AI/ML in Wire-Cell

Haiwang Yu (BNL) for the Wire-Cell Team

BNL Physics fourth joint meeting on AI/ML

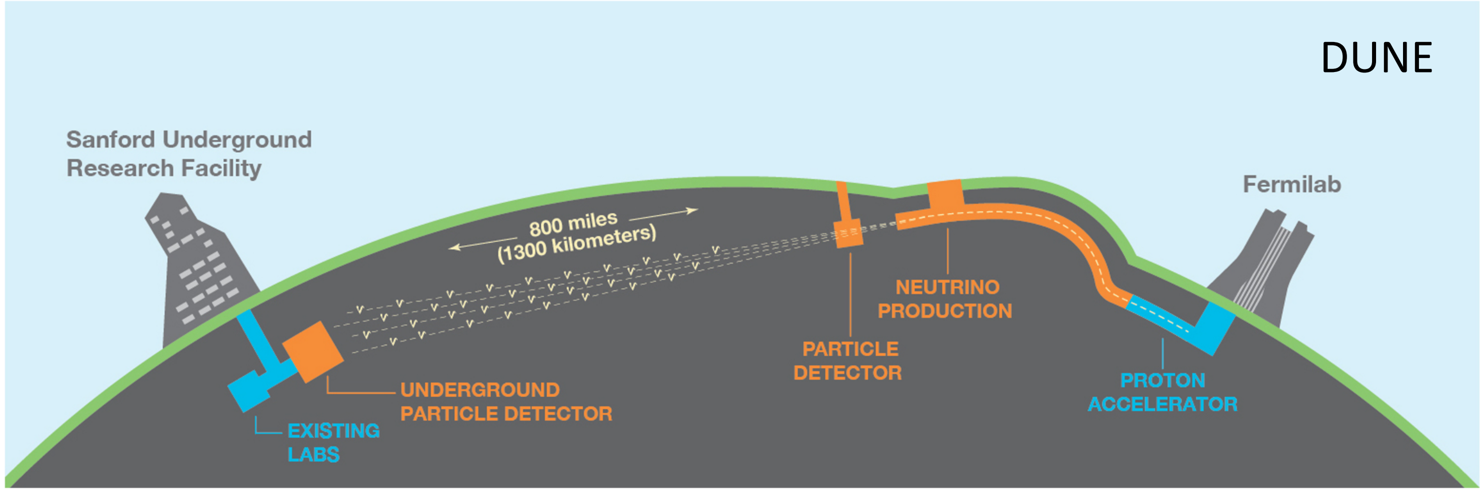
23 July 2021

Brookhaven National Laboratory

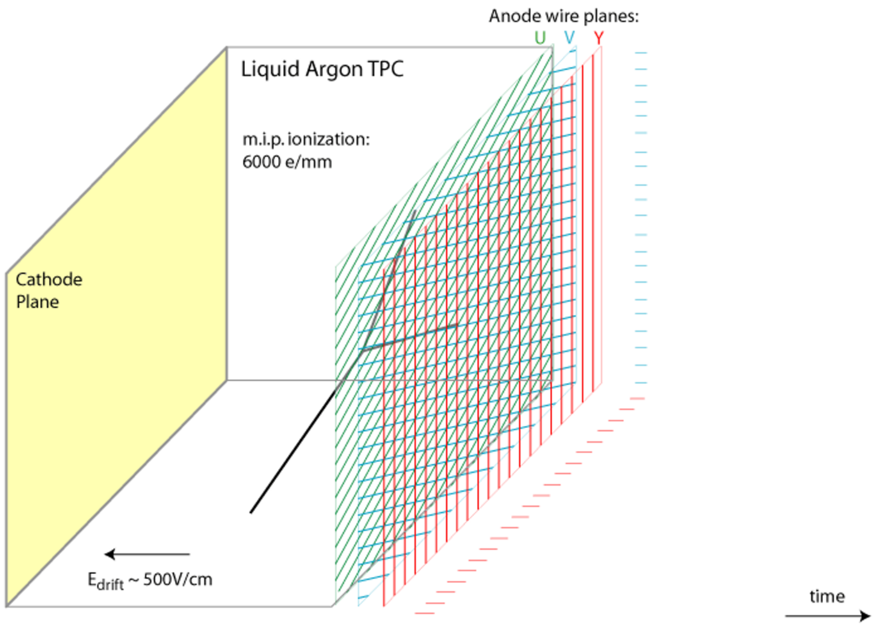
Wire-Cell

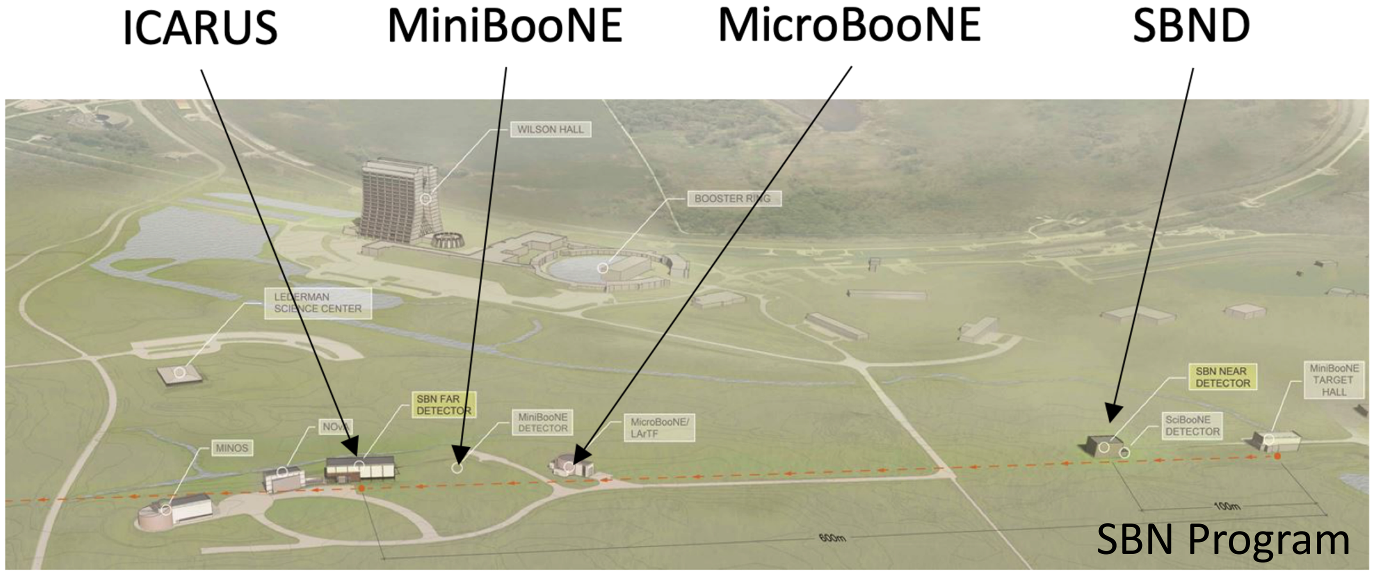
LArTPC is a key detector technology for many next-gen neutrino experiments

- calorimetry + rich and precise topology info.



LArTPC Signal Formation by B. Yu (BNL)



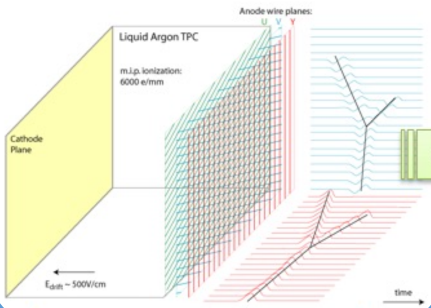


Wire-Cell reconstruction for LArTPC

Wire-Cell

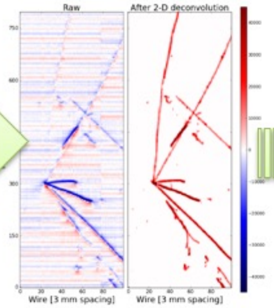
https://lar.bnl.gov/wire-cell/

LArTPC Signal Formation



deconvolving

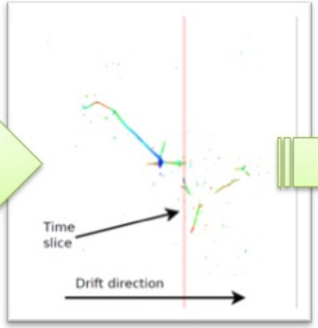
Noise filtering and Signal Processing



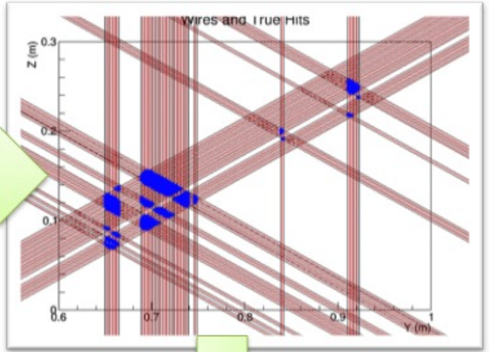
JINST 12 P08003
JINST 13 P07006, P07007

slicing

3D Imaging

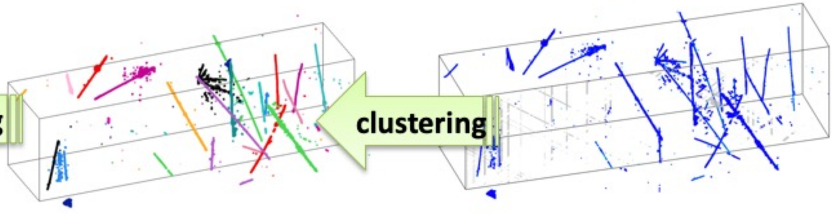


tiling



solving

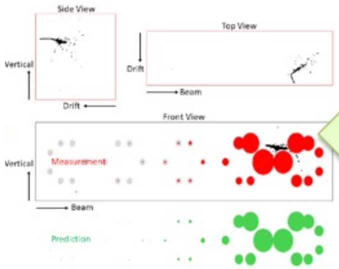
clustering



JINST 13, P05032
3D imaging principle

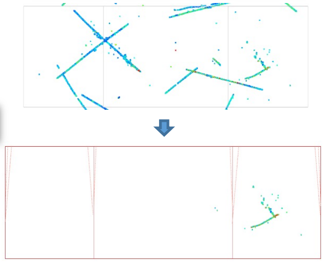
matching

TPC/Light Matching



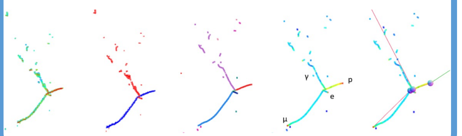
arXiv:2011.01375
(MicroBooNE imaging, clustering, matching, accepted by JINST)

