Fermilab U.S. DEPARTMENT OF ENERGY Office of Science

End to End Real Time Supernova Pointing

Speaker: Maira Khan for the DUNE Collaboration

Fast ML Conference 2023
27 September 2023

In partnership with:

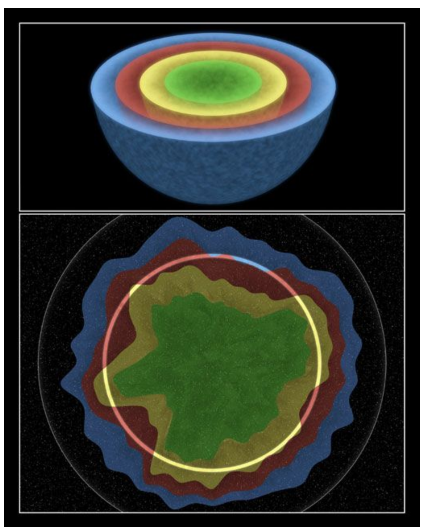
Duke University Columbia University DUNE

Supernovas and Disclaimer!

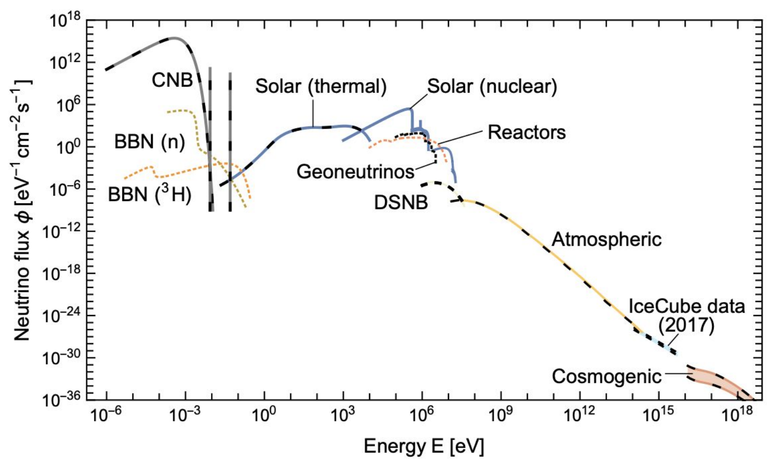
A core collapse supernova is a huge energetic release that is highly luminous and gives us much opportunity to learn about neutrino physics.

During this collapse, ~99% of the gravitational binding energy is released, giving us bursts of neutrinos of all flavors.

Disclaimer: Timing, depends on who you ask (neutrino physicists, astronomers)

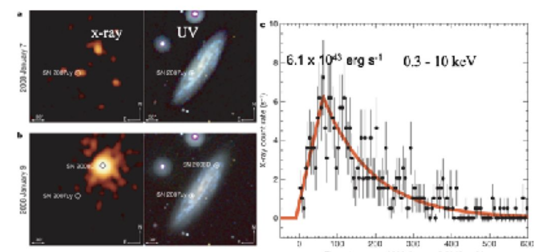


Core collapse SN burst: NASA



9-10 order of magnitude increase in diffuse supernova neutrino background

Grand unified neutrino spectrum at Earth: Sources and spectral components." *Reviews of Modern Physics* 92.4 (2020): 045006.

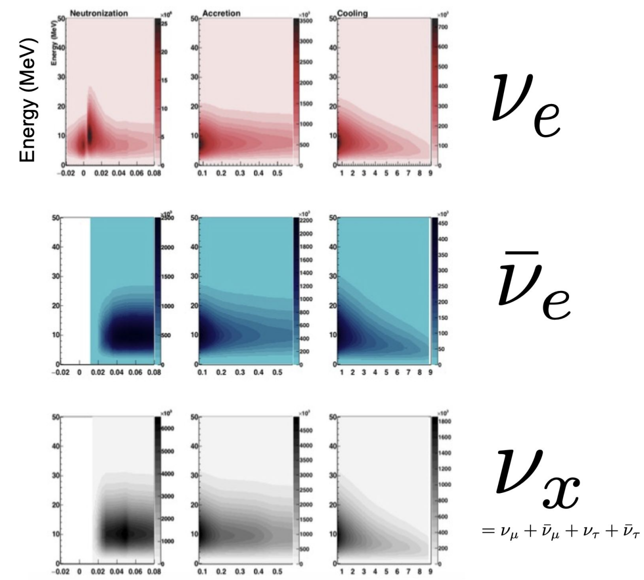


Soderberg et al 2008, *Nature*, 453, 469

Shock breakout in Type Ib supernova SN 2008D (serendipitous discovery while observing SN 2007uy)

