AI/ML Efforts at Fermilab and Plans for Future Accelerator Operations

Tia Miceli (Accelerator Controls | AI/ML Coordinator)

Invited Presentation at the Workshop: AI/ML for Particle Accelerator, X-Ray Beamlines and Electron Microscopy at Argonne National Laboratory

November 1-3, 2021

Fermilab's scientific program

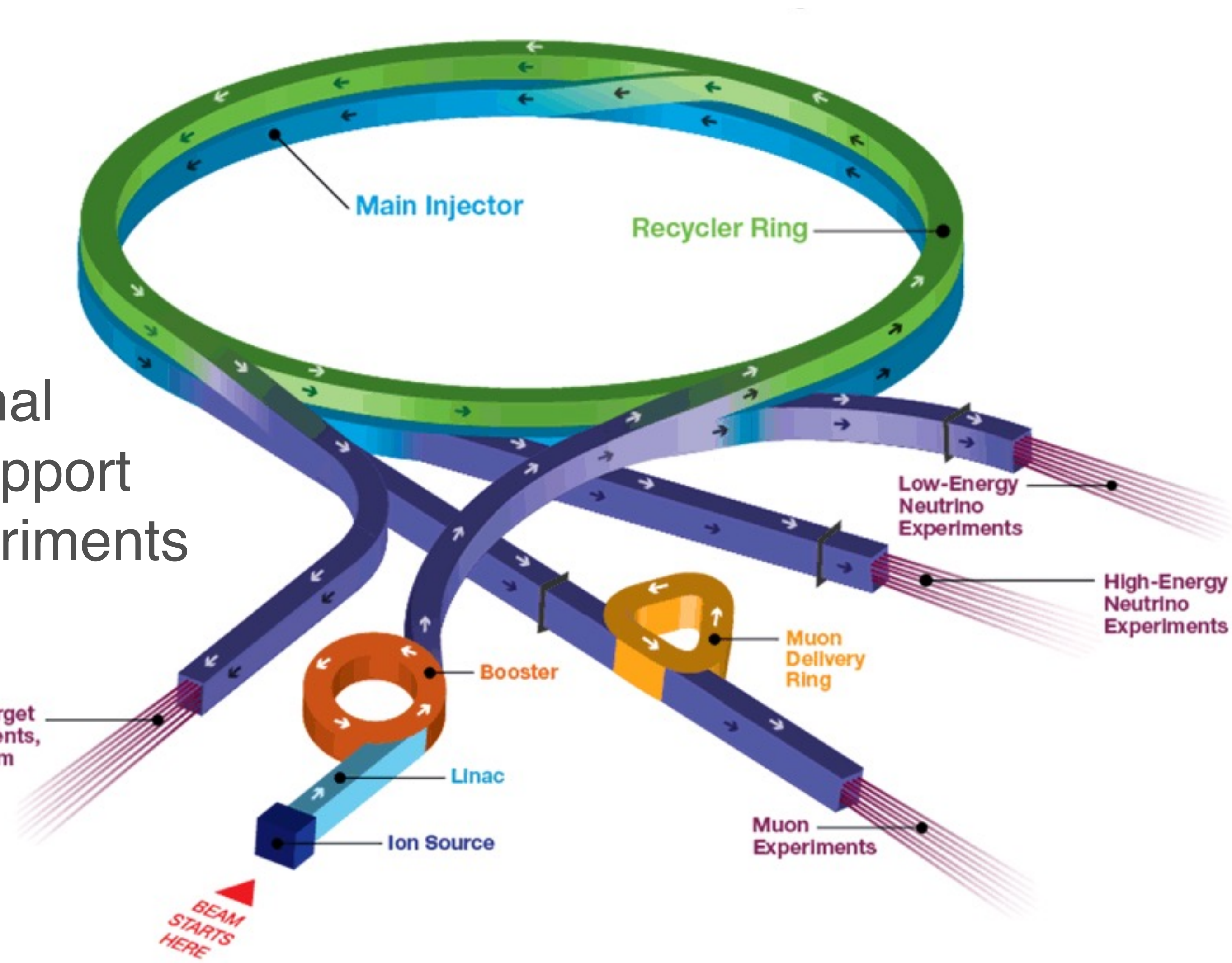
- International short and long-baseline neutrino beamlines
  - Booster Neutrino Beamline [p@8GeV → ν@~1GeV]
  - Neutrinos at the Main Injector [p@120GeV → ν@~8 GeV]
  - New: Long-Baseline Neutrino Facility [same or 60GeV, high flux]
    - Made possible by new linac from Proton Improvement Plan II.
- Muon beamline enhancements
  - Supported g-2 experiment
  - New: Mu2e experiment

The picture can't be displayed.

Fermilab’s accelerator complex

Fermilab's accelerator complex

4 main external beamlines support multiple experiments

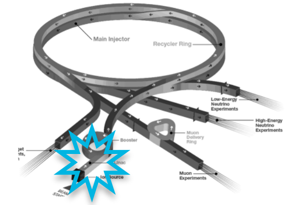


A vision for the future

- Improve beam optimization and automate operations.
- AI / ML will help make this a reality.
  - The following slides are an overview of our latest efforts.

Image Credit: Jacobs Engineering - Pre-conceptual design for a new Main Control Room in a future Center for Accelerator Science and Technology (CAST).

