Summary of Computing+Machine Learning Parallel Session

CPAD 2018

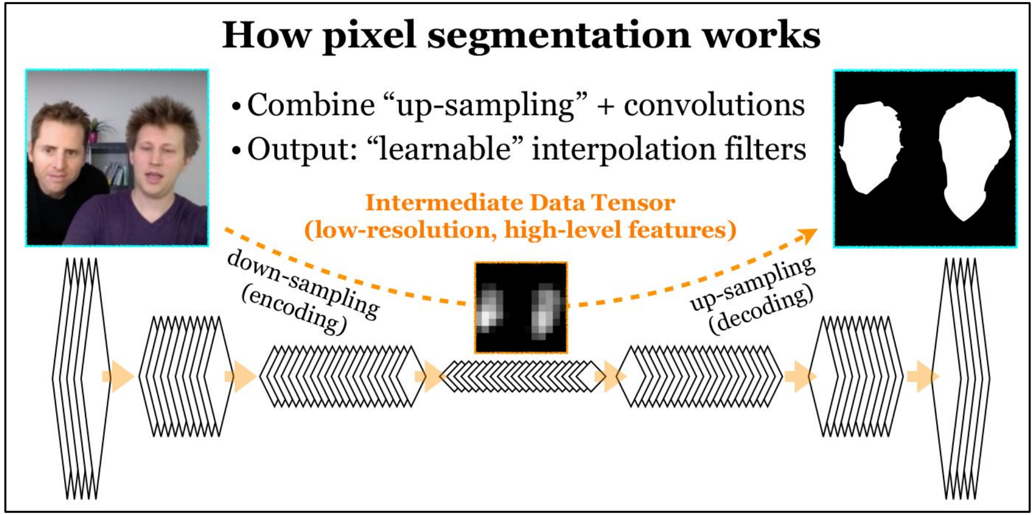
T. Wongjirad (Tufts)
on behalf of the session conveners
S. Gleyzer, O. Gutsche

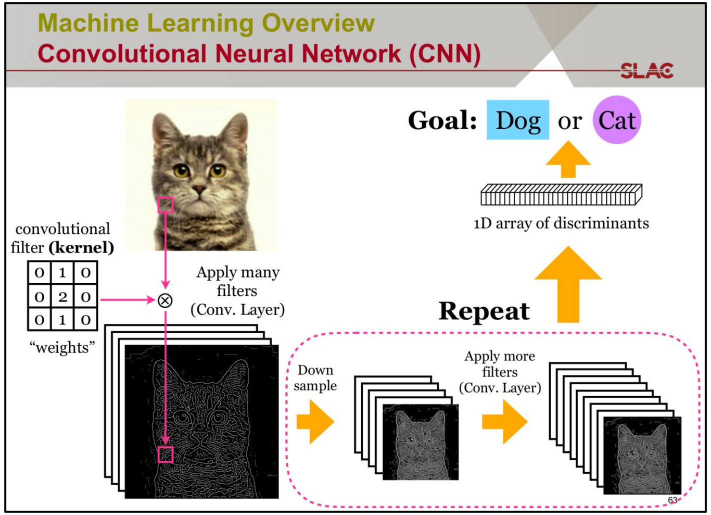
Computing & Machine Learning

- Three parallel sessions
- 12 Talks (+1 plenary session)
- Many thanks to all of the speakers for presenting an exciting diverse set of talks
- And thanks to the audience for questions and discussions
- Due to time -- please forgive me if I missed names and slide links

Adoption of Deep Neural Networks

- Some quick descriptions of these algorithms (backups have more details) can be found in K. Terao’s talk
- https://indico.fnal.gov/event/18104/session/23/contribution/77





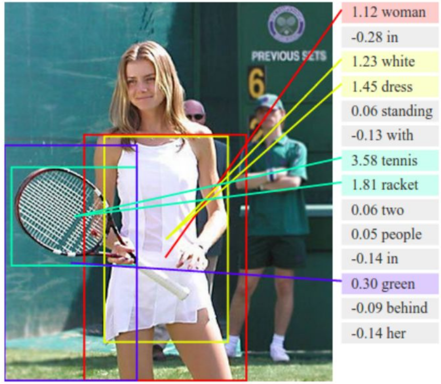
Deep Neural Networks

- Deep neural networks featured in majority of talks
- Such algorithms, e.g. convolutional neural networks, have enabled advances in fields such as computer vision -- bringing techniques to particle physics

Object detection and instance segmentation

Caption generation/scene parsing

"girl in pink dress is jumping in air."

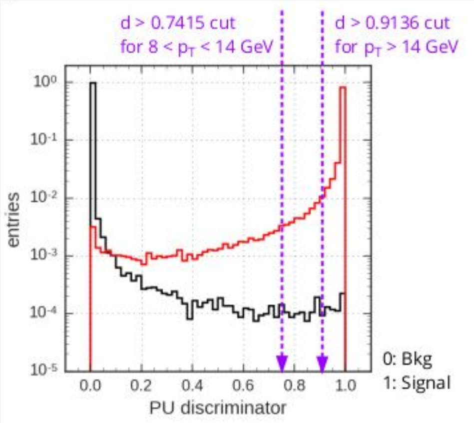


Themes

- Algorithm development in a variety of contexts
  - LHC: Jet ID, track recon.
  - Liquid argon neutrino experiments
  - Cherenkov/scintillator neutrino experiments
  - Adoption of machine learning techniques
- Deployment of algorithms beyond CPUs
  - Deploying Machine Learning Networks on FPGAs
  - Track reconstruction using quantum annealers
- Bringing Everything Together
  - Deploying into online systems
  - Organizing training in computing, ML, etc. in order to maintain the rapid pace of development of new tools