Cosmic Ray Removal



arXiv:2012.07928
Phys. Rev. Applied 15, 064071

3D Pattern Recognition



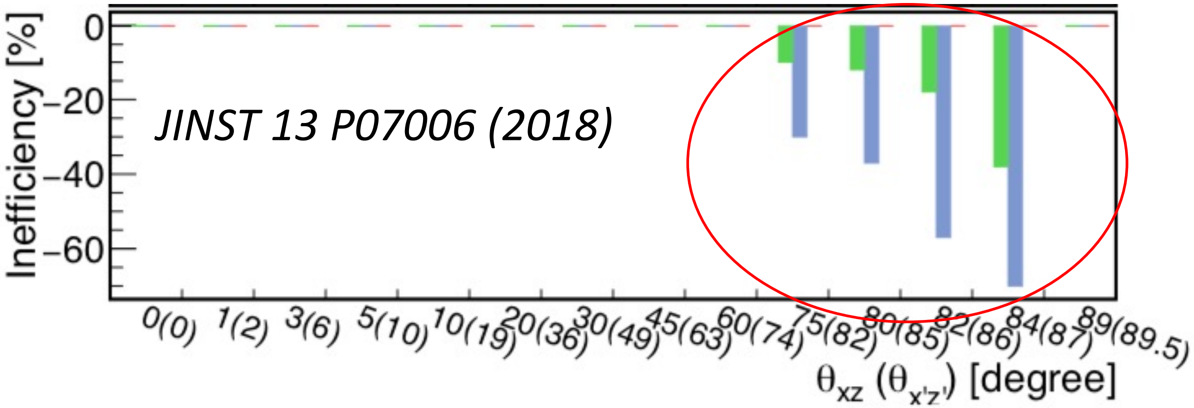
JINST in preparation

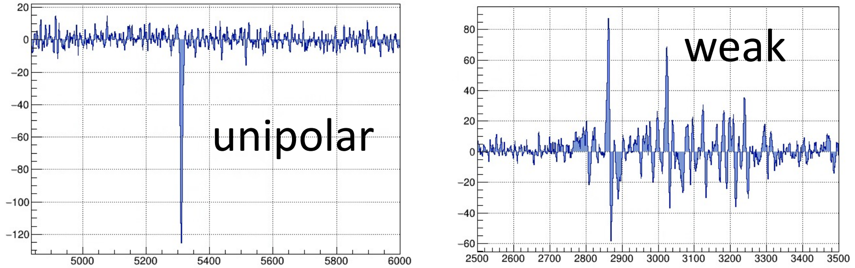
Event Taggers

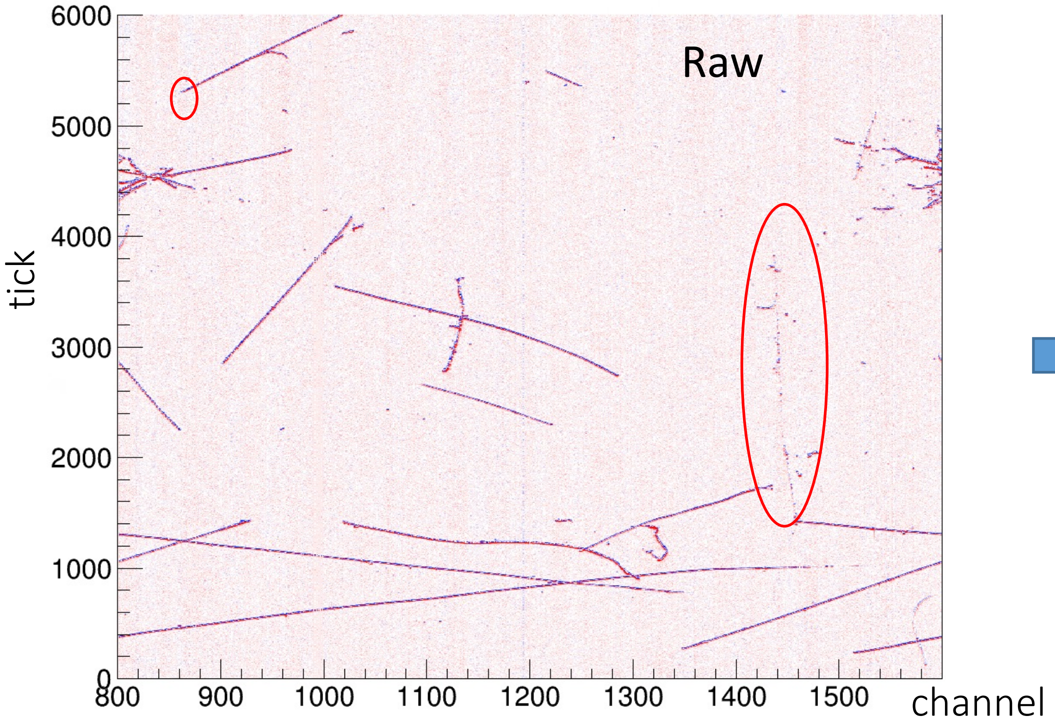
Multiple Physics Papers based on it

DNN ROI finding to improve LArTPC Signal Processing

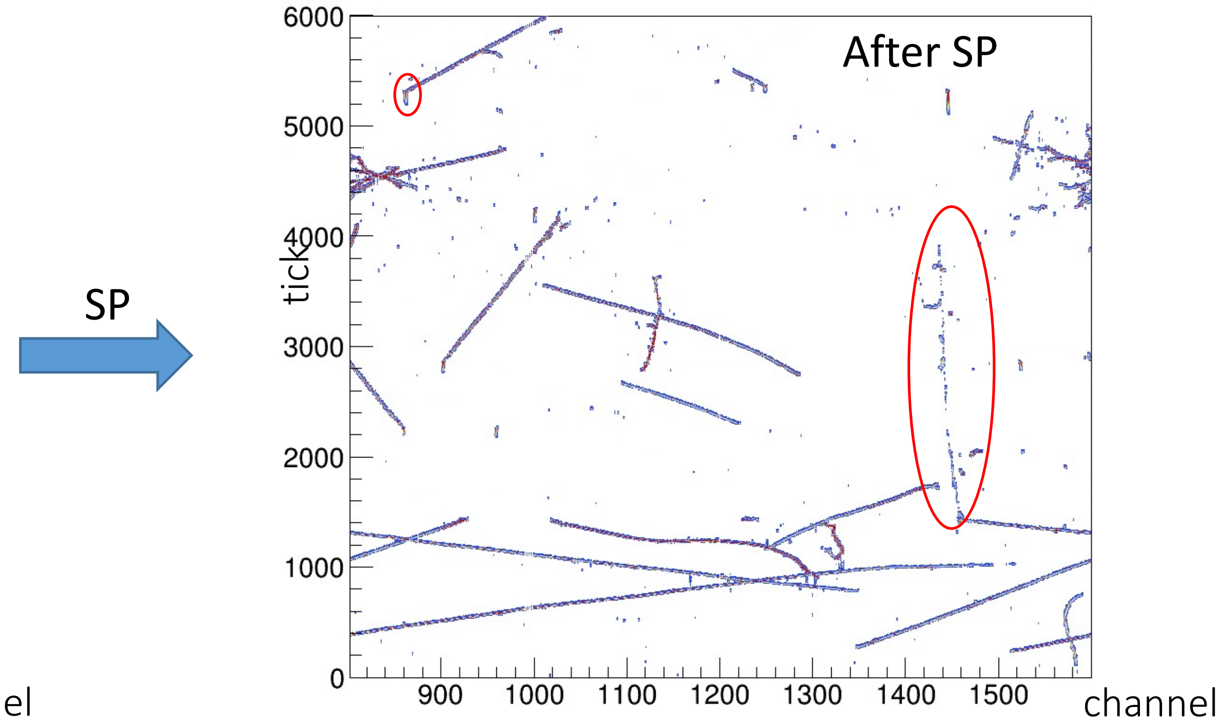
- “Prolonged Track” – weak signal
- “Tear Drop” - distorted waveform
- Noisy dots - noise