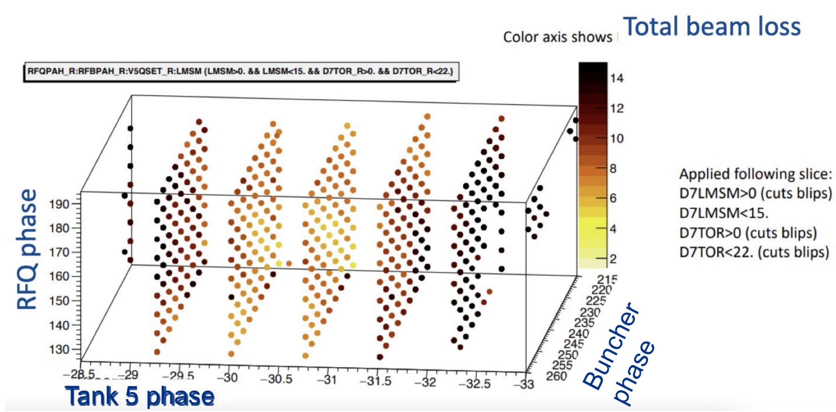
Linac RF phase optimization



Goal: Reduce longitudinal beam emittance and lower overall losses in Linac.

– today: 93% efficiency → push to > 95%

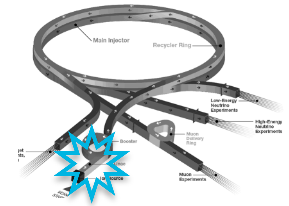
3-phase scan 06/29/2021: D7LMSM



Approach: Build a model that can be inferenced a few times a day using loss and current monitors to adjust phases.

Status: Shallow NN regression model being prototyped.

L-CAPE: Linac conditional anomaly prediction of emergence in collaboration with PNNL



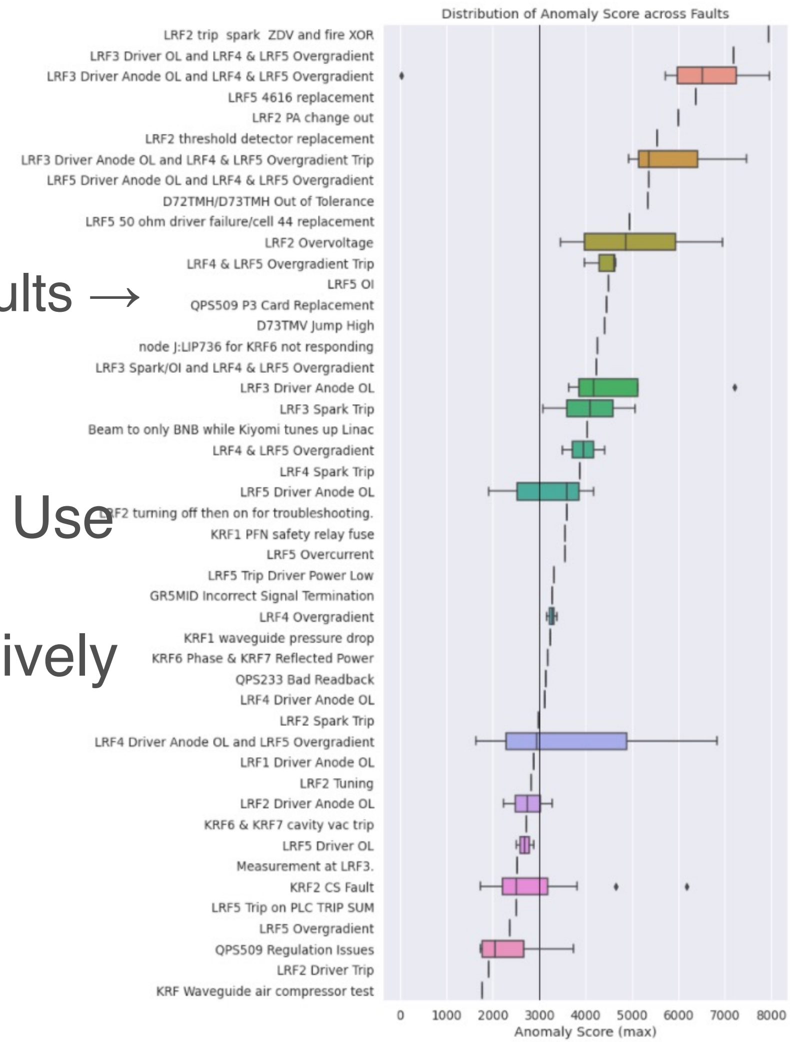
Goal: Improve accelerator operations efficiency to reduce energy use, plan maintenance, and monitor hardware lifetime.

– Today: 8% energy consumed during faults → push to < 4%

Approach: Build time-series models to predict thousands of device readings. Use deviation from observation to form a combined anomaly metric at progressively early lead-times.

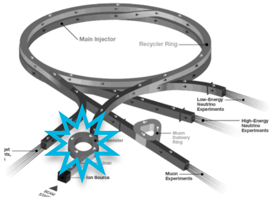
Status: Model pipeline complete (data extraction, transformations, models, performance evaluation).

– Model iterations in progress.



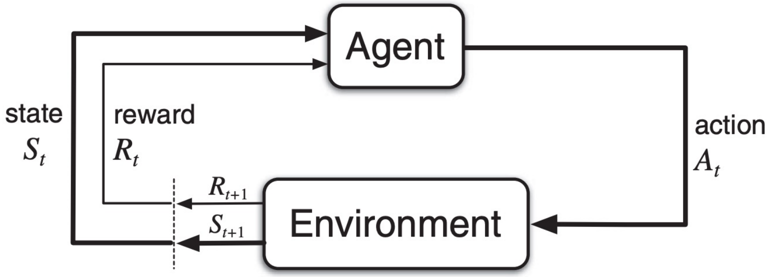
Fermilab

AI GMPS: Precision regulation for the Booster main bends in collaboration with JLab, PNNL, UCSD, and Columbia U



Goal: Precisely regulate gradient magnet power supply to minimize injection field error.

- Today: 1:3,000 variation in current → push to 1:5000+



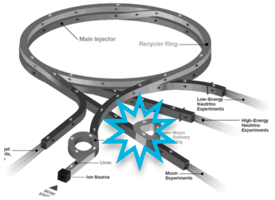
Approach: Train a NN via reinforcement learning using a digital twin to mimic accelerator responses. Deploy on FPGA and replace conventional PID regulation system.