LHC Applications

- Particle ID: quark/gluon, electron/photon, and more.
- L1 muon reconstruction using NN on FPGAs

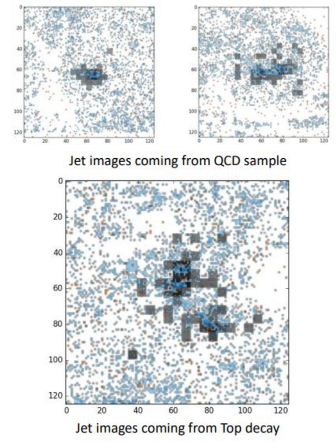


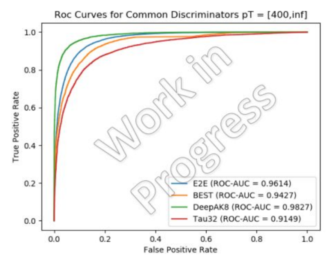
Jia Fu Low et al.
(Monday afternoon)

- From simulation studies, we have achieved roughly 4x rate reduction with the addition of the new muon detectors. The overall efficiency has also been improved.
  - Still work in progress.

https://indico.fnal.gov/event/18104/session/23/contribution/71/material/0/0.pdf

B. Burkle et al. (Tues morning)



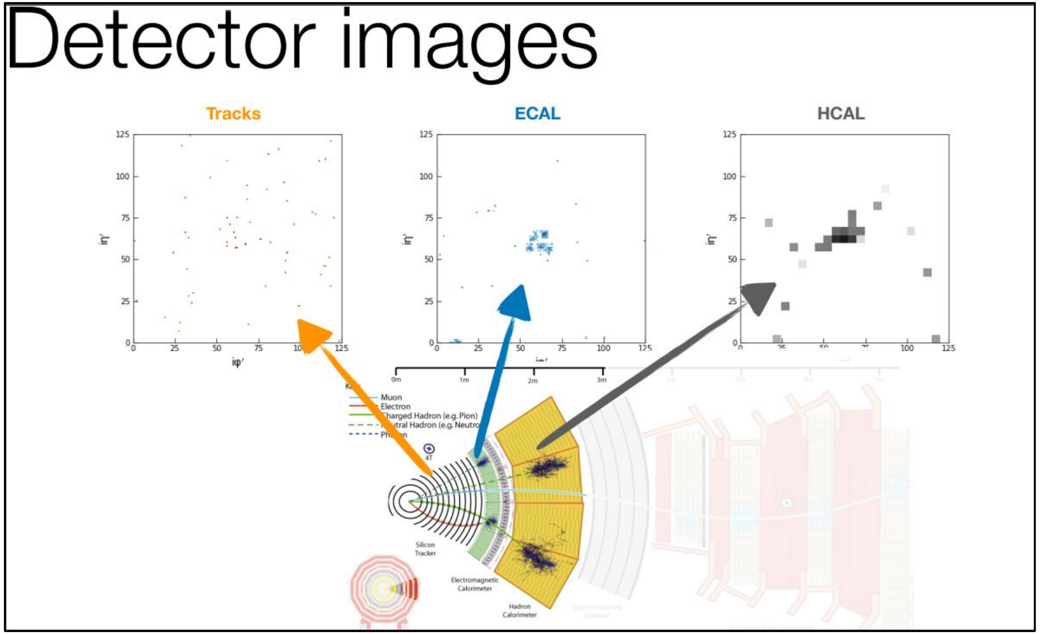


https://indico.fnal.gov/event/18104/session/23/contribution/147/material/0/0.pdf

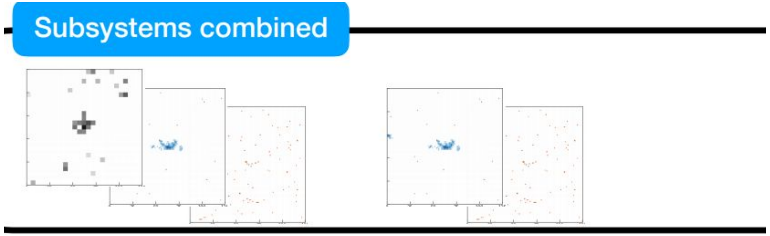
Example: End2End method

Simultaneously provide images of energy depositions from several subsystems to convolutional neural networks

E Usai et al (Monday afternoon)



