nEXO's Machine Learning Approaches

Jason Brodsky, LLNL
on behalf of the nEXO collaboration

August 24, 2023

LLNL-PRES-853345

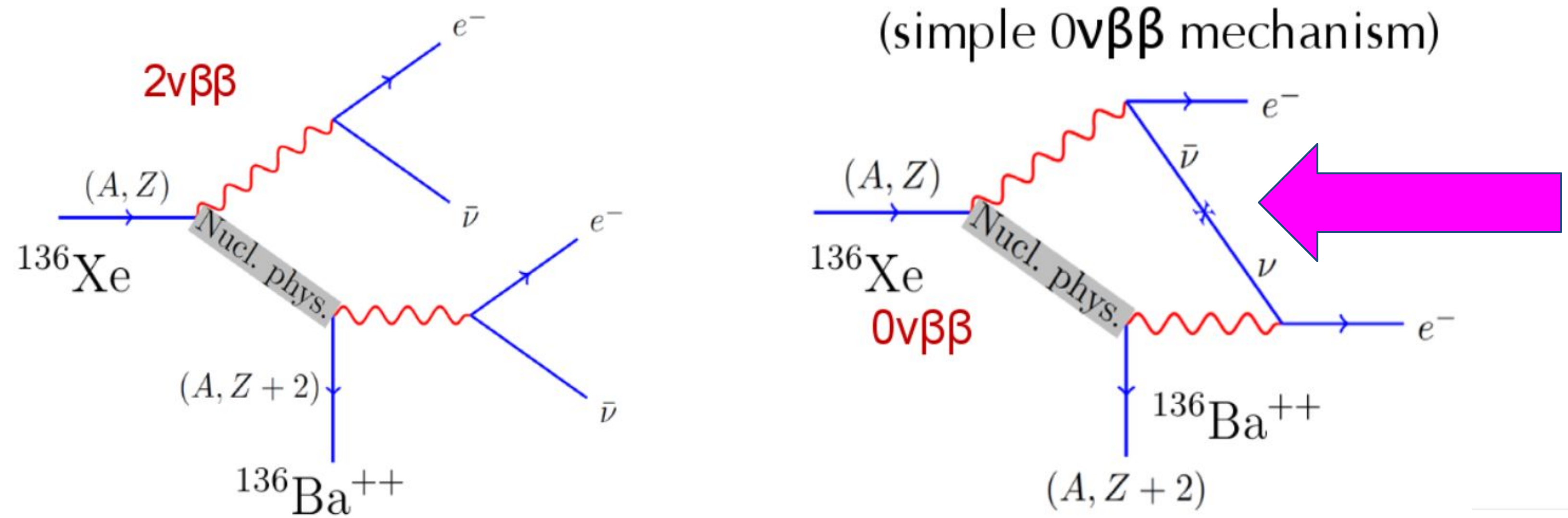
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Neutrinoless Double-Beta Decay

Hypothetical decay that violates lepton number: antineutrinos annihilate each other

Possible if neutrinos have a Majorana mass–and Majorana mass is a very appealing way to fit massive neutrinos into the SM.

