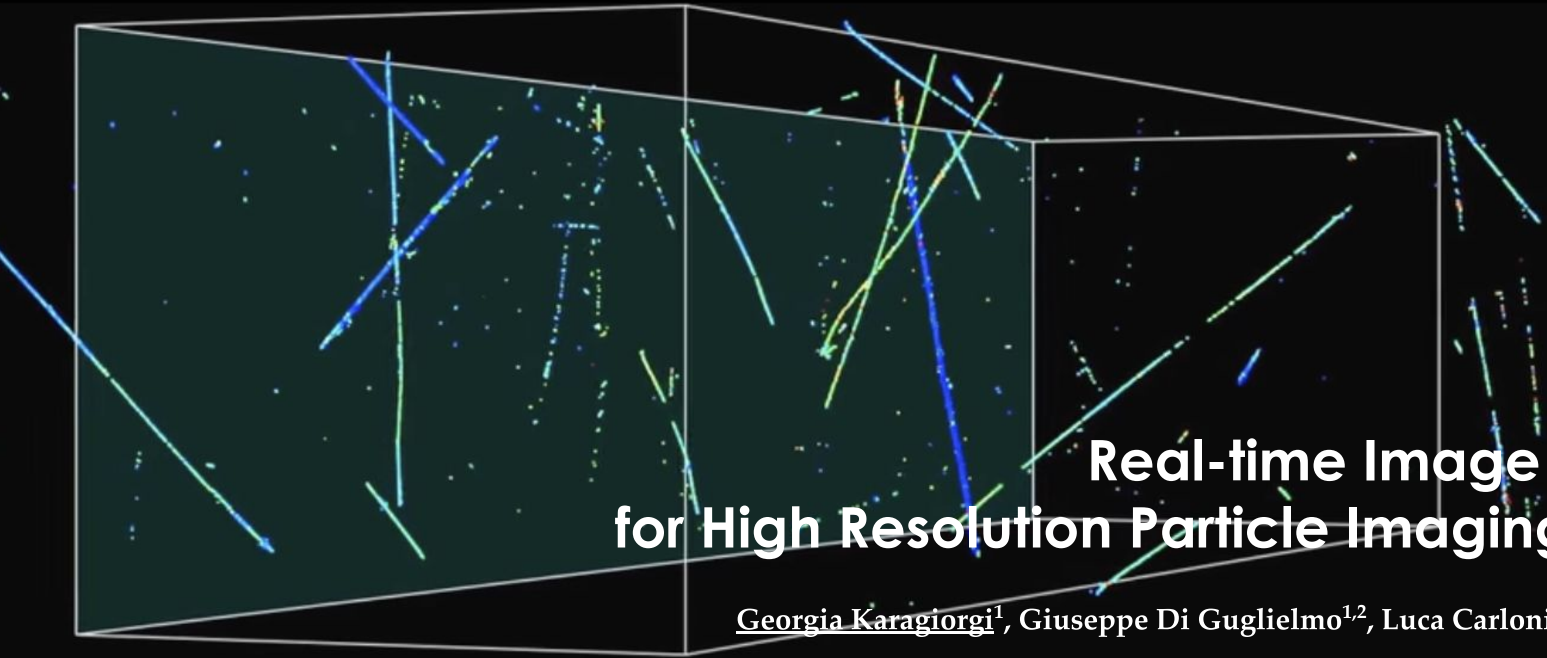
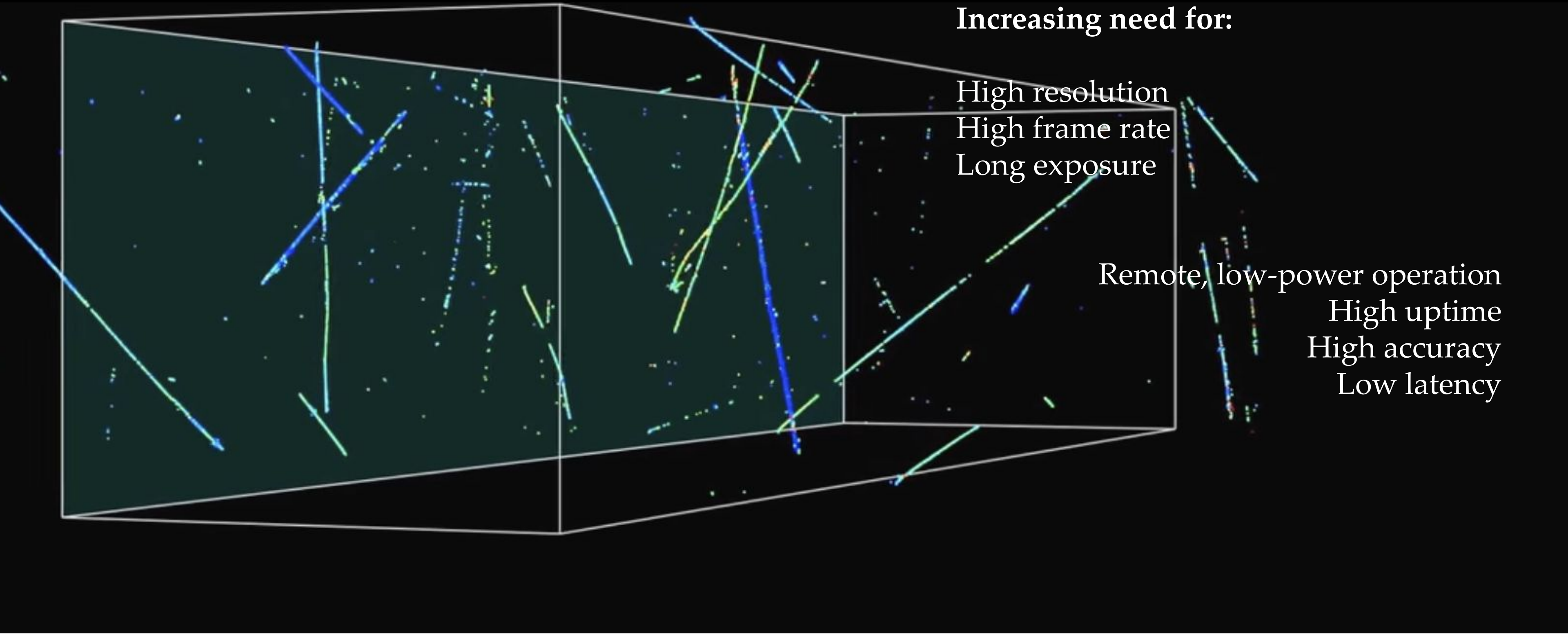


Real-time Image Processing for High Resolution Particle Imaging Detectors

Georgia Karagiorgi[1], Giuseppe Di Guglielmo[1,2], Luca Carloni[1], Nicholas Kasseinov[1]
[1]Columbia University
[2]Fermi National Accelerator Lab

Fast ML for Science Workshop 2022

Modern-day particle detectors

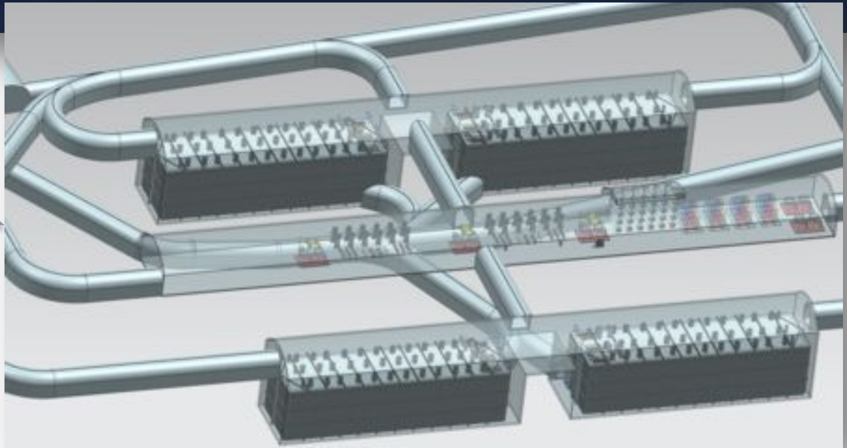


Increasing need for:

High resolution
High frame rate
Long exposure

Remote, low-power operation
High uptime
High accuracy
Low latency

Case study: Deep Underground Neutrino Experiment (DUNE)



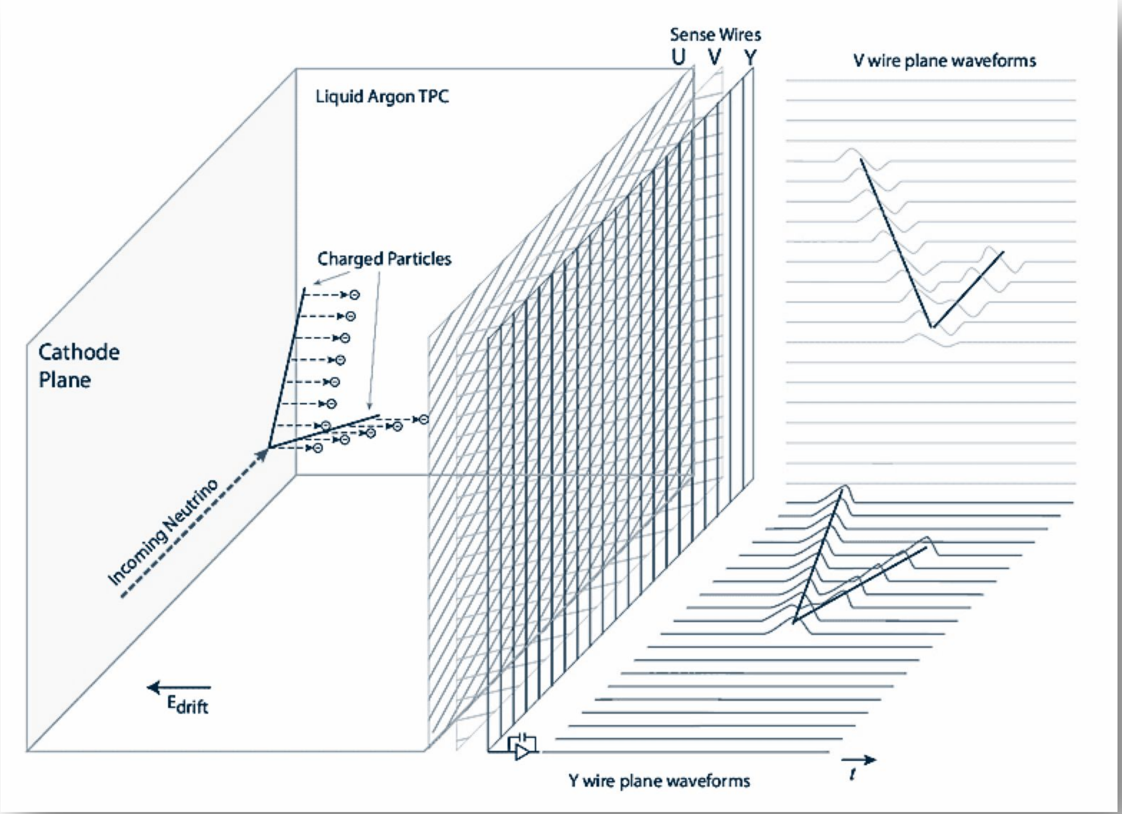
4 neutrino detectors
1 mile underground
17 ktons each
mm-scale resolution
O(kHz) imaging rate

DEEP UNDERGROUND NEUTRINO EXPERIMENT

Operations start by end of current decade
https://www.dunescience.org

ALFRED P. SLOAN FOUNDATION

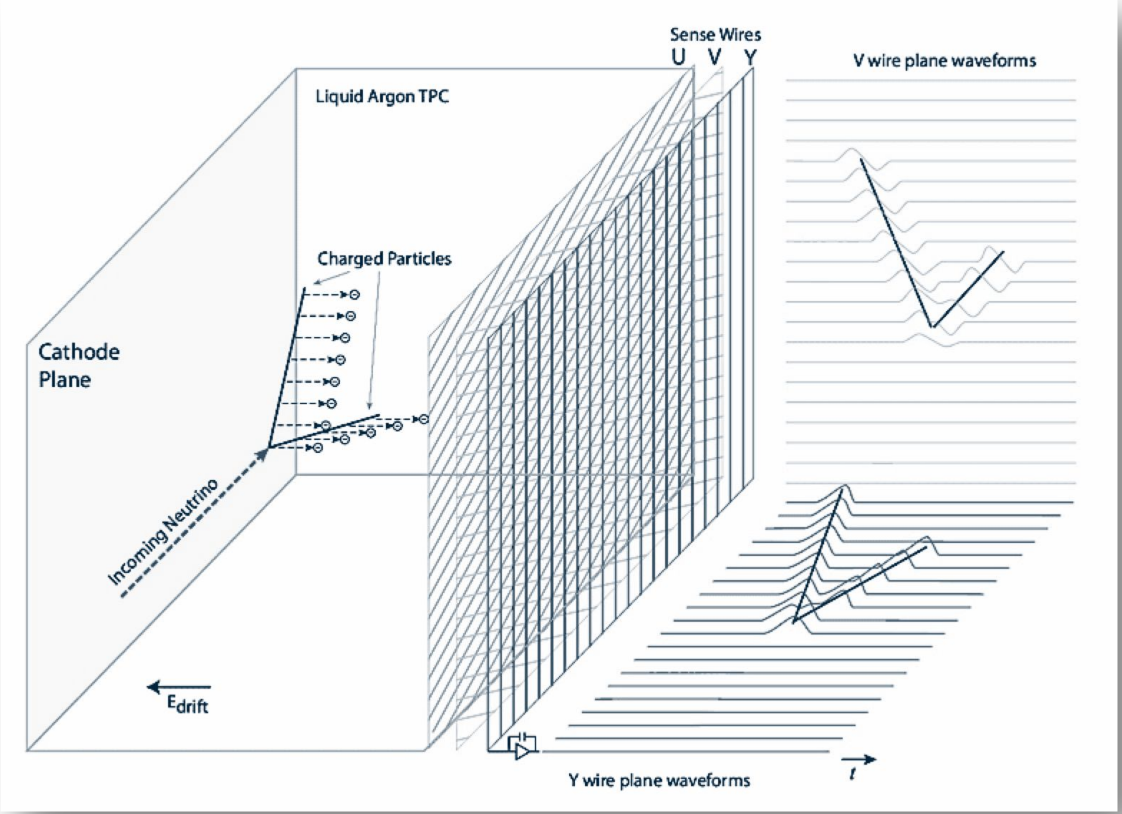
DUNE: a 3D stereoscopic imaging device



- Liquid Argon Time Projection Chamber (LArTPC): charged particle-imaging detector
- stereoscopic "video capture" of activity within detector volume with sub-mm spatial resolution