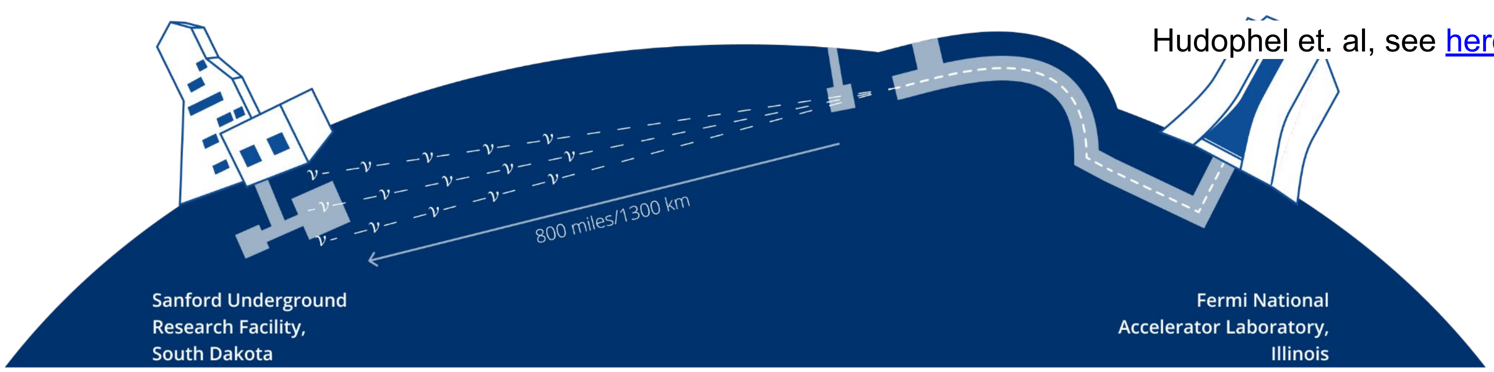
DUNE and Supernova Neutrino Physics

- Galactic Supernova events have much to tell us about neutrino physics.
- Time scale on O(seconds) gives us the opportunity to observe neutrinos of different flavors in the low energy range
- We saw dozens of neutrino interactions in 1987A, now we can potentially do better with resolution in DUNE
- Talk discussed two main aims
  - Reduction of radiological background
  - How we can suppress electronics noise and radiological background to obtain low energy signals from scattering interactions?



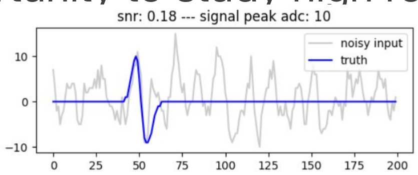
Neutrinos per cm^2 per bin (per ms per 0.5 MeV)

Hudophel et. al, see [here]

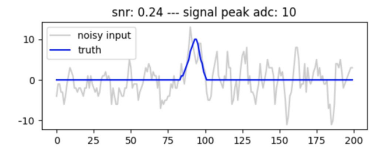


Unique Opportunity with LArTPC Detector

- DUNE's LarTPC detector gives is the opportunity to study high resolution images.
  - up to 4x200 cell volumes
  - 11.5 megapixel frames per 2.25ms
  - 14-bit resolution (raw waveforms)
  - 40 TB/s data



Plane U (collection)



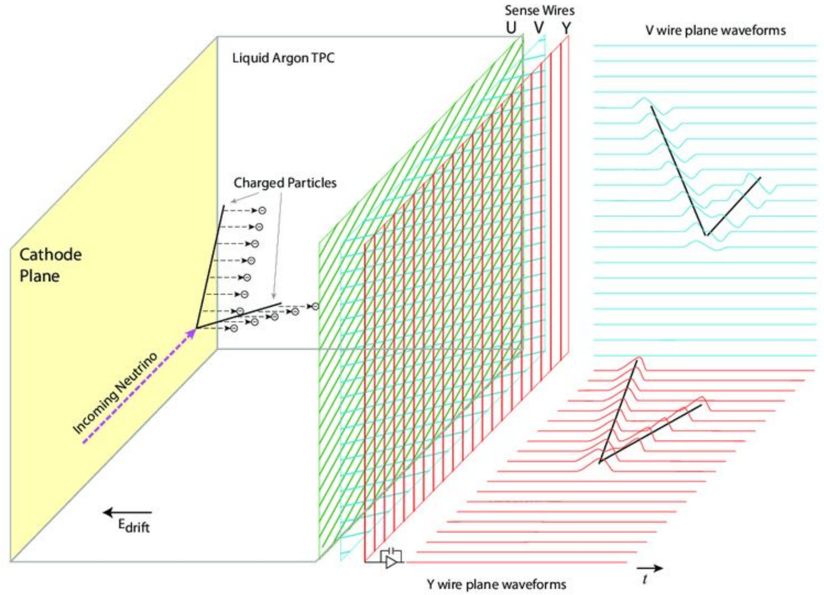
Plane Z (induction)

- 3 plane geometry, can study individual wires and plane projections.
- 70kt (4 modules) 17kt (1 module) horizontal drift
- LAr sensitive to nuCC interactions and eES, but eES are harder to detect (10:1 nuCC:eES ratio)
- Seems like an appropriate candidate for image based techniques