Major cosmological implications for matter-antimatter asymmetry

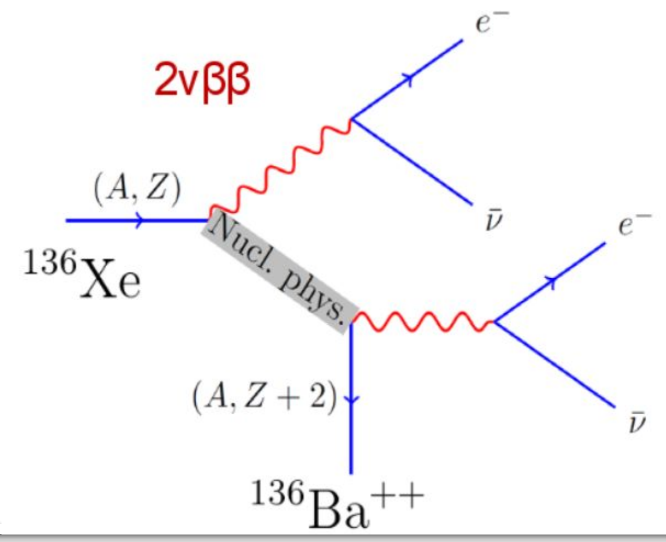


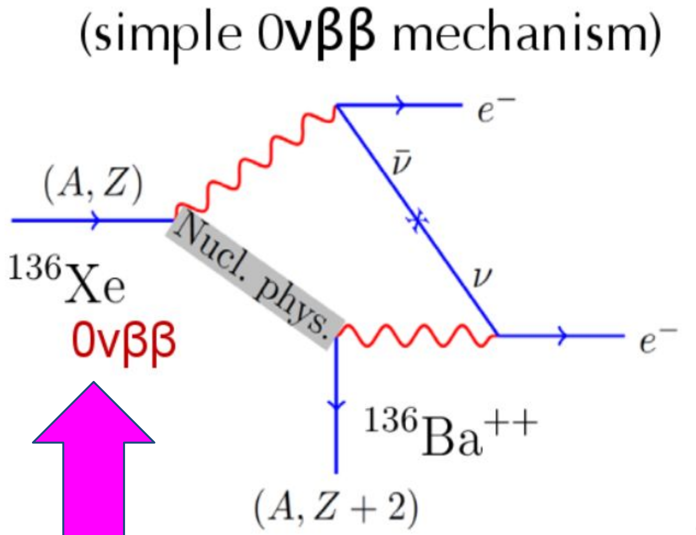
Neutrinoless Double-Beta Decay in Xenon-136

Ordinary double-beta decay first observed in EXO-200