ALFRED P. SLOAN FOUNDATION

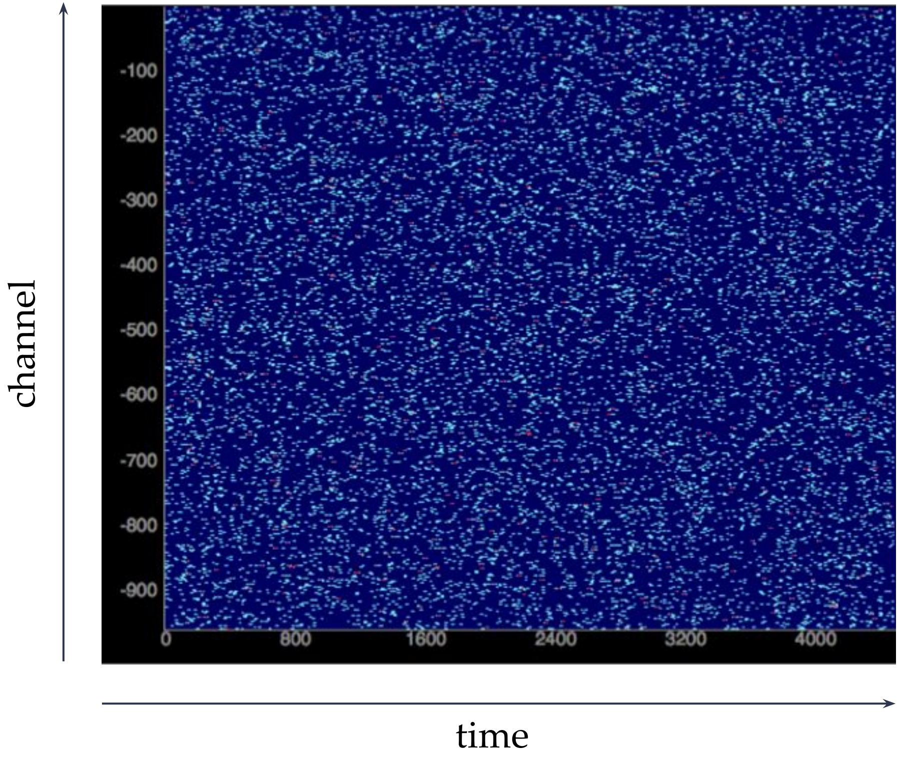
DUNE: a 3D stereoscopic imaging device



- Liquid Argon Time Projection Chamber (LArTPC): charged particle-imaging detector
- stereoscopic "video capture" of activity within detector volume with sub-mm spatial resolution
- **DUNE** high-resolution "video" stream:
  - up to 4x150 cell volumes
  - 11.5 megapixel frames per 2.25ms
  - 12-bit resolution
  - a total of **~40 terabits/s**
- **continuous operation for more than a decade**

ALFRED P. SLOAN FOUNDATION

DUNE images



Example of single frame from high-resolution video: One of three 2D views from one of hundreds of cells in the detector

Color scale represents energy deposition (due to ionization) in the detector

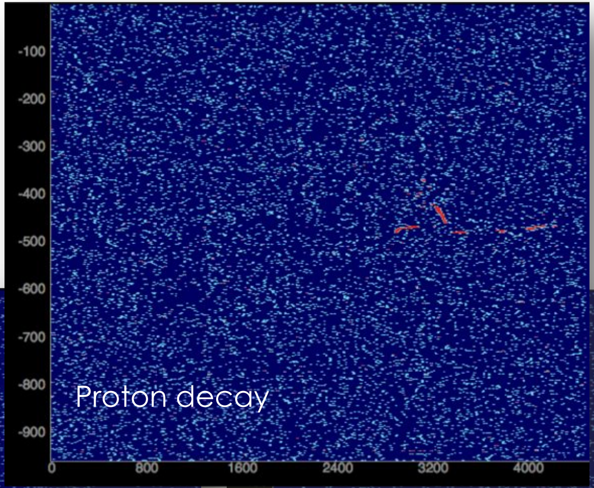
In this example: electronics noise and small energy deposits from ambient radiological impurities in the detector

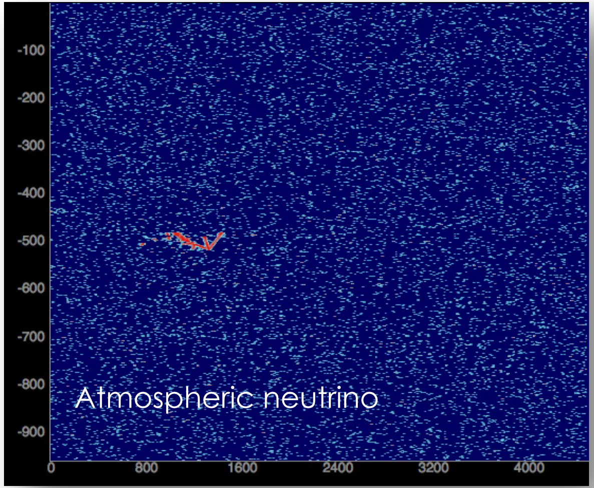
[simulation]

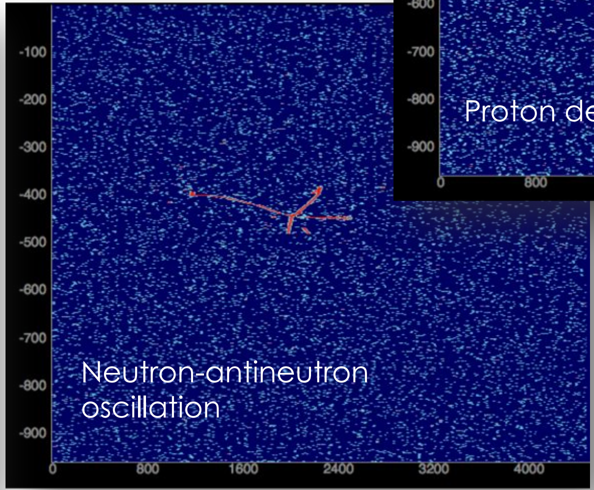
ALFRED P. SLOAN FOUNDATION

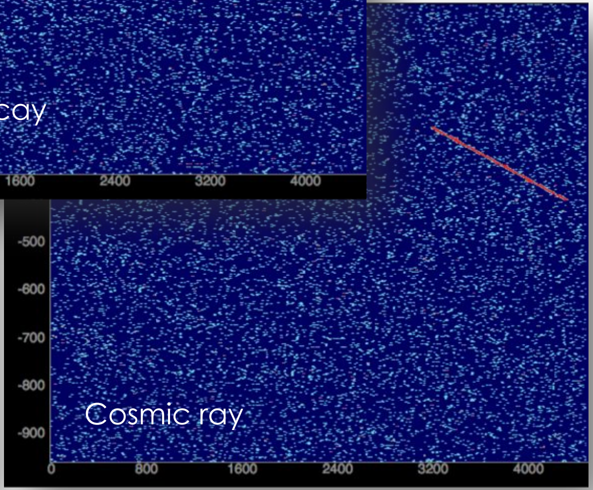
DUNE images

Rare interactions of interest in DUNE:







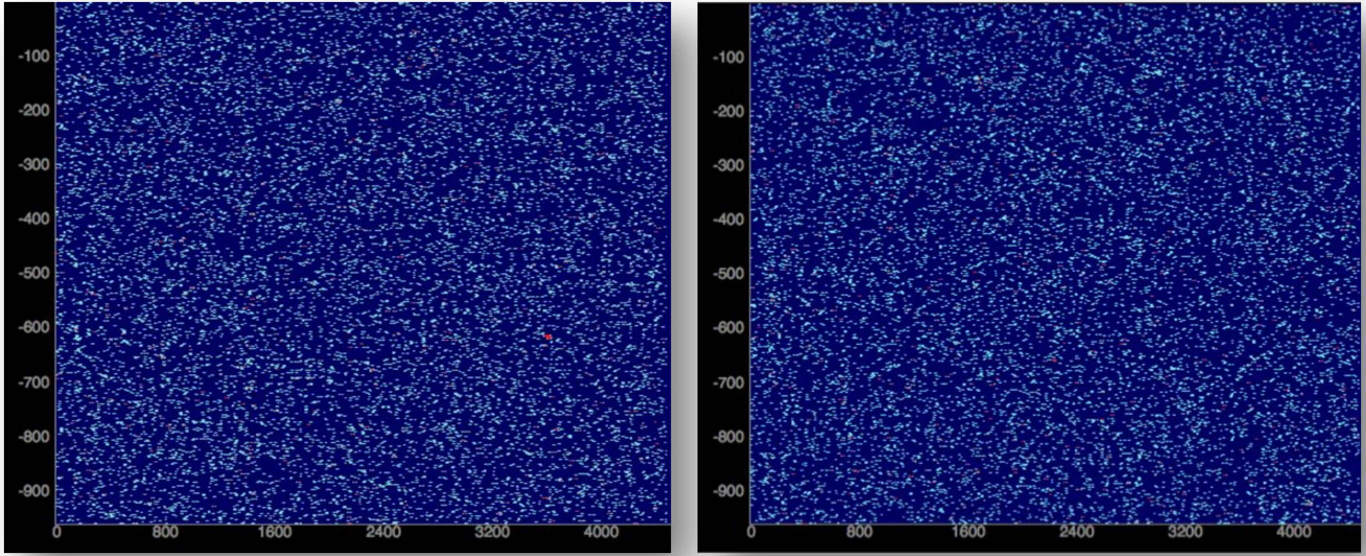


Easy to pick out from noise, but difficult to differentiate from each other
On average, can be differentiated based on their energy (pixel intensity) and topology characteristics (spatial extent, shape, e.g. tracks vs. showers and multiplicity, connected vs. detached, …)

ALFRED P. SLOAN
FOUNDATION

Biggest challenge:

Rare interactions of interest in DUNE: Neutrinos from supernova core collapse!
Very low energy and small (in extent) topology, similar to ambient radiological background activity:



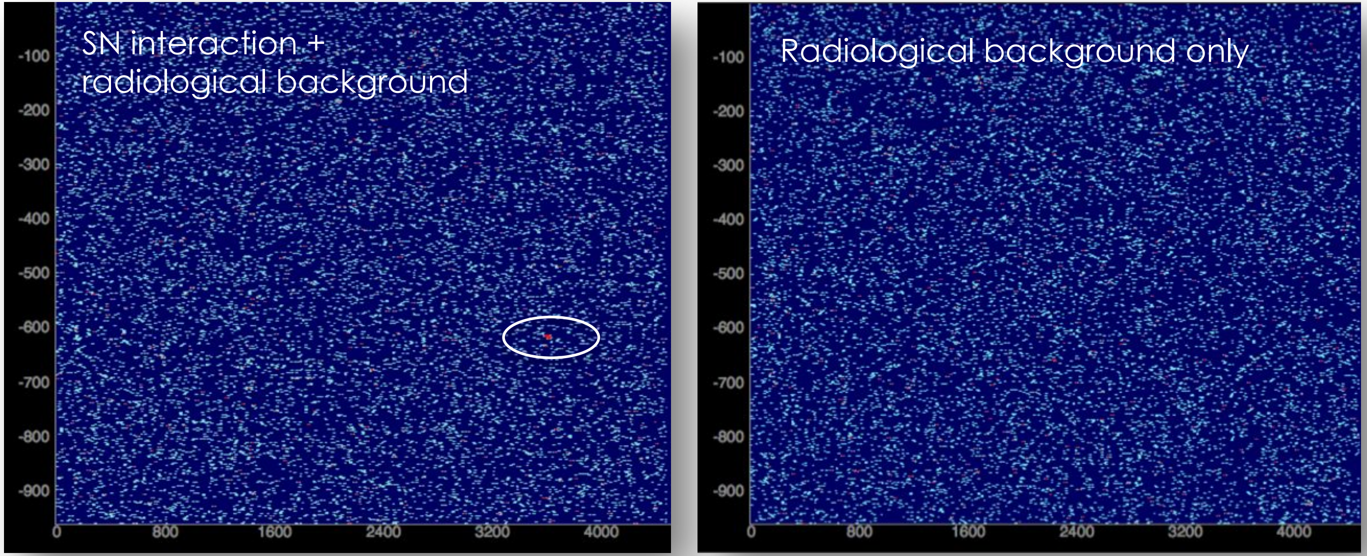
Requires $O(10^4)$ background suppression, while maintaining high efficiency to a frame containing a supernova neutrino interaction

ALFRED P. SLOAN FOUNDATION

Biggest challenge:

Rare interactions of interest in DUNE: Neutrinos from supernova core collapse!

Very low energy and small (in extent) topology, similar to ambient radiological background activity:

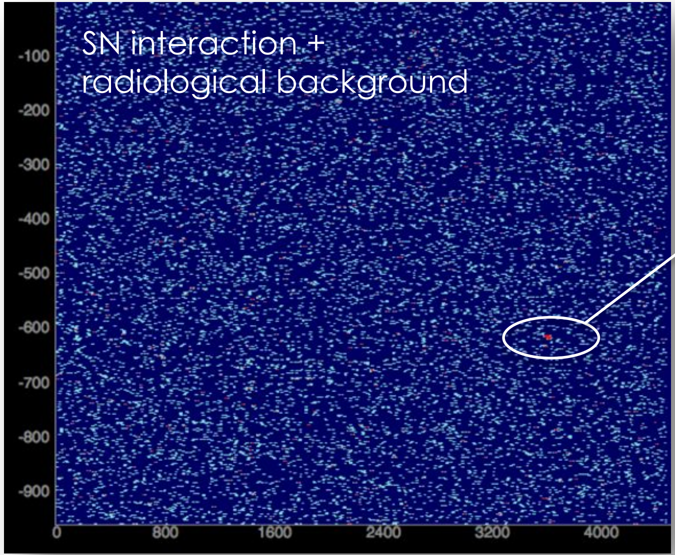


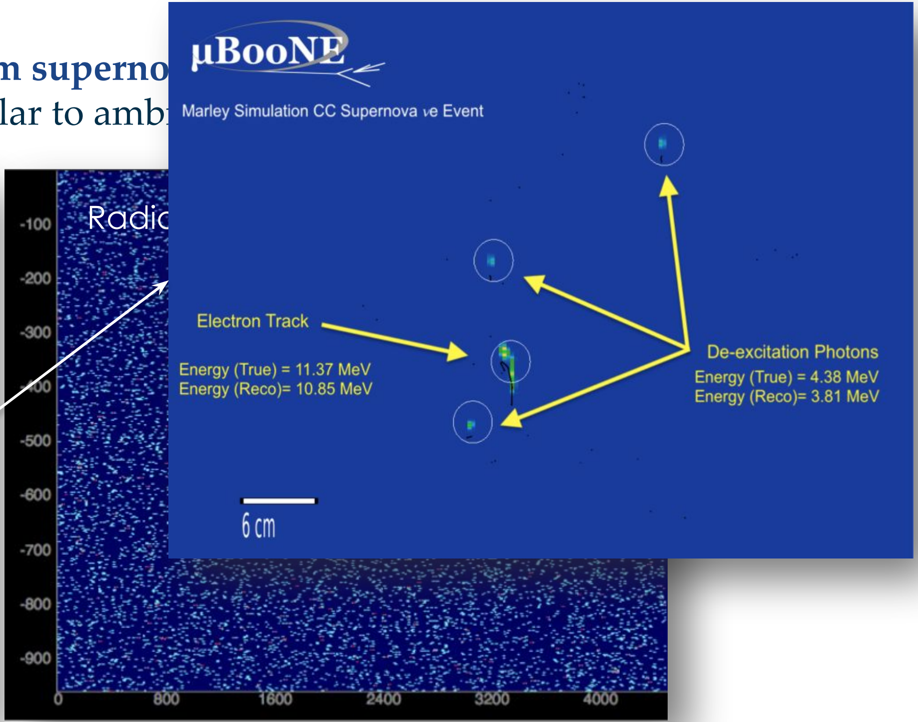
Requires O(10^4) background suppression, while maintaining high efficiency to a frame containing a supernova neutrino interaction

Biggest challenge:

Rare interactions of interest in DUNE: Neutrinos from superno

Very low energy and small (in extent) topology, similar to amb





Requires O(10^4) background suppression, while maintaining high efficiency to a frame containing a supernova neutrino interaction

DUNE **cannot afford to record all its streaming data** (at 40Tb/s) for 10 years! (Zettabytes of data!)

