Artificial Intelligence at Fermilab

Aleksandra Ciprijanovic - SCD

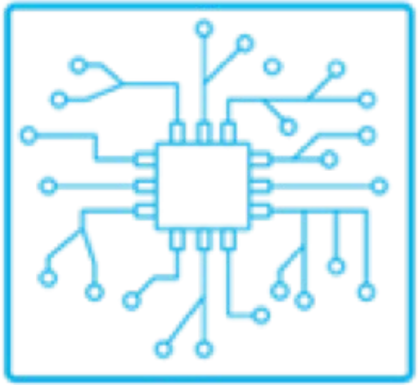
Research Associate

53^{rd} Annual Users Meeting

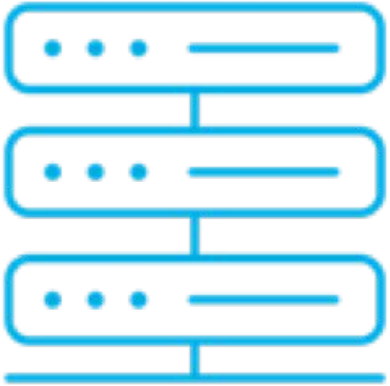
10-13 August 2020.

“Defining the Decade” - Artificial Intelligence

There are two major reasons for AI boom in the last decade:



Computing Power

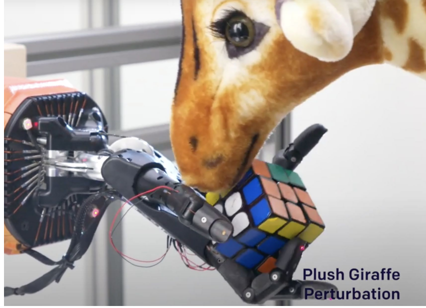


Data Availability

2010

2014

2018



2019

2011

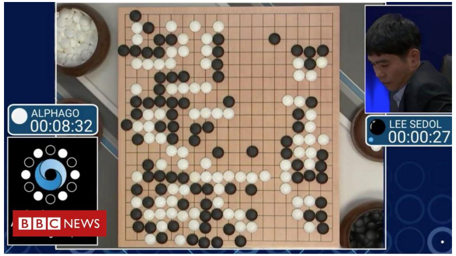
IBM - Watson

2015

TensorFlow

2016

PyTorch

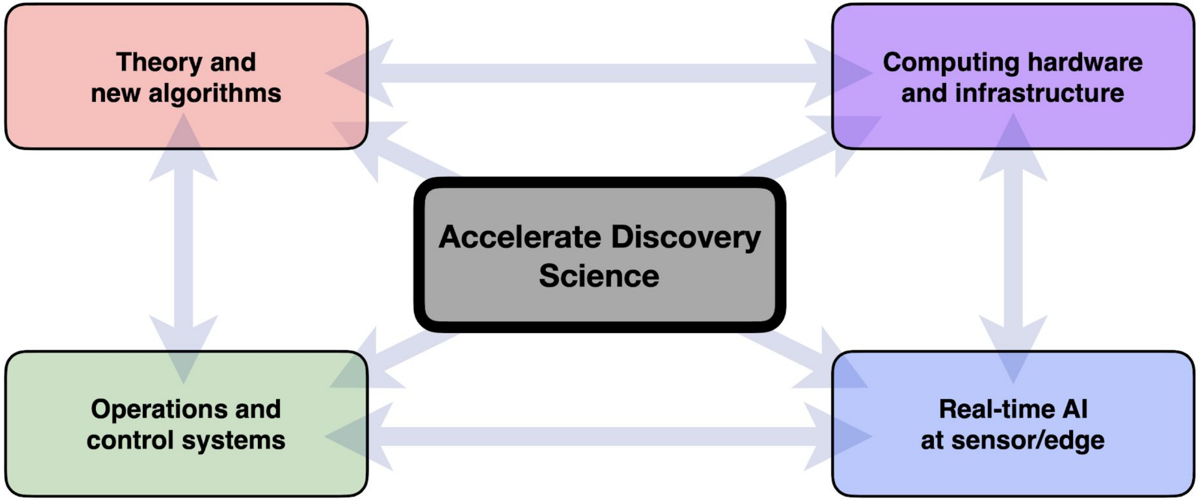


2020

AI capabilities and focus areas

Uncertainty quantification
Learning on graphs

Coprocessors for fast inference
Distributed training



Self operating telescopes
Experiment controls

hls4ml and FPGAs
ASIC codesign

Fermilab - long AI history

FermiNews

November 18, 1988 Vol. XI, No. 21 Fermi National Accelerator Laboratory

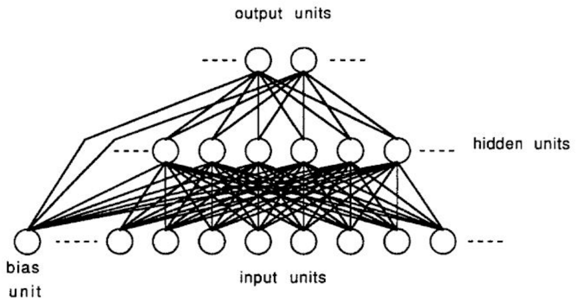
Neural Network Startup

In the past few years, there has been a tremendous resurgence in research on neural networks, the name given to arrays of single-bit, quasi-digital processors whose high level of interconnectivity resembles that of nerve cells in the brain. Neural nets seem to be good at problems that humans solve easily, but that conventional computers are notoriously bad at, such as pattern recognition and decision making based on incomplete or faulty data.

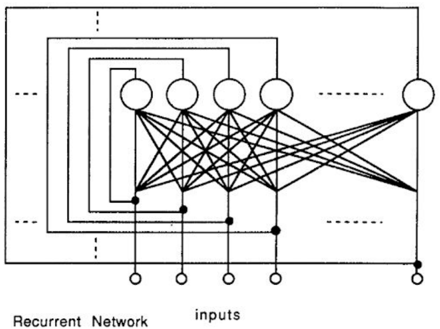
Bruce Denby, who has recently joined the Lab as a Wilson Fellow based in the Computing Department, is beginning a project to explore the possibility of using artificial neural networks and other fine-grained SIMD architecture devices in experimental triggers or offline pattern recognition engines.

Networks implemented in VLSI have demonstrated enormous speedups over conventional microprocessors for certain applications. Also, because of the high redundancy in the interconnection network, neural sets are relatively insensitive to localized faults caused by point defects in silicon substrate or by errors in the data input.

Persons wishing to find out more about neural networks should contact Bruce Denby at FNAL::DENBY or drop a note to him at MS 120. If there is sufficient interest, regular discussion sessions can be set up.



Feed forward neural network



Recurrent Network

B. Denby, “Neural Network Tutorial for High Energy Physicists”, FERMILAB-Conf-90/94, May 1990

The AI Initiative

AI Project

Project office

Primary for strategy and day-to-day operations

Nhan Tran (coordinator), Farah Fahim, Burt Holzman, Brian Nord, Gabe Perdue

ex-officio: Daniel Elvira (AISP), Charles Thangaraj (IARC)

Tech Division	Quantum	Detectors
Acc. Division	Cosmic	Theory
WDRS/FESS/...	Intensity/v	Communications
IARC	Energy	
Users/UEC	SCD/Infra.	

+

Liaisons as links across the laboratory

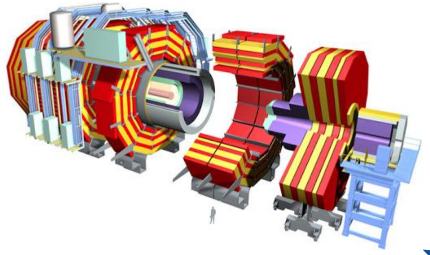
Scientific Computing Division

Cross-cutting projects and initiatives

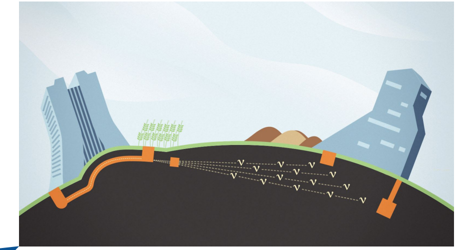
AI Projects

Formal home in SCD but engaging the entire laboratory.

