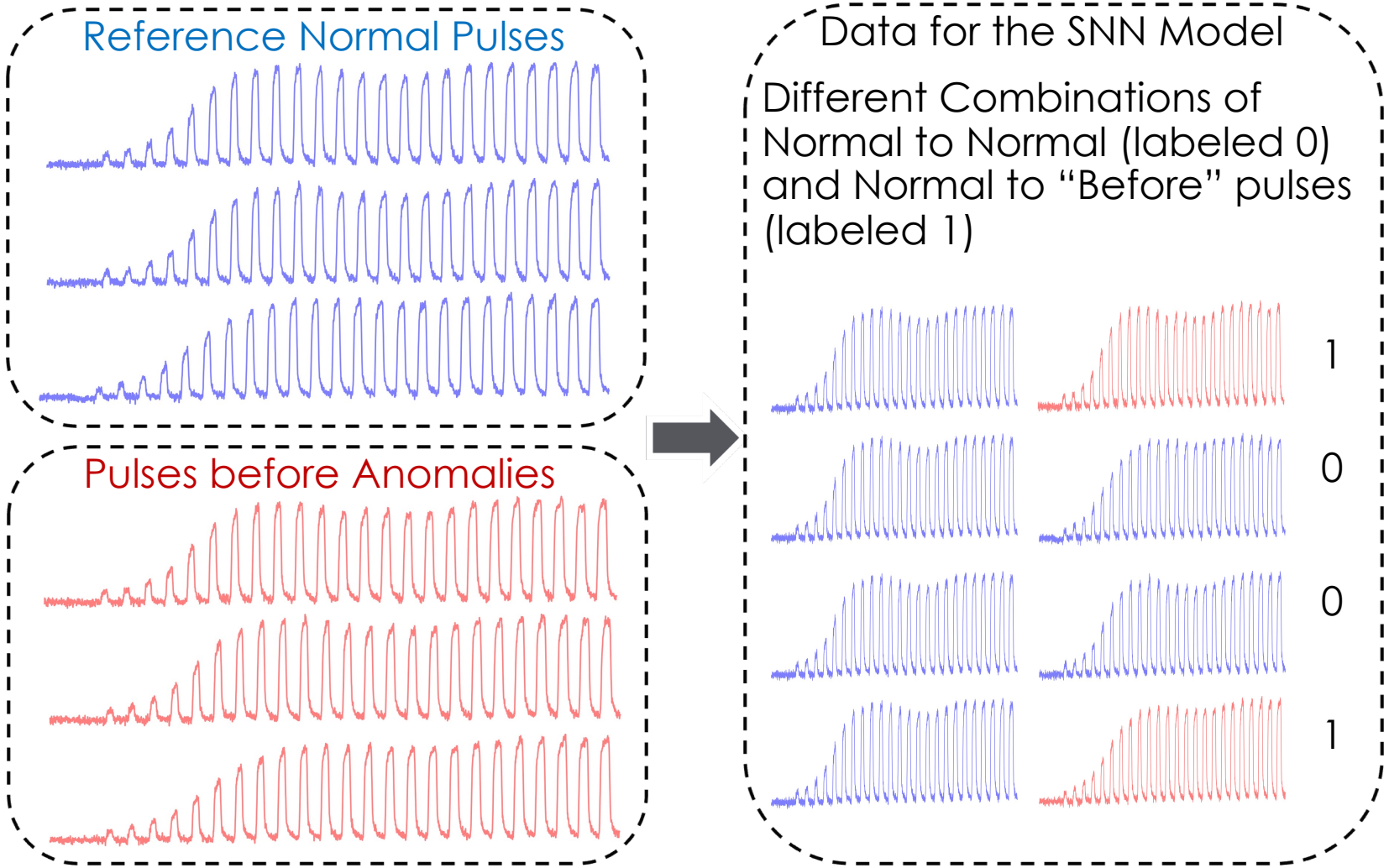
Traditional classification models vs Siamese model

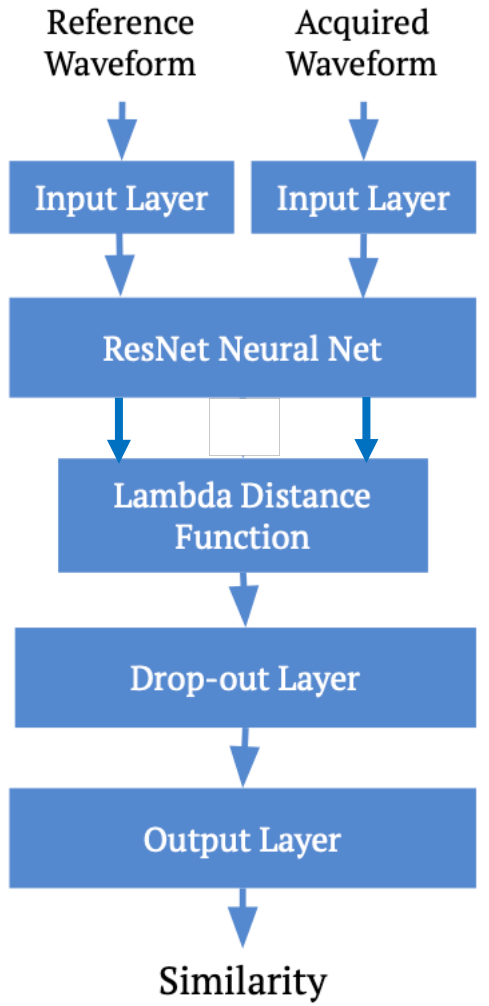
- Traditional DL classification models fails to identify unseen anomalies (OOD)
- Similarity based models can correctly classify unseen anomalies. Ex Siamese model
- Siamese model does not explicitly model the classification but focuses on the similarities
- It learns twin embedding models to transform inputs into a latent space
- Distance measures are applied at latent space to compute the similarity