136Xe has many useful properties for this search, including ease of isotopic enrichment and high Q-value.

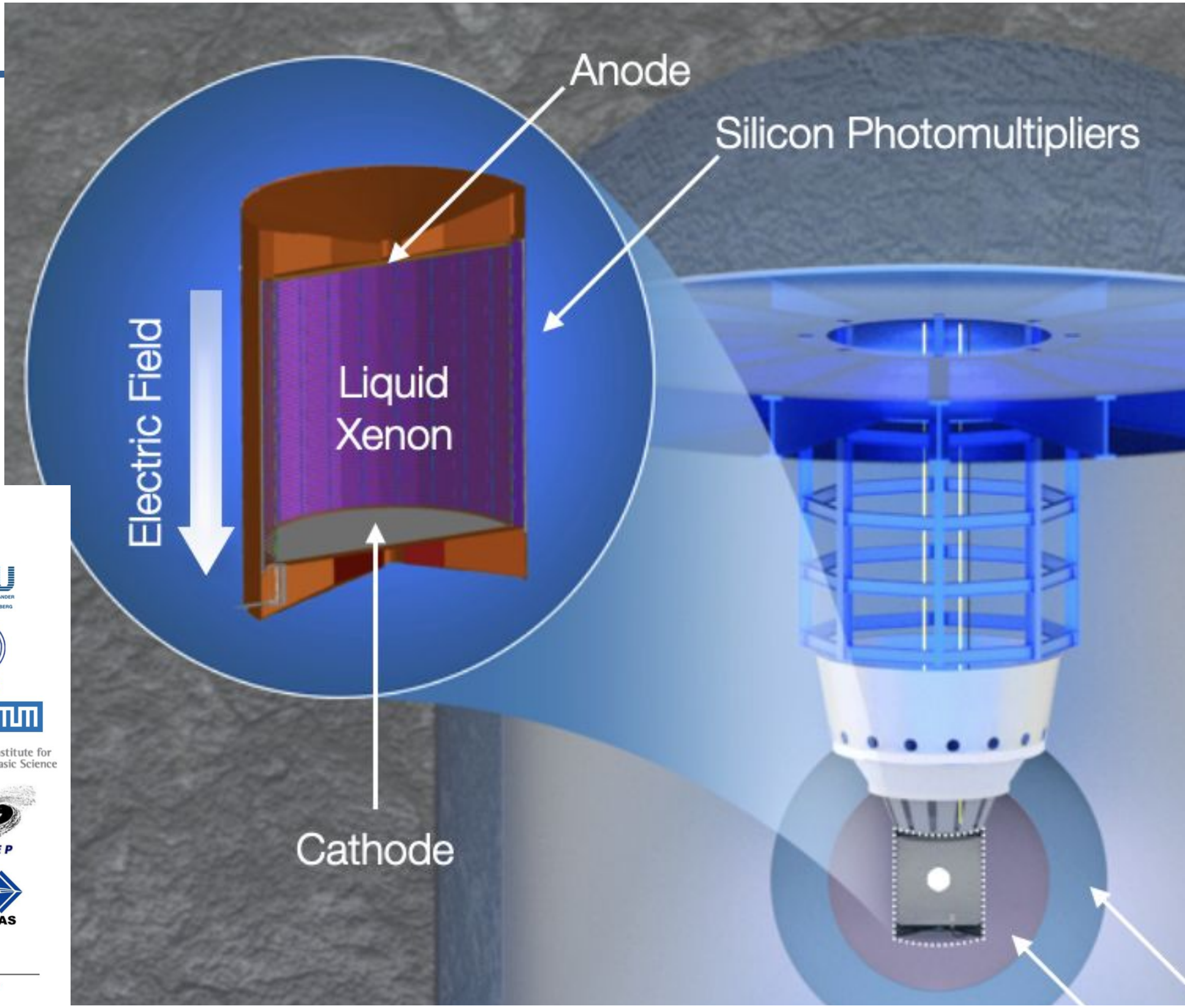
Xenon is the source of the decay and also the detection medium





