HARDWARE AWARE ARTIFICIAL INTELLIGENCE (HAAI): PROGRESS ON REALTIME BEAM TOMOGRAPHY RECONSTRUCTION FOR THE FERMILAB RECYCLER USING MACHINE LEARNING AND EDGE PROCESSING

K. Hazelwood*, R. Ainsworth, J. Berlioz, K. Danison-Fieldhouse,
C. Gonzalez-Ortiz, W. Leinonen†, Li Xiang, M.-J. Yang
Fermi National Accelerator Laboratory, Batavia‡, United States
N. Isenberg, M. Jain, J. Strube
Pacific Northwest National Laboratory, Richland, United States

Abstract

The resistive wall current monitor (RWCM) data from the Fermilab Recycler Ring (RR) is used to reconstruct the longitudinal profile of proton beams circulating in the machine. This procedure, commonly referred to as tomography, has proved to be invaluable in tuning the machine. In 2013 Recycler was re-purposed as a proton stacker and charged with implementing slip-stacking to double the intensity for the Main Injector (MI). With two slipping beams, the RWCM data from Recycler is difficult to differentiate and the tomography procedure is slow to compute using traditional means. Building upon past efforts, the Hardware Aware Artificial Intelligence (HAAI) project aims to develop a Machine Learning (ML) model to reconstruct the Recycler beam tomography in real-time and deploy this model on edge hardware. Once developed, this new streaming virtual diagnostic would be used to better track the beam parameters over larger time spans, attribute settings to beam effects, tune the machine, and provide an input into other future automatons of the machines.

INTRODUCTION

Longitudinal tomography is the process of reconstructing the longitudinal phase space profile of bunch(es) of beam using successive turn measurements from a resistive wall current monitor (RWCM).

TARDIS

In 2013 a software client affectionately named TARDIS (Tomography And Related Diagnostics In Synchrotrons) was developed to reconstruct the beam tomography in both Main Injector and Recycler [1]. TARDIS has become an invaluable tool for tuning the machines, giving relatively quick insight into important beam measurements such as injection phase offset, injection energy offset, and overall bunch size and distribution (Fig. 1). While TARDIS has greatly improved our monitoring of the beam longitudinal profile, it does leave some things to be desired. TARDIS relies on RWCM data collected by an oscilloscope. This can take up to a minute, while typical Recycler beam cycles only last a fraction of a second. Furthermore, while pre-made configuration files exist for most beam scenarios, their values often need to be checked against the current running state of the machine, otherwise incorrect measurements occur.

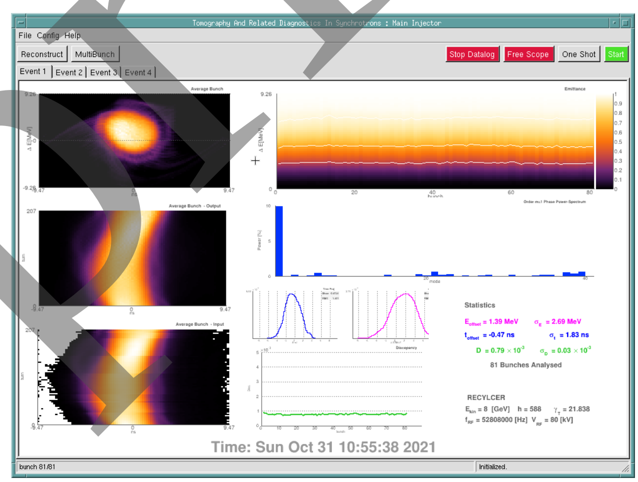

Figure 1: Example of TARDIS output. Left-Center shows the processed averaged bunch RWCM data, Left-Top shows the longitudinal tomography reconstruction.

Recycler Slip-Stacking

The most problematic deficiency of TARDIS is its inability to distinguish between the two slipping beams from the RWCM data. Recycler is now used as a proton stacker for injection into Main Injector and ultimately extraction to the NuMI experiments. Recycler utilizes slip-stacking [2], a process in which batches of beam are injected, decelerated, and then more batches are injected. At the end of slip-stacking, the two momentum beams are quickly accelerated to form double intense batches for extraction to Main Injector. This means that the RWCM at the ideal Recycler frequency samples the in-bucket motion of the on momentum bunch, as well as the streak of decelerated bunches, as the on momentum beam slips past them (Fig. 2).

* kjh@fnal.gov
† Also at University of Glasgow, Glasgow, United Kingdom
‡ This work was supported in part by the U.S. DOE Office of Science, Office of High Energy Physics, under award 84399: "Hardware-aware Artificial Intelligence". Fermi National Accelerator Laboratory is managed by Fermi Forward Discovery Group, LLC, for the U.S. Department of Energy under under Contract No. 89243024CSC000002. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC05-76RL01830.