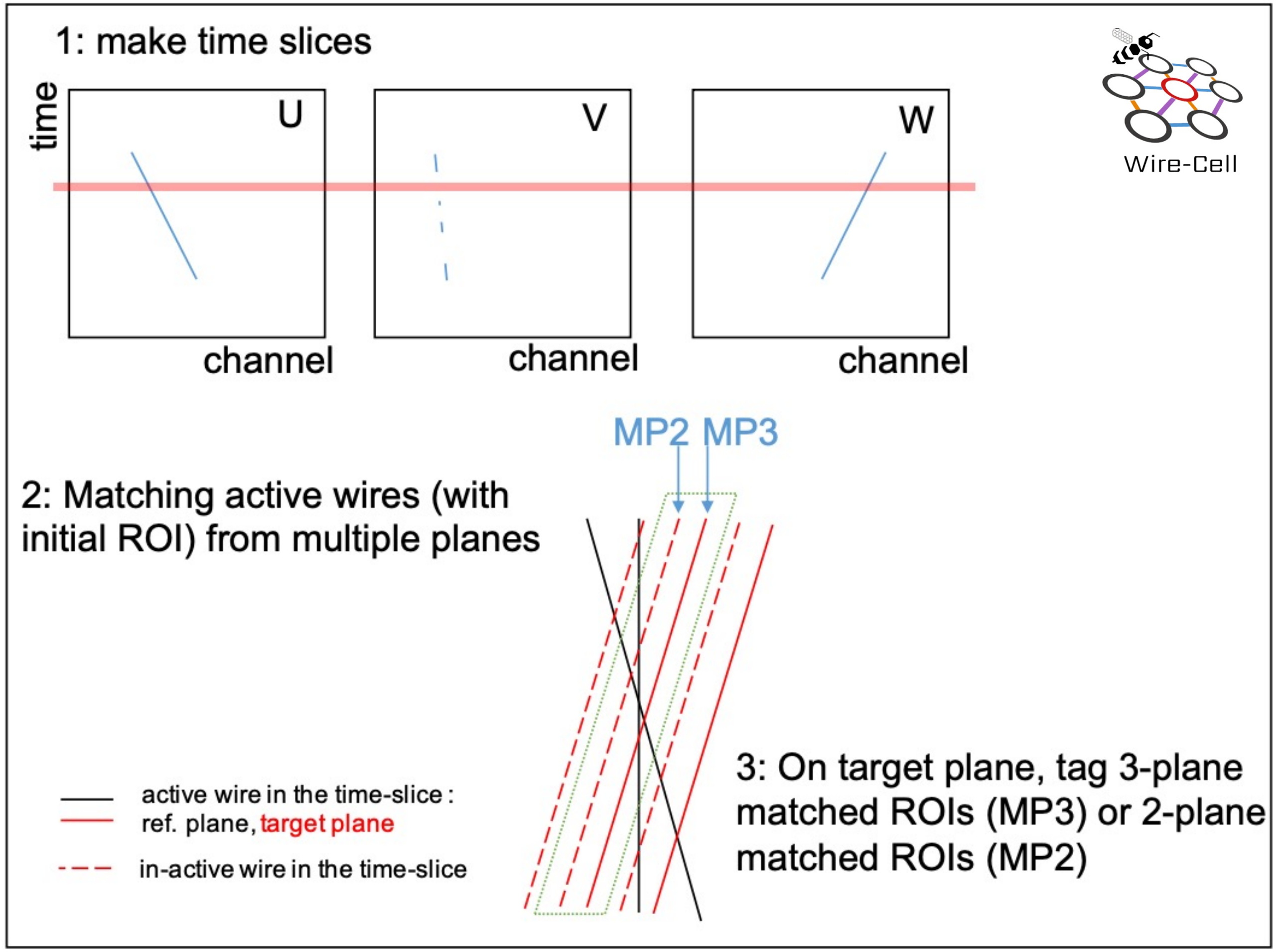
SP



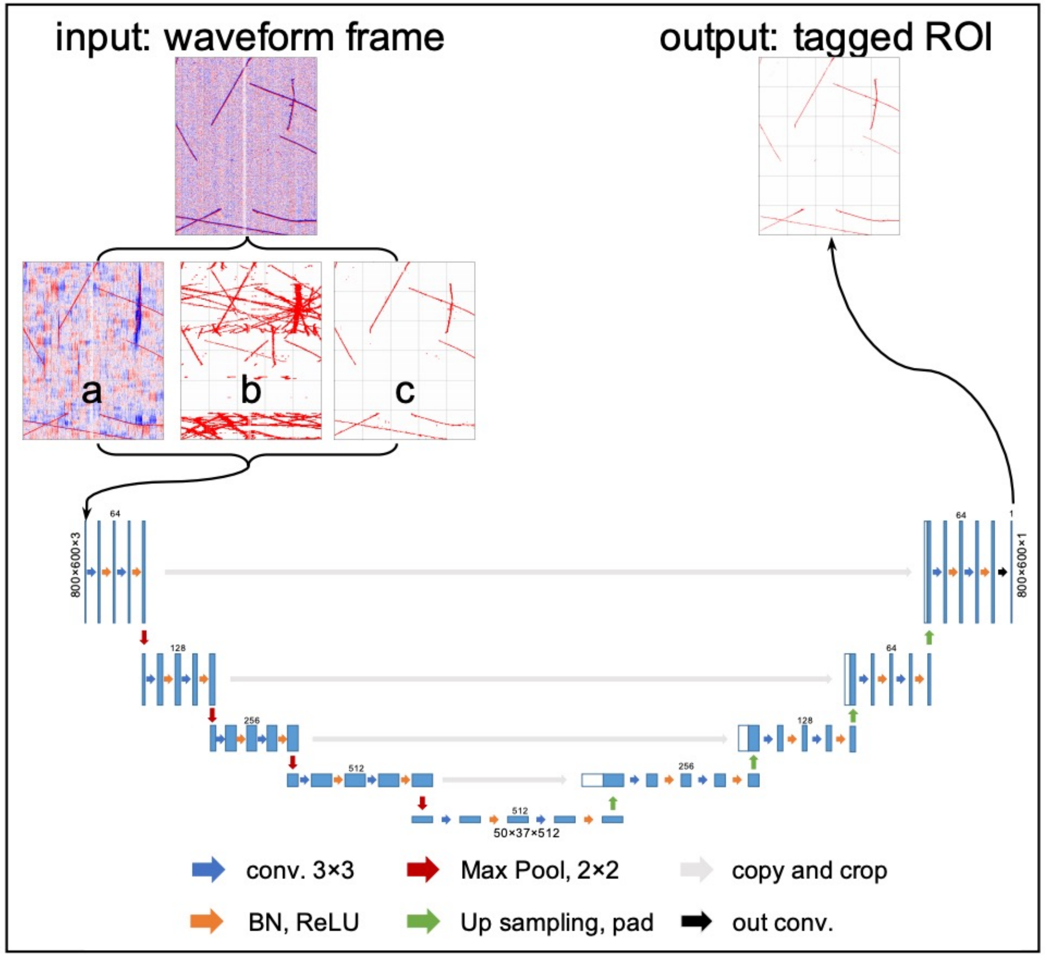
DNN ROI finding with multi-plane information

JINST 16 (2021) 01, P01036

Multi-plane information in Signal Processing

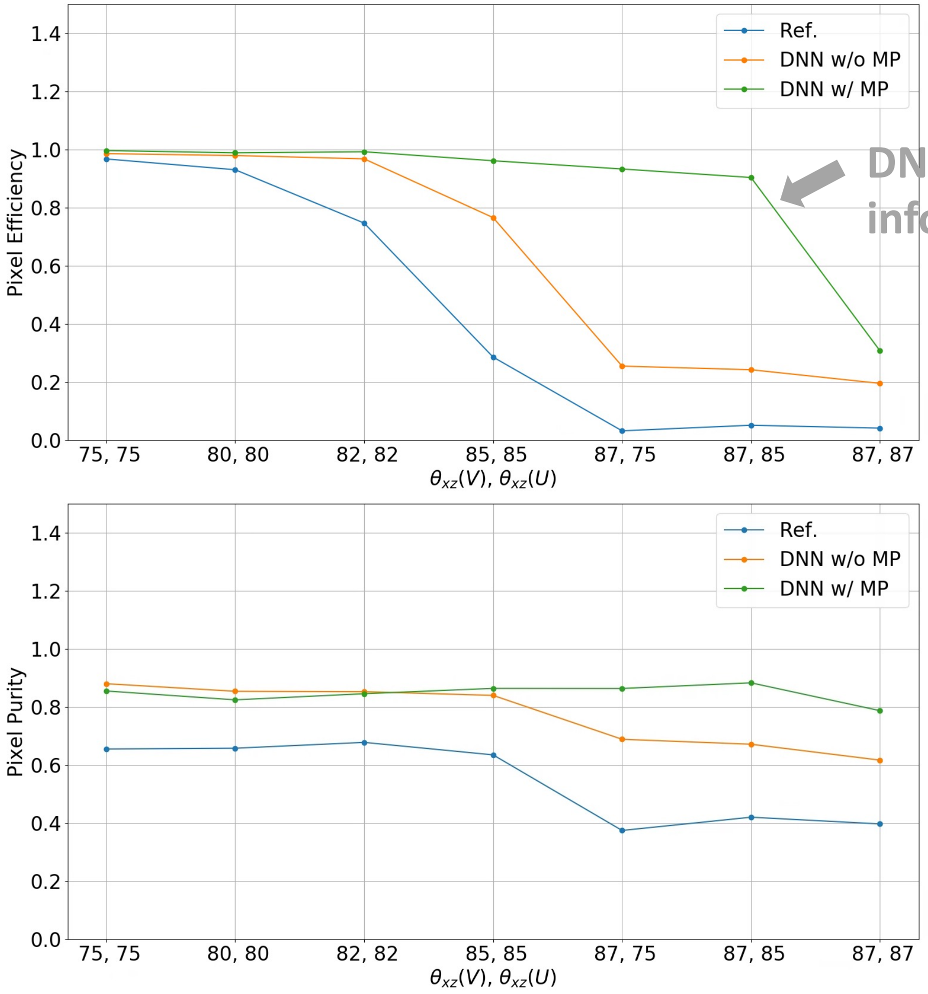


DNN ROI finding with multiple input channel



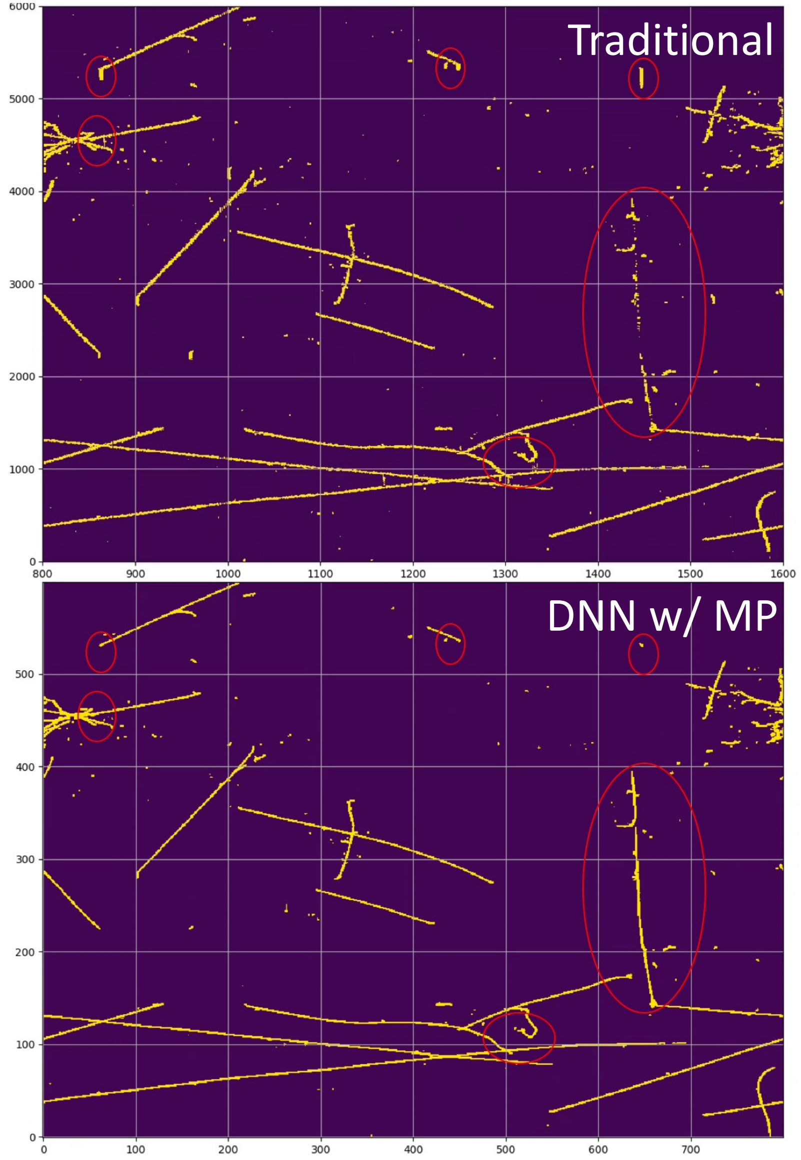
DNN ROI finding with multi-plane information

ProtoDUNE simulation
ROI finding on V plane (2nd induction)



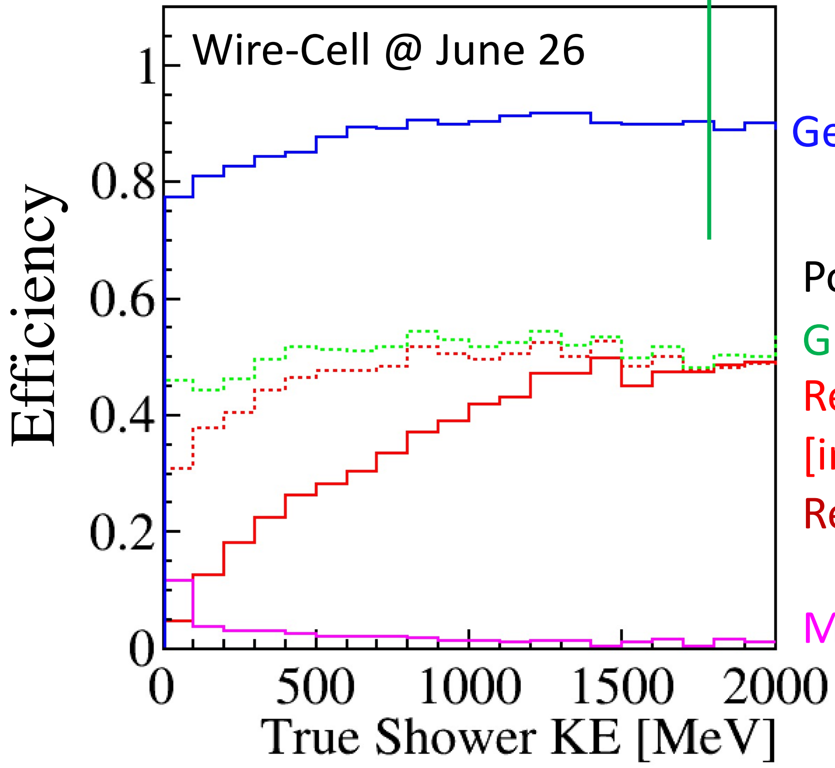
DNN With 3-plane information

tested on ProtoDUNE data



Improve Vertexing for νe-cc tagging

Enhance the vertexing efficiency using deep learning based technique ⟶ increase of νe-cc tagging efficiency



Generic neutrino selection [finalized]

Post generic neutrino selection:

Green dotted: Good nu vtx [overall 50%]

Red dotted: Good nu vtx & good primary shower vtx [improved vs. Jay's talk DocDB#30948]