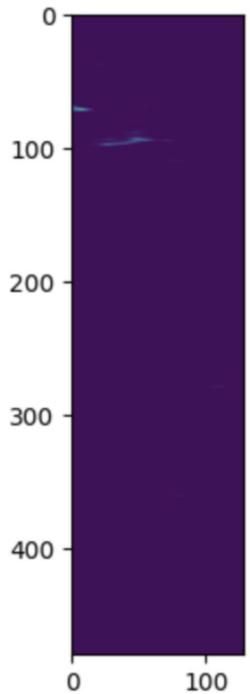
Channel	Events "Livermore" model	Events "GKVM" model
$\nu_e + ^{40}\mathrm{Ar} \rightarrow e^- + ^{40}\mathrm{K}^*$	2720	3350
$\overline{\nu}_e + ^{40}\mathrm{Ar} \rightarrow e^+ + ^{40}\mathrm{Cl}^*$	230	160
$\nu_x + e^- \rightarrow \nu_x + e^-$	350	260
Total	3300	3770
	no oscillations	*collective effects*

Ratio of nuCC:eES scattering interactions

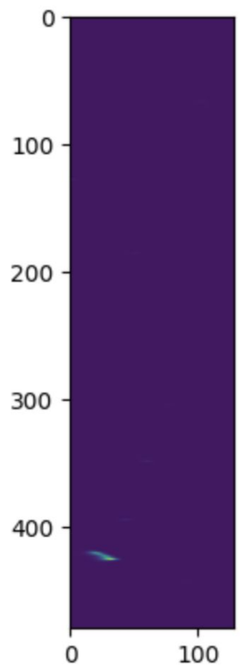
LBNF and DUNE CDR Volume 2: The Physics Program for DUNE at LBNF (arXiv:1512.06148)



Challenges of our LE Physics Data

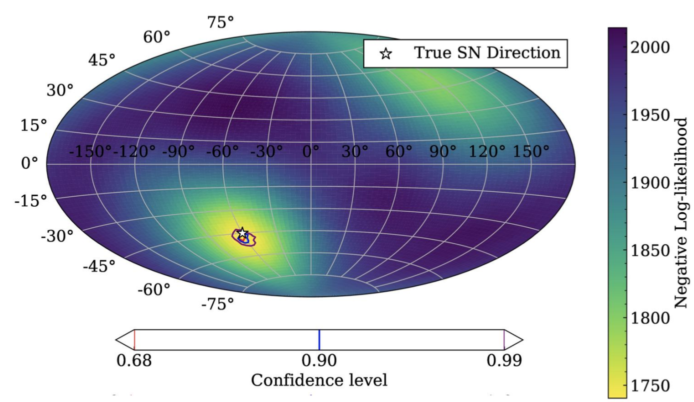


Simulated nuCC event with Ar-39 background



Simulated eES event with Ar-39 background

- Sparse images with radiological background events to suppress
- Pure background frames give no energy event display
- We need $O(10^4)$ background suppression to find our interaction events for SN Pointing
- We should retain a representation of the data that allows us to get directional information from eES events to point to a direction on our SkyMap



See DUNE-Doc-23154

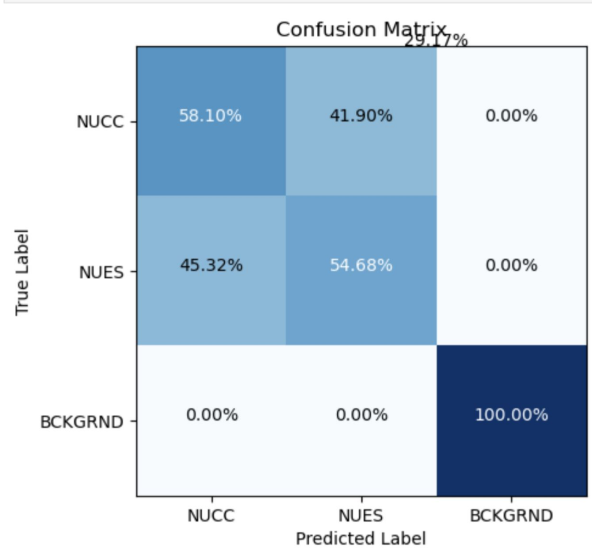
Data Reduction Using CNN

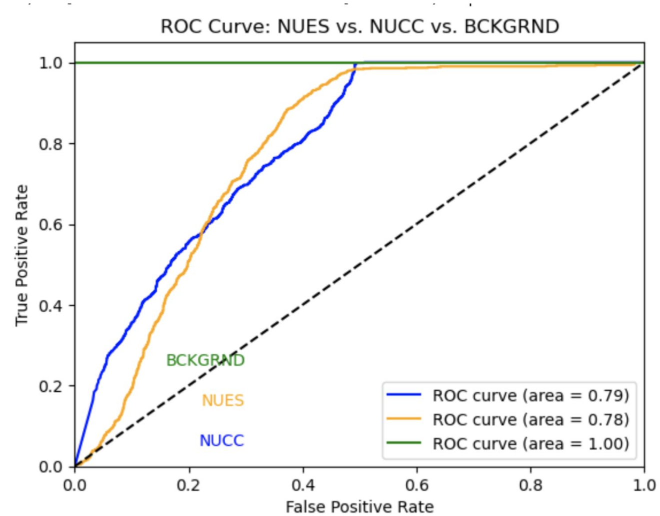
We must remove pure radiological background frames, and retain those only with eES or nuCC events that contain directional information.