Data Preparation for SNN Model Training

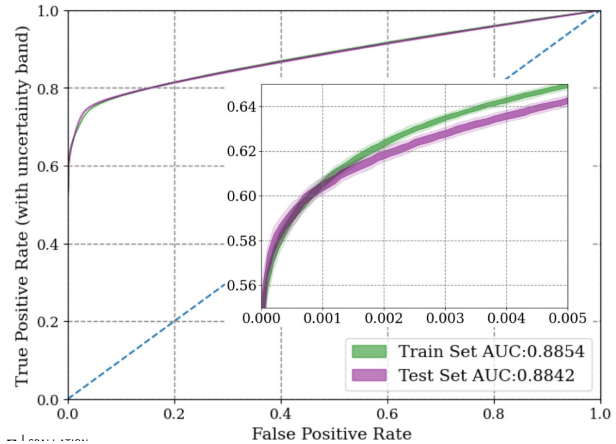


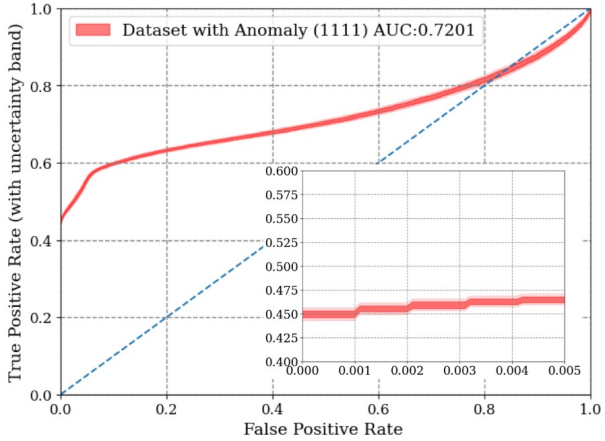


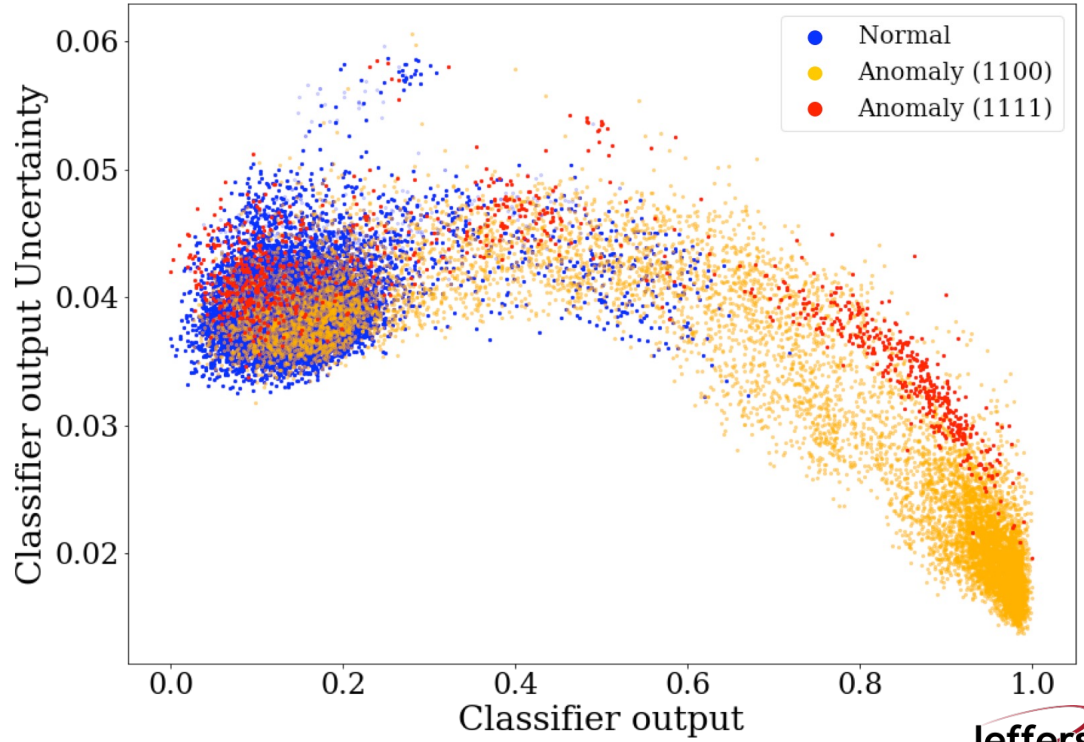
Uncertainty aware Siamese model

- We enhanced our Siamese model by adding GP layer providing an 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 shows 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)

  After a fault is predicted, is it possible to associate with a particular equipment failure?