Red solid: νe-cc tagged

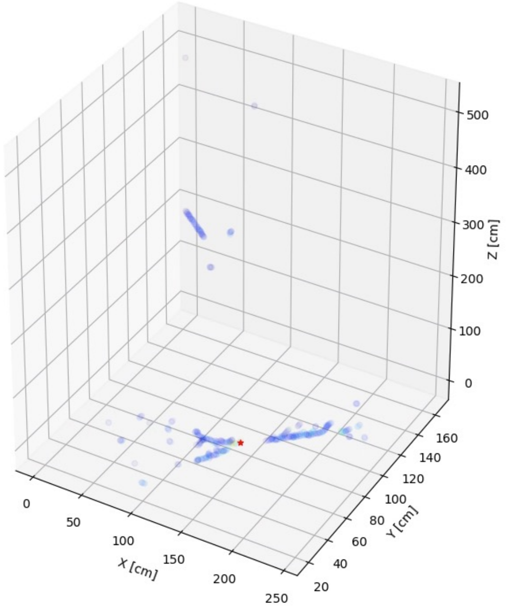
Magenta: no primary shower (>70MeV) reconstructed

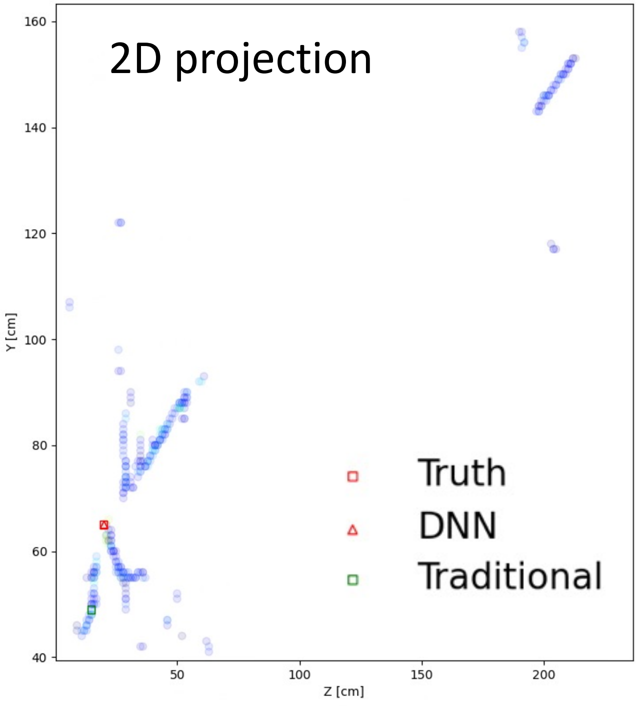
Vertex finding as regressional segmentation with U-Net

Current strategy: vertex finding as regressional segmentation with UNet

- U-Net: efficiently use geometry info which is critical
- regressional loss on distance based "confidence map" to use a region of points instead of only one (next slide)

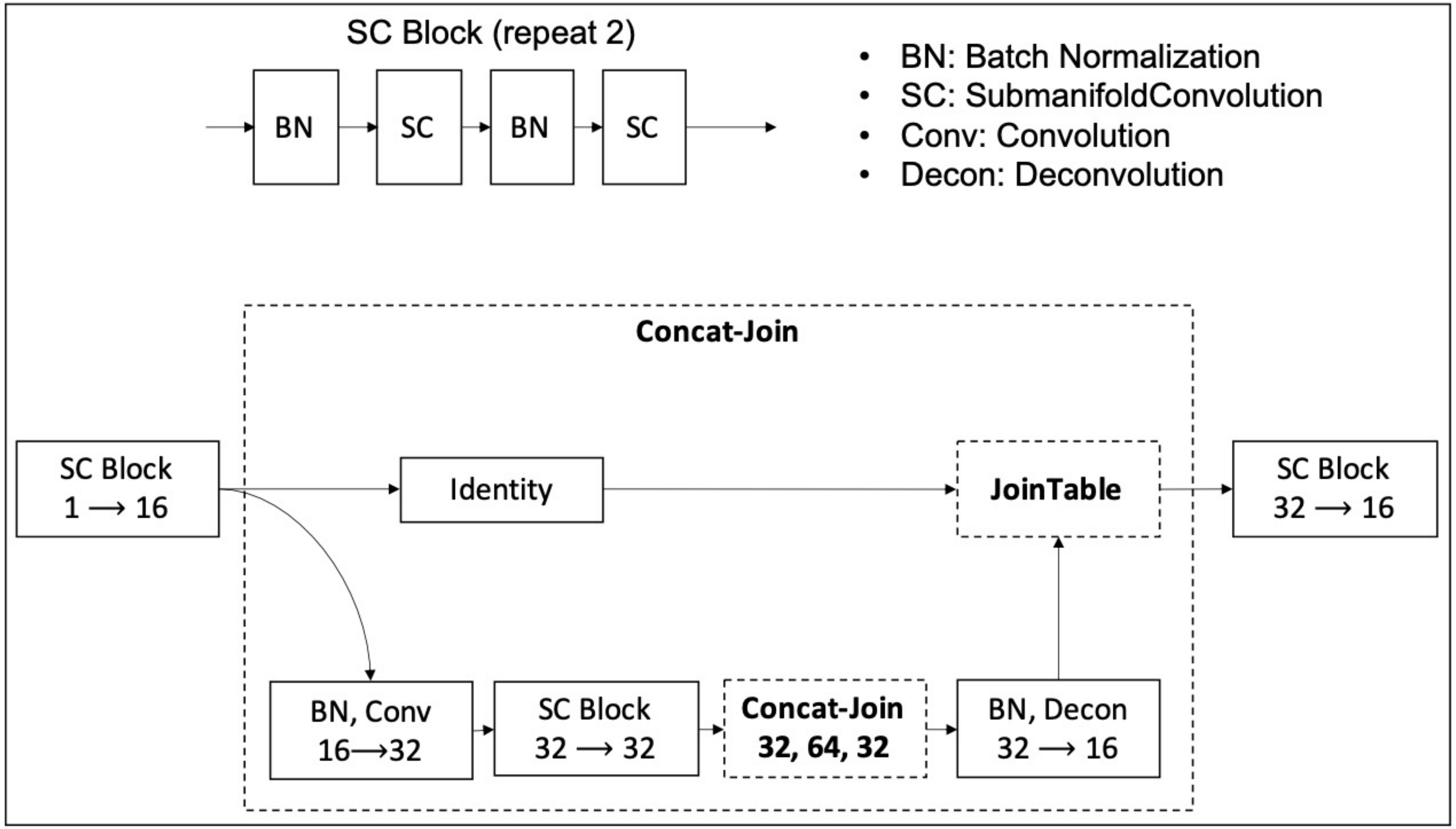
3D points from Wire-Cell



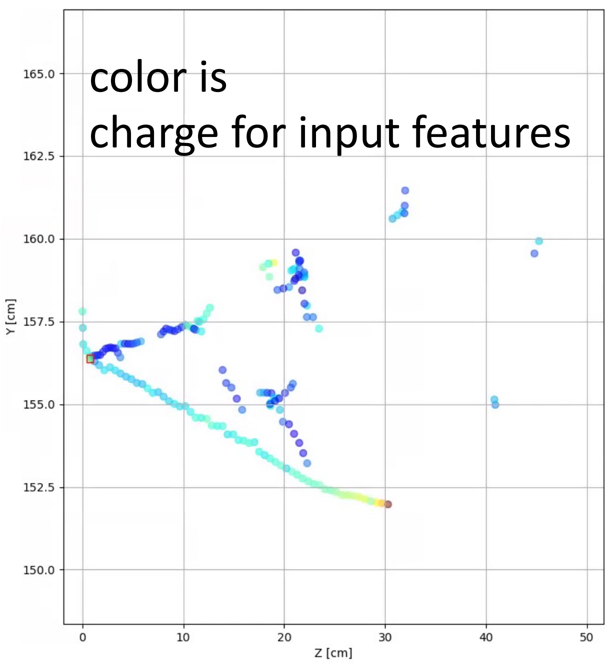


OpenPose: https://arxiv.org/pdf/1812.08008.pdf

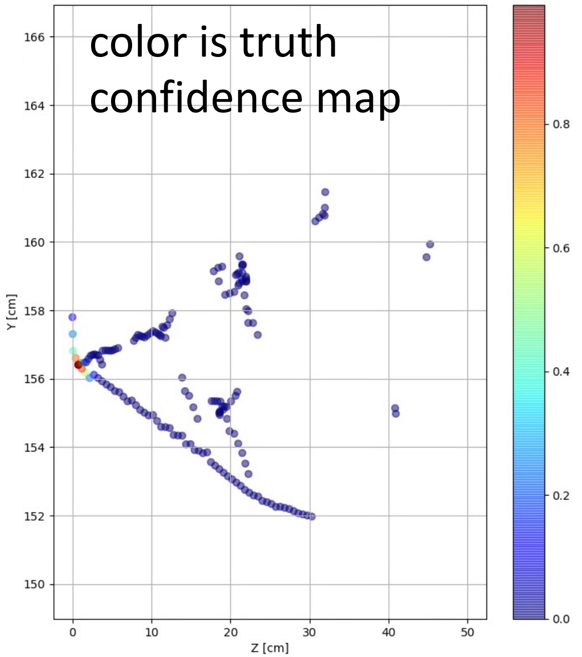
DL-Vertexing: network structure and data preparation



input



label



Used *SparseConvNet* to realized 3D sparse conv. DNN
https://github.com/facebookresearch/SparseConvNet
This work: https://github.com/HaiwangYu/nue-cc

DL-Vertexing: initial results on MicroBooNE νe-cc

Current best model evaluated on νe-cc test samples:

- with **1cm cut**: Trad 50.1% DNN 67.6% Hybrid 65.2%
- relative 30% improvement compared with traditional Wire-Cell vertexing