On a O(10^4) frames, we can remove 100.0% of pure background frames of size 480 by 128 for pixel frames with a 12 bit resolution.

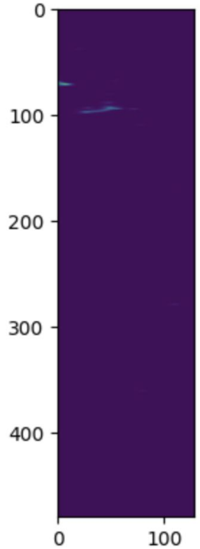
Many results about fast implementation of CNN in hardware as seen [here]

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape
=================================================================
 zero_padding2d (ZeroPaddin  (None, 512, 128, 1)
 g2D)

 conv2d (Conv2D)             (None, 510, 126, 32)

 average_pooling2d (Average  (None, 127, 31, 32)
 Pooling2D)

 conv2d_1 (Conv2D)           (None, 125, 29, 64)

 average_pooling2d_1 (Avera  (None, 62, 14, 64)
 gePooling2D)

 dropout (Dropout)           (None, 62, 14, 64)

 conv2d_2 (Conv2D)           (None, 60, 12, 64)

 average_pooling2d_2 (Avera  (None, 30, 6, 64)
 gePooling2D)

 conv2d_3 (Conv2D)           (None, 28, 4, 32)

 average_pooling2d_3 (Avera  (None, 14, 2, 32)
 gePooling2D)

 dropout_1 (Dropout)         (None, 14, 2, 32)

 flatten (Flatten)           (None, 896)

 dense (Dense)               (None, 8)

 dropout_2 (Dropout)         (None, 8)

 dense_1 (Dense)             (None, 3)