OAK RIDGE National Laboratory | SPALLATION NEUTRON SOURCE

Jefferson Lab

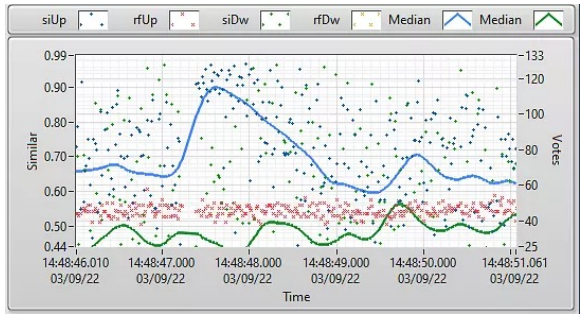
Online results

DCML:

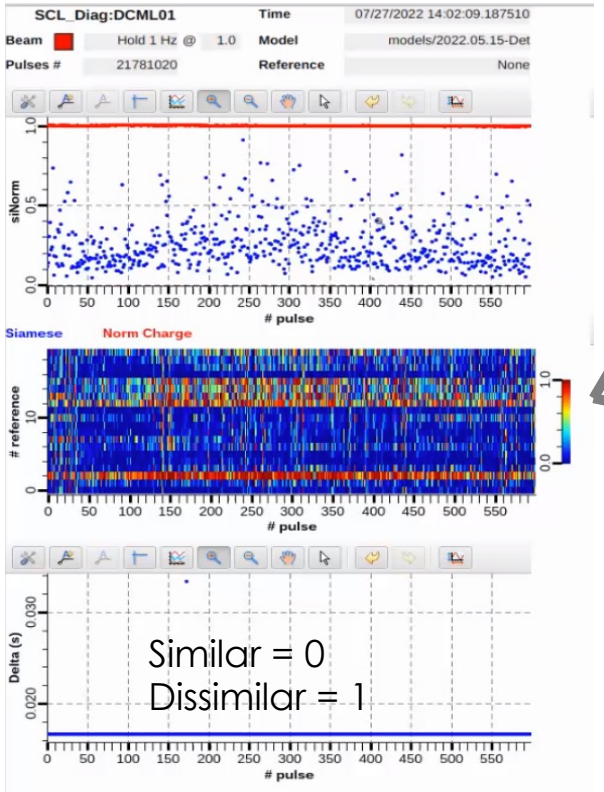
- Can run up to 4 deterministic SNN inferences

ML Server:

- Can run 20 deterministic inferences per pulse at 60 Hz to compare incoming waveform with multiple references (can be normal or abnormal)
- Create average similarity to improve results
- Presents results over EPICS

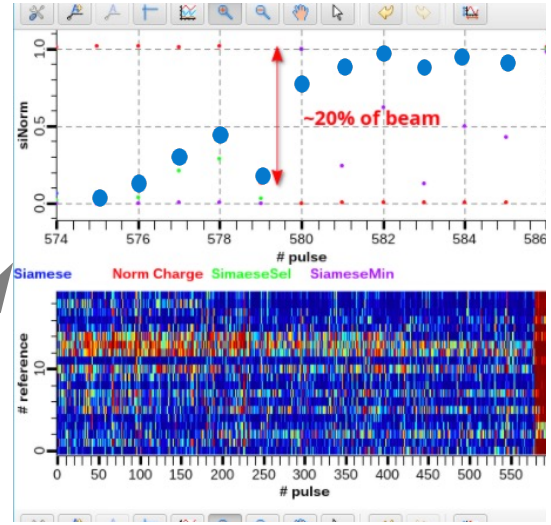


DCML live results (Siamese/RF upstream/downstream

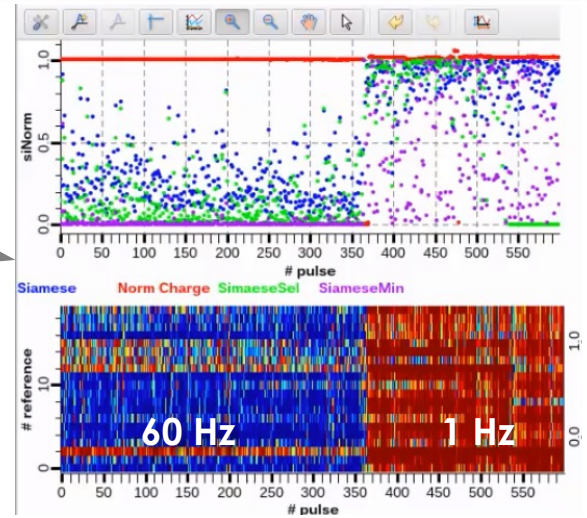


ML Server Results Control Room Screen

examples



Chopper partial failure is seen as abnormal beam



1 Hz beam (instead of 60 Hz) is seen as abnormal

Path Forward

- Replace deterministic SNN model with Uncertainty Aware SNN for online system
- Include beam configuration to the SNN model as conditional inputs