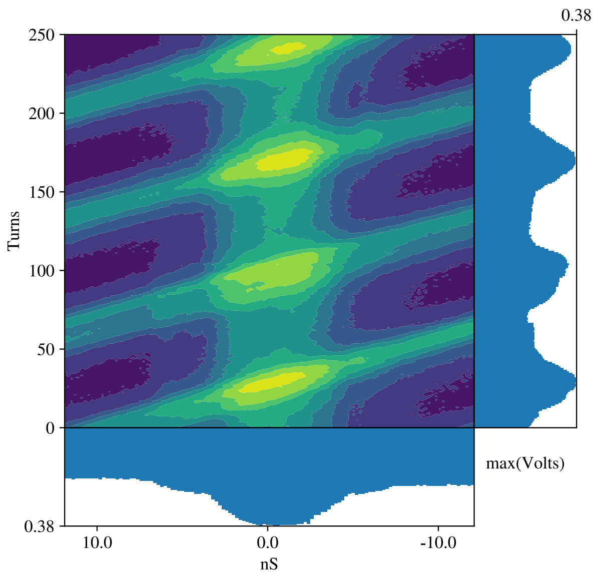

Figure 2: Example Recycler RWCM slipping beams machine data.

TOMOGRAPHY USING MACHINE LEARNING

The motivation for developing ML methods to replace the traditional approach to tomography are three-fold:

1. ML models learn models directly from data, avoiding the need for calibrations
2. Actively developed libraries for ML reduce the time-to-solution, and the maintenance overhead of the developed tool
3. Tooling to deploy ML models on edge hardware like FPGAs is comparatively mature and reduces the need to write specialized code by hand.

Simulations

To create a training dataset from machine data alone that captures most beam conditions of the Recycler would involve prohibitive machine study time and unnecessary beam losses. This fact, coupled with our current tomography systems inability to reconstruct slipping beams, meant beam simulation was the most feasible approach for creating sufficient training data. The simulations were done using BLonD, a beam longitudinal dynamics simulation package developed at CERN. Using BLonD, thousands of combinations of beam parameters and initial beam distributions were simulated for 500 Recycler beam revolutions (turns). The output of the BLonD simulations were transformed into virtual RWCM data with a corresponding first turn tomography image as the truth label.

Prior Work

An earlier attempt at creating a machine learning model to perform tomography on RR slip stacked beam treated the simulated RWCM data as images and trained convolutional neural networks (CNNs) to detect longitudinal properties of the beam. On test simulation data, the model was able to predict the longitudinal reconstruction to the extent that it would be useful for course machine tuning and diagnostics (Fig. 3). However, the CNN size was large, and while it could reconstruct slipped beams, for non-slipped bunch data, it was not markedly faster than the existing TARDIS reconstruction. Also, when used on machine data, the model provided predictions that did not track relative changes applied to the machine. The model would also occasionally hallucinate "ghost" bunches where no beam existed. The effort was stopped during the COVID-19 pandemic.

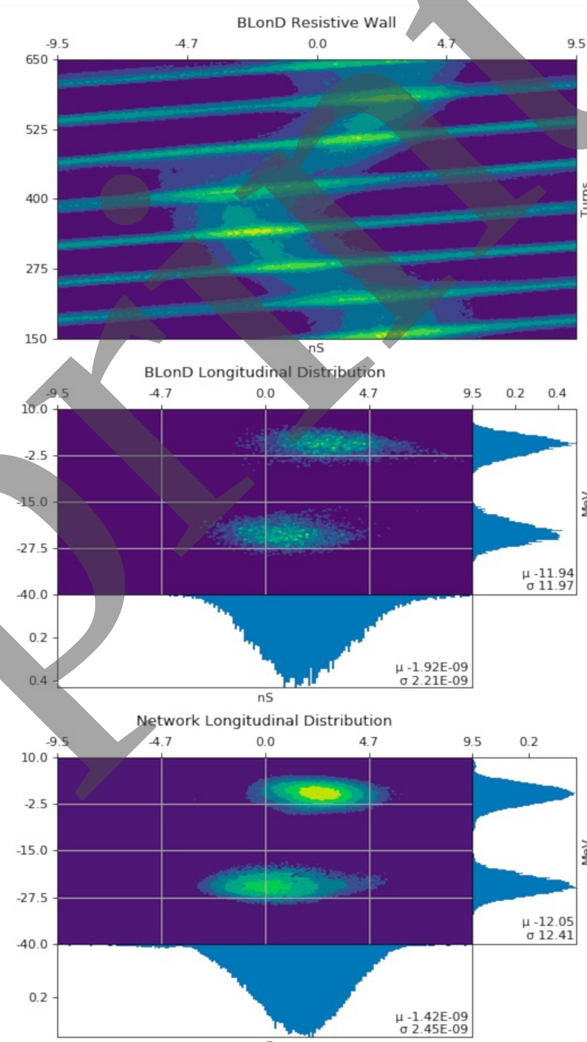

Figure 3: Prior CNN performance

Renewed Effort

Our approach is to treat the inverse problem of reconstructing the longitudinal distribution as a parameter estimation problem that we address with Bayesian methods. As a first step, we train a neural network to learn a reduced representation of RWCM data (using 30 principal components) from a given longitudinal distribution of a single beam, which we simplify to 5 ellipse parameters. Using Bayesian Inference, the network output can be inverted to reconstruct the longitudinal distribution from a given RWCM measurement. Some