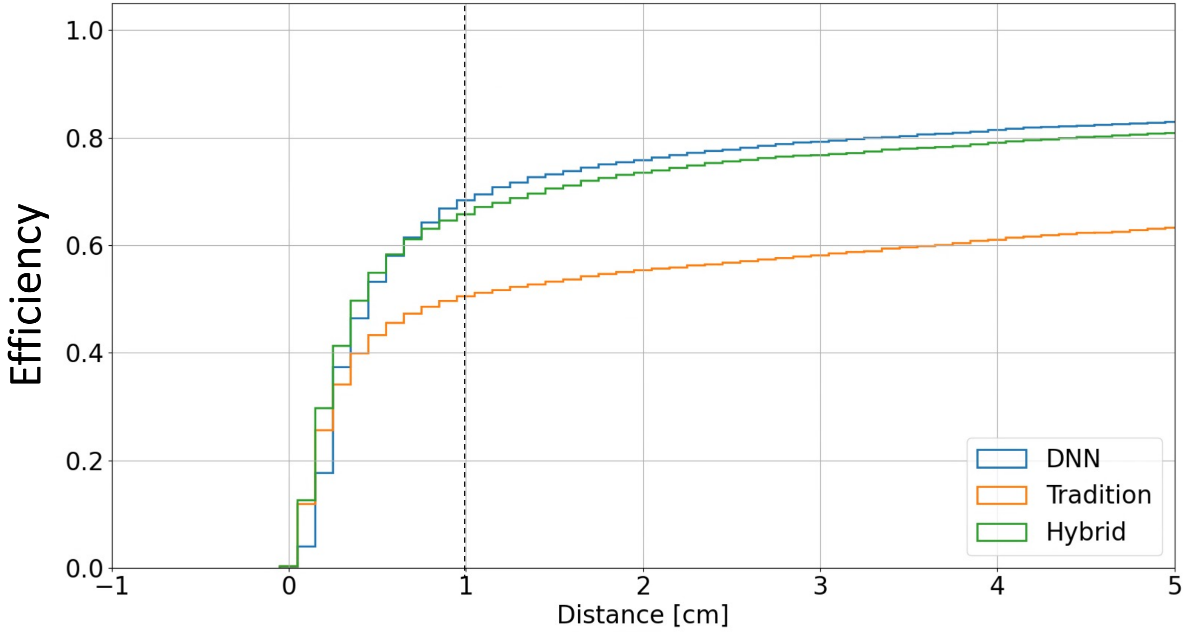
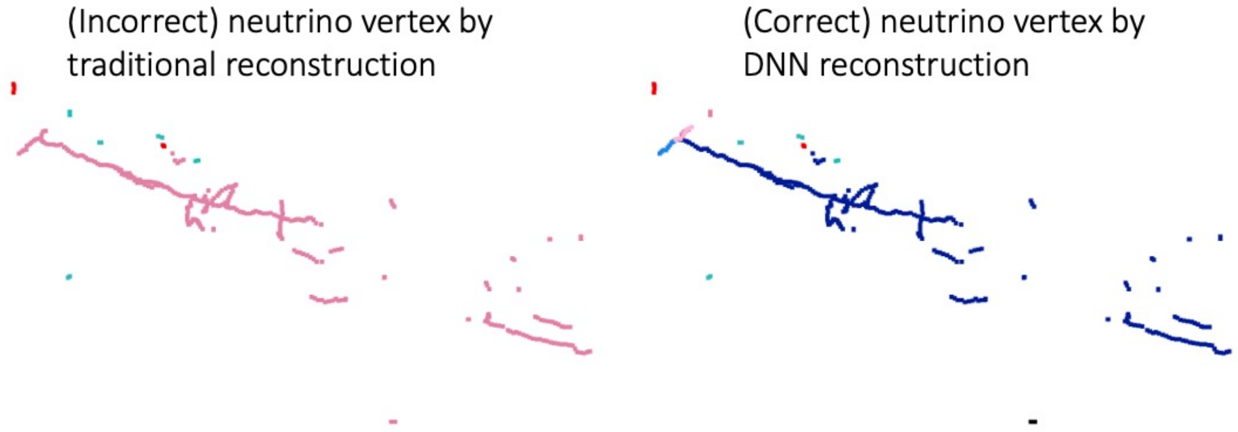


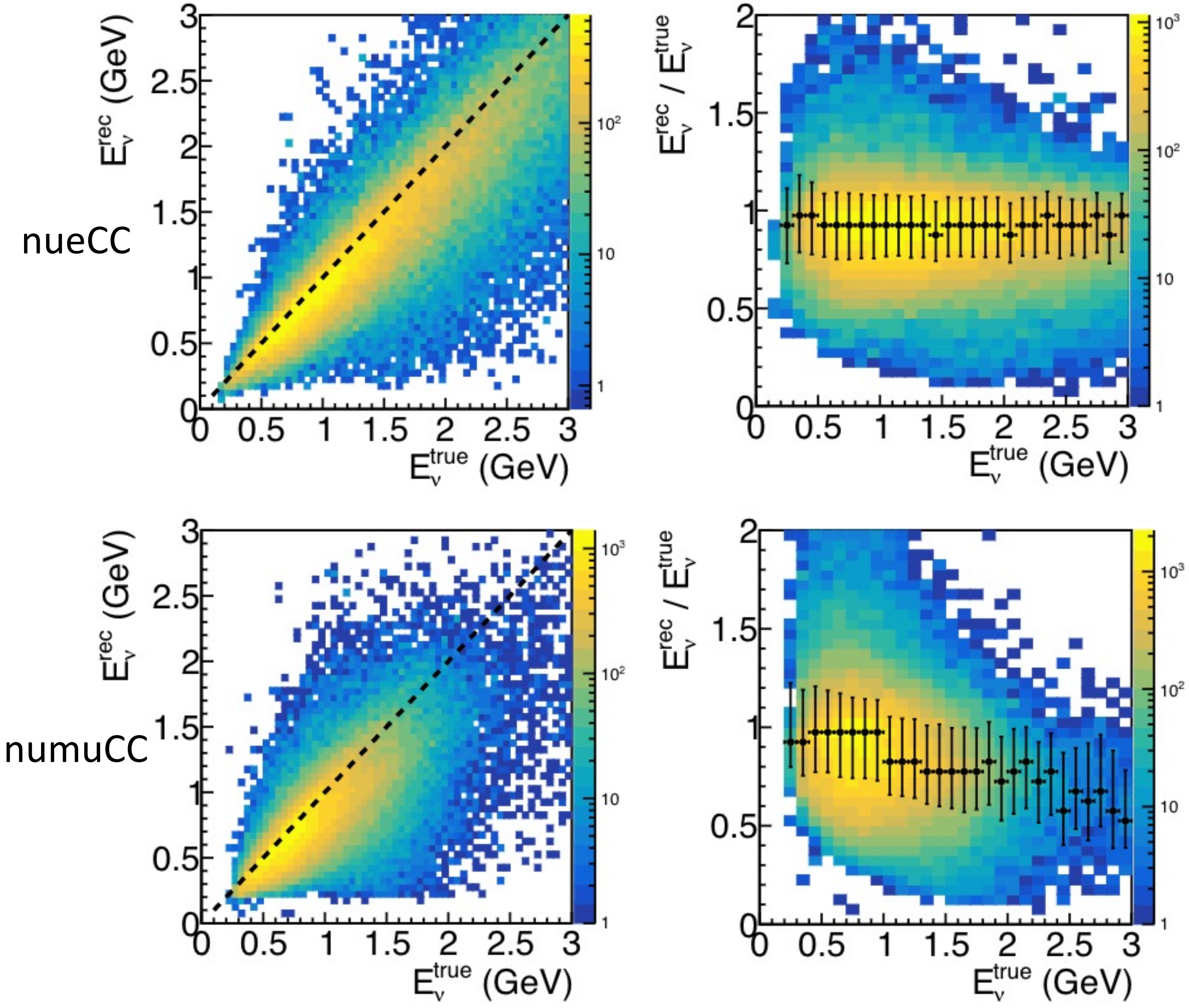
Illustration of impact of vertex ID on the full event reconstruction



Neutrino Energy Estimator (EE) could be improved

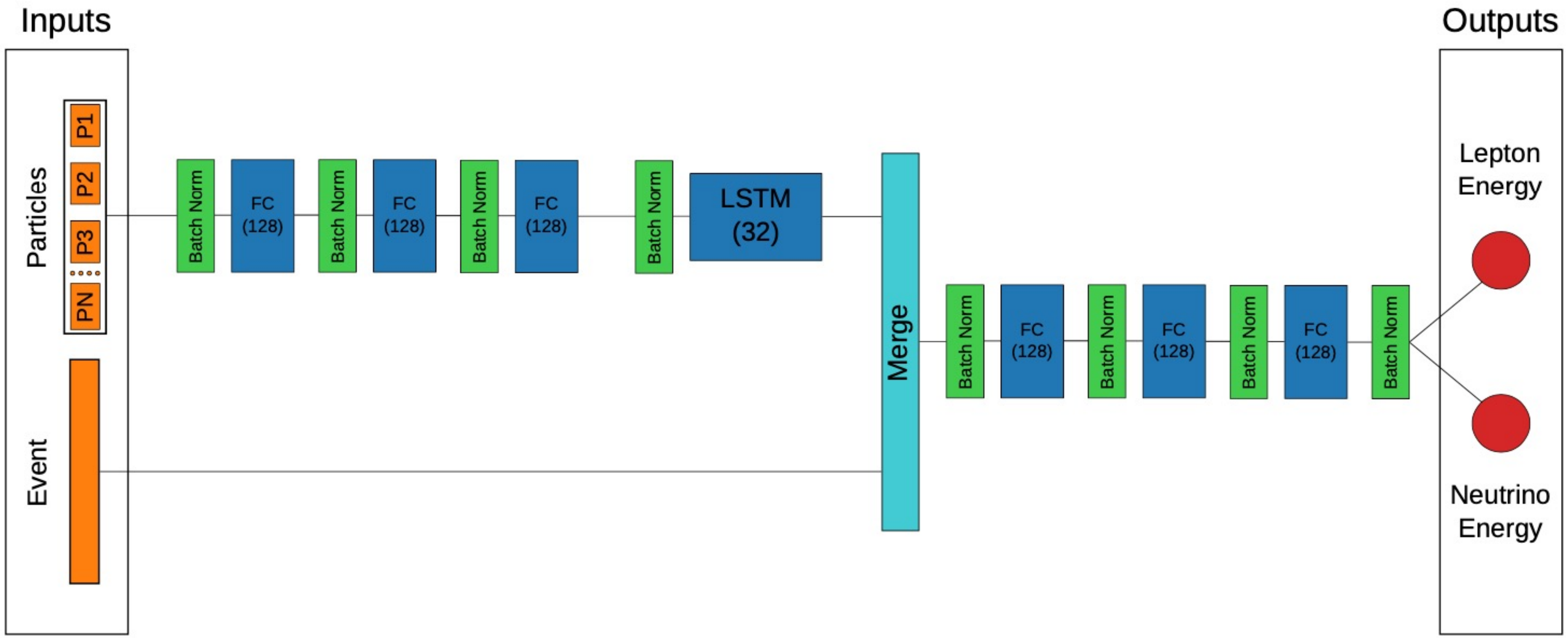
Current Wire-Cell EE: (Calorimetric) Reco neutrino energy = kinetic energy of leptons and hadrons + muon/pion masses + binding energy (8.6 MeV) associated with each visible proton

- straightforward, decent performance
- could be improved by utilizing the underlying correlation between event topology and reco-truth bias
    - better LEE sensitivity
    - smaller systematic err. for unfolding xsec analysis.