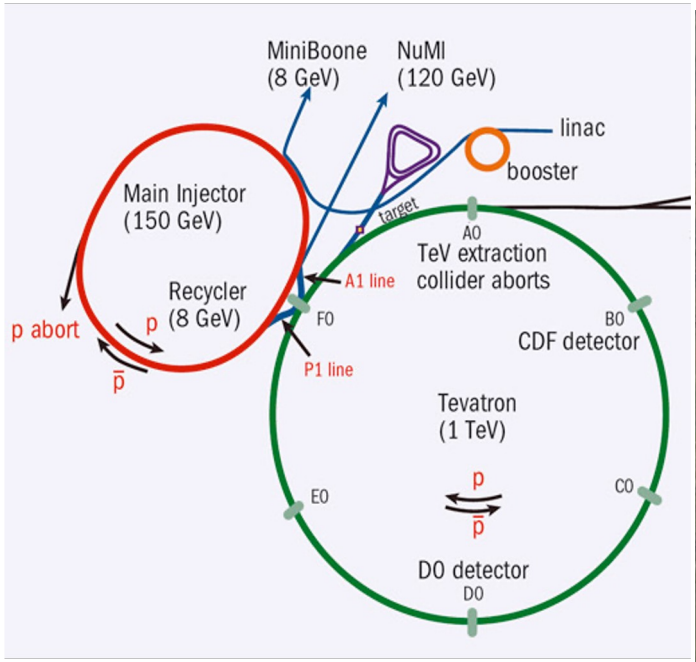
OUTLINE

- Uncertainty Quantification in Deep Learning
- Uncertainty Aware Errant Beam Prediction at SNS
  - Uncertainty Aware Siamese Classifier
- Uncertainty Aware Booster Surrogate for FNAL
  - Uncertainty Aware Deep Regression with single inference

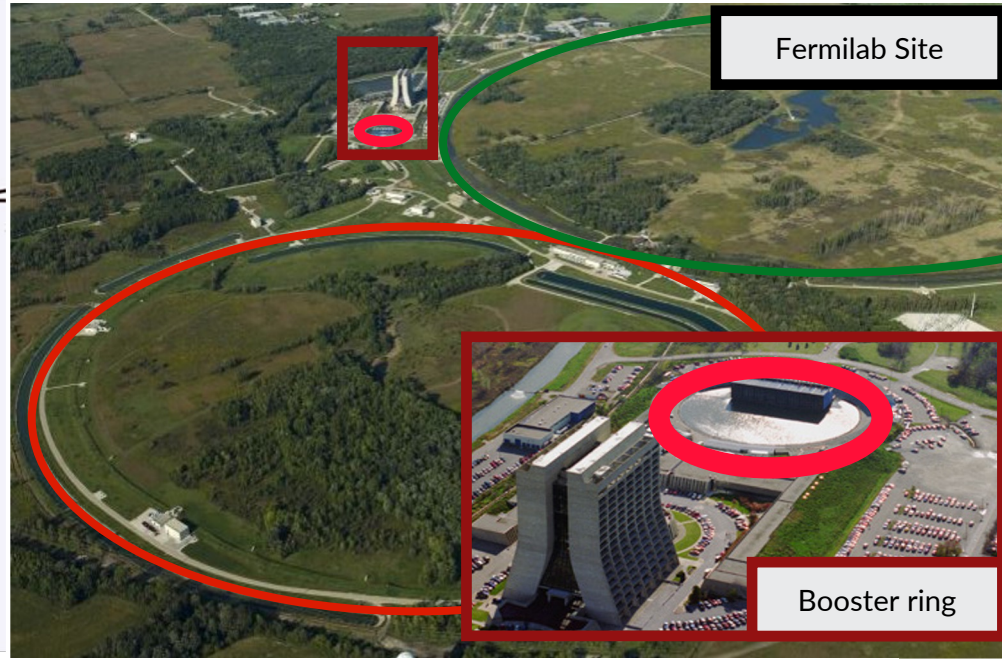
UNCERTAINTY AWARE BOOSTER SURROGATE

Aim:
Reduce beam losses in the FNAL Booster by developing a Machine Learning (ML) model that provides an optimal set of actions for GMPS regulator

FNAL Accelerator Complex:



Courtesy: Christian Herwig



The Booster receives the 400 MeV (kinetic energy) beam from the Linac

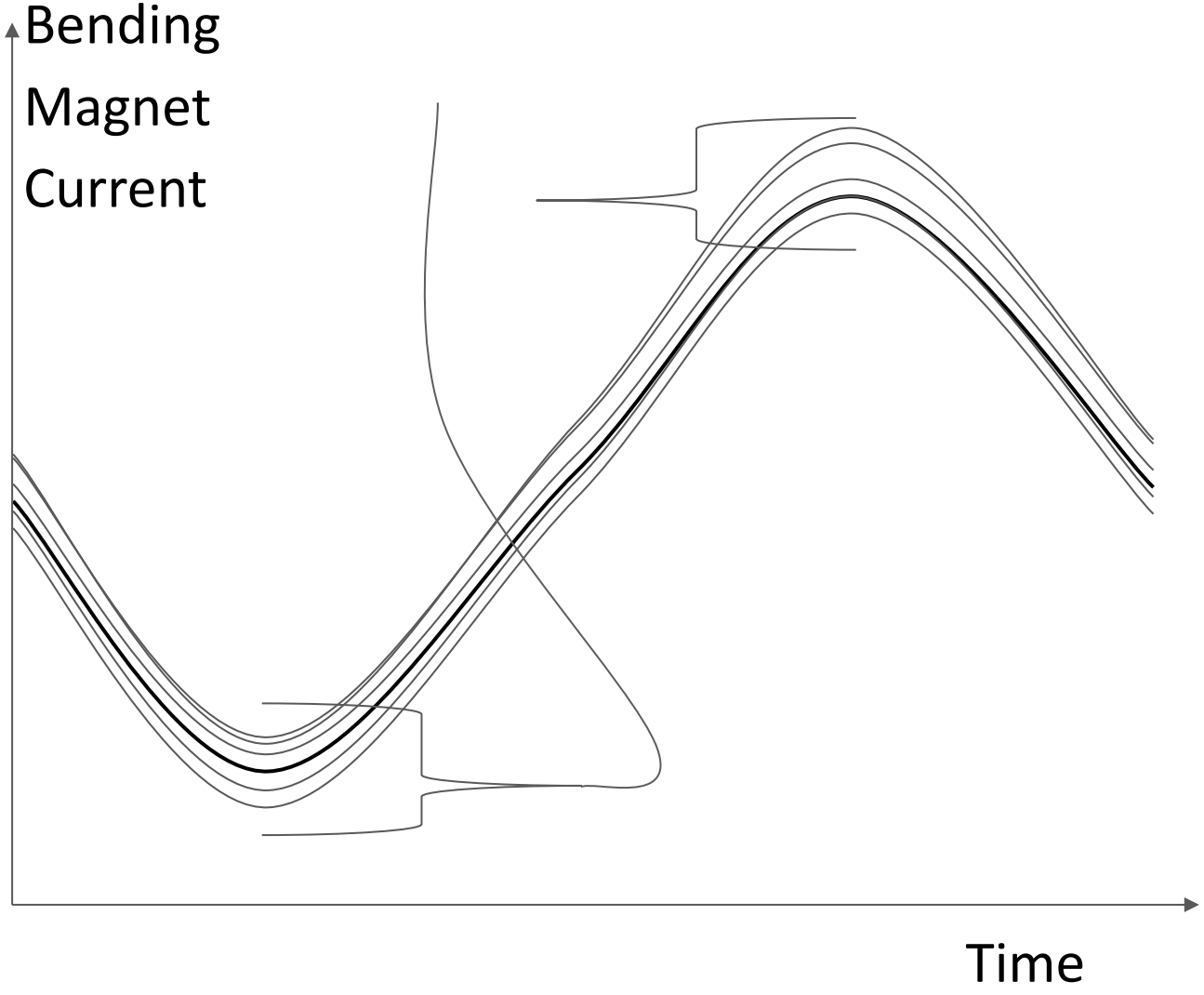
It is then accelerated to 8GeV with the help of booster cavities and Combined-function bending and focusing electromagnets known as gradient magnets.