Example: quark vs. gluon jet

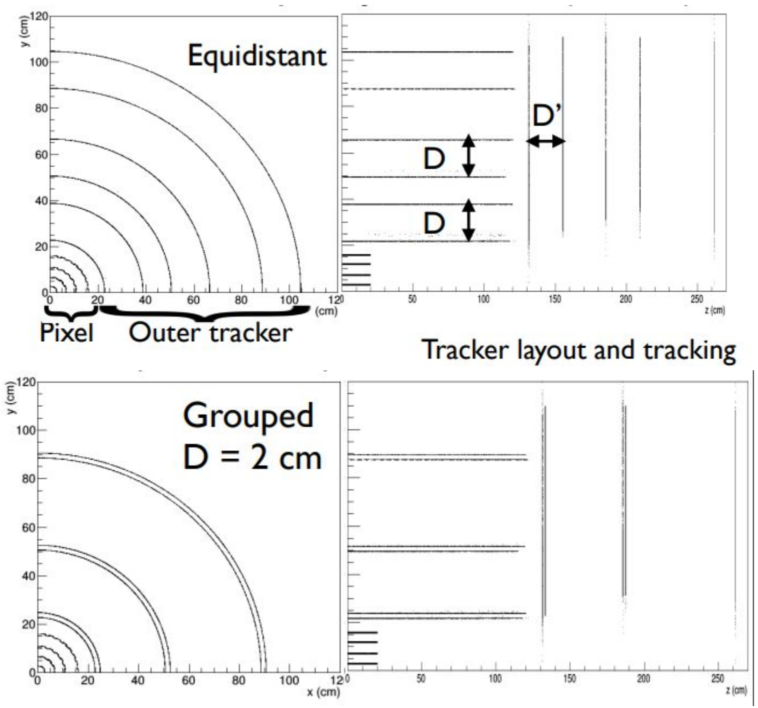


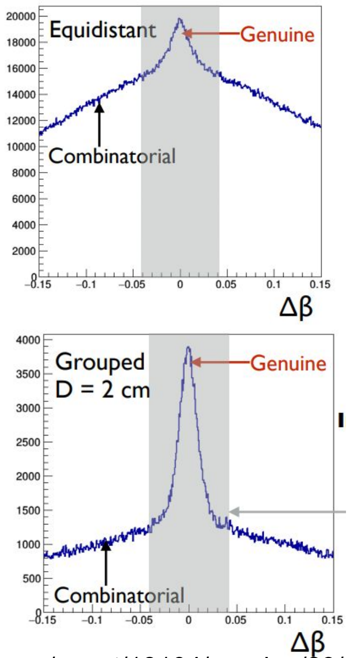
	ROC AUC
E2E image, ECAL+HCAL+Tracks	**0.8077 ± 0.0003***
RecNN, ascending-p_T	0.8017 ± 0.0003*
RecNN, descending-p_T	0.802
RecNN, anti-k_T	0.801
RecNN, Cambridge/Aachen	0.801
RecNN, no rotation/reclustering	0.800
RecNN, k_T	0.800
RecNN, k_T-colinear10-max	0.799
RecNN, random	0.797

https://indico.fnal.gov/event/18104/session/23/contribution/75

LHC Applications -- Not all machine learning

- Studying the impact of the tracker detector layout on CPU usage on reconstruction time and software





V. KRUTELYOV et al.
(Monday)

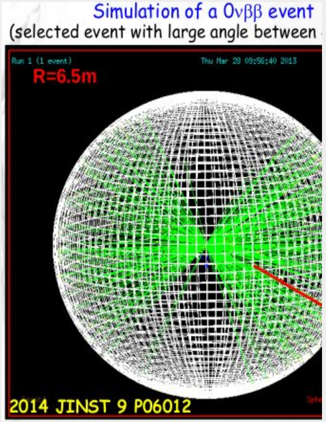
One example:
false tracks reduced

Translates to reduced CPU wall time

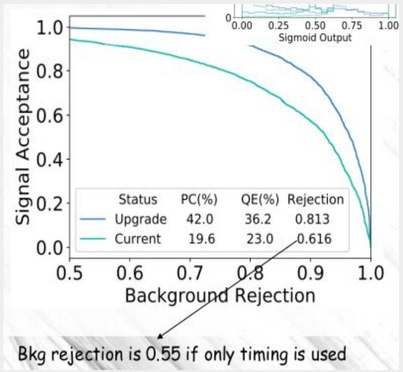
https://indico.fnal.gov/event/18104/session/23/contribution/50/material/0/0.pdf

Neutrino Experiments

- Heard from applications to liquid argon TPCs (GeV events) to Chenenkov/Scintillator detectors (MeV events)

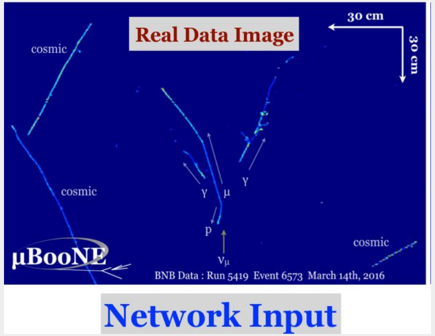


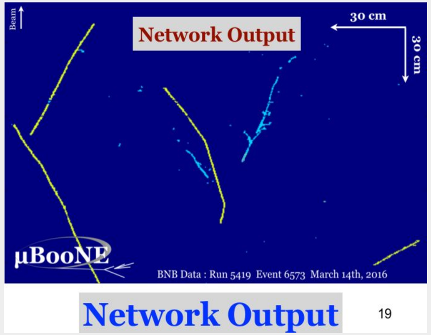
A. Elagin (Sunday)