Status:

- Model created: Phys. Rev. Accel. Beams **24**, 104601 – Published 18 October 2021
- Waiting for FPGA to do pre-deployment tests.

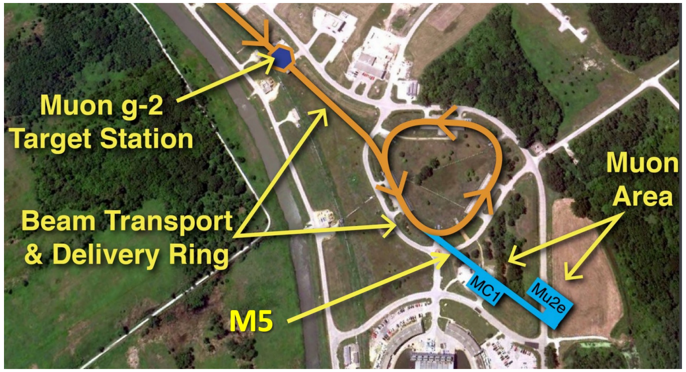
READS: real-time edge AI for distributed systems

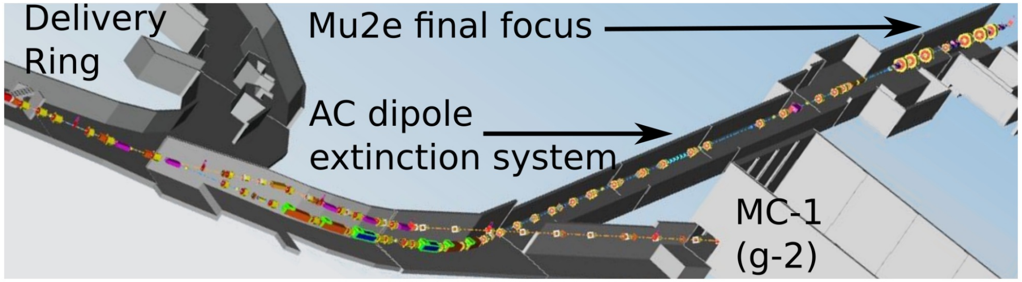
Optimizing mu2e spill regulation system algorithms



Goal: Control beam emittance in real-time (~ms) to smooth the slow extraction spill profile for the Mu2e experiment.

– Today: 50% spill flatness (as designed) → push to > 75%





Approach: Train simple NN in-situ to set optimal gains.

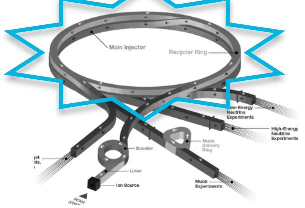
Status:

– Completed simulations (noise reduced up to 66%).

– IPAC2021 Proceedings THPAB243

– Ready for FPGA to test.

READS: real-time edge AI for distributed systems

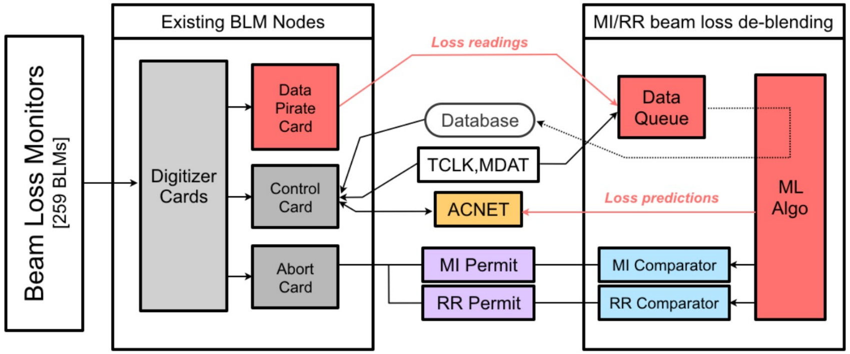


Beam loss de-blending for the Fermilab Main Injector and Recycler

Goal: Attribute beam loss to Main Injector or the Recycler Ring in real-time (~ms). Improve beam up-time.

– Today: 3-5% cycles lost / day → push to 1-2%

Approach: Train NN to be deployed on a central node accepting fast pirate signals from beam loss monitors. Connect to monitoring and revise thresholding.



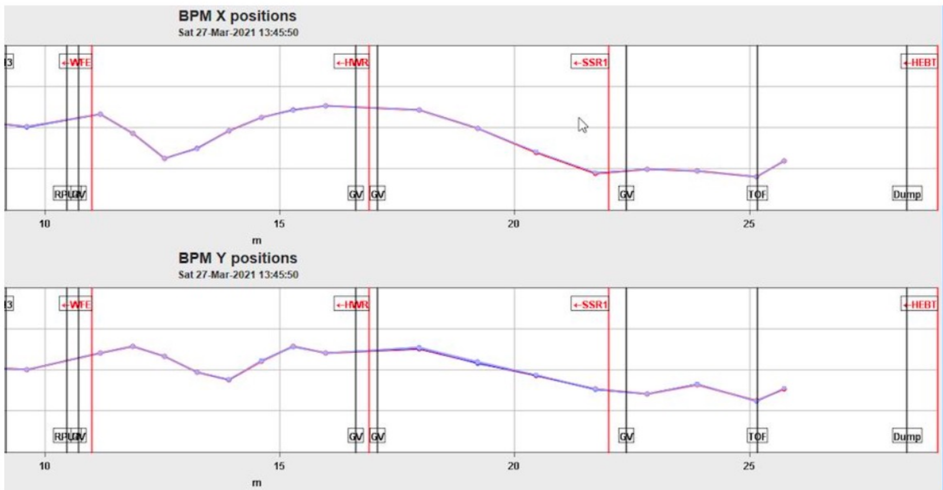
Status: Blended datasets are being formed. Models are being compared.