Fermilab



Accelerators

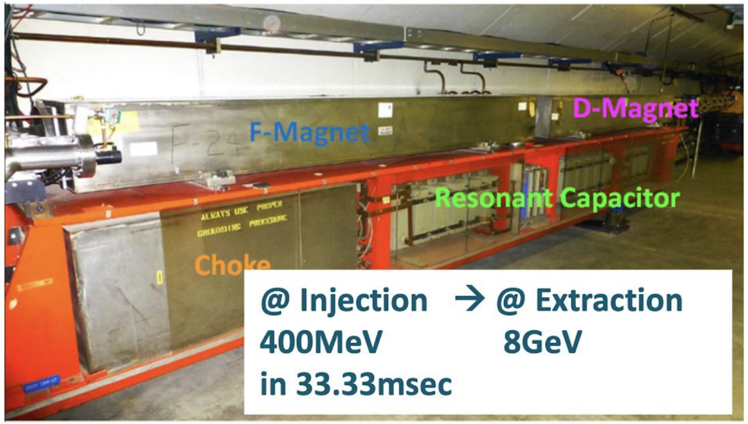


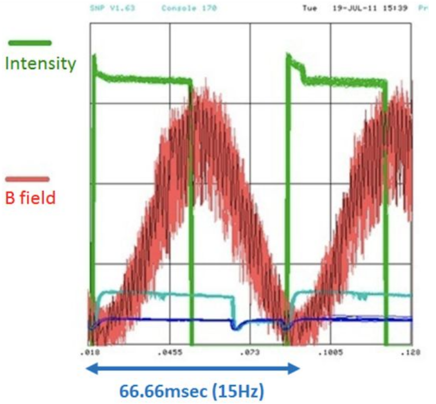
Netrunos

Astronomy

Quantum Science

AI for Accelerator Control





- **Goal** - reduce beam losses in the Booster.
- **Approach** - using a reinforcement learning algorithm on a custom FPGA board to control the gradient magnet power supplies (GMPS).
- **The scope** - single crate control system (one board with back-up), but the project lays the foundation for a more ambitious future program.
- **Achievements so far**:
  - Significant progress in adopting **hls4ml** tools to Intel chipsets (optimizing resource allocation on the FPGA).
  - Proof-of-principle "agent" trained a reproduction of the current (accelerator domain expert-built) system using a surrogate model trained using historical data from the accelerator complex.

Quantum Computing for ML

- QML is the **use of quantum resources for machine learning problems**.
- Studying quantum algorithms for object classification in astronomy.
- **Goals -** hardware co-design, building up the software stack, characterizing the performance of modern hardware, etc.
  - Lockheed Martin: “Studying the use of quantum annealers for machine learning”
  - https://arxiv.org/abs/1911.06259

FERMILAB-PUB-19-546-QIS-SCD

Restricted Boltzmann Machines for galaxy morphology classification with a quantum annealer

João Caldeira,[1,*] Joshua Job,[2,*] Steven H. Adachi,[2] Brian Nord,[1,3,4] and Gabriel N. Perdue[1]

[1] *Fermi National Accelerator Laboratory, Batavia, IL 60510*
[2] *Lockheed Martin Advanced Technology Center, Sunnyvale, CA 94089*
[3] *Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637*
[4] *Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637*
(Dated: February 17, 2020)

  - Alphabet X (with U. Waterloo): studying gate-based ("universal") quantum kernel methods using QC built by Google (their supremacy chip)
- *Funded through DOE QuantISED*