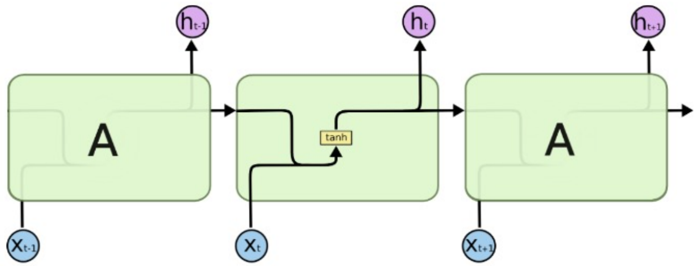


RNN Energy Estimator: variable length list of particles ⟶ energy

D. Torbunov



Simple RNN



RNN EE

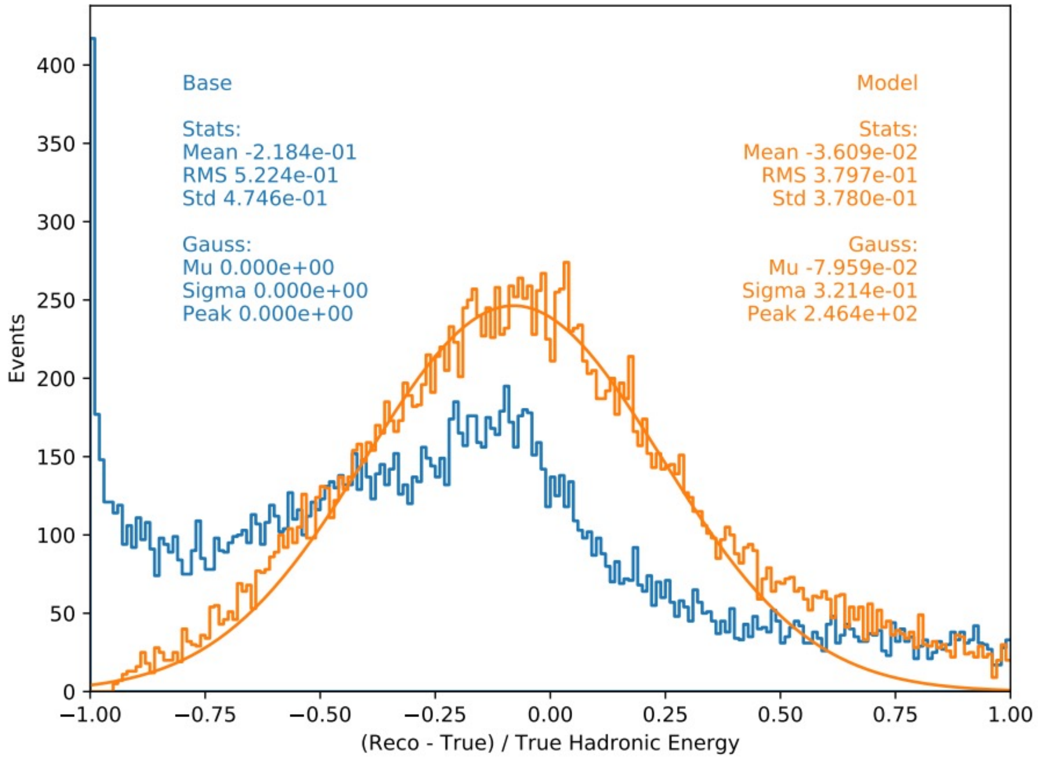
- Extracts information from each particle
- Aggregates it with a help of an LSTM neural network
- Then combines aggregated information with event level variables and predicts energy of neutrino and energy of the primary lepton.

Initial results on MicroBooNE

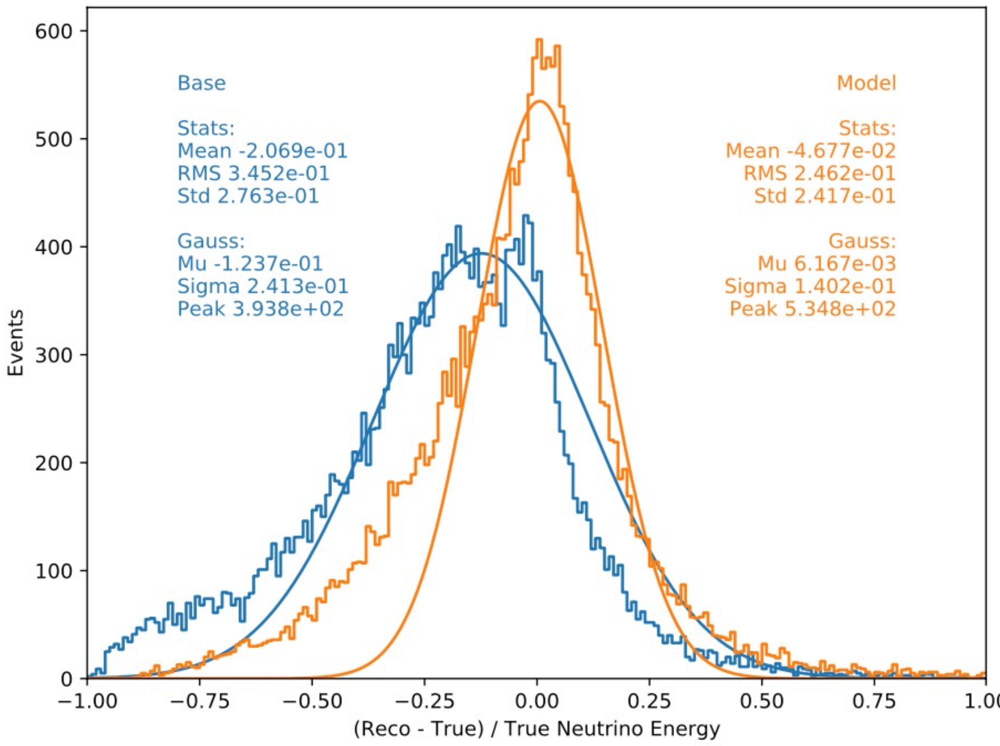
D. Torbunov

New RNN EE improved the neutrino energy reconstruction with first try:
- resolution: 24% ⟶ 14%
- bias: -12% ⟶ 0.6%

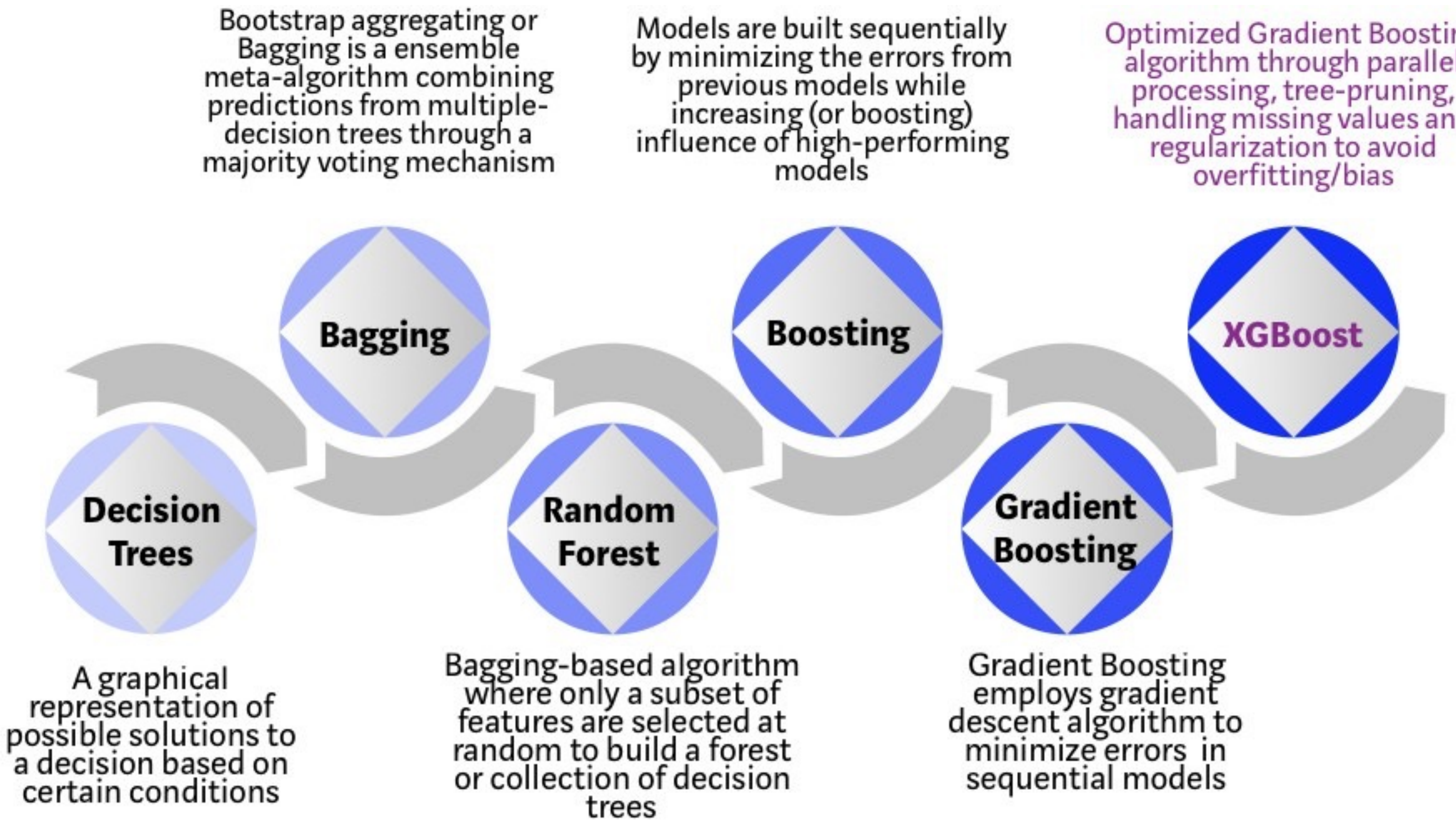
Hadronic energy reco: traditional vs. RNN-EE