– IPAC2021 Proceedings MOPAB288

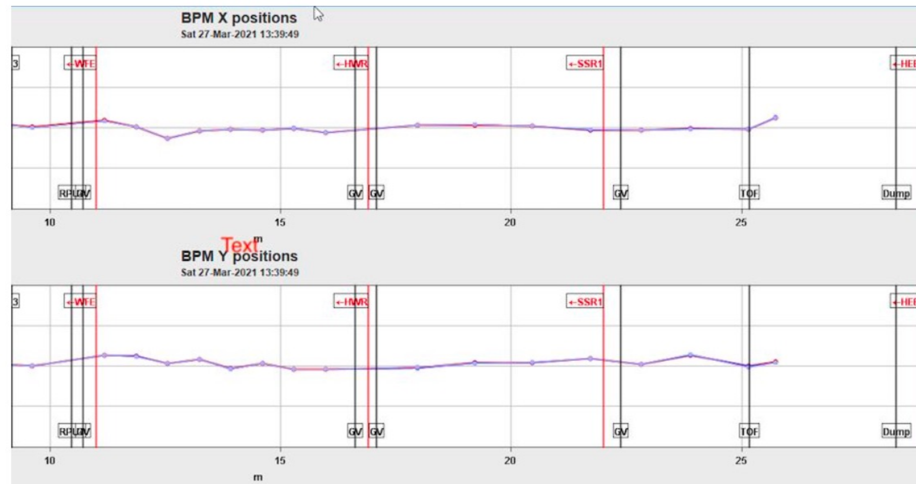
PIP-II Linac beam trajectory correction

Goal: Optimize beam energy and loss quickly, automatically.

Before Optimization



After Optimization



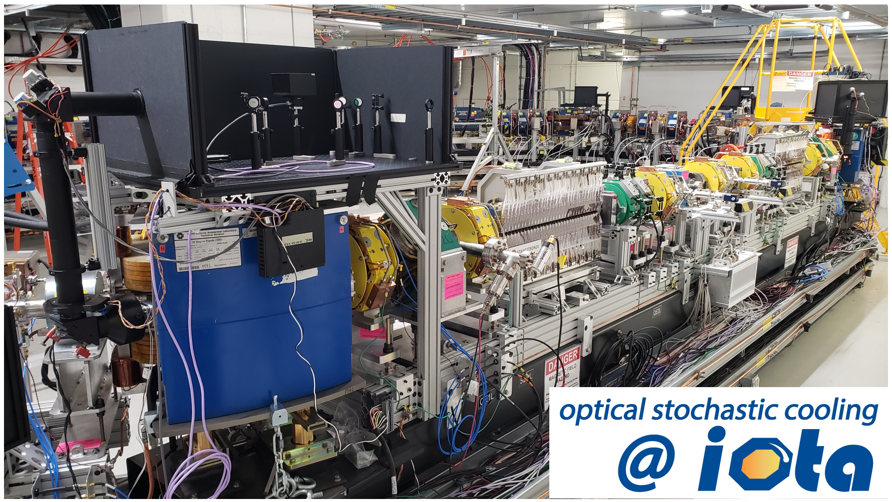
Approach: Bayesian optimization

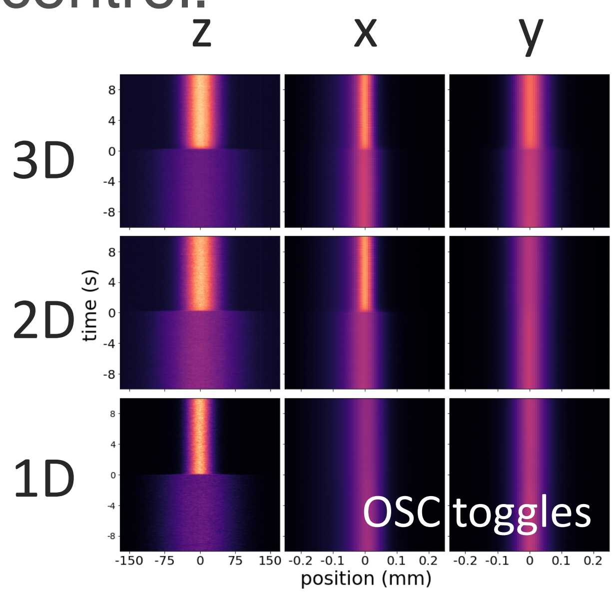
Status: Algorithm tested on TraceWin virtual accelerator.

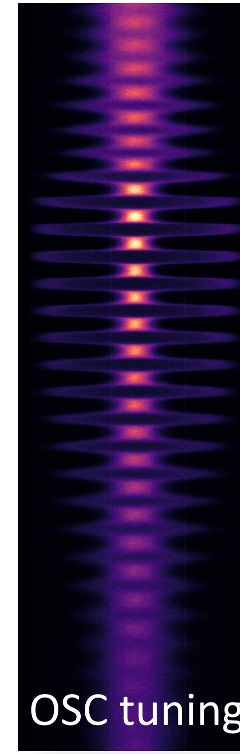
– Ready for hardware testing in this coming quarter.

Next-generation beam cooling and control with optical stochastic cooling (OSC) @ IOTA

Goal: Demonstrate high-gain OSC and use as a flexible, high-bandwidth tool for beam control.