Its data acquisition system (DAQ) **must make intelligent and prompt decisions** on which data to keep (signal) and which data to reject (background):

"data selection" or **"trigger decision"**

Can we adopt and adapt advances in ML/AI for real-time data processing and data selection?

Real-time data processing using ML

A promising solution!

Accelerating Deep Neural Networks for Real-time Data Selection for High-resolution Imaging Particle Detectors

Yeon-jae Jwa
Dept. of Physics
Columbia University

Giuseppe Di Guglielmo
Dept. of Computer Science
Columbia University

Luca P. C
Dept. of Compu
Columbia U

DOI: 10.1109/NYSDS.2019.8909784

frontiers | Frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 18 May 2022
doi: 10.3389/frai.2022.855184

Check for updates

Real-Time Inference With 2D Convolutional Neural Networks on Field Programmable Gate Arrays for High-Rate Particle Imaging Detectors

Yeon-jae Jwa, Giuseppe Di Guglielmo, Lukas Arnold, Luca Carloni and Georgia Karagiorgi*

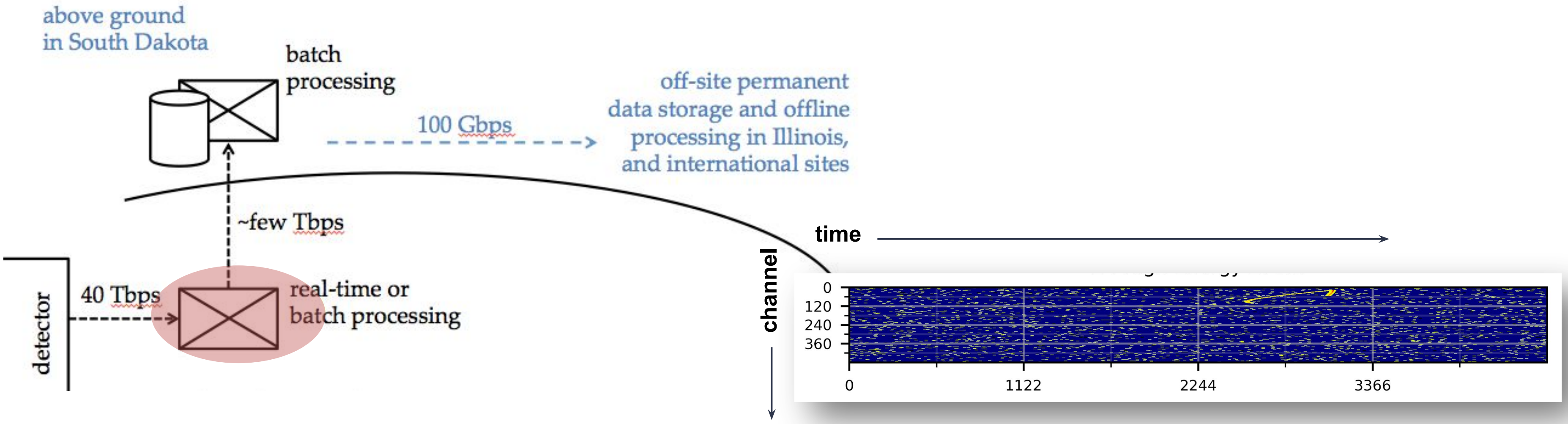
Columbia University, New York, NY, United States

This talk: Ongoing R&D to demonstrate such solution for DUNE

See also talk by B. Hawks on Monday

Front.Artif.Intell. 5 (2022) 855184

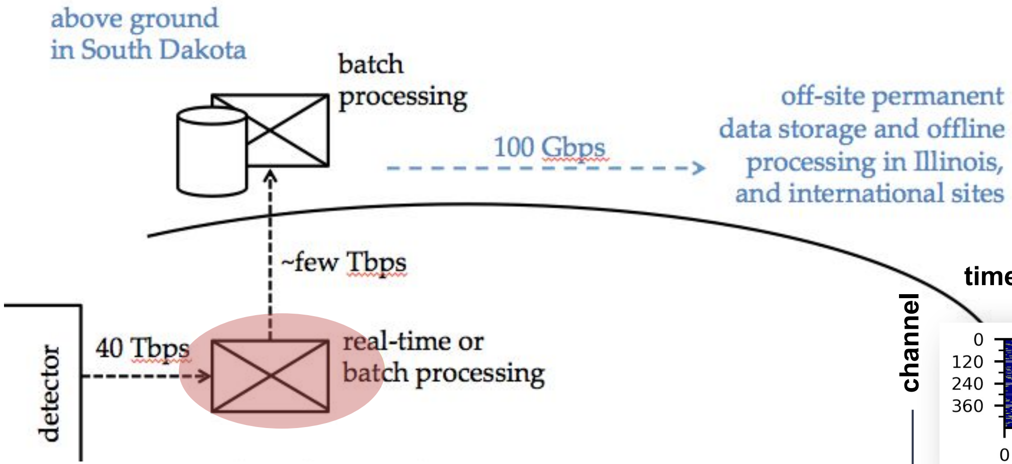
DUNE DAQ: ML-based Data Selection (Trigger)



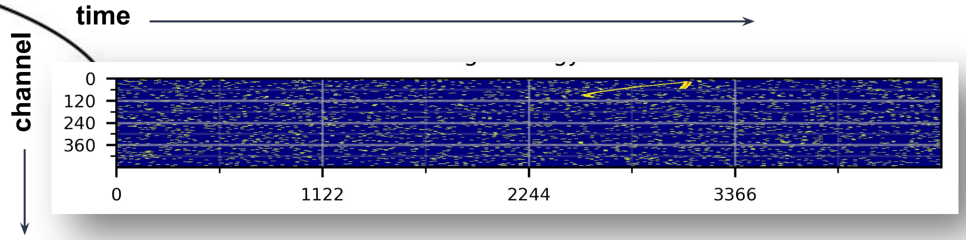
Underground:

- Pre-processing of streaming data in FPGAs (Xilinx UltraScale+ XKCU115)
- Low-level trigger in COTS CPUs; GPUs are a possibility, but important considerations are power constraints underground, and data I/O.

DUNE DAQ: ML-based Data Selection (Trigger)



Can real-time ML-based data selection be employed as part of the "low-level trigger"?



Underground:

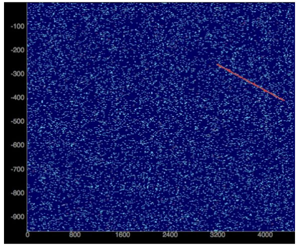
- Pre-processing of streaming data in FPGAs (Xilinx UltraScale+ XKCU115)
- Low-level trigger in COTS CPUs; GPUs are a possibility, but important considerations are power constraints underground, and data I/O.

Can FPGA resources underground be used for real-time ML-based data selection?