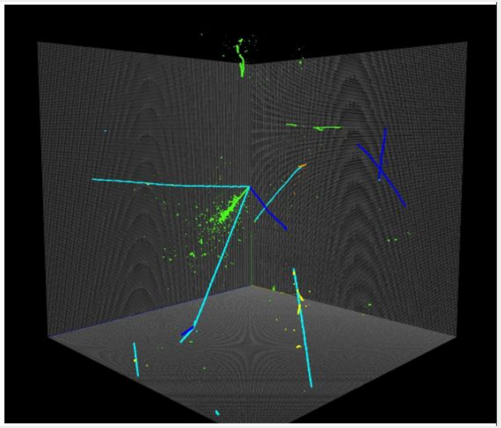
CNN improves rejection of certain class of backgrounds (photons) for neutrinoless double-beta searches

https://indico.fnal.gov/event/18104/session/23/contribution/68



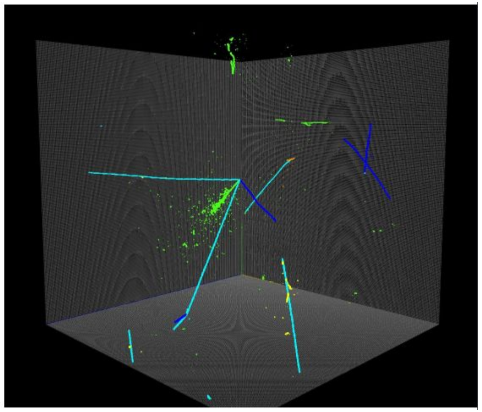


Pixel-level particle ID (on data)
For LArTPC images (uBooNE)



3D hit-level particle ID for LArTPCs. ~99% accuracy demonstrated on simulation

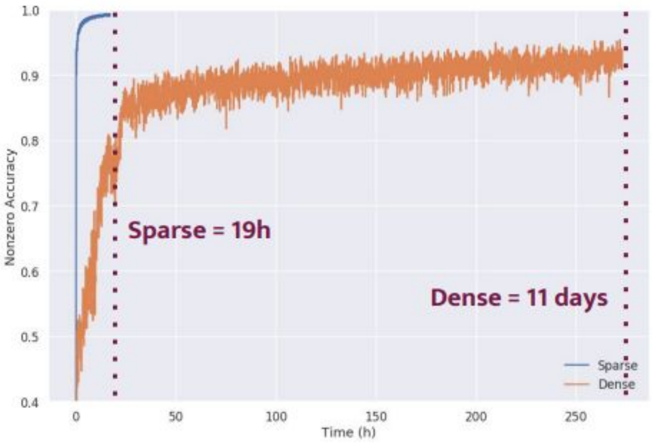
Sparse Operations



- CNN operation on 3D space-points enabled by **using sparse matrix operations rather than dense operations** -- much better fit to liquid argon TPC data (and particle physics data)

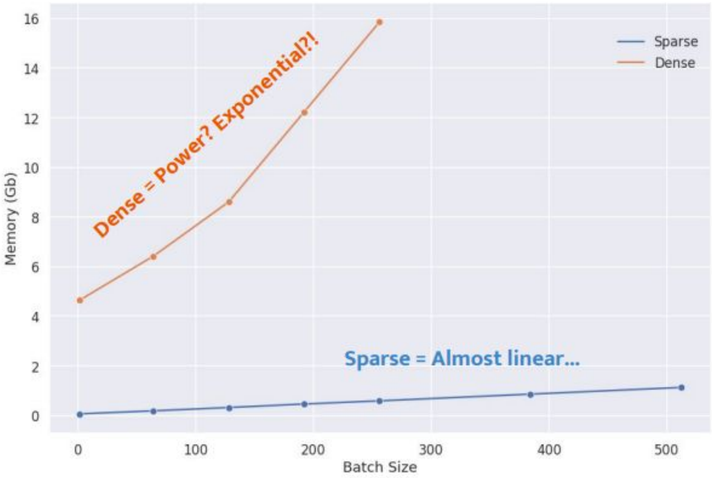
L. Domine et al (Tues Morning)