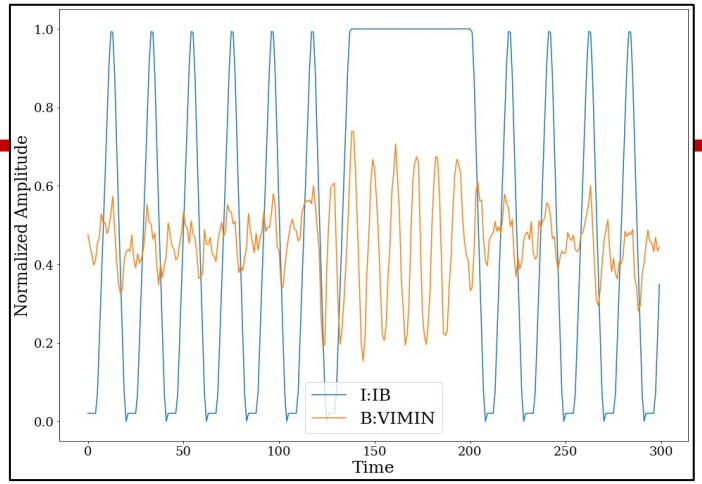
These magnets are powered by the gradient magnet power supply (GMPS)

Jefferson Lab

PROBLEM DEFINITION

- Observed $\delta I/I$ for min and max currents: ~10^{-3} each

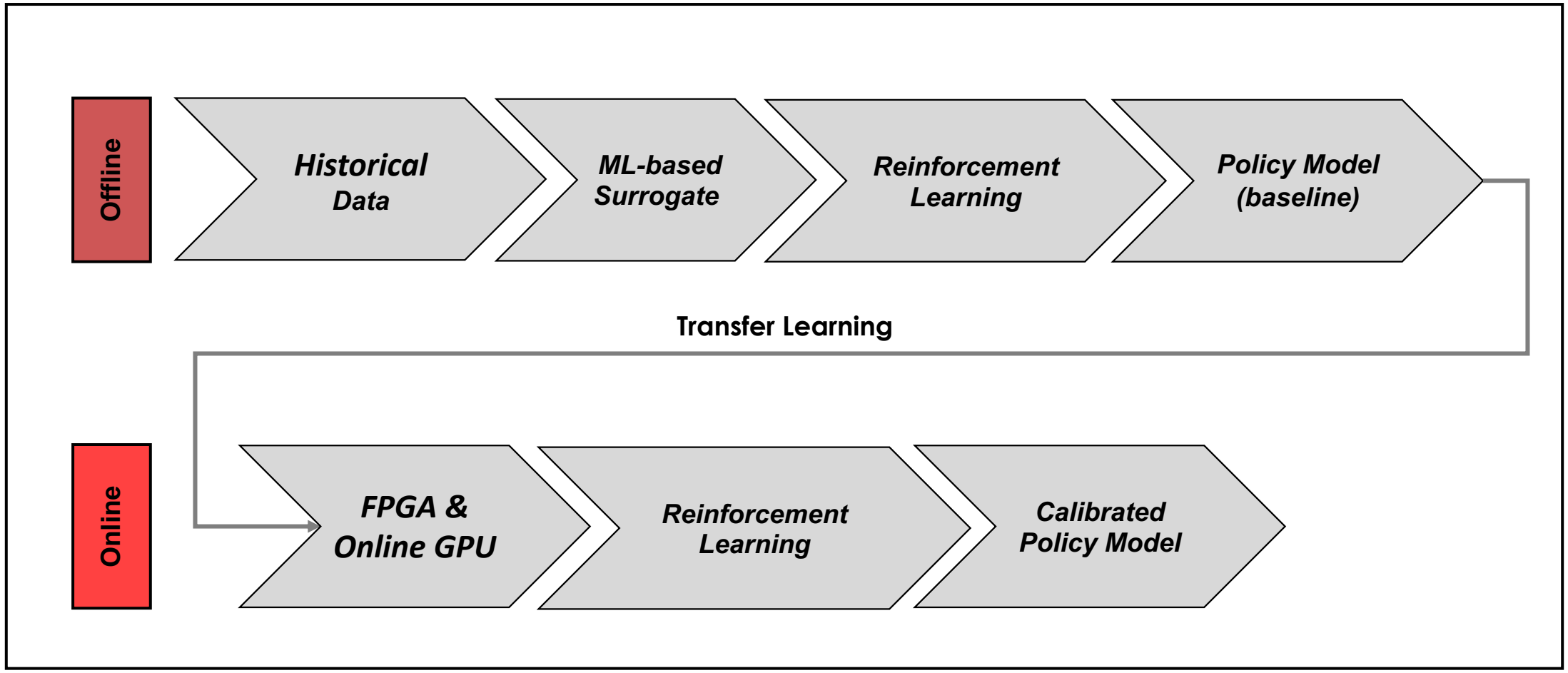




- Perturbing influences:
  - Recent corrections made
  - Other nearby synchrotrons
  - Fluctuation of 60 Hz power
  - Temperatures, etc
- Available data mostly with the current PID regulator
- Spread in B-field degrades beam quality, degrades repeatability, & contributes to losses

Jefferson Lab

WORKFLOW


Offline
Historical Data
ML-based Surrogate
Reinforcement Learning
Policy Model (baseline)
Transfer Learning
Online
FPGA & Online GPU
Reinforcement Learning
Calibrated Policy Model

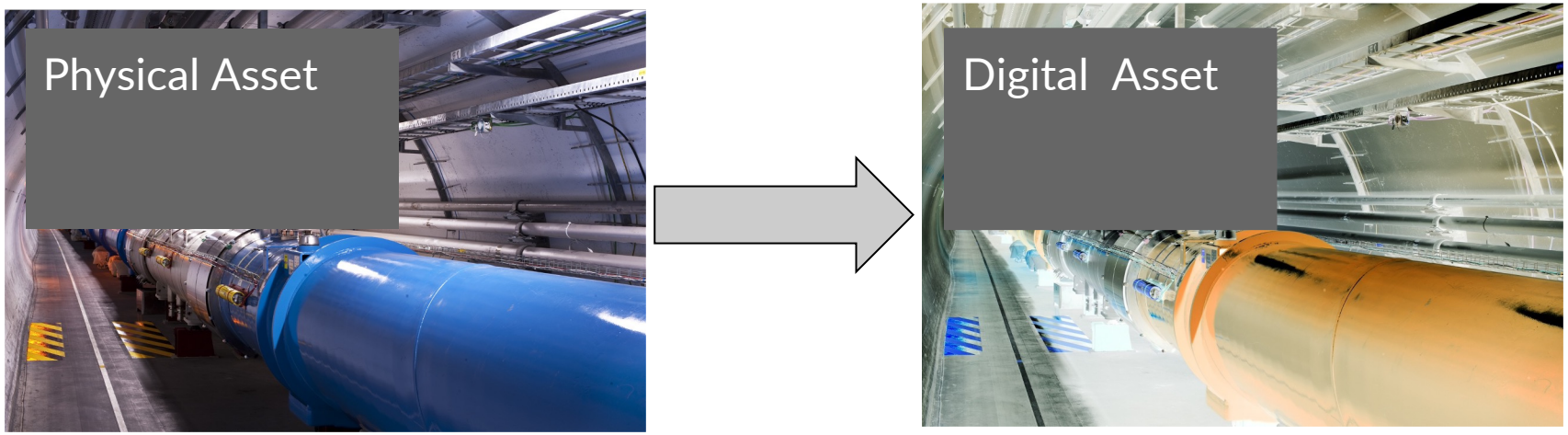
DATA DRIVEN ML-BASED SURROGATE MODEL

Scope and usage for surrogate model:

- Provide accurate predictions of future time for key variables to be used by the reinforcement learning framework

Dataset provided:

- Historical temporal information from key variables was available based on subject matter expert input
- Caution:
  - Data did not include detailed history on commissioning, maintenance, etc.
  - Should conduct a full data inventory assessment