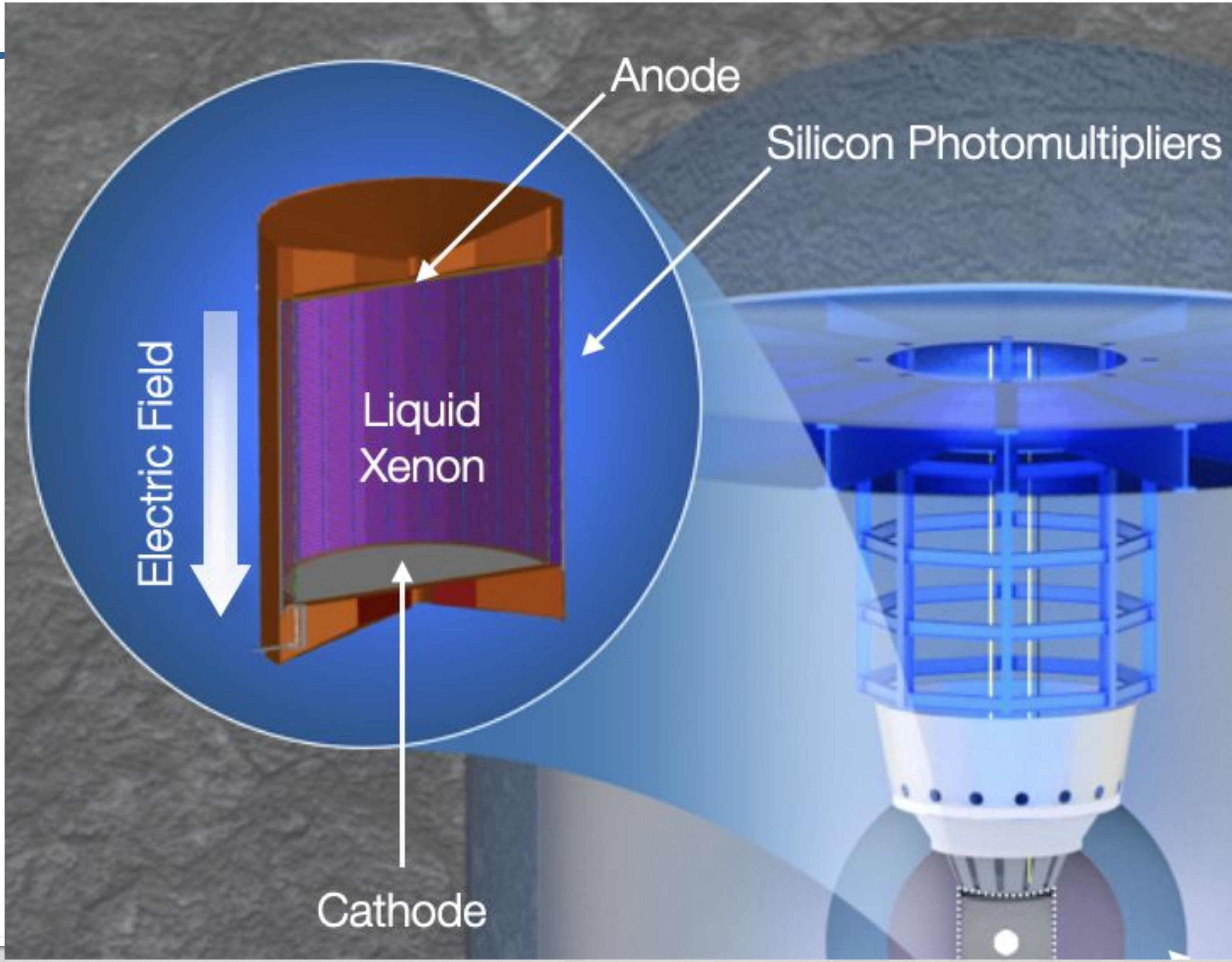
The nEXO Detector

- Xenon-136-enriched time projection chamber
- 5 tonne xenon, pushing sensitivity out to $>10^{28}$ year half life
- Conceptual design stage of project