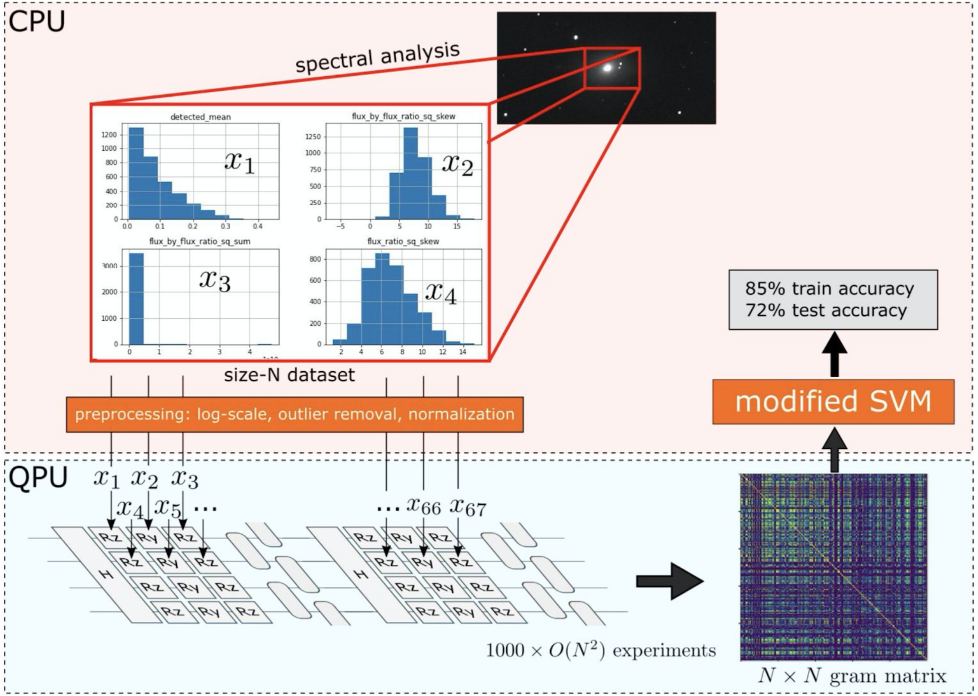
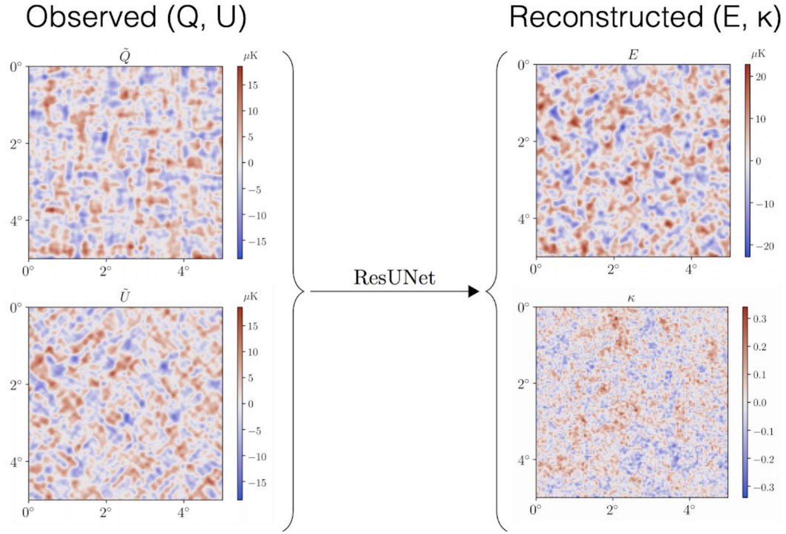


Figure by E. Peters (Waterloo)

AI for Astronomy



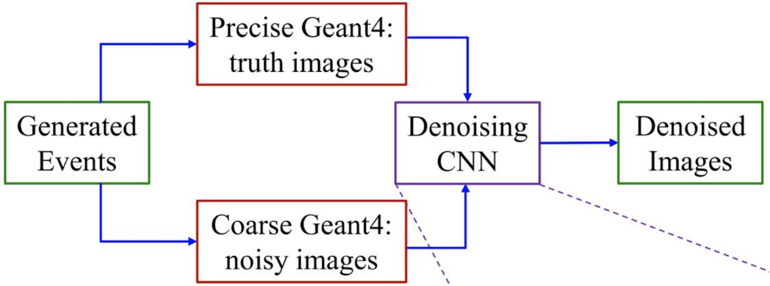
Noise and other foregrounds obfuscate primordial GW signatures
Pioneered use of **Residual UNets** to separate lensing signals (κ) from CMB polarization map (E)

Caldeira et al. 2019, Astronomy & Computing, 28, 100307

Many projects, joint with several other institutions (UChicago, ORNL, UMichigan, Argonne):

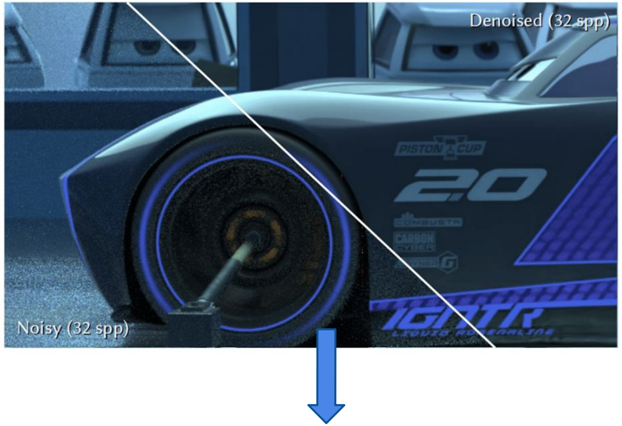
- **ML helping study astronomical objects in the big data era** (large surveys - DES, SDSS, LSST, SPT etc.) - strong lens finding and measuring; galaxy mergers: classification and domain adaptation; CMB lensing; galaxy-cluster-finding; solar flare classification; Extrasolar planets.
- **ML helping automation**: self-driving telescope; artifact classification in DECam images.
- **ML and uncertainty quantification** in astronomy.