=================================================================
Total params: 81411 (318.01 KB)
Trainable params: 81411 (318.01 KB)
Non-trainable params: 0 (0.00 Byte)
```



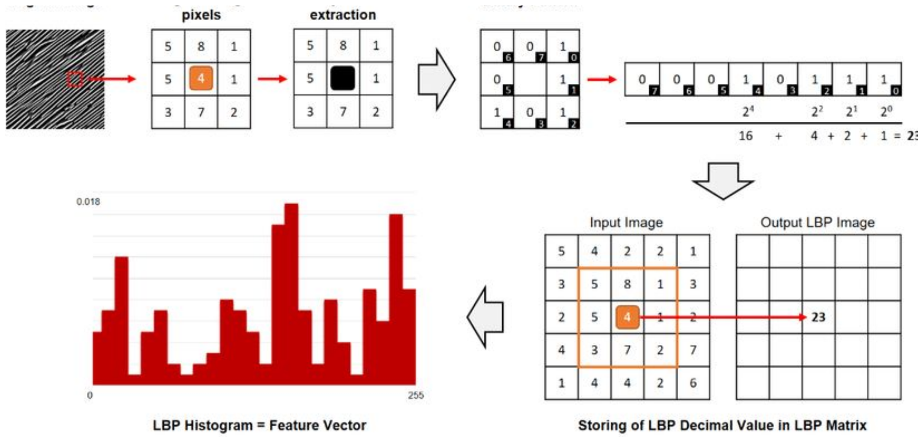


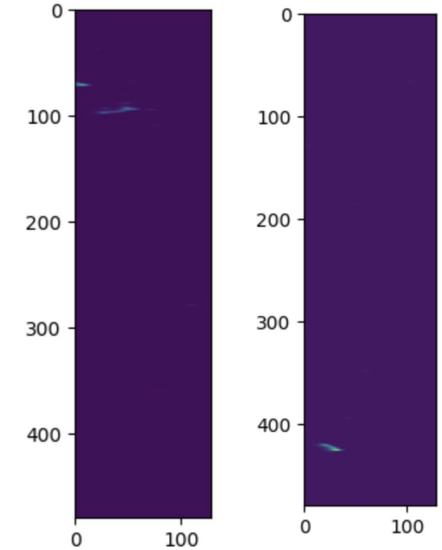
Distinguishing eES vs nuCC



Raw ADC value

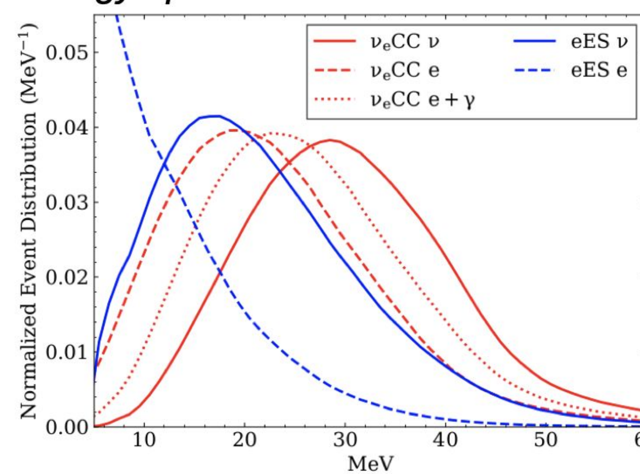
Binary Pattern Extraction





nuCC event vs. eES event

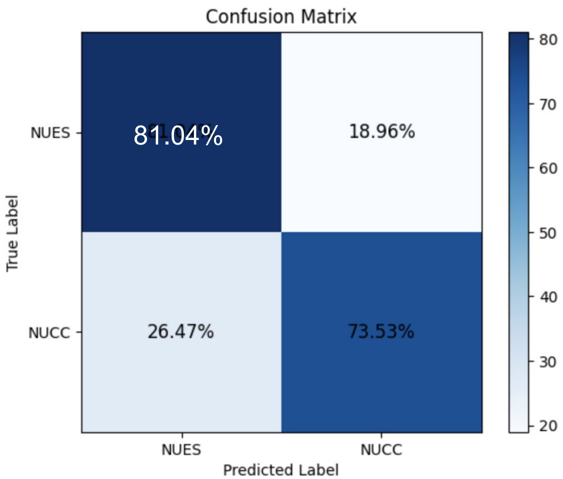
energy spectra: See DUNE-Doc-23154

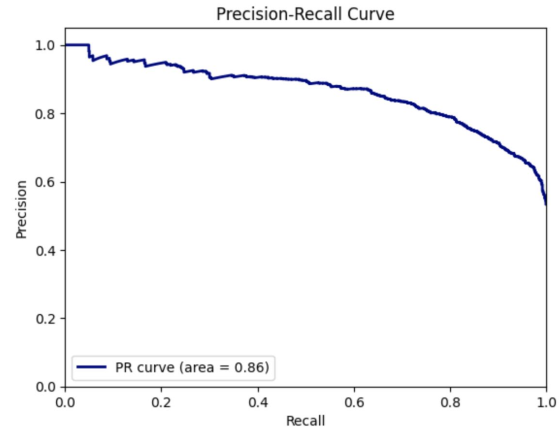


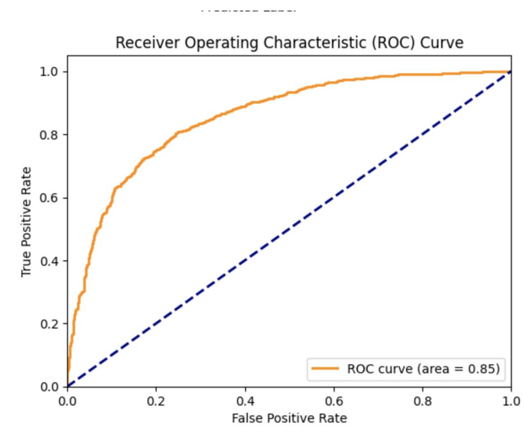
- Data is sparse and standard image based filtered using power spectral density don't work
- However weight distribution for sparse data makes CNN optimizations easier
- We still must allow model to take in local features but how?
- Solution: Local Binary Pattern fed into a dense architecture
- Algorithm optimizable for hardware implementation due to existence of binary codes

Fermilab

ES vs. CC Interactions in Frames



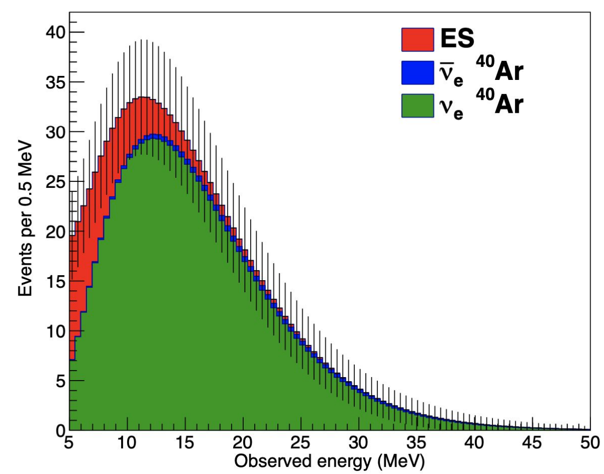




We can achieve ~81% ES->ES identification in a given frame which allows us to increase the spike in our likelihood function

Such a model uses a feature extraction method, and feeds into a dense fully connected network for binary classification.

This gives us a resolution of *~0.007* of the Sky Fraction



Hardware Implementation of 2D CNN

- Implemented on xcvu9p-flgc2104-2L-e FPGA
- Preliminary latency of *33ms w*hich is sufficient for our desired data reduction

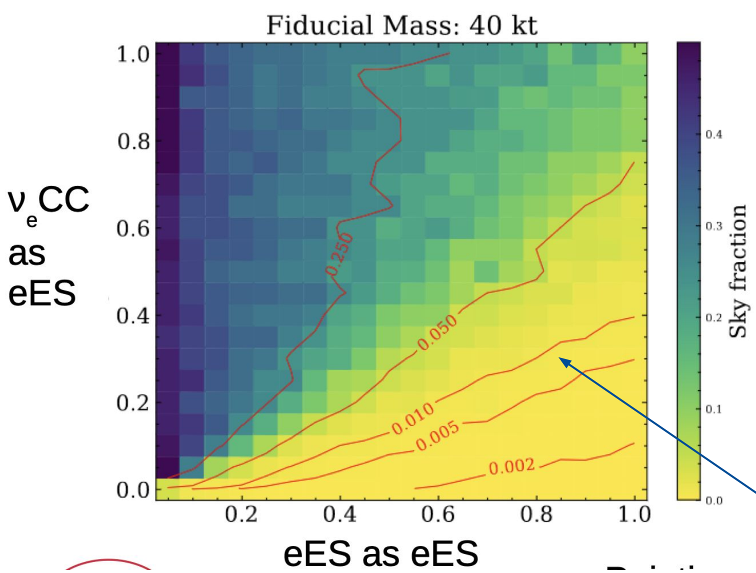
	BRAM(%)	DSP (%)	FF (%)	LUT (%)	URAM (%)
Default Buffer Size	30	8	8	12	0
FIFO Buffer Optimization	26	8	8	12	0
Using Minimal Pipelines	~0	8	9	13	0

Data Reduction Estimates

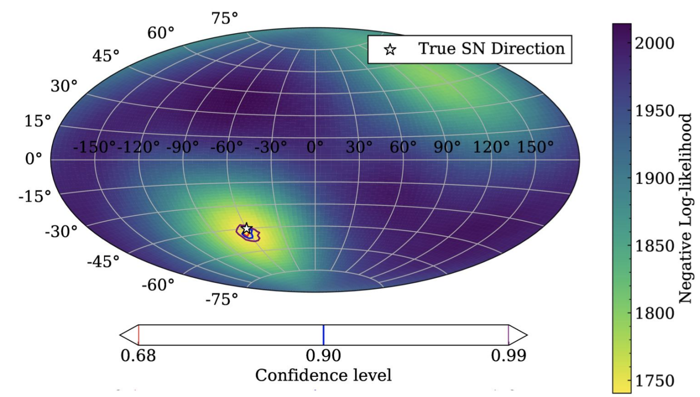
- Data rate out of one detector module