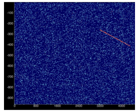
ALFRED P. SLOAN FOUNDATION

DUNE DAQ: ML-based Data Selection (Trigger)

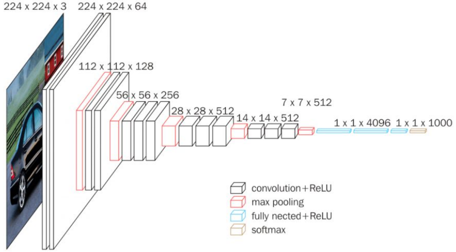
Real-time, frame-by-frame processing/classification (and selection):



raw image
input
(~4500x480 px)



pre-processing,
resizing (e.g. 64x64)



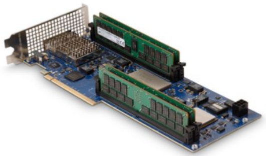
CNN
classification

- Background
- Supernova neutrino
- Other interaction

selection (e.g., lowest background
class score)

hls4ml

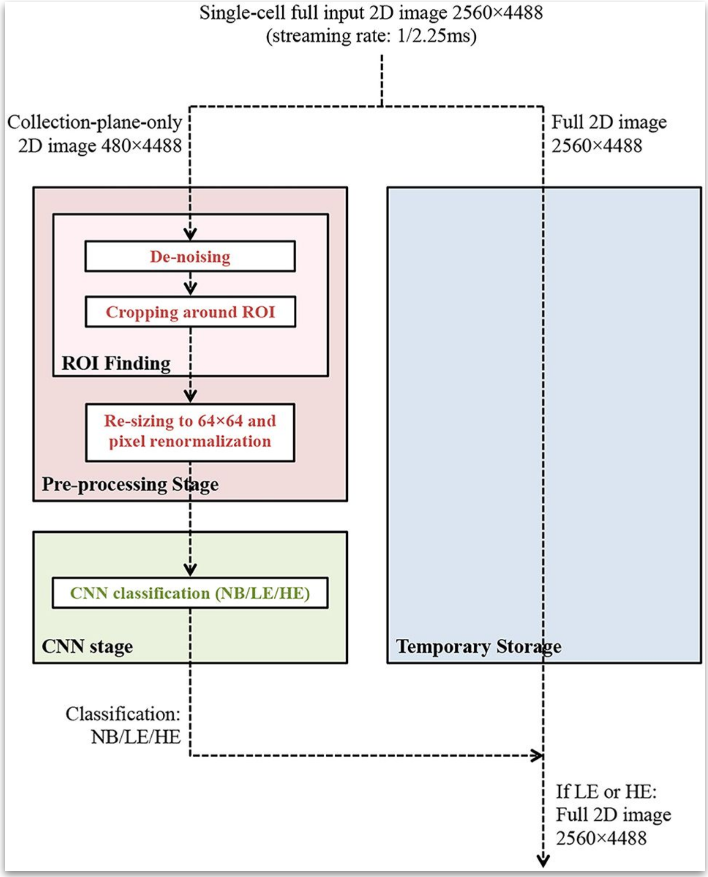
Micron



ALFRED P. SLOAN
FOUNDATION

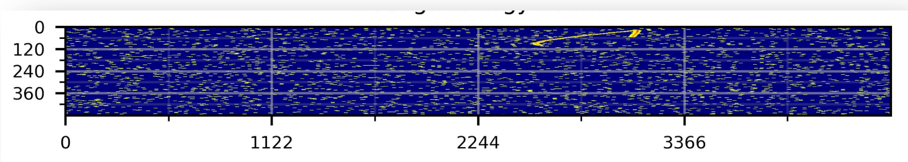
Viability of implementation in DUNE?

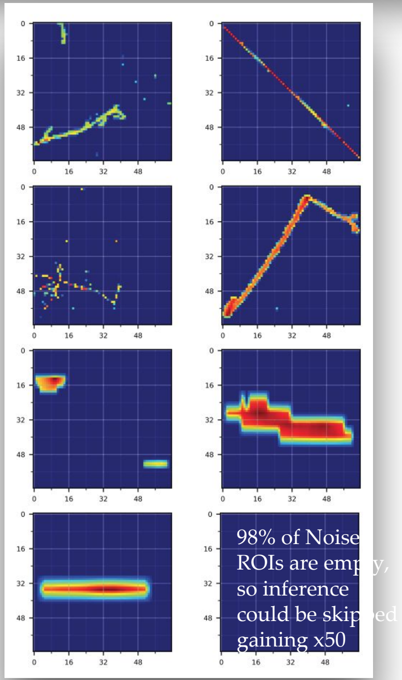
FPGA processing scheme:



ALFRED P. SLOAN FOUNDATION

Viability of implementation in DUNE?