AI for CMS

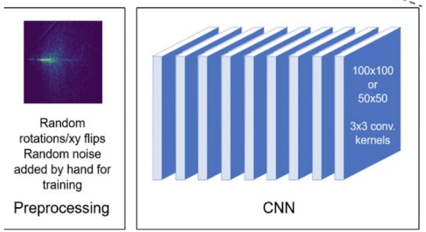


Pixar

ACM Trans. Graph 36 (2017) 97

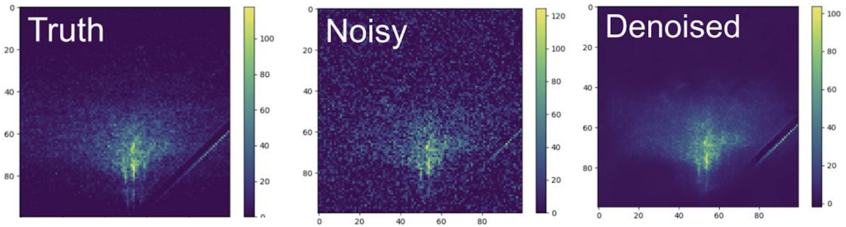


- Detector simulation w/ Geant4 is accurate but computationally intensive.
- Tune parameters for faster, but less accurate output.
- Use Denoising CNN to restore output quality.



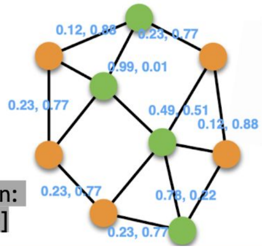
Target: 1 color = energy deposited (normalized: $c' = (c - \mu_c)/\sigma_c$)
CNN: 9 layers, 100 features/layer
Patch-based ℓ_1 loss function

CMS



- Event reconstruction using Graph NNs:
arXiv:1801.07829
arXiv:2003.11603
arXiv:1902.07987

Edge classification: [is sigal, is noise]



Fermilab

AI for Neutrinos

- ML for **faster workflow and enhanced event reconstruction.**
- A publication is being prepared: "GPU-accelerated machine learning inference as a service for accelerator neutrino experiment computing".
- In the future implement the process of FPGAs.

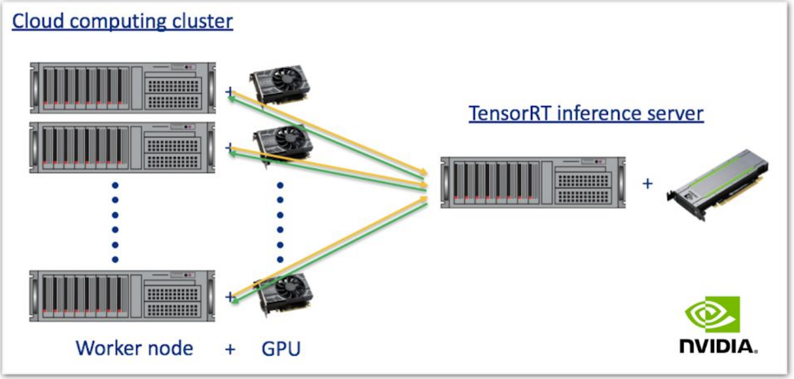
Original Article | Published: 14 October 2019

FPGA-Accelerated Machine Learning Inference as a Service for Particle Physics Computing

Javier Duarte, Philip Harris, Scott Hauck, Burt Holzman, Shih-Chieh Hsu, Sergo Jindariani, Suffian Khan, Benjamin Kreis, Brian Lee, Mia Liu, Vladimir Lončar, Jennifer Ngadiuba, Kevin Pedro, Brandon Perez, Maurizio Pierini, Dylan Rankin, Nhan Tran, Matthew Trahms, Aristeidis Tsaris, Colin Versteeg, Ted W. Way, Dustin Werran & Zhenbin Wu

Computing and Software for Big Science **3**, Article number: 13 (2019) | Cite this article