  - 14 bits x 2 MHz = 3.5 MB/s per channel
  - 150 APAs x 2560 channels/APA x 3.5 MB/s/channel ~ 1.3 TB/s
- Buffered supernova data in one detector module
  - DUNE will buffer 100s of data in SN buffer: 130 TB
  - Useful SN directionality information is in first 10s: 13 TB
- Estimated SN interactions per module
  - 3300 CC + 330 eES -> 3630/4 = 907 SN nu / module
- Assume able to identify SN nu and reject background
  - 14 bits/tick x 128 ticks x 2560 channels x 907 ~ 496 MB

What about the rest of the pipeline?

- If our background rejection scales accordingly, we may get relevant pointing information from the 496MB of data
- On each frame, we can apply the LBP algorithm to identify eES events and obtain about a *0.007* Sky fraction resolution



See DUNE-Doc-23154

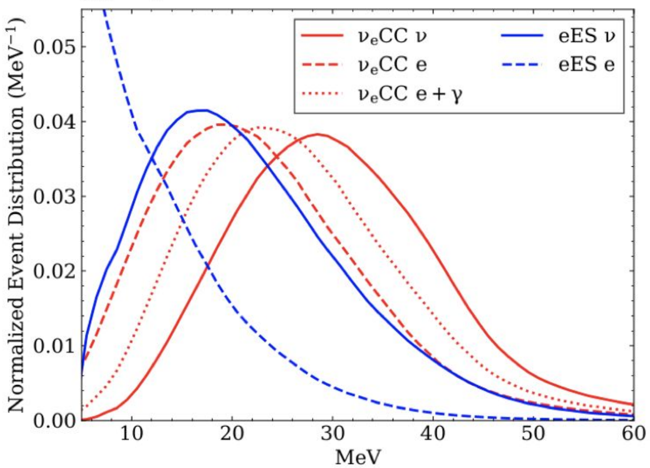


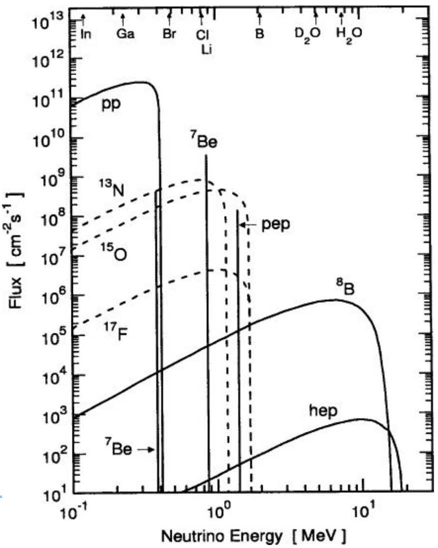
See DUNE-Doc-23154

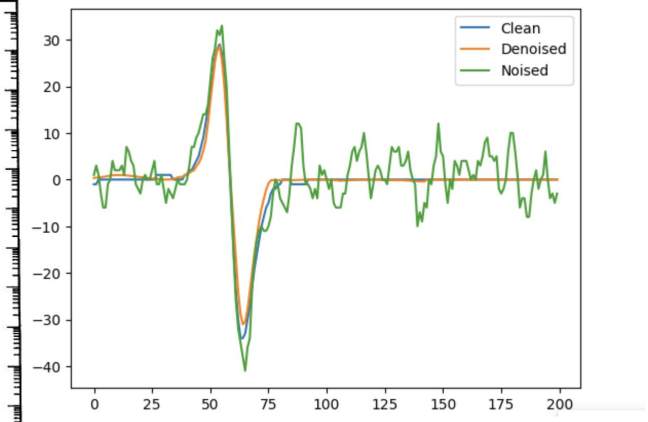
Where we can be with just ML data reduction

Digging for Low Energy Signals

energy spectra:







The more events we may identify upstream, the more pointed the likelihood function becomes.

This might not be ideal for pointing, but what about calibrating our 14-bit ADCs as front end electronics change.

If we can lower our threshold to search for signals in our buffered data we also have potential to identify solar neutrinos (middle)

Solar Neutrino energy spectrum

As discussed yesterday by Jovan's denoising AE talk, and by V.T.B Lian at A3D3 we can potentially dig out low energy signals using an auto-encoder

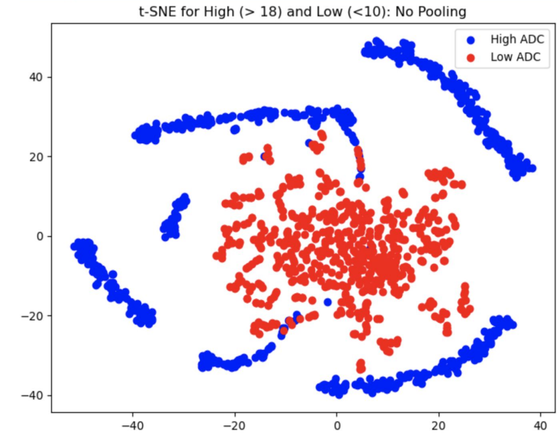
Comparison of Signal Efficiency in AI vs. TP algorithms

We can achieve ~15% improvement in signal efficiency per plane using the denoising auto-encoder, which gives us the opportunity to lower our current ADC threshold for a better signal search. This might allow us to capture signals from other neutrinos such as solar neutrinos