Neutrino energy reco: traditional vs. RNN-EE

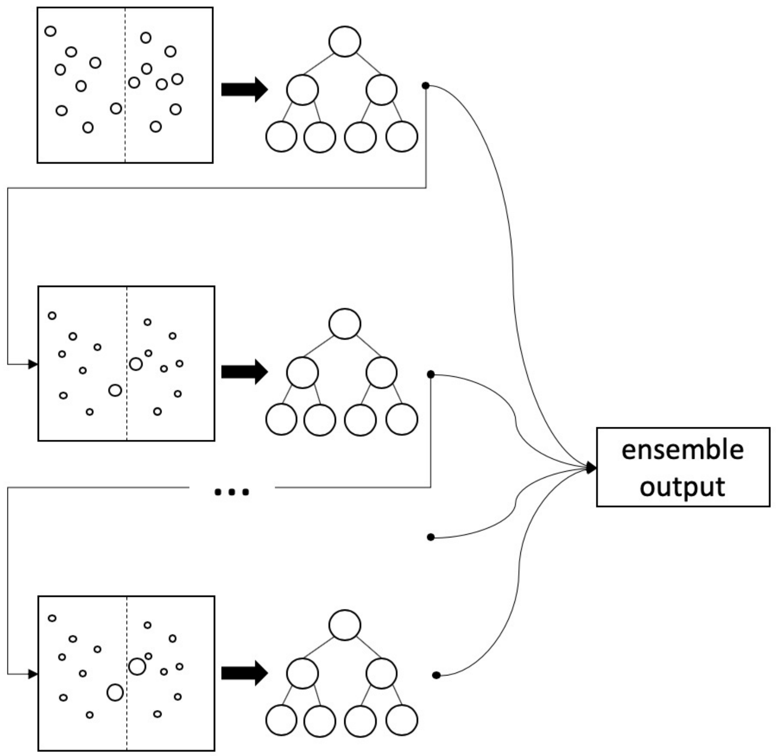


Boosted Decision Trees (BDT) for neutrino event selection



https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d

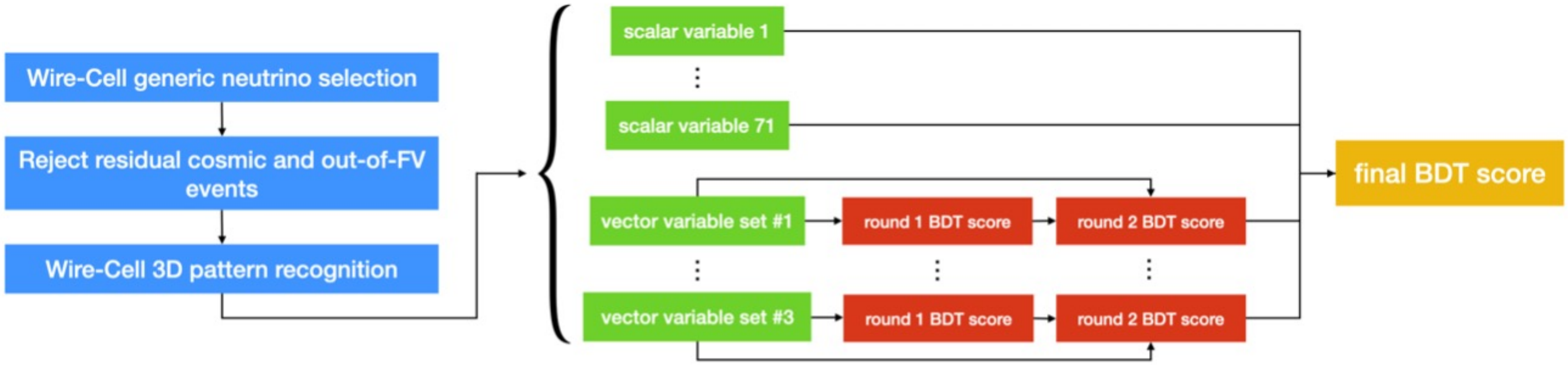
Gradient Boosting



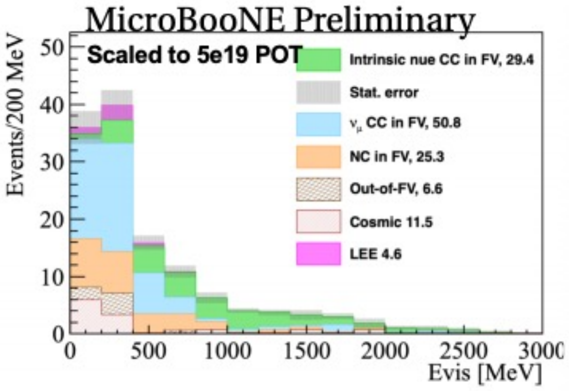
BDT for neutrino event selection

MicroBooNE Wire-Cell Team

Hierarchical BDT

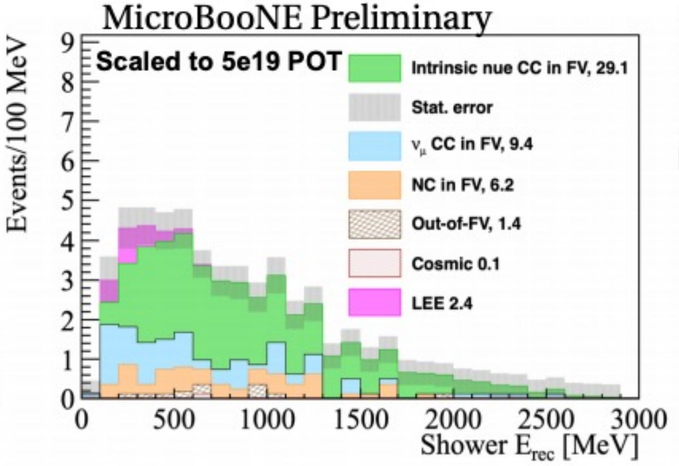


Cut-based



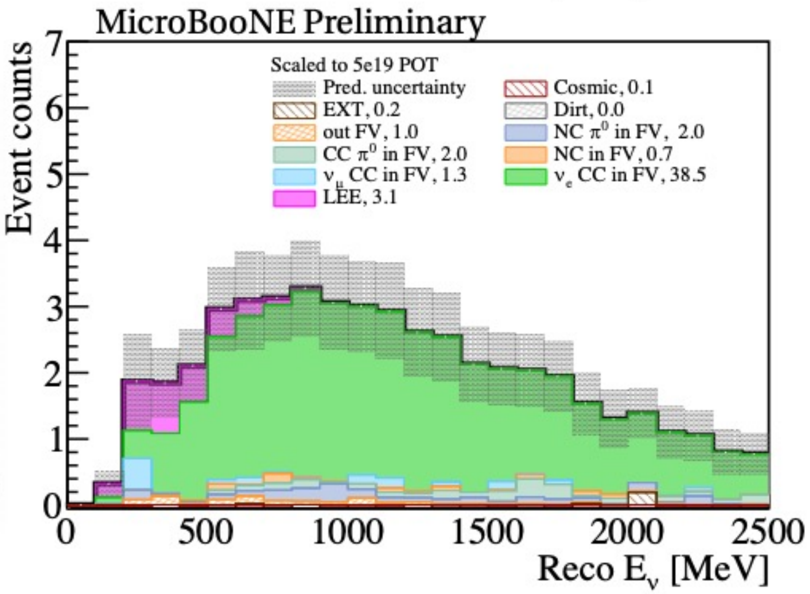
Poor purity
32% nueCC eff

TMVA BDT



63% nueCC purity
32% nueCC eff

XGBOOST BDT (final)

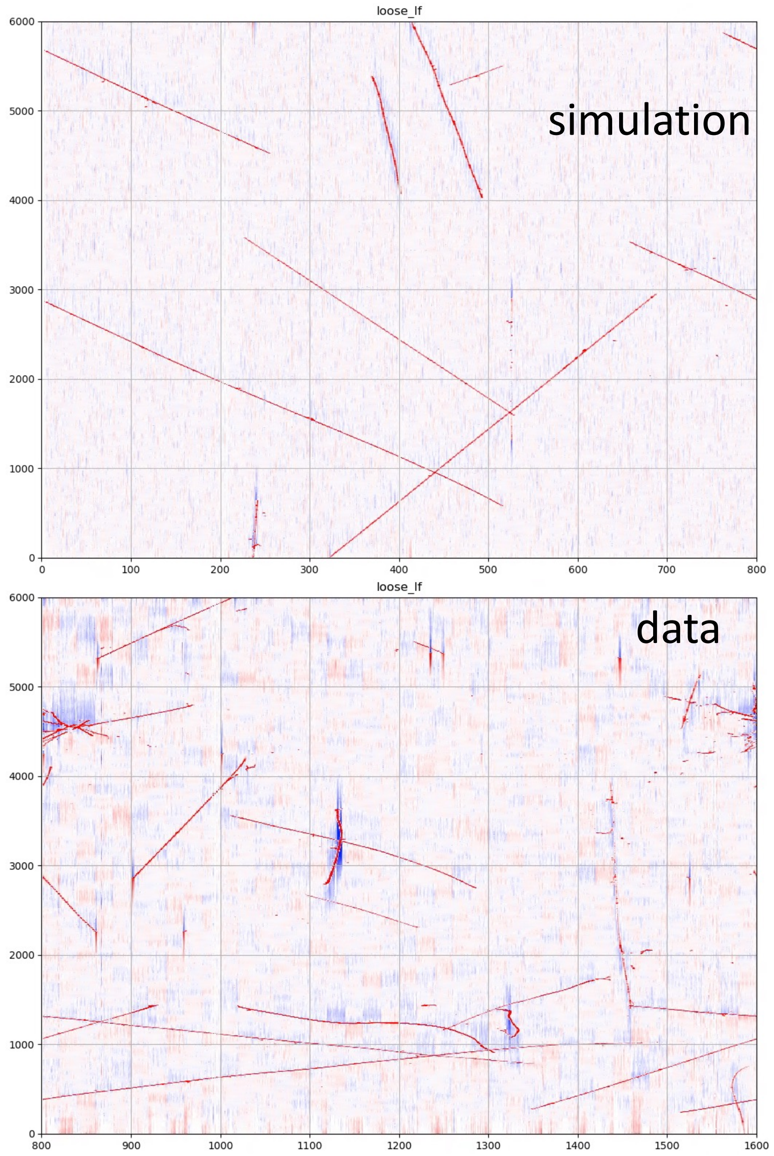


83% nueCC purity
42% nueCC eff

MICROBOONE-NOTE-1095-PUB

Large Scale Scientific Simulation Systematics GAN (LS4GAN)

Collaborating with BNL CSI

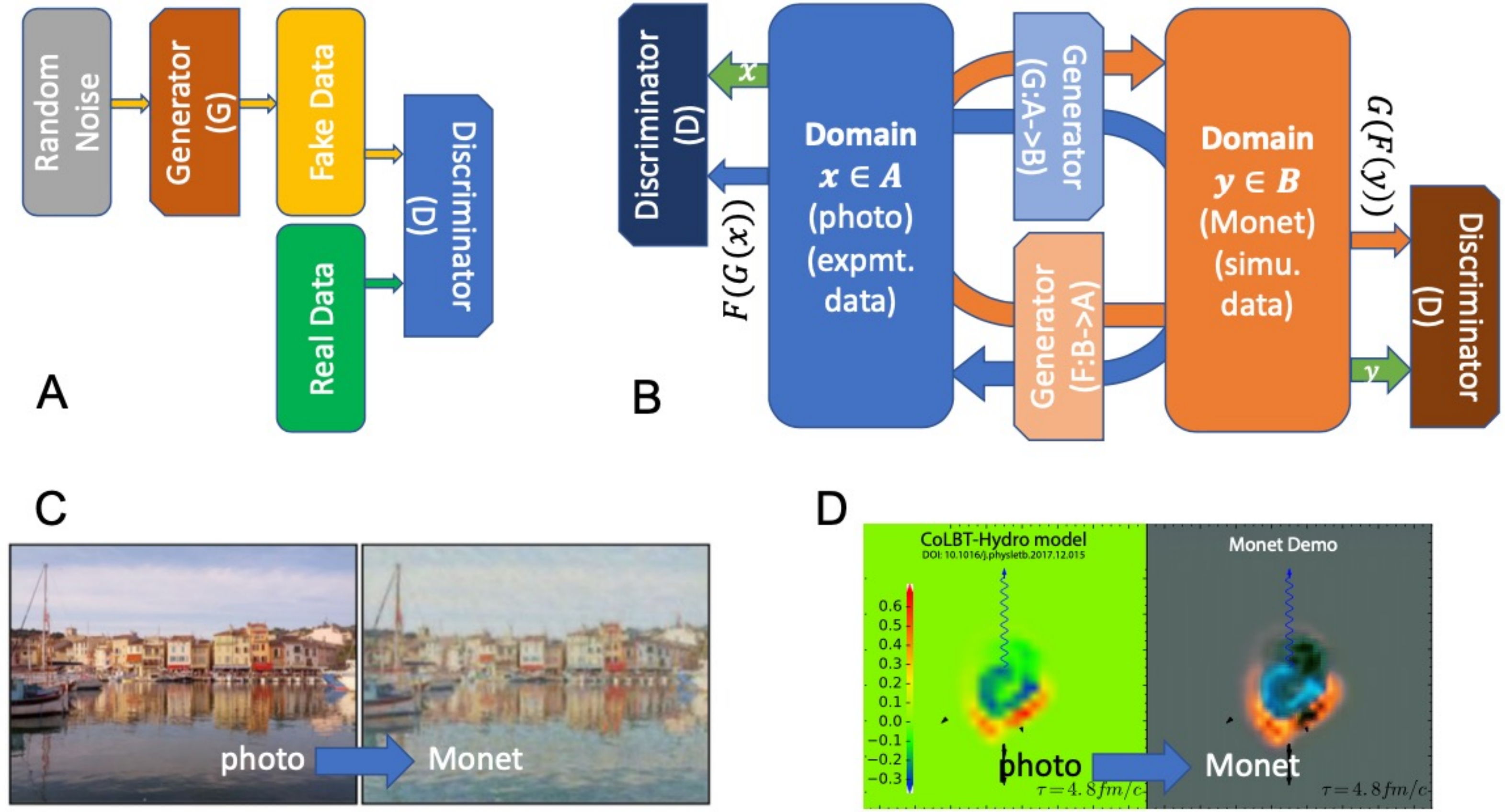


deconvoluted waveform with loose low frequency filter

LS4GAN: To precisely handle of systematic differences between simulation and reality using GANs