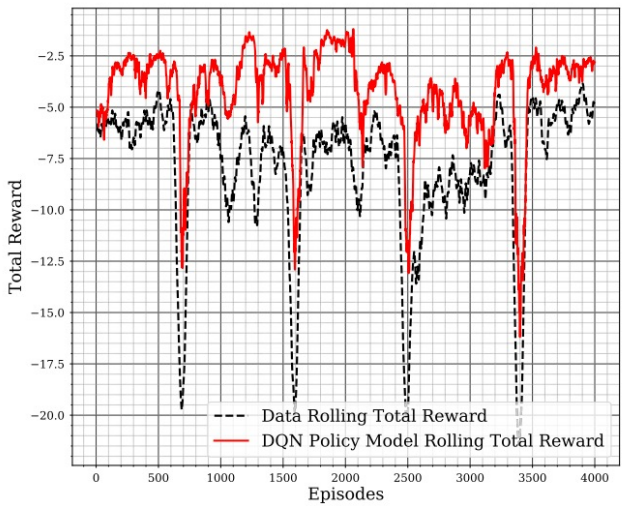
SURROGATE MODEL

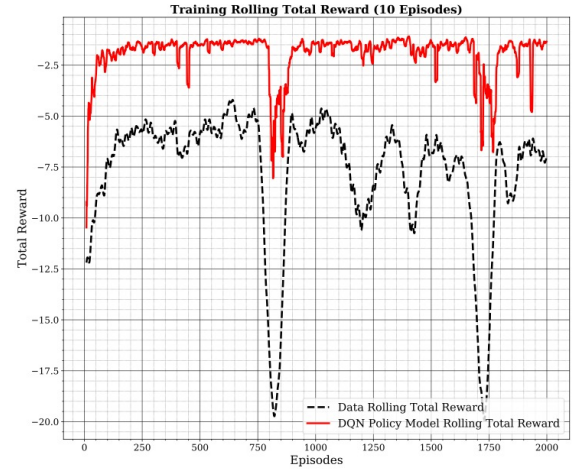
- Physical surrogate models
  - Generally well defined boundary condition for areas of application
  - Example: Newtonian physics vs special relativity

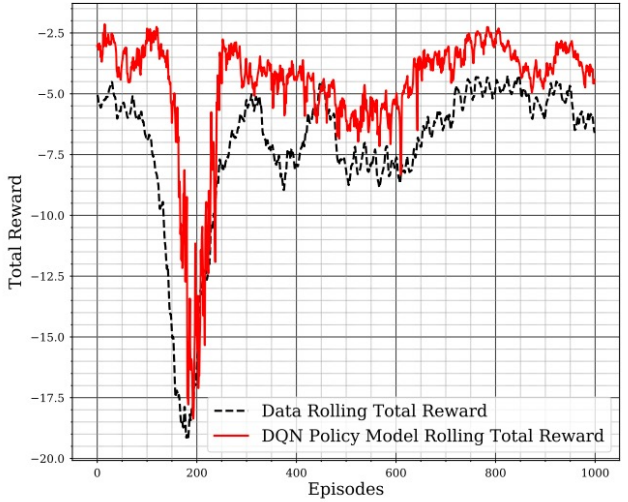
- Data Driven ML-based surrogate model
  - No clear physics boundary to avoid
  - Must include include UQ that accounts for OOD to avoid RL agents exploring areas that are poorly modeled by the surrogate model

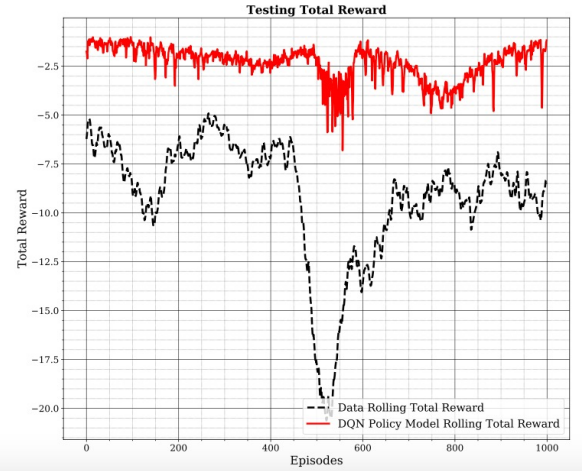
RESULTS

- We used a DDQN agent
- Original results used a stacked LSTM model yielding ~2x improvement over the original control system
  - Real-time artificial intelligence for accelerator control: A study at the Fermilab Booster (https://journals.aps.org/prab/abstract/10.1103/PhysRevAccelBeams.24.104601)
- Second study used a BNN to incorporate uncertainty quantification (calibrated) and showed improved results and stability:
  - Developing Robust Digital Twins and reinforcement learning for accelerator control systems at the Fermilab Booster (https://arxiv.org/abs/2105.12847)