Training speed up



Nonzero Accuracy (training) vs Wall Time

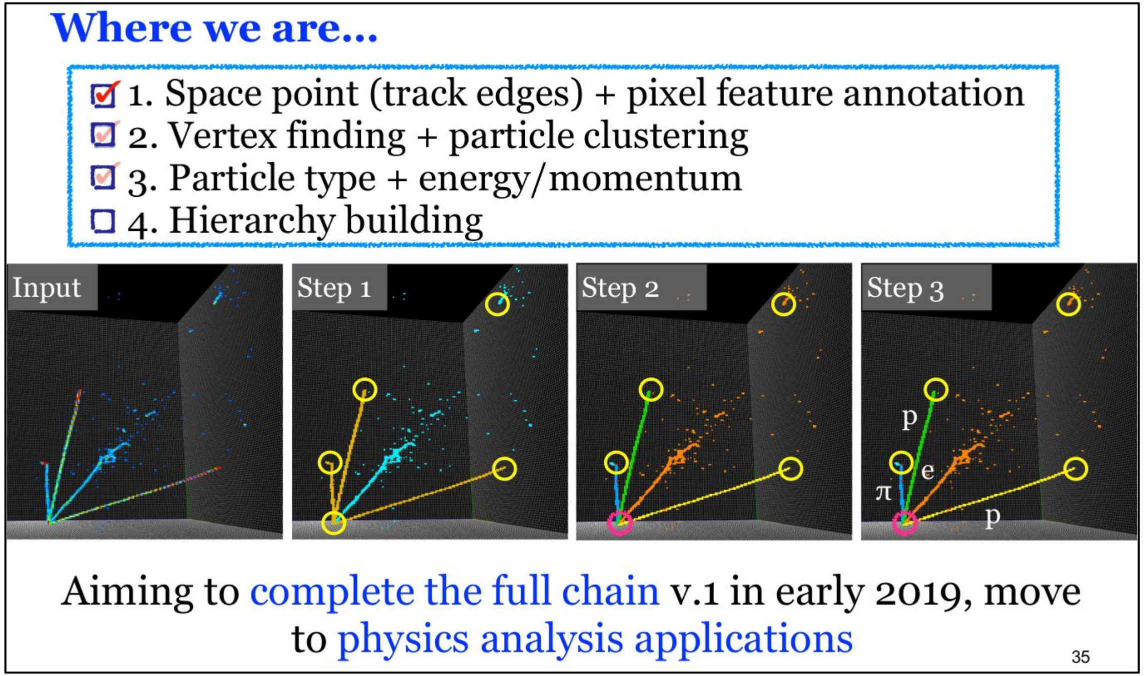
Much less memory usage

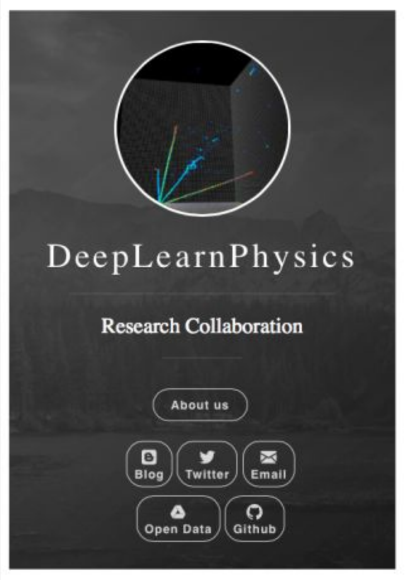


https://indico.fnal.gov/event/18104/session/23/contribution/78

Online/Offline deployment of full reco. chain

- Such technical developments allow for fast training AND more efficient deployment/inference
- Allow us to reach eventual, ever-growing ambitions





Code, tutorials, data sets
http://deeplearnphysics.org/

[K. Terao et al. Sunday]()

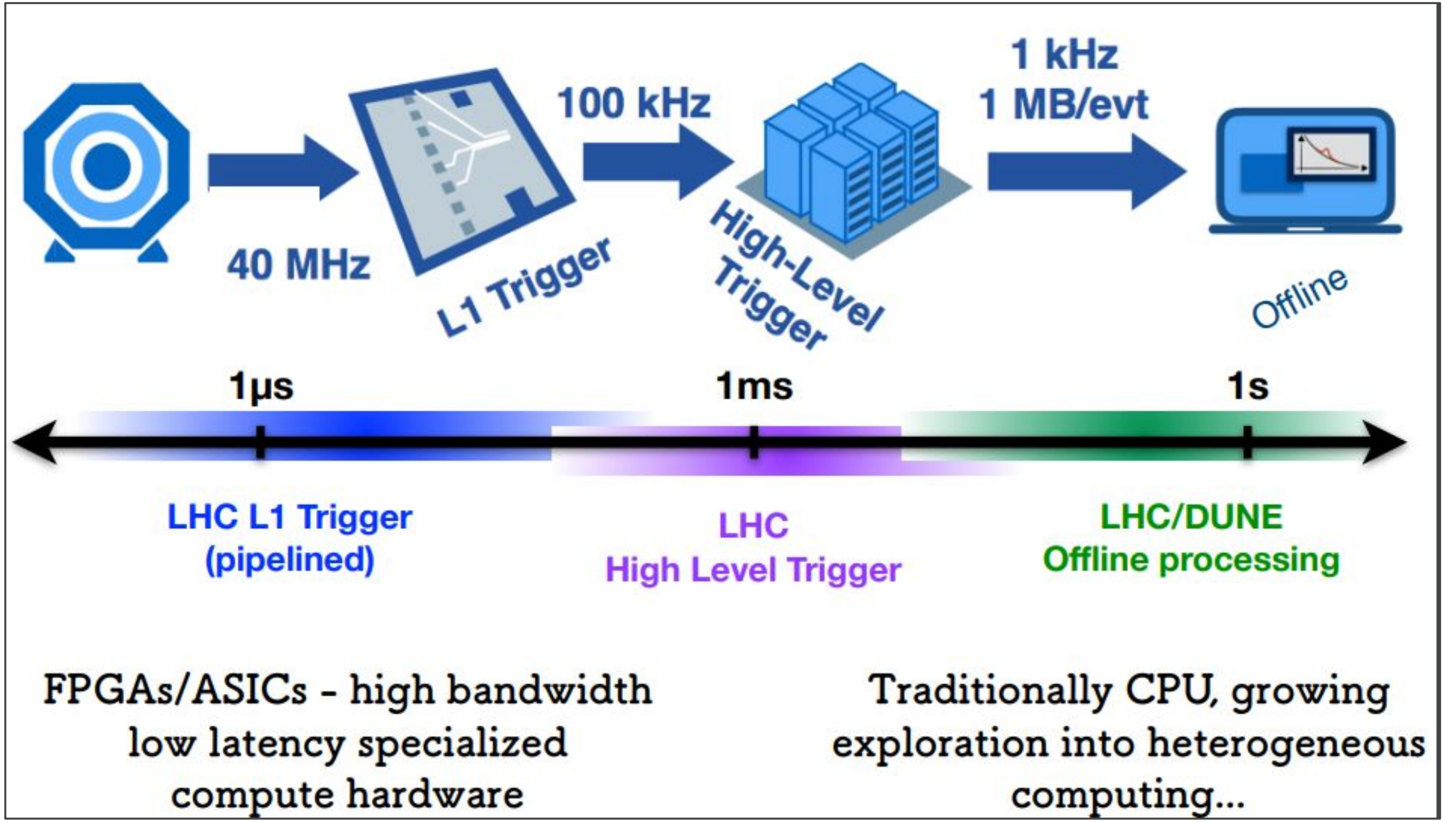
Online/Offline deployment of reco. chain