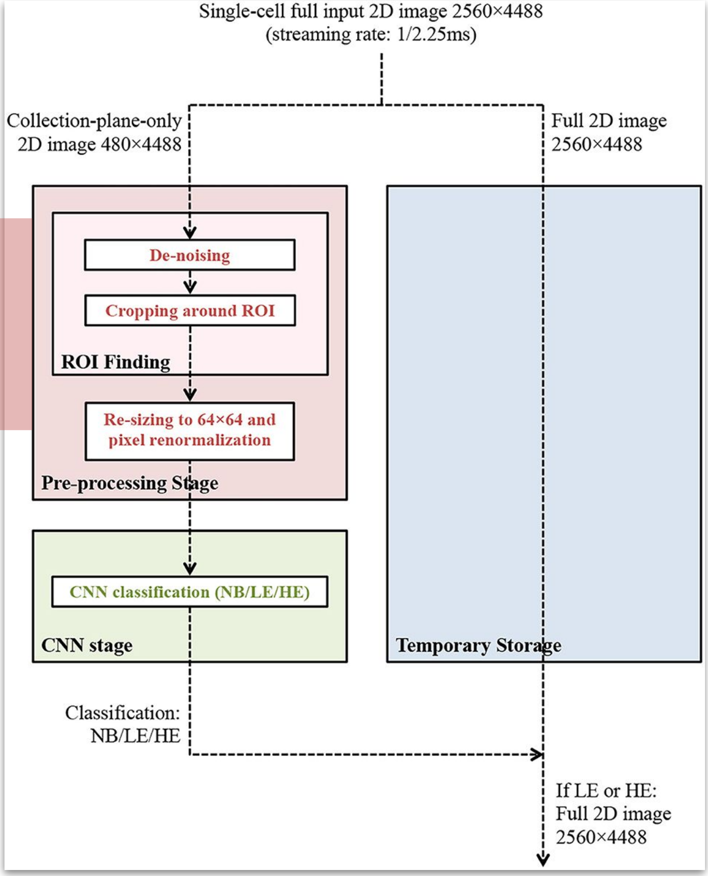




Other interactions

Supernova neutrino interactions

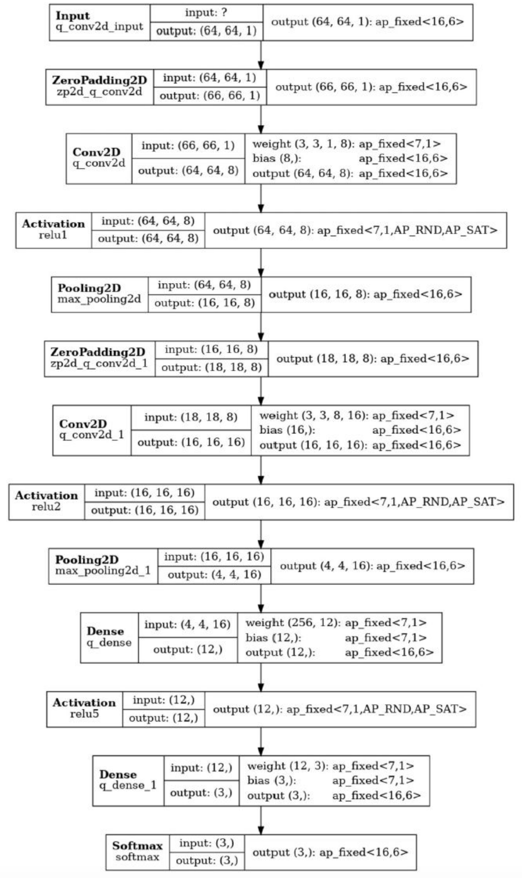
Background

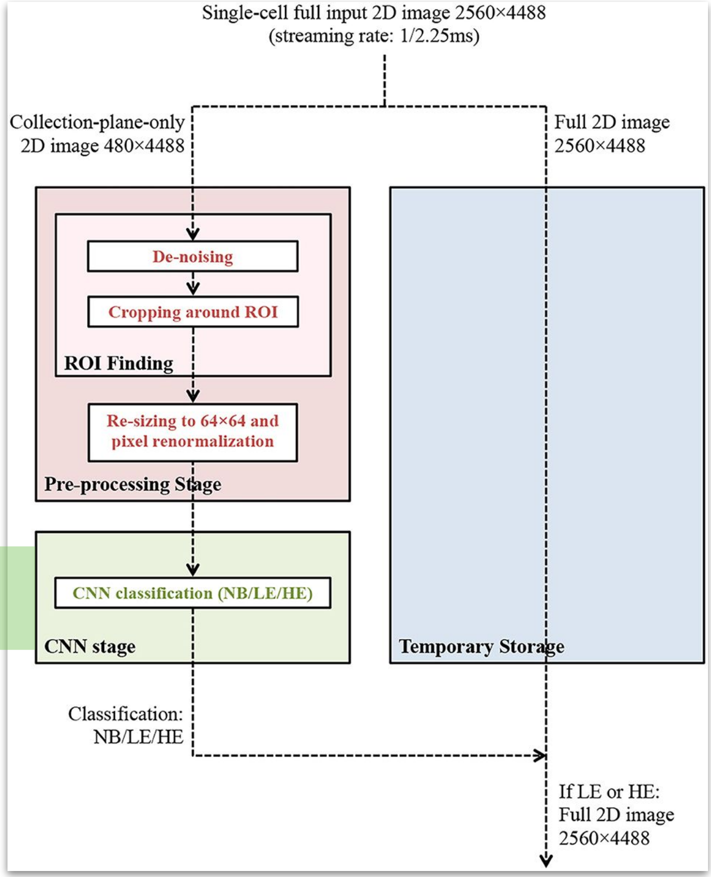


ALFRED P. SLOAN FOUNDATION

Viability of implementation in DUNE?

Small CNN, downsized, optimized to reduce computational footprint, and trained with quantization awareness





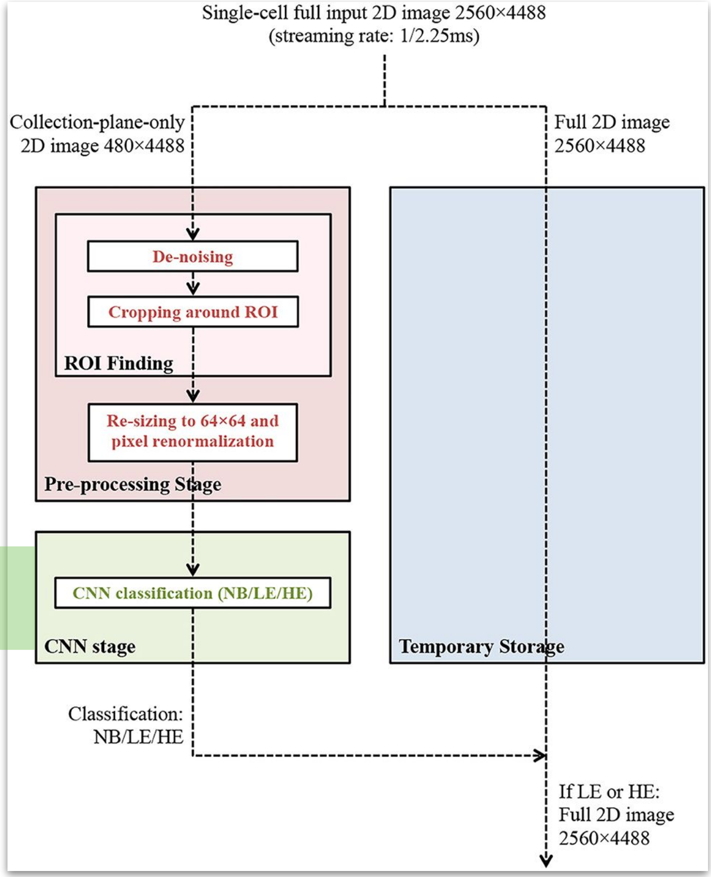
ALFRED P. SLOAN FOUNDATION

Viability of implementation in DUNE?

Achieved CNN accuracy (on non-empty image inputs):

Fixed-point (QAT)	NB	LE	HE
true_NB	99.68%	0.32%	0%
true_LE	3.90%	94.69%	1.41%
true_HE	3.25%	6.35%	90.40%
total accuracy		95.16%	

[Y. Jwa, *et al.* arXiv:2201.05638]



ALFRED P. SLOAN FOUNDATION

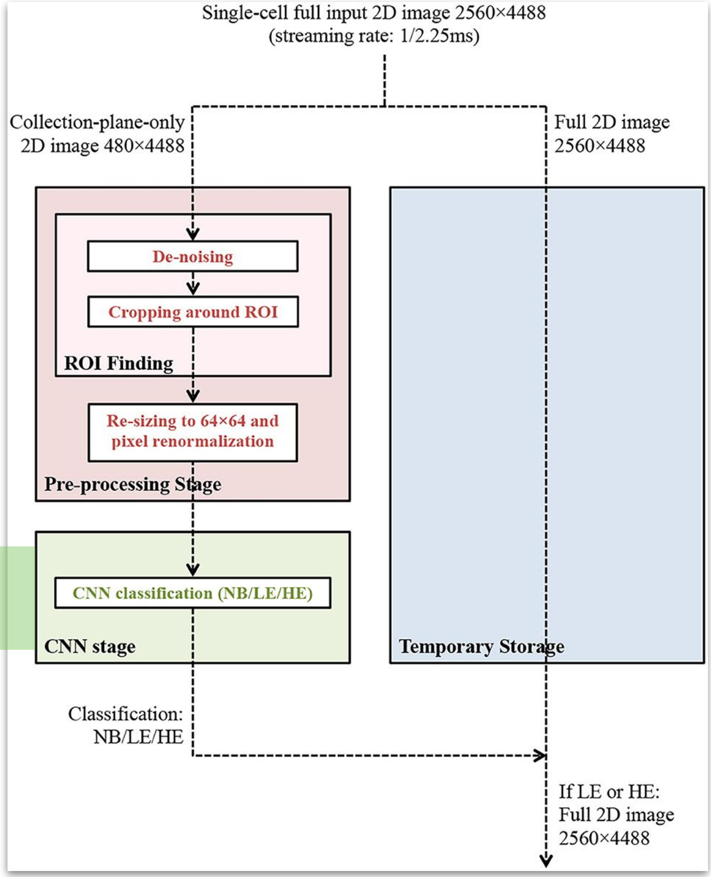
Viability of implementation in DUNE?

Estimated CNN resource utilization (from Vivado HLS) on Xilinx UltraScale+ (XKCU115) FPGA:

	Block RAM	DSP Units	Flip Flops	Look-up Tables
Available	4320	5520	1326720	663360
CNN02-DS-OP (PQT)	331 (7%)	4309 (78%)	226982 (17%)	163460 (24%)
Q-CNN02-DS-OP (QAT)	187 (4%)	2106 (38%)	142128 (10%)	138715 (20%)

[Y. Jwa, *et al.* arXiv:2201.05638]