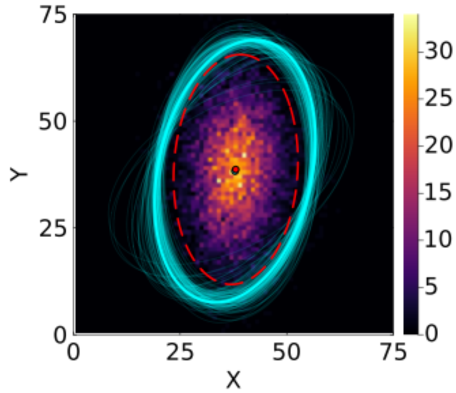

Figure 4: Longitudinal distribution of a simulated beam in the Recycler. Overlaid in a red dashed line is the ellipse fit to this distribution. The predictions from a Bayesian estimator are shown as turquoise circles.

initial results are shown in Figure 4, where the longitudinal distribution is shown together with the ellipse fit and a scan of predictions from a Bayesian estimator, which agree well with each other and show only a small deviation from the ground truth.

Our current ability to predict the 5 ellipse predictors for the center (x_0, y_0), the major an minor axes a and b, and the angle with respect to the x-axis θ is summarized in Table 1. While we are able to predict the center with good accuracy, the predictions for the other ellipse parameters are still outside of an acceptable range for machine tuning.

Table 1: Current Status of the Reconstruction of the Initial Distribution for a Single Beam

	R	RMSE
x_0	0.86	0.74
y_0	0.85	1.60
a	0.50	6.40
b	0.60	4.0
θ	0.02	1.30

Next Steps

To improve the physics relevance and interpretability of the neural network, we will replace the principal component representation of the RWCM with physics-based variables that capture the oscillation of the bunch around the equilibrium in the RF bucket. We expect this to further reduce the number of parameters required to be learned and hence improve the Bayesian Inference of the longitudinal distribution.

MACHINE DATA

To readout the RWCM signal, a Quantum Information Control Kit (QICK) has been purchased and installed [3]. The QICK system allows the RWCM data to be digitized at 5 GS/s and triggered samples to be continuously streamed. The data is written to S3 object stores for model train and test datasets.

Edge Processing

Once an adequate tomography ML model has been developed, plans are to synthesize and deploy this model on an FPGA to provide continuous tomography inferences [4]. The beam distributions and/or scalar measurements from the tomography could be logged over time to track changes in the beam longitudinally and attribute those changes to machine settings or other readings in Recycler Ring and upstream accelerators.

CONCLUSION

Recent work shows that Recycler Ring beam tomography can be solved as as a parameter estimation problem by modeling the longitudinal distribution as a compact set of ellipse descriptors and inferring those parameters from the RWCM data. A neural network surrogate trained on BLonD forward simulations enables fast approximate inversion, and avoids the full reconstruction problem.

Future plans are to extend these methods to simulated Recycler Ring slip stacked data as well as machine data provided by the new QICK readout system.

REFERENCES

[1] N. J. Evans, S. E. Kopp, P. Adamson, and D. J. Scott, "A New Tool for Longitudinal Tomography in Fermilab's Main Injector and Recycler Rings", in *Proc. IPAC'13*, Shanghai, China, May 2013, paper MOPWA061, pp. 816–818. `https://jacow.org/IPAC2013/papers/MOPWA061.pdf`

[2] R. Madrak, "A New Slip Stacking RF System for a Twofold Power Upgrade of Fermilab's Accelerator Complex", *Nucl. Instrum. Meth. A*, vol. 758, pp. 15–25, 2014. doi:10.1016/j.nima.2014.05.004

[3] "Quantum information control kit (QICK)", `https://safran-navigation-timing.com/product/quantum-instrumentation-control-kit/`,

[4] KJ. Hazelwood *et al.*, "Disentangling Beam Losses in The Fermilab Main Injector Enclosure Using Real-Time Edge AI", in *Proc. 19th Int. Conf. Accel. Large Exp. Phys. Control Syst. (ICALEPCS'23)*, Cape Town, South Africa, Oct. 2023, pp. 273–279. doi:10.18429/JACoW-ICALEPCS2023-TU1BCO06