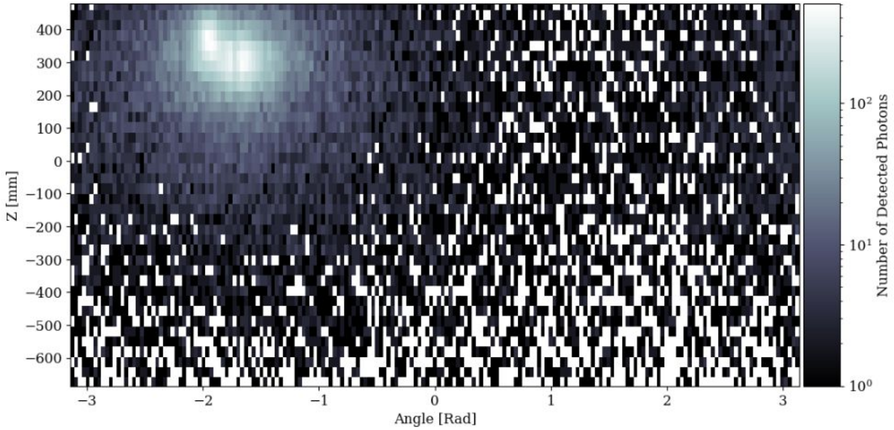


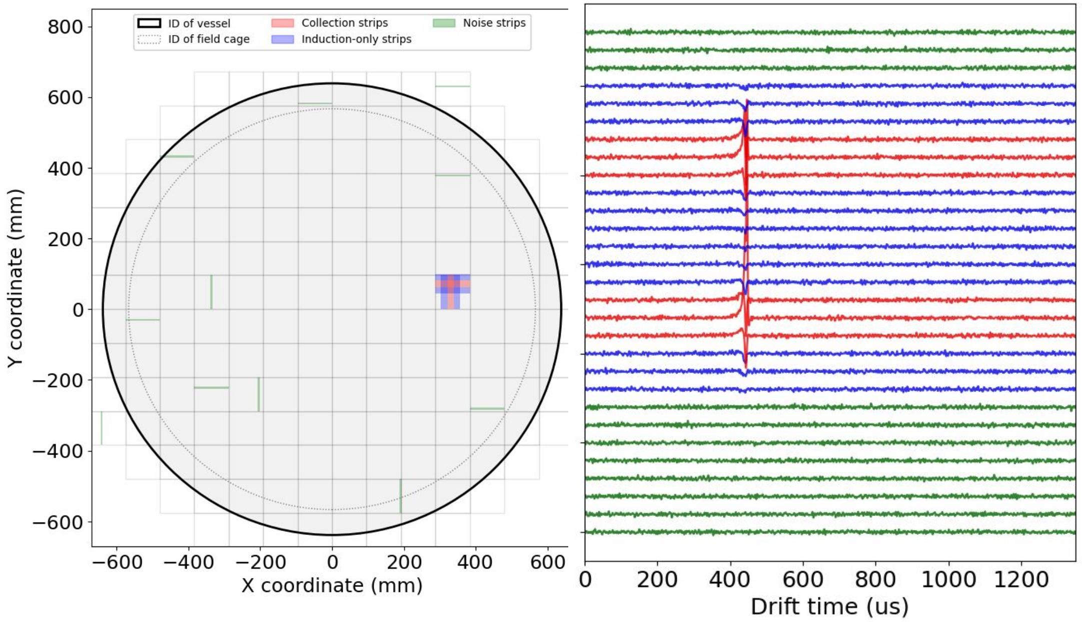
The nEXO Detector

- Powerful multi-dimensional signal/background discrimination
- SiPM light detection on the “barrel”
- Single phase TPC, collecting charge on electrode strips