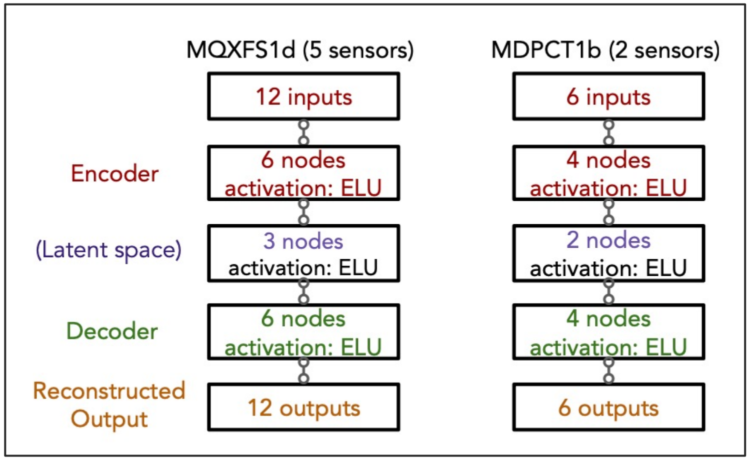


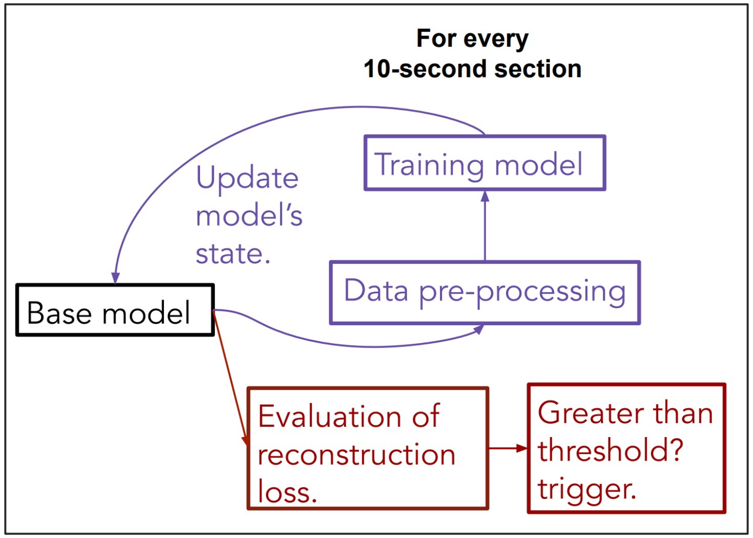
Approach: Turn-by-turn modulations of OSC as an RL action space for developing beam-control policies

Status: World's first OSC demonstrated in 04/21; high-gain OSC program underway with experiments in ~2023

Intelliquench: Superconducting quench prediction

Goal: Use acoustic sensors to predict quenches.





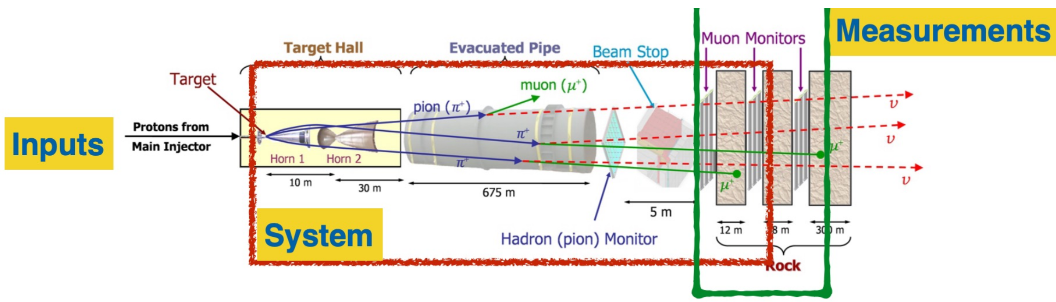
Approach: Train autoencoder on magnet current and acoustic sensors dynamically during current ramps.

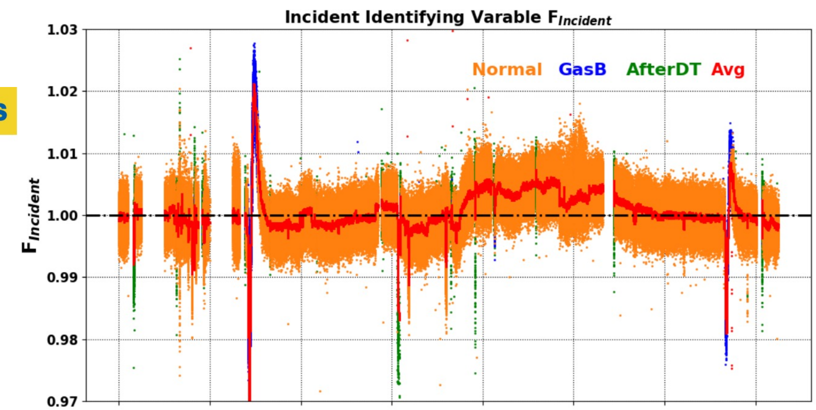
Status: Able to detect 77% of quenches within 15 seconds.

– IEEE Transactions on Applied Superconductivity vol 31, no 5,p1-5, Aug 2021

Neutrino beam monitoring and systematics

Goal: Monitor the NuMI beamline to detect various incidents and understand neutrino beam quality.





Approach: Train simple neural network on expected muon signals. Form an ID variable: ratio with measured signals.

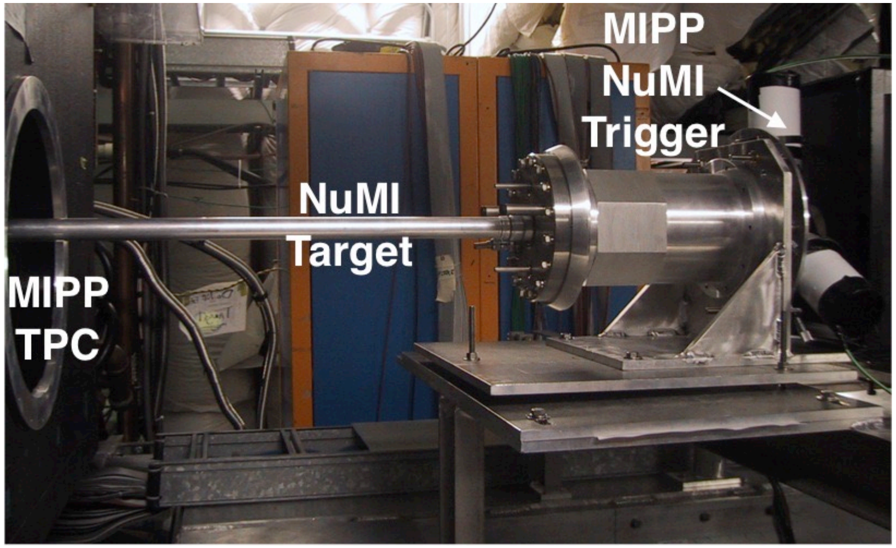
Status: Shown to work, ready to optimize and test.