Inference latency: < 25 μs



ALFRED P. SLOAN FOUNDATION

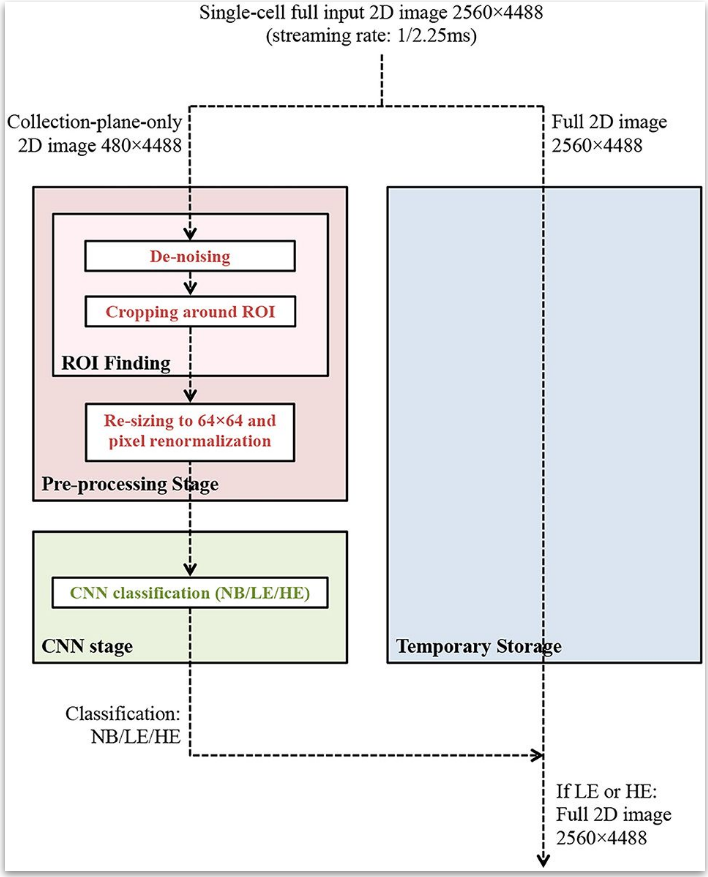
Viability of implementation in DUNE?

Proposed data processing pipeline can meet accuracy requirements for DUNE data selection.

At least CNN stage of processing can meet computational resource constraints, and latency constraints.

Next steps:

- Pre-processing resource and latency evaluation
- Full chain implementation and demonstration
- Evaluation of alternate pre-processing/full ML processing approaches



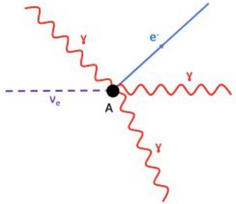
ALFRED P. SLOAN
FOUNDATION

Future directions

The ability to select a large sample of supernova neutrino candidates in real time offers the possibility of additional, prompt studies of directionality → DUNE could become an important player in multi-messenger astrophysics!

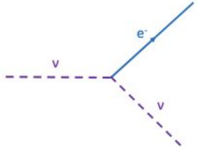
Charged Current

$$\nu_e \ {}^{40}Ar \rightarrow e^- \ {}^{40}K^*$$

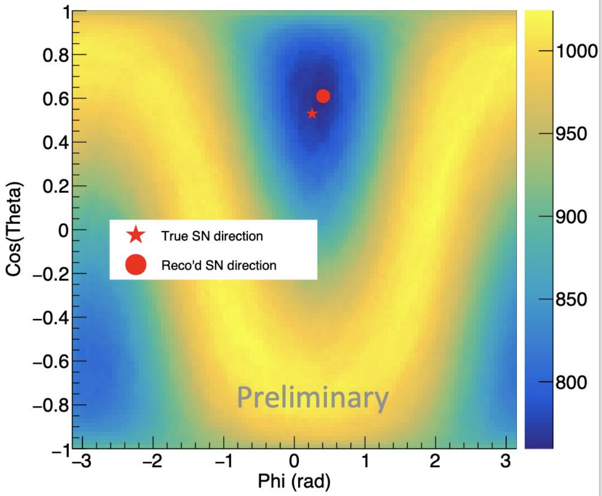


Elastic Scattering

$$\overset{(-)}{\nu} \ e^- \rightarrow \overset{(-)}{\nu} \ e^-$$



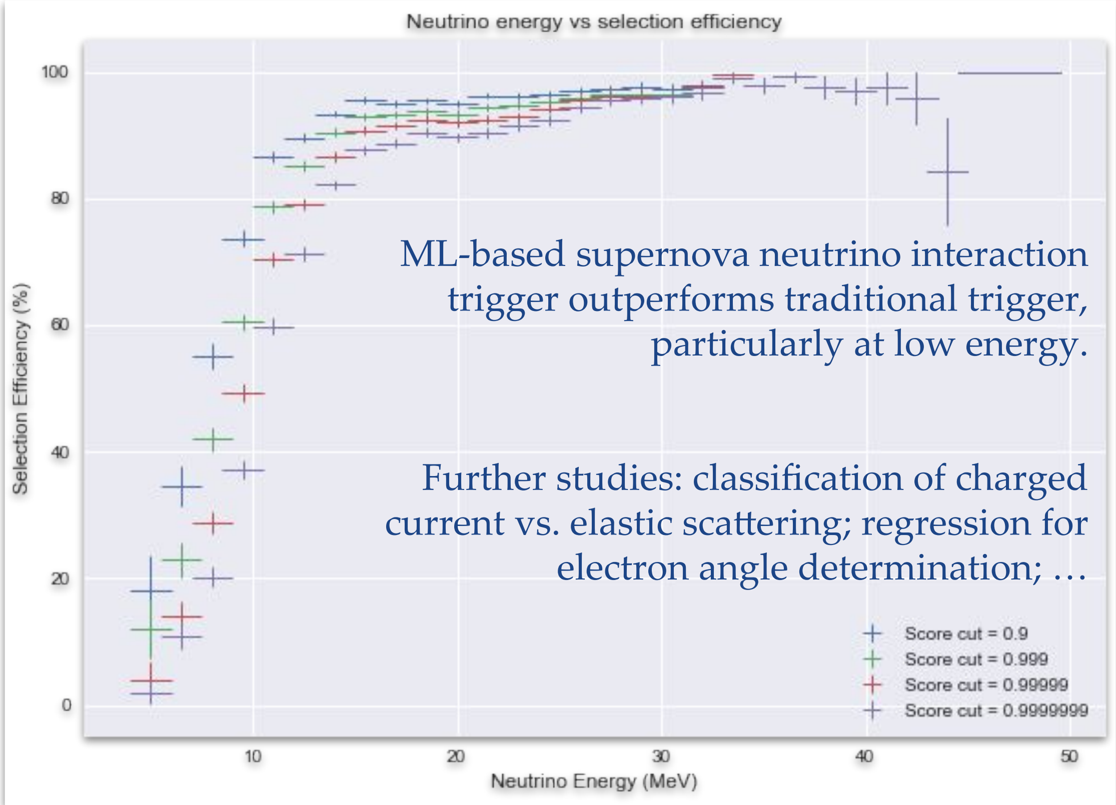
Electron direction in elastic scattering much more correlated with incoming neutrino direction.
Accurate and precise reconstruction of outgoing electron direction provides opportunity for supernova "pointing"!



Example of likelihood function values for 1 supernova sample of elastic scattering events, with radiological backgrounds and noise

[A.J. Roeth, ICHEP 2020]

Future directions



ML-based supernova neutrino interaction trigger outperforms traditional trigger, particularly at low energy.

Further studies: classification of charged current vs. elastic scattering; regression for electron angle determination; …

ML-based data selection efficiency performance bodes well for use of ML for additional data processing

ALFRED P. SLOAN FOUNDATION

Summary

Today's particle detectors stream raw data with increasingly higher resolution, volume, and complexity, posing a data processing (and analysis) challenge!

A particular detector technology (LArTPC) employed in the future DUNE experiment is unique: The large, relatively uniform, high-resolution, sparsely occupied with activity, and translationally invariant neutrino interactions in raw DUNE image-like data make it **ideal for applications of ML for image analysis for the purposes of data selection/triggering.**

Ongoing developments demonstrate the **viability of ML-based online/real-time triggering in DUNE,** and invite further exploration of such applications.

Fast ML can enable a deeper, more detailed look into particle physics data sets, and fundamental physics discoveries in astro/particle physics!

Thank you!

@GKaragiorgi
georgia@nevis.columbia.edu

ALFRED P. SLOAN FOUNDATION