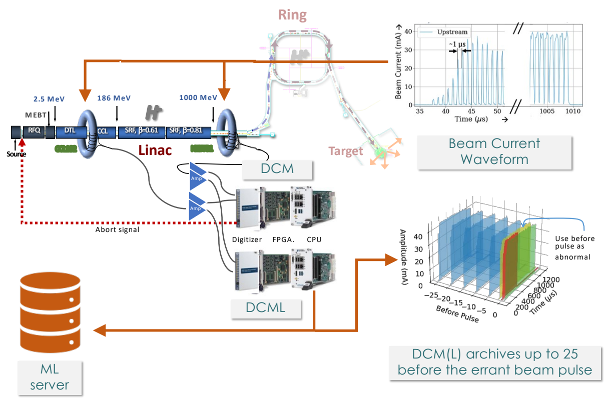
Online Machine Learning Version of the SNS Differential Beam Current Monitor

Willem Blokland*, David Brown, Frank Liu, Narasinga R. Miniskar, Charles Peters, Pradeep Ramuhalli, Aaron R. Young, Yigit A. Yucesan , Alexander Zhukov (ORNL, Oak Ridge, TN)
Kishansingh Rajput, Malachi Schram (JLAB, Newport News, VA)

Introduction

Applying Machine Learning (ML) algorithms to off-line data from the Differential beam Current Monitor (DCM) showed the existence of precursors to errant beam pulses[1,2].

Duplicating the DCM into the DCML allows us to update and run the ML algorithms during operations without affecting the beam abort functions.



Beam Current Waveform

DCM

DCML

ML server

DCM(L) archives up to 25 before the errant beam pulse

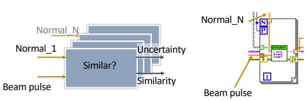
- The DCML runs a Random Forest model on its FPGA and a Siamese Twin model on its real-time CPU
- The ML Server receives a live data-stream containing the waveforms and can archive all data for a limited time (~6 hrs)

Machine Learning Methods

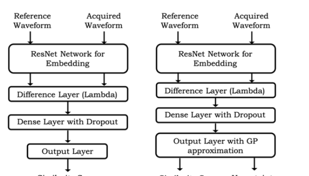
Siamese Neural Network (SNN) Model

Determine similarity between normal and unknown sample.

Gaussian Process (GP) approximation is introduced for uncertainty quantification (UQ) including out-of-distribution detection.



Siamese model on DCML



Deterministic[1] Uncertainty aware [1]

Real-time Siamese

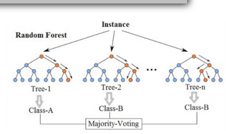
- Makes inferences before the next pulse
- Can abort before the next beam pulse

What performance is needed?
March 2021, production was 26.4 days, 1.5% beam lost
- 0.22% beam lost due to SCL beam loss
- 1.30% beam lost due to truncated beam

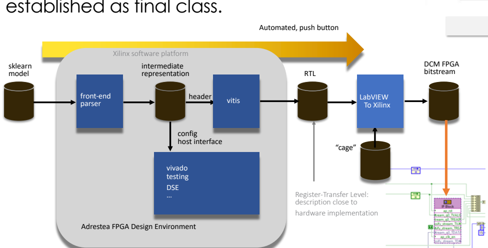
Predict a reasonable fraction of the errant pulses: TPR ≈ 50% and don't add much down-time due to false positives → 0.2% but each abort holds off beam for 4 pulses → we want to achieve a False Positive Rate ≈ 0.05%

Random Forest (RF) Model on FPGA

RF consists of multiple decision trees to use distinct features in a waveform to vote whether normal or errant. Majority vote established as final class.



Random Forest[2]



RF FPGA Design Flow[3]

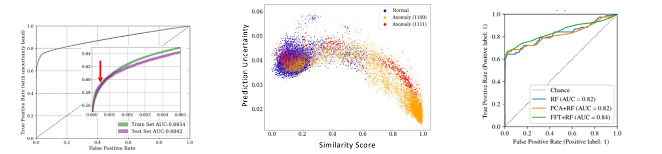
FPGA RF

- Makes an inference << 1 μs after the data is available (100 μs used of 1 ms long pulse)
- Can abort within the beam pulse

Results

Offline training and testing: Receiver Operator Characteristics (ROC) curve from training and test data illustrate binary classification performance → RF and Siamese have < 0.05% FPR with FPR between 50-60%.

Uncertainty-aware SNN provide uncertainty for prediction to distinguish unseen anomaly types.

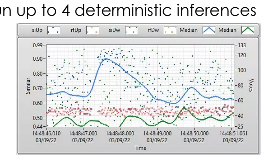


Siamese model ROC Curve[1] Siamese model uncertainty[1] Random Forest ROC Curve[2]

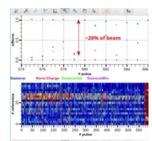
Implementation: We apply trained models, implemented in C++ on DCML and Phyton on ML server, to incoming accelerator data during production.

DCML:

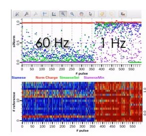
- Can run up to 4 deterministic inferences



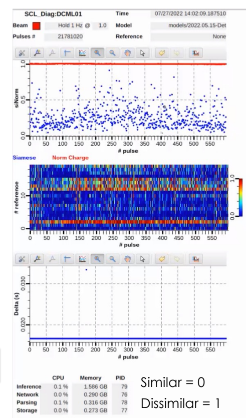
DCML live results (Siamese/RF upstream/downstream)



Chopper partial failure is seen as abnormal beam



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



Similar = 0
Dissimilar = 1

ML Server Results Screen

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

- Initial results during production showed some trends and detected errant beam pulses but not good enough to execute an abort during operations. The model and references samples were from data in the past.
- We can now have several sets of 2 hours of all beam pulses (@60 Hz) to analyze

Future

- Finish framework to quickly train new models
  - Support incremental training
- Further improve models and add equipment fault classification
- Abort beam pulses to determine if the errant pulse is avoidable

References

1. Blokland, W., Ramuhalli, P., Peters, C., Yucesan, Y., Zhukov, A., Schram, M., ... & Jeske, T. (2021). Uncertainty aware anomaly detection to predict errant beam pulses in the SNS accelerator. arXiv preprint arXiv:2110.12006.
2. Miha Reščič et al. "Improvements of pre-emptive identification of particle accelerator failures using binary classifiers and dimensionality reduction." In: Nuclear Instruments and Methods in Physics Research Section A. issn: 0168-9002. doi: https://doi.org/10.1016/j.nima.2021.1
3. Narasinga Miniskar, Aaron Young, Frank Liu, Willem Blokland, Anthony Cabrera and Jeffrey Vetter, "Ultra Low Latency Machine Learning for Scientific Edge Applications", Proceedings of the 32nd IEEE International Conference on Field-Programmable Logic and Applications (FPL), pp 411—417, Belfast, UK, Aug 2022, DOI 10.1109/FPL57034.2022.00068.

Acknowledgement

The authors acknowledge Sarah Cousineau for making this grant work possible. The authors are grateful for support from the Neutron Sciences Directorate at ORNL in the investigation of this work. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The Jefferson Science Associates (JSA) operates the Thomas Jefferson National Accelerator Facility for the U.S. Department of Energy under Contract No. DE-AC05-06OR23177. This research used resources at the Spallation Neutron Source, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory. The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan)..

* Contact: Willem Blokland (blokland@ornl.gov)