Robots carrying ML/AI sensors (seeking collaborators)

Goal: Use robots for inspections in locations of radiation or oxygen deficiency hazards.

- visual
- audio
- radiological

Approach: Use computer vision, audio models to detect leaks or strained pumps.



Status: SPOT purchased from Boston Dynamics.

- Determining location precision

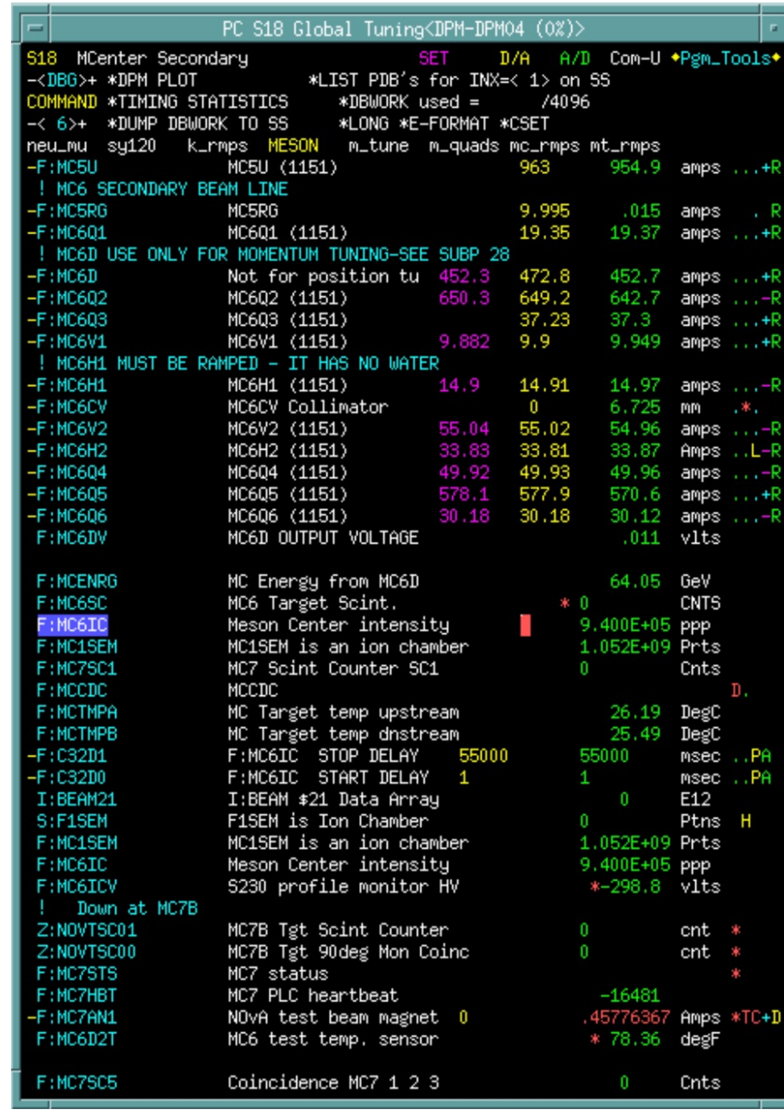
Fermilab

NOICE: Neural Optical Image Categorizer for the E-log

Goal: Search for device readings and settings from screen captures saved in accelerator operations electronic logbook.

Approach: Use transfer learning on an established OCR NN (optical character recognition).

Status: Started summer pre-covid, waiting for new summer students to complete.



Fermilab

Transition from R&D to future operations

- Common hardware needs
  - FPGAs to deploy predictive models in-situ
  - Ability to collect more data for AI applications

- Common data needs
  - Range of sampling frequencies needed
  - Database of settings and beam characterizations
  - Dataset version control

- Common security needs
  - Validated software packages
  - Protected repositories
    - Settings, models, code

Photo credit: Roxanna Tessman of Ashland submitted this photo of an acorn woodpecker for the 2014 Oregon Outdoors Wild Bird Photo Contest.

ACORN Vision Statement

The ACORN Project will modernize the accelerator control system and replace end-of-life accelerator power supplies to enable future operations of the Fermilab Accelerator Complex with megawatt particle beams. The control system will be a unified system that satisfies user requirements, is integrated with and supports operating experiments, is maintainable and adaptable to future needs, and empowers users to achieve their research goals.

Conclusion of Fermilab's Accelerator AI

- We have a variety of efforts underway, and in planning.*
- Collaboration is welcome!

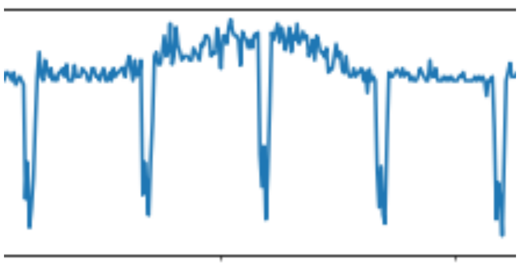
Image Credit: Jacobs Engineering - Pre-conceptual design for a new Main Control Room in a future Center for Accelerator Science and Technology (CAST).

* One of which deserves and explicit call out: Update and deploy Auralee's PIP-II RFQ Resonance Control from the PIP-II test stand.