- to augment simulation
- to help understanding of sim-data difference
- https://ls4gan.github.io/

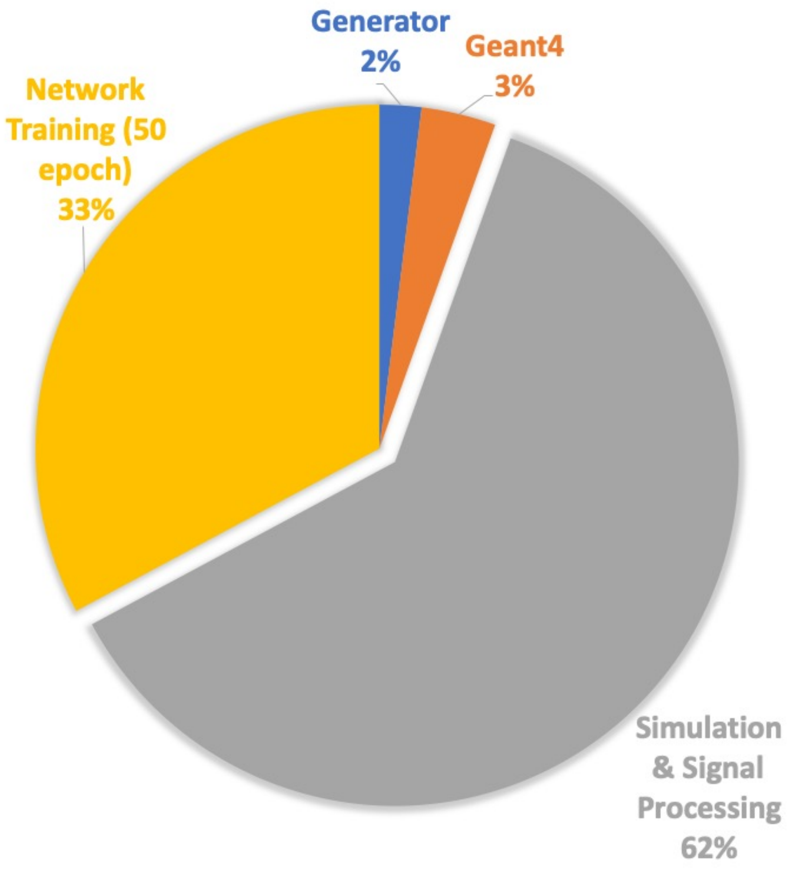
Main idea: Domain translation without paired data, e.g. CycleGAN



LArTPC simulation acceleration with portable solutions

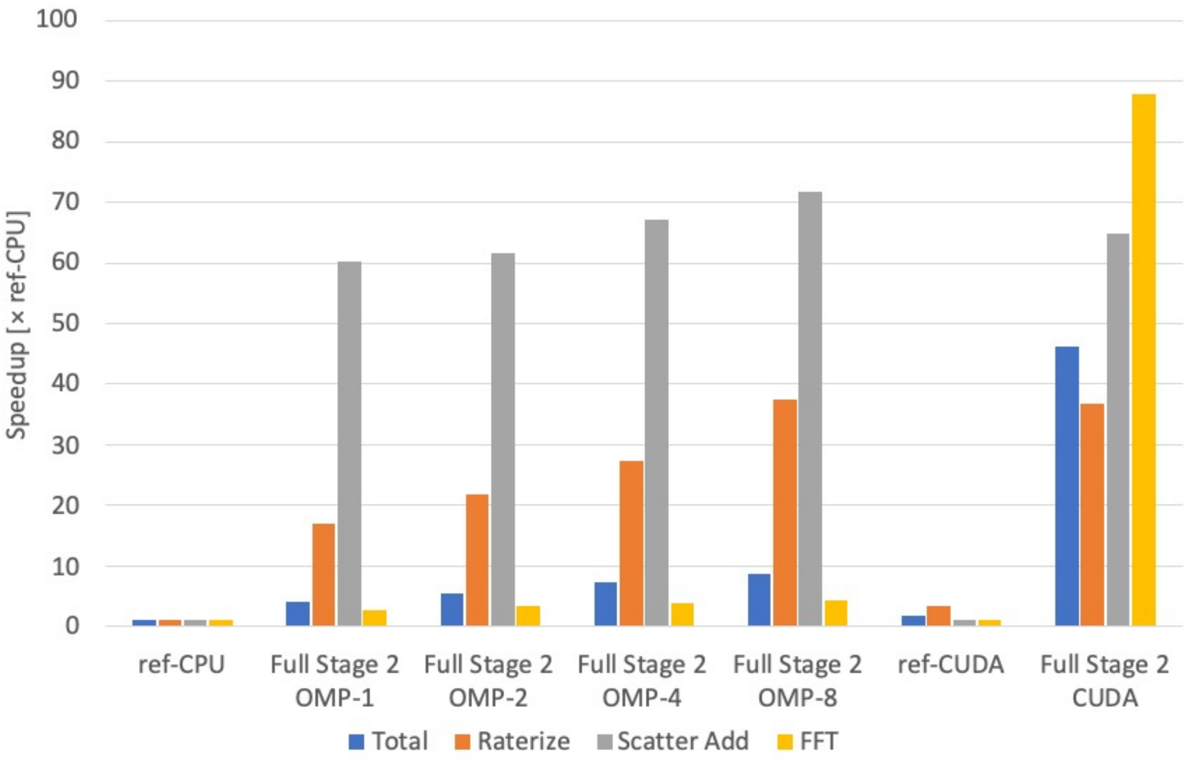
Collaborating with BNL CSI

Computing time breakdown for the DNN ROI finding task



Wire-Cell 2D Conv. Simulation Kokkos Porting (unfinished) Speedup

- H. Yu, Wire-Cell PPS talk at vCHEP21
- Intel i9-9900K, NVIDIA RTX 2080Ti



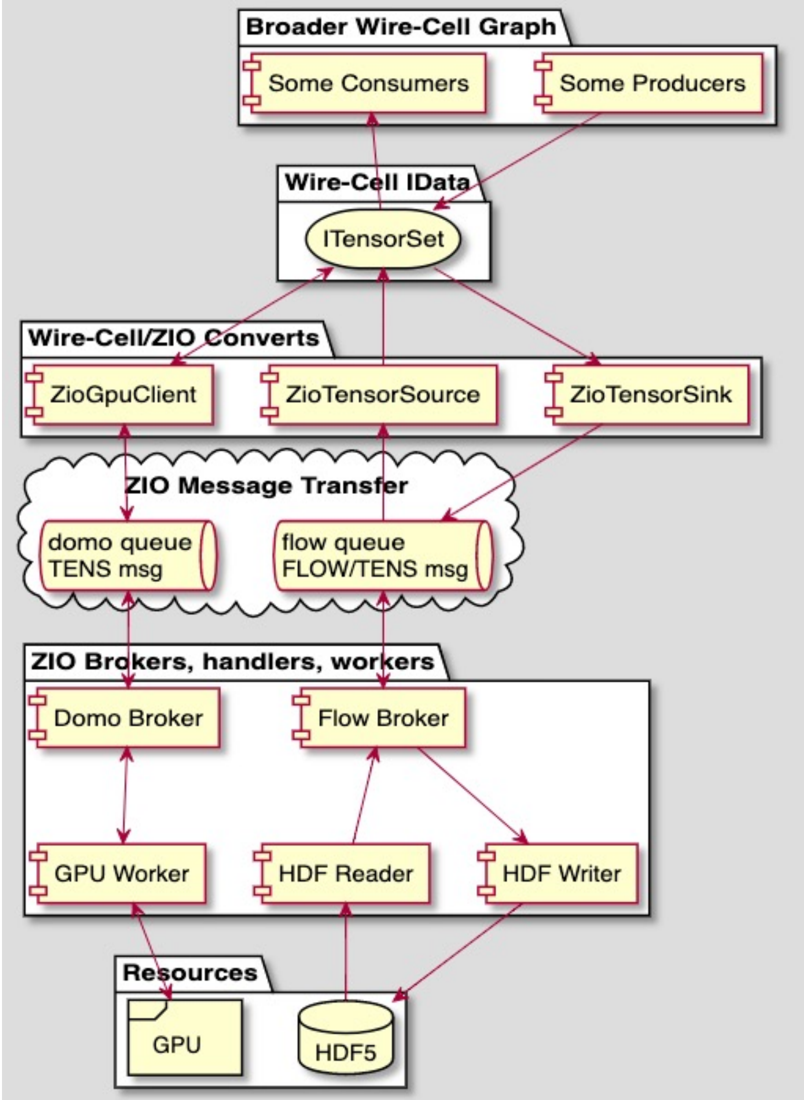
Inferencing with ZIO

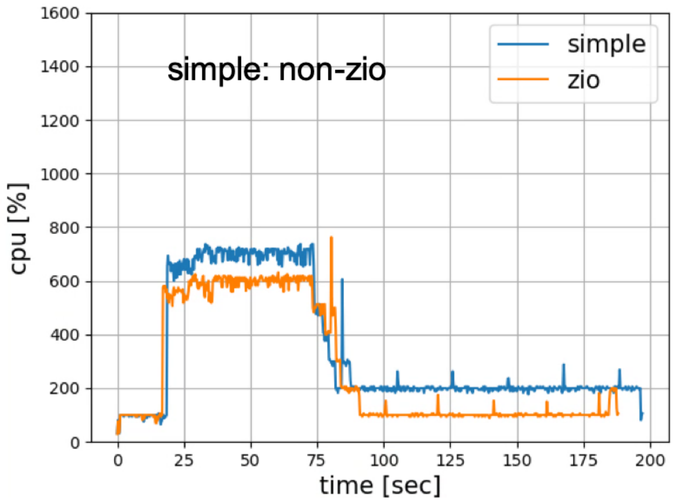
B. Viren

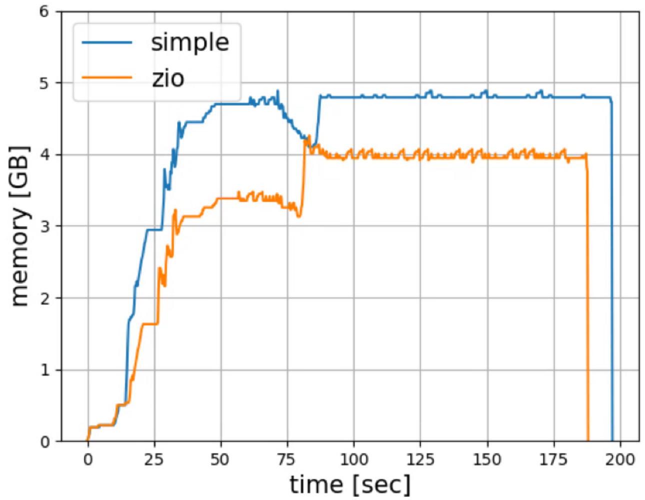
ZIO: ZeroMQ based distributed computing system developed by Brett

ZIO could be used to build neutral network inferencing services

Brett: **ZIO design discussion**
https://brettviren.github.io/zio/whytos.html







Summary

Wire-Cell is a project targeting LArTPC and beyond

- Wire-Cell Toolkit data flow programing framework
- ZIO distributed computing framework
- LS4GAN project
- ...

Improving LArTPC reconstruction takes both human learning and machine learning

- learn hardware features; explore most efficient reco. algorithms
- some issues suit AI/ML better
    - more complicated correlations
    - cover phase space faster

AI/ML

- + domain knowledge ⟶ efficient
- + validation ⟶ reliable

https://lar.bnl.gov/ml/