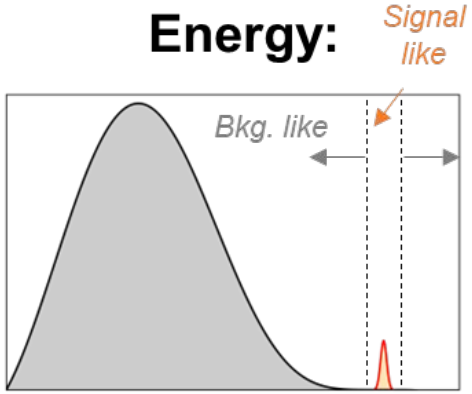
Signals in nEXO

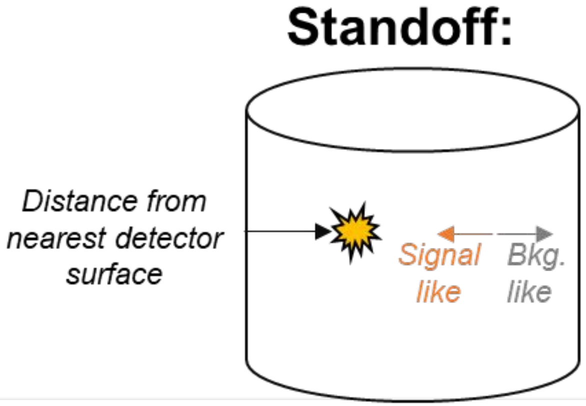


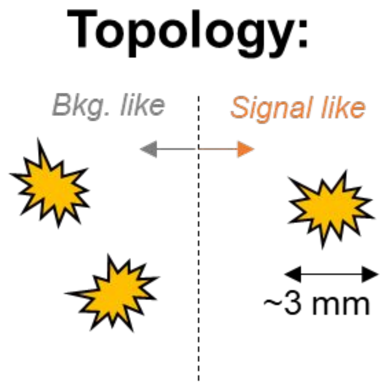


Scintillation + Drift Charge

Measures

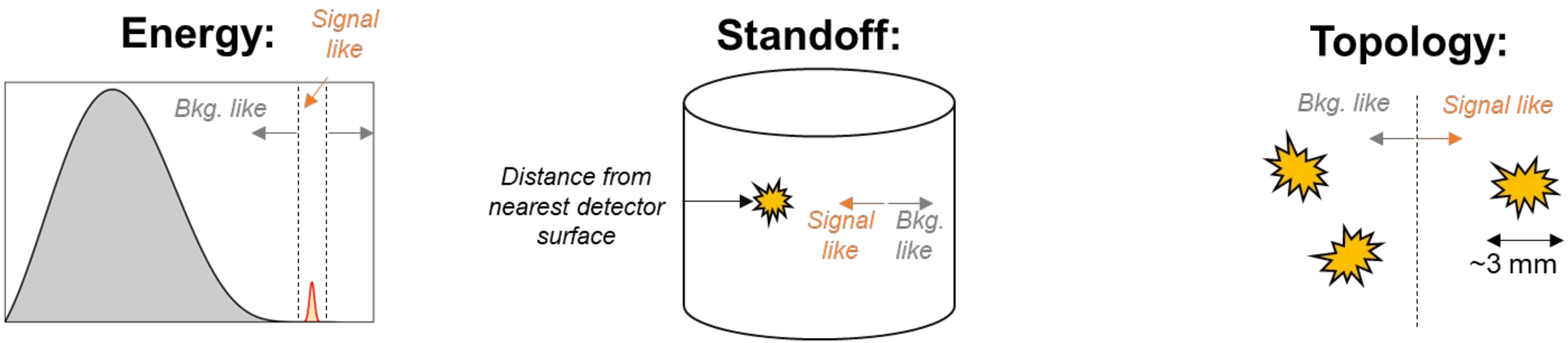




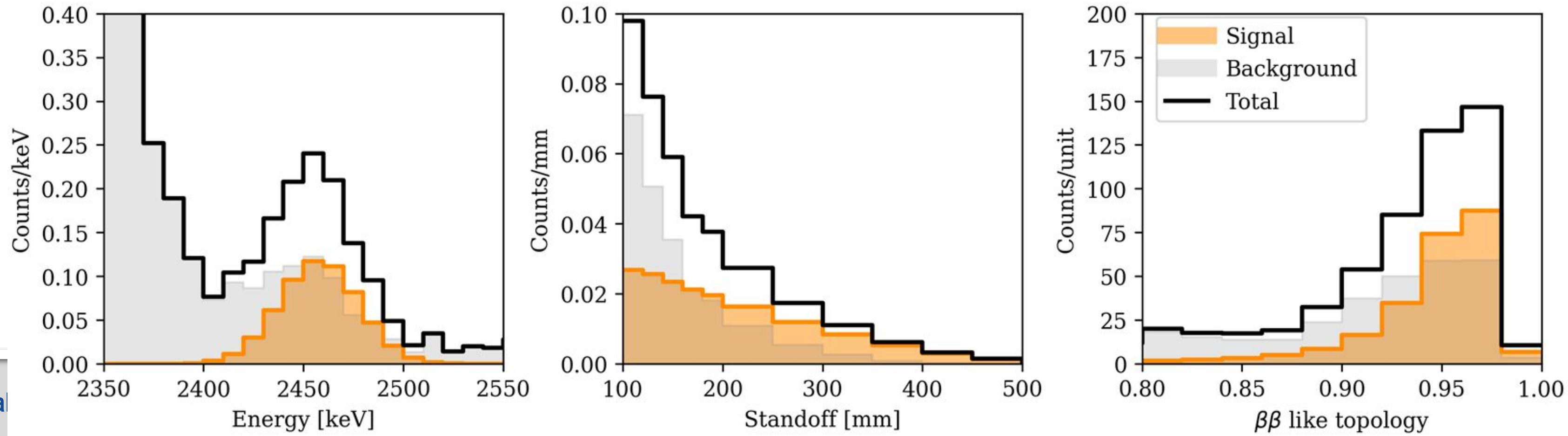


Signals and Backgrounds in nEXO