Tested on 100 Marley-generated ES and CC events

nu-ES	U-plane		V-plane		Z-plane	
	Sig eff	Bkg rej	Sig eff	Bkg rej	Sig eff	Bkg rej
TP (FIR)	--	--	--	--	0.59	0.998
TP (AbsRS)	0.63	0.998	0.68	0.998	0.62	0.998
1DCNN	0.74	0.999	0.75	0.998	0.70	0.998

nu-CC	U-plane		V-plane		Z-plane	
	Sig eff	Bkg rej	Sig eff	Bkg rej	Sig eff	Bkg rej
TP (FIR)	--	--	--	--	0.43	0.998
TP (AbsRS)	0.48	0.998	0.50	0.998	0.49	0.999
1DCNN	0.59	0.999	0.60	0.998	0.61	0.998



Autoencoder latent space clustering for events above and below TP threshold

Conclusions and Further Investigations

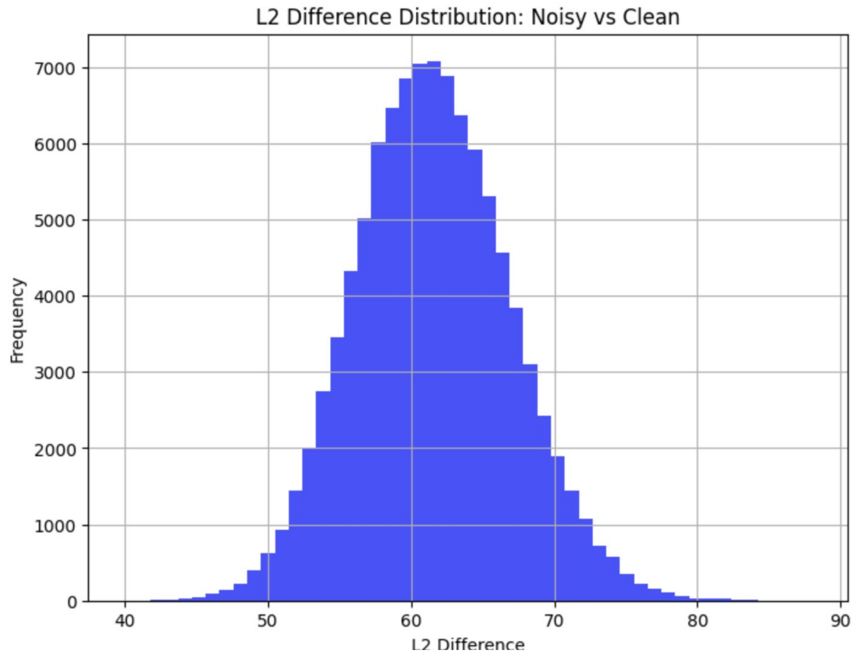
- **General performance is promising**
  - We see that we are able to to use a 2D CNN and DNN to extract features that allow us to obtain ~81% correct eES identification in extremely low energy signals, achieving a pointing resolution of ~0.007 of the sky fraction
  - Rejected pure background and electronics noise samples at ~99.8% gives us significant background rejection that may be buffered on disk
- **Improving the 1D Autoencoder**
  - We are investigating alternative training methods such as (GANs) where the output of the autoencoder is iteratively updated for improved performance
  - Same hardware implementation, but unclear hardware implementation will account for same signal efficiency gains of GAN
  - Better LE resolution with the changing electronics of DUNE gives the opportunity to push identification at the end of the low-energy spectrum.
- **Finding the optimal MLE**
  - Can we use RL to optimize our parameters in our MLE estimation?
  - Can we use existing tools from traditional algorithms to regress better on path for pointing GNNs (from [Cerati et. al, 2022](#))