M. Liu
Mon Plenary

These algorithms most efficient on specialized devices -- which are not always as readily available as CPU cluster

(surge in field enabled by GPUs developed to run increasingly demanding video games)

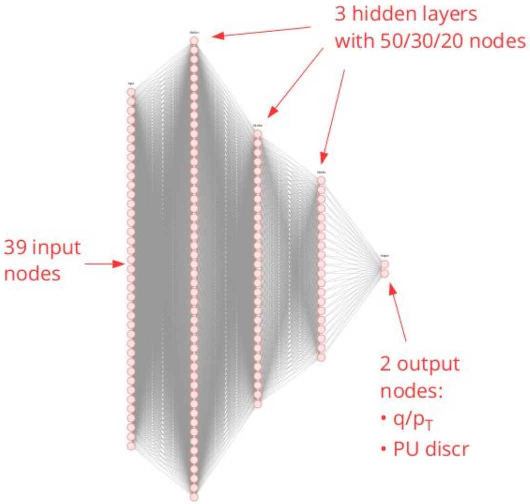
Solving same problems as industry -- opportunities for co-development



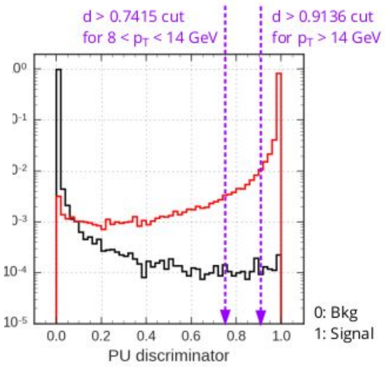
https://indico.fnal.gov/event/18104/session/8/contribution/16/material/0/0.pdf

Deployment of NN for L1 Muon Selection on FPGAs

- From simulation studies, we have achieved roughly 4x rate reduction with the addition of the new muon detectors. The overall efficiency has also been improved.
  - Still work in progress.



Jia Fu Low et al.
(Monday afternoon)



https://indico.fnal.gov/event/18104/session/23/contribution/72

- NN understudy deployed using HLS2ML -- exciting tool to ease the deployment of networks on FPGAs

D. Rankin et al
(Monday afternoon)
https://indico.fnal.gov/event/18104/session/23/contribution/73

Problems for all next-gen experiments: EIC

Analyze larger, more complex datasets a challenge for everyone
Talk from Electron Ion Collider tackling realtime processing (and beyond)

M. DIEFENTHALER et al. (Monday afternoon)

Streaming Readout III: Prototype DAQ systems being discussed

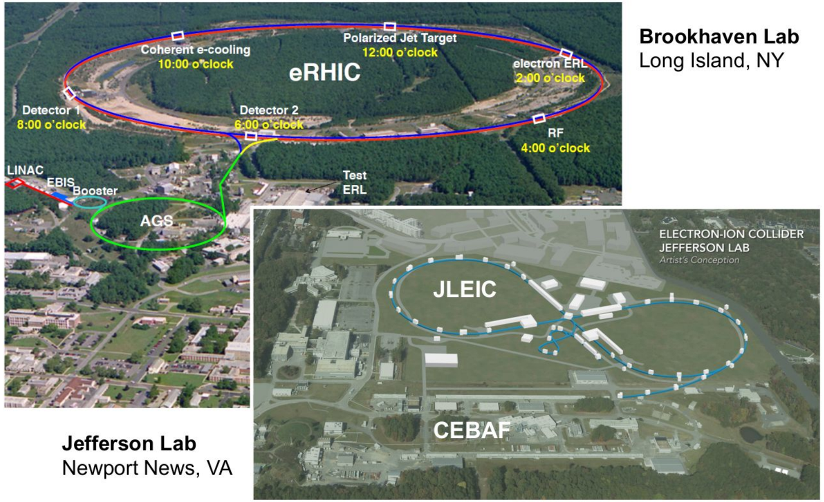
BDX dark matter experiment at JLAB
- digitization: INFN "wave board" digitizer (250 MHz, 14 bit, 12 ch)
- online event reconstruction: TRIDAS system from KM3NeT
- ongoing data validation of prototype syste,

CBM upcoming fixed-target heavy-ion experiment at FAIR
- event rates up to 10MHZ (current heavy-ion experiments 100Hz – several kHz)
- no hardware trigger, real-time data selection exclusively on CPU (under development)
- validation with detector prototypes (eTOF@STAR) and with full-system tests (mCBM@GSI)