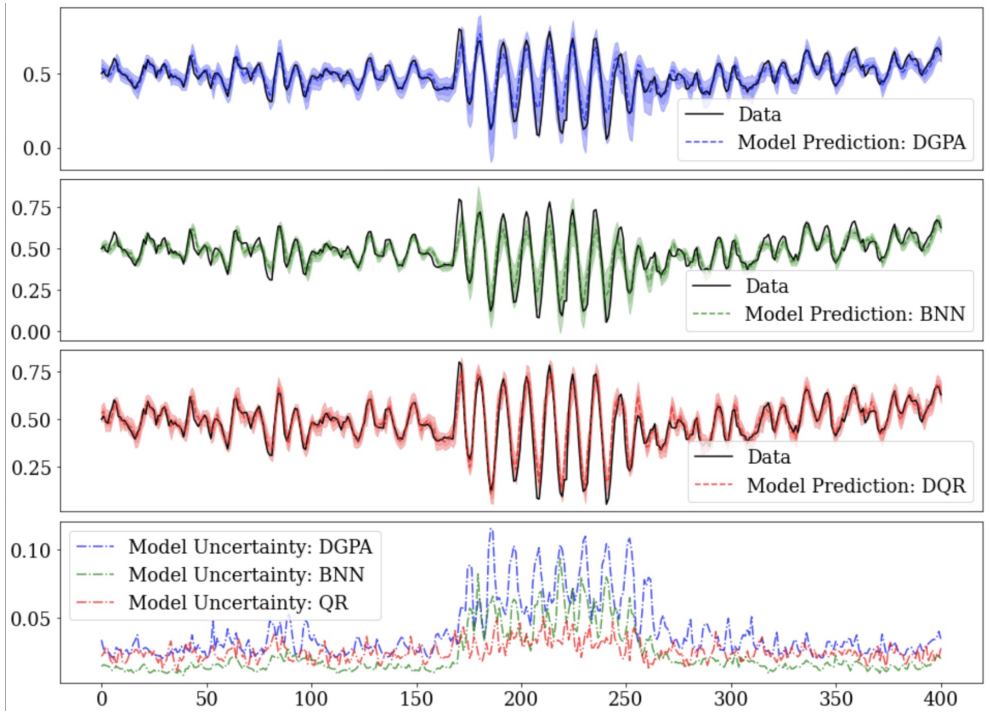
Jefferson Lab

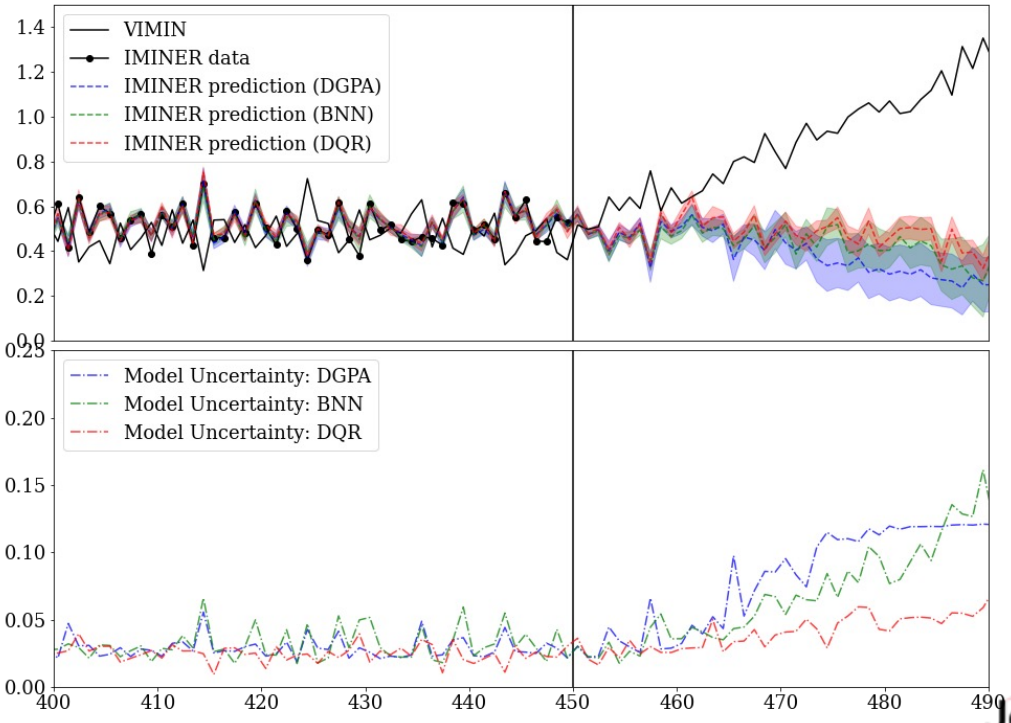
UNCERTAINTY AWARE DL REGRESSION MODEL

Why uncertainty quantification is important in Digital Model?

- Uncertainty Quantification can help determine how well a region of a phase space is modeled by the surrogate

- Gaussian Process Approximation (DGPA) method can quantify the regression uncertainties for a DL model
- Unlike most other methods, DGPA does not require multiple inferences and does not require offline calibrations making it easy to deploy in online settings





Jefferson Lab

PATH FORWARD

- Study high input dimension UQ for Deep Learning
- Explore online system requirements and approximation trade-offs
- Expand to other areas of Deep Learning application

THANK YOU!

References:

"Uncertainty aware anomaly detection to predict errant beam pulses in the Oak Ridge Spallation Neutron Source accelerator"
Willem Blokland, Kishansingh Rajput, Malachi Schram, Torri Jeske, Pradeep Ramuhalli, Charles Peters, Yigit Yucesan, and Alexander Zhukov, Phys. Rev. Accel. Beams 25, 122802

"Uncertainty aware machine-learning-based surrogate models for particle accelerators: Study at the Fermilab Booster Accelerator Complex" **Malachi Schram, Kishansingh Rajput, Karthik Somayaji NS, Peng Li, Jason St. John, and Himanshu Sharma**
Phys. Rev. Accel. Beams 26, 044602

"Uncertainty Aware Deep Learning for Particle Accelerators", **Rajput, Kishansingh*; Schram, Malachi; Somayaji, Karthik**
Machine Learning and the Physical Sciences NeurIPS 2022

"Real-time artificial intelligence for accelerator control: A study at the Fermilab Booster"
Jason St. John, Christian Herwig, Diana Kafkes, Jovan Mitrevski, William A. Pellico, Gabriel N. Perdue, Andres Quintero-Parra, Brian A. Schupbach, Kiyomi Seiya, Nhan Tran, Malachi Schram, Javier M. Duarte, Yunzhi Huang, and Rachael Keller
Phys. Rev. Accel. Beams 24, 104601

"Developing Robust Digital Twins and Reinforcement Learning for Accelerator Control Systems at the Fermilab Booster"
Diana Kafkes, Malachi Schram, https://arxiv.org/abs/2105.12847