```
fChannel(channel)
fStartTick(start_tick)
fEndTick(end_tick)
fPeakTime(peak_time)
fSigmaPeakTime(sigma_peak_time)
fRMS(rms)
fPeakAmplitude(peak_amplitude)
fSigmaPeakAmplitude(sigma_peak_amplitude)
fSummedADC(summedADC)
fIntegral(hit_integral)
fSigmaIntegral(hit_sigma_integral)
fMultiplicity(multiplicity)
fLocalIndex(local_index)
fGoodnessOfFit(goodness_of_fit)
fNDF(dof)
fView(view)
fSignalType(signal_type)
fWireID(wireID)
```

Potential to regress on information from standard reconstruction algorithm downstream

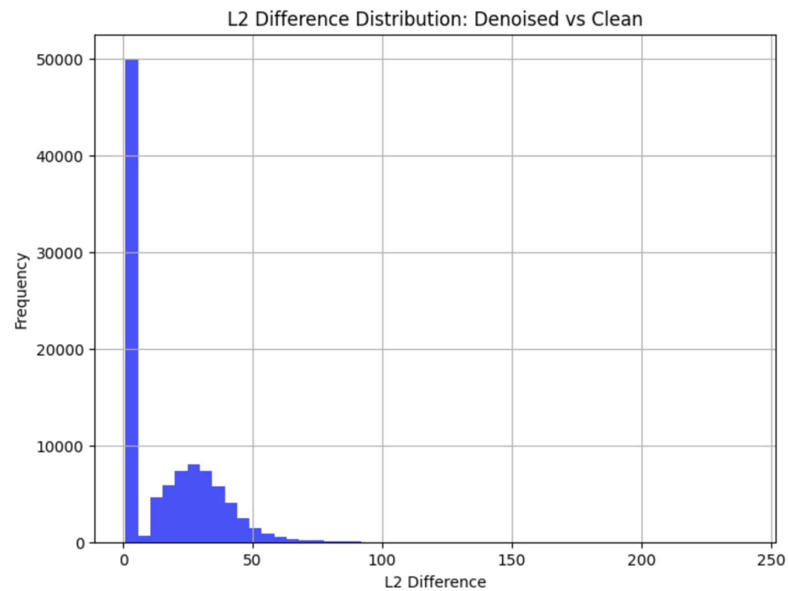
Backup slides

Brief Comments on Alternate Training Methods



L2 Loss on Noisy Data

Results from training using a GAN to improve weights of the auto-encoder.



Adversarial loss for GAN training attempt to improve AE

$$\sum_{n=1}^{N}(\log D(G(I^{input})) + \log(1 - D(I^{original})))$$