Tingjun Yang, Maria Acosta, Phil Harris, Ben Hawks, Burt Holzman, Jeff Krupa, Kevin Pedro, Nhan Tran, Mike Wang



Tot: 227 s/evt
EmTrkMichelId: 142 s/evt
GausHit: 13.9 s/evt

→

Tot: 98.9 s/evt (**~2.3x speedup**)
EmTrkMichelId: 10.1 s/evt (**~14x speedup**)
GausHit: 14.0 s/evt

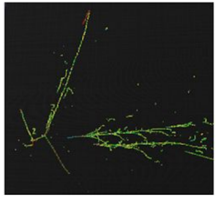
https://arxiv.org/abs/1912.10133

Enhancing Neutrino Event Reconstruction with Pixel-Based 3D Readout for Liquid Argon Time Projection Chambers

C. Adams[a] M. Del Tutto[b,c] J. Asaadi[d] M. Bernstein[c] E. Church[e] R. Guenette[c] J. M. Rojas[c,1] H. Sullivan[d] A. Tripathi[d]

[a]*Argonne National Laboratory, Lemont, IL*
[b]*Fermi National Accelerator Laboratory, Batavia, IL*
[c]*Harvard University, Cambridge, MA*
[d]*University of Texas-Arlington, Arlington, TX*
[e]*Pacific Northwest National Laboratory, Richland, WA*

E-mail: corey.adams@anl.gov

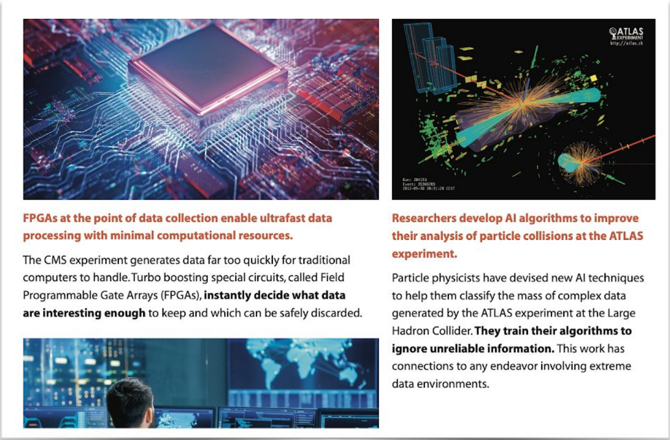


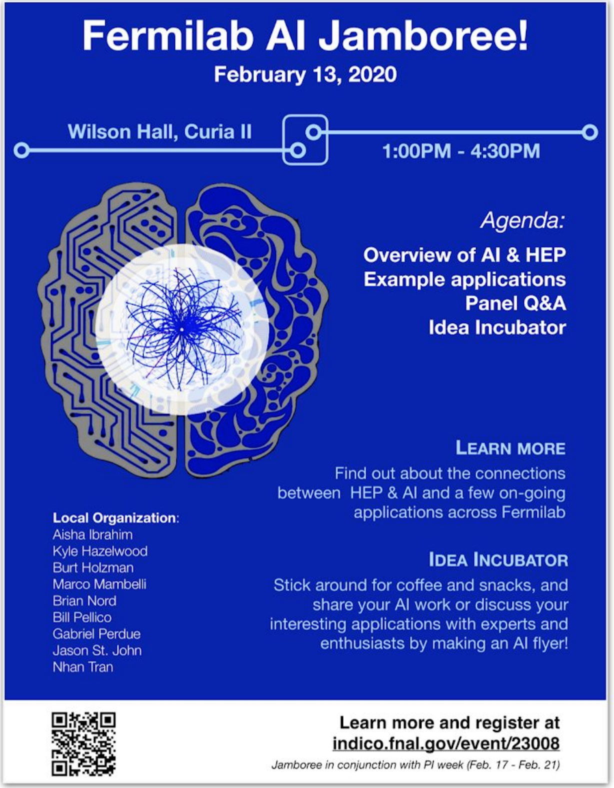
Fermilab

Other activities

- First community event - AI Jamboree
- Collected 20+ AI “flyers” on ideas
- Hands on ML tutorials

- AI seminar series
- Webpage redesign





Thank you!

Home in SCD but involves **all science groups**

Real-time AI - fast triggering, fast control, nanosecond scale inference

Collaboration with other labs

AI on **ASICs** and FPGA

Working with **huge datasets**

Cross-cutting new technologies

Discovery science

Theory and **new algorithms**