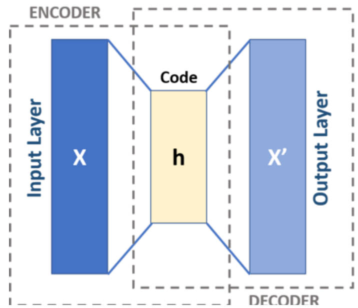
Extra

L-CAPE

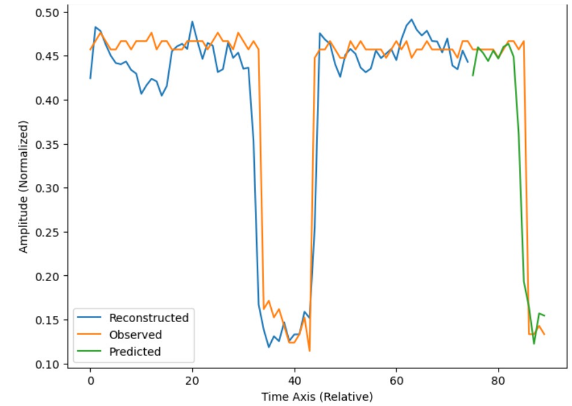
Time-series data from devices

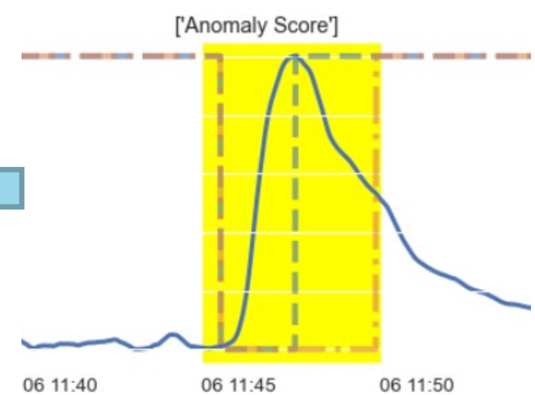


LSTM autoencoder

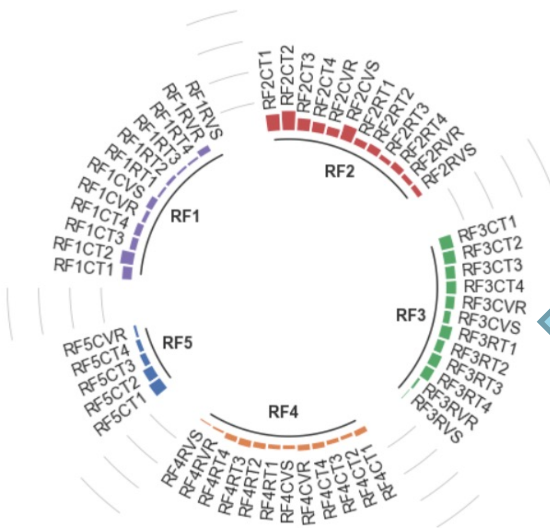


Reconstruction error

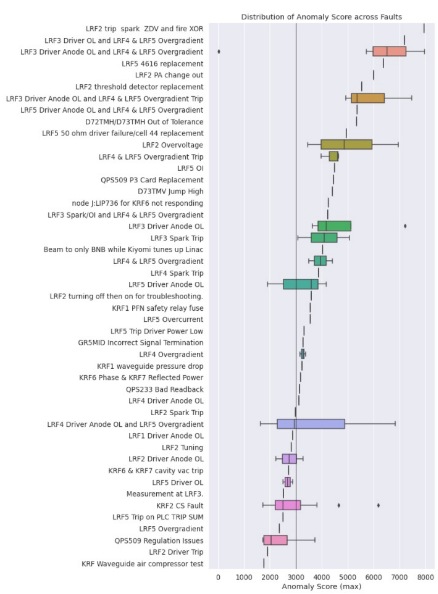




Combined anomaly score from all devices



Breakdown of devices contributing to a given anomaly



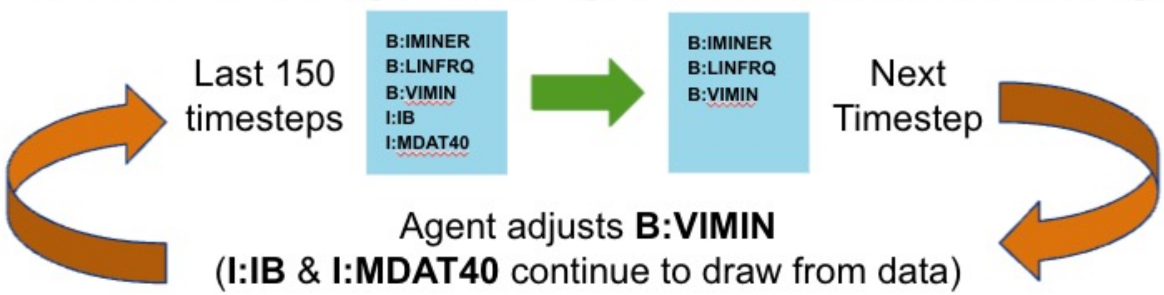
Thresholding to catch anomalies

Fermilab

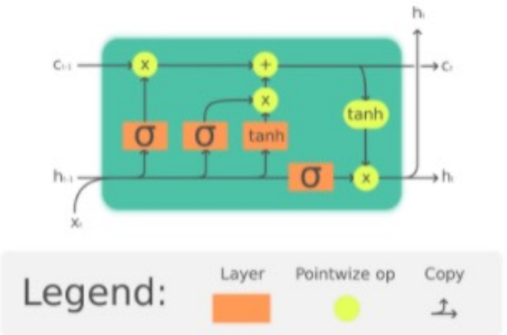
GMPS AI: Generative Multivariate LSTM as Digital Twin

Trained an LSTM to accurately predict next time step.