JLAB streaming readout for upcoming TDIS, SoLid, and EIC experiments
- build various generic streaming DAQ using existing hardware at JLAB
- could serve as an upgrade of existing DAQ systems at JLAB
- gain valuable experience for R&D for future hardware

PHENIX heavy-ion experiment and upgrades at BNL
- 15kHz signal collisions, 1.4M channels streaming

sPHENIX upgrade of PHENIX experiment
- sPHENIX TPC: 160k channels 10b flash ADC @ 20MHz with SAMPA ASIC -> 2 Tbit/s stream rate.
- BNL-712/FELIX-type DAQ with data rate of 200 Gbit/s



Brookhaven Lab
Long Island, NY

Glimpse of the future

- Demonstration of track reconstruction using a quantum annealer
- Exciting step in quantum computing

<u>L. Linder et al. (Sunday afternoon)</u>

Map the problem of selecting true trajectory components from set of possible to the minimization of energy of a quantum system

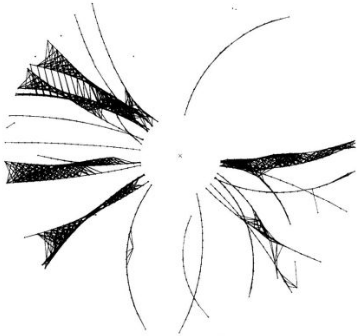


Fig. 1. Display of all generated lines for a real $Z^0 \to$ hadrons (XY projection).

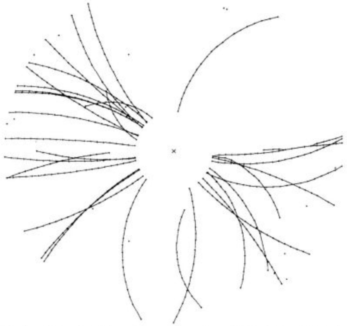


Fig. 3. Display of the activated lines after convergence for a real $Z^0 \to$ hadrons (XY projection).

source: fast track finding with neural nets

quantum machine instruction (QMI)
objective function:

$$O(a;b;q) = \sum_{i=1}^{N} a_i q_i + \sum_{i}^{N} \sum_{j}^{N} b_{ij} q_i q_j \quad q_i \in \{0,1\}$$

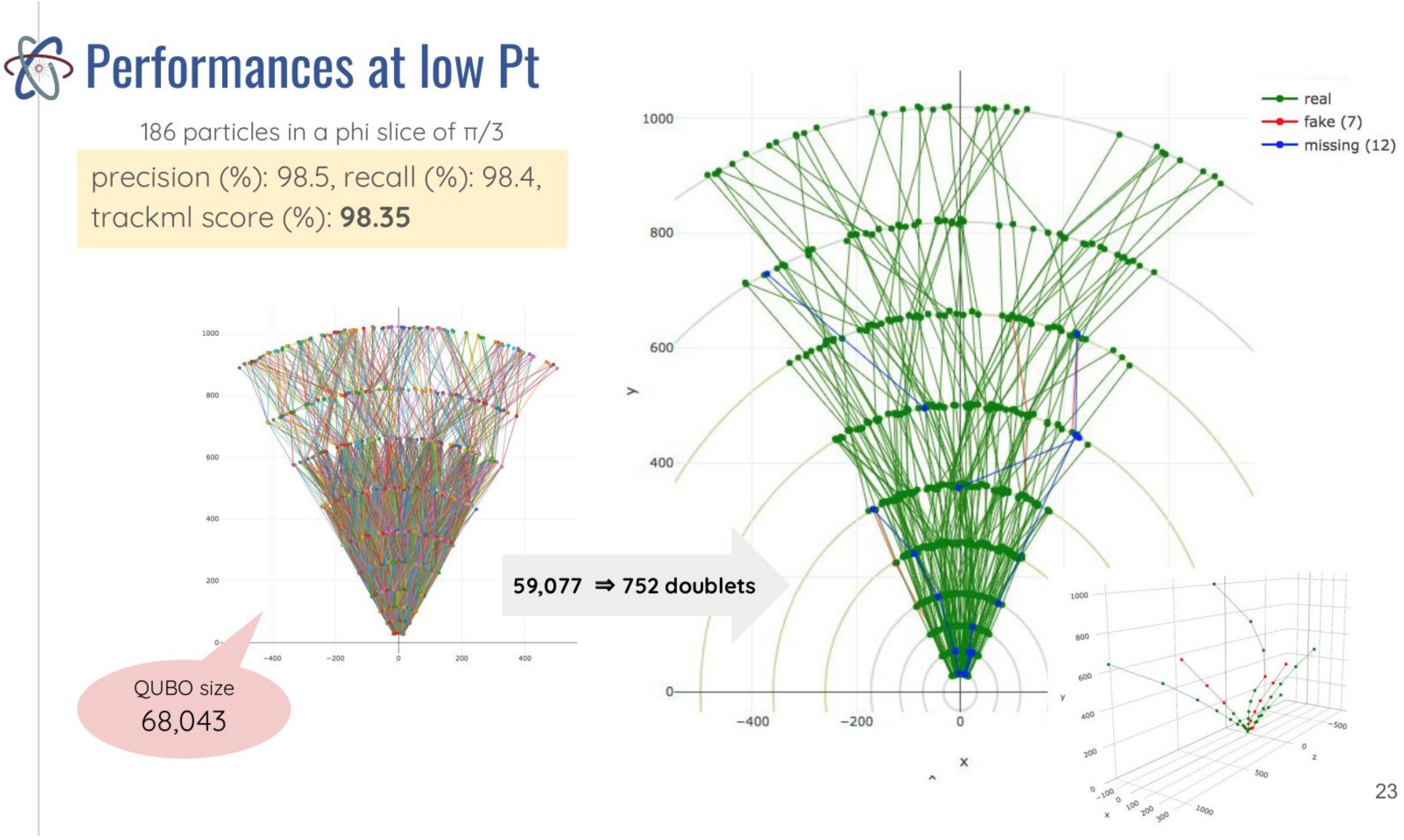
QUBO
Quadratic Unconstrained Binary Optimisation