In "Ouroboros" configuration, this reproduces the learned multivariate dynamics, providing an offline environment to train a control agent through Reinforcement Learning.



Agent adjusts **B:VIMIN**
(**I:IB** & **I:MDAT40** continue to draw from data)



Long Short-Term Memory: A family of Recurrent Neural Network architectures featuring an hidden state, giving ability to learn long-timescale behaviors from data

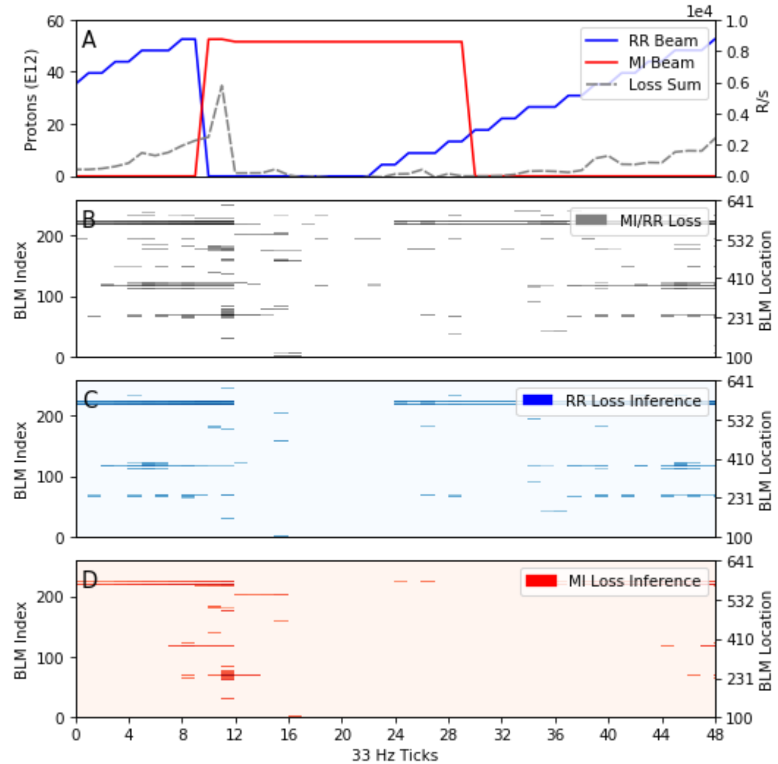
READS: Beam Loss Deblending

Real-time Main Injector and Recycler machine beam loss attribution using Machine Learning

Status

- Sample dataset collected year-round from machine operations and studies data (15-33 Hz)
- Various model architectures are under investigation.
  - Preliminary models have been trained from sample dataset
  - Many show great promise
- FPGA based VME bus reader cards (Pirate Cards) designed and manufactured to stream remote BLM readings at much higher rates.
  - Currently being tested
  - Will ultimately provide final training dataset
- Central real-time FPGA based inference node under development
- Published work progress at IPAC'21, paper MOPAB288

NORTHWESTERN UNIVERSITY



Example of preliminary model inference / beam loss attribution

Fermilab

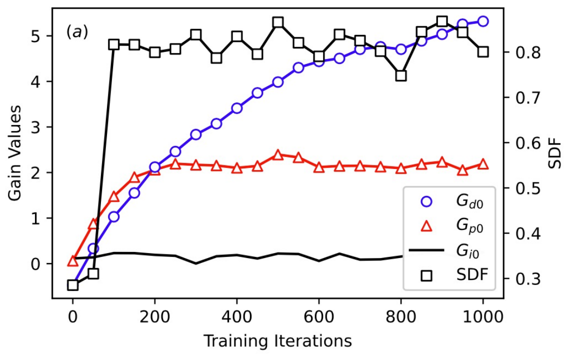
READS: Slow Spill Regulation

NORTHWESTERN UNIVERSITY

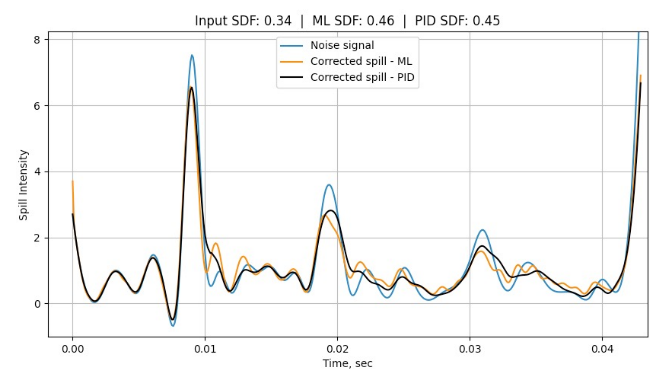
Future Mu2e slow spill regulation using Machine Learning

Status

- Using differentiable slow spill simulation code to generate simulated extractions and corrections
- Have explored two ML regulation schemes
  - PID gain optimization using ML
    - Preliminary model converges on optimal PID gain values
    - Consistently achieves high spill duty factors in simulation
  - Direct ML regulation
    - Preliminary model ingests last n corrected spill observations and generates the next correction
    - Currently performs on-par with standard PID method
- Published work progress at IPAC'21, paper THPAB243



PID gain optimization: Evolution of PID gains during training



Direct ML regulation: Optimized PID regulation vs ML regulator

Fermilab

Outline

- Intro: Accelerator current state (supported beamlines)
- Motivation: Where we want to go (PIP-II, LBNF)
- Goal: What we want operations to look like
  - semi-autonomous operations
  - preemptive mitigation of losses
  - etc?
- Example R&D projects (answer why and how and who)
  - READS shout out to Nhan, the Fermilab AI community leader & ML->FPGA deployment and Northwestern
  - L-CAPE collaboration with Malachi, then Jan et al at PNNL
  - Booster GMPS
  - NuMI beam systematics
  - Robots & 5G
  - FAST / IOTA
- How we get there