1. generate the set potential doublets (apply early cuts)
2. *binary classification task* to determine which doublets should be kept in the solution

https://indico.fnal.gov/event/18104/session/23/contribution/61

Track reco. w/ quantum annealer

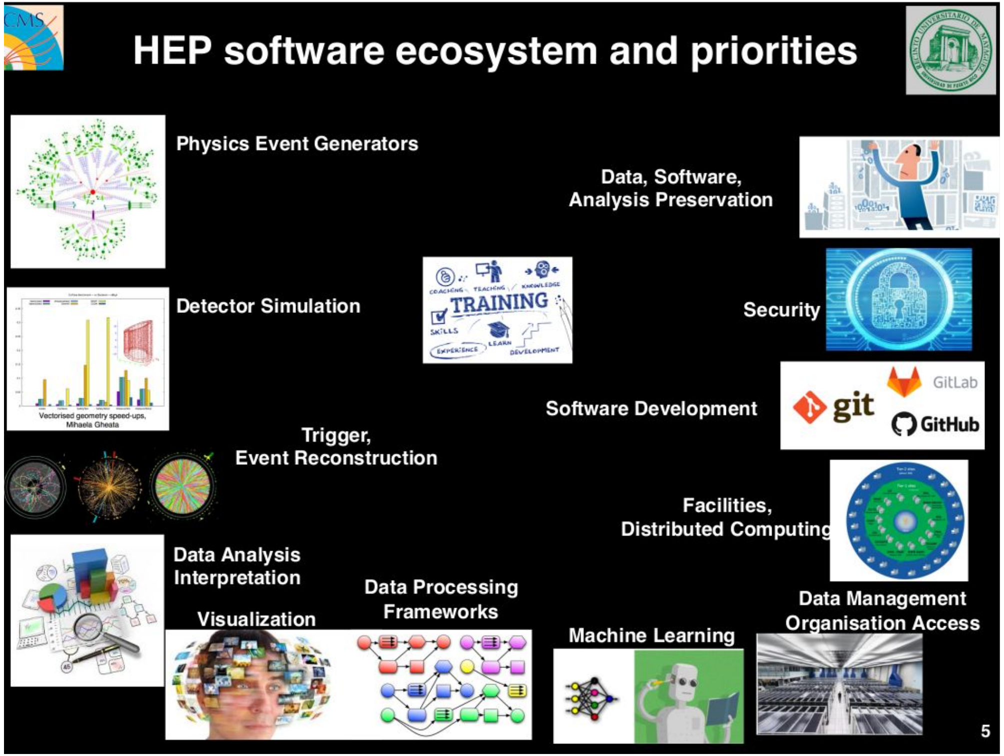
L. Linder et al. (Sunday afternoon)



Training *in vivo* networks

S. Malik et al. (Tues morning)

- Accelerating R&D requires training all those are interested in joining in (not just software, detector development as well)
- Often overlooked, but critical
- Organize (often redundant) materials from different experiments for common techniques - disseminate widely, host workshops
- For ML:
  - Produce more users and developers
  - *Spur community critiques as well*



https://indico.fnal.gov/event/18104/session/23/contribution/19

Get involved!

- IRIS-HEP website http://iris-hep.org/
  - Jobs on IRIS-HEP and Collaborating Projects http://iris-hep.org/jobs
  - General public announcement mailing list for IRIS-HEP events, talks, meetings, workshops, opportunities for training and job opportunities (subscribe to)
  - announcements@iris-hep.org

- HSF (HEP Software foundation) - https://hepsoftwarefoundation.org
  - General Information about HSF (subscribe to): hsf-forum@googlegroups.com
  - Discussions and activities in the HEP Software Foundation mailing lists can be found here (General and Dedicated Forums): https://hepsoftwarefoundation.org/forums.html
  - You can contribute https://hepsoftwarefoundation.org/cwp/cwp-working-groups.html
  - HSF Events/Workshops - https://hepsoftwarefoundation.org/events.html

- FIRST-HEP website http://first-hep.org
  - Funding for participants and lecturer support for Training

Summary

- A great collection of new, promising developments
  - Theme across domains: often can (should?) employ information closer to the raw data of our detectors -- make use of the increasingly precise information coming from new detectors
  - (can provide means to preliminary performance estimates to inform hardware design as well)
- Many directions to explore -- way more than people to work on them
  - Please join!
- Clear that communities should continue to be in close communication to share advances
  - Shared code
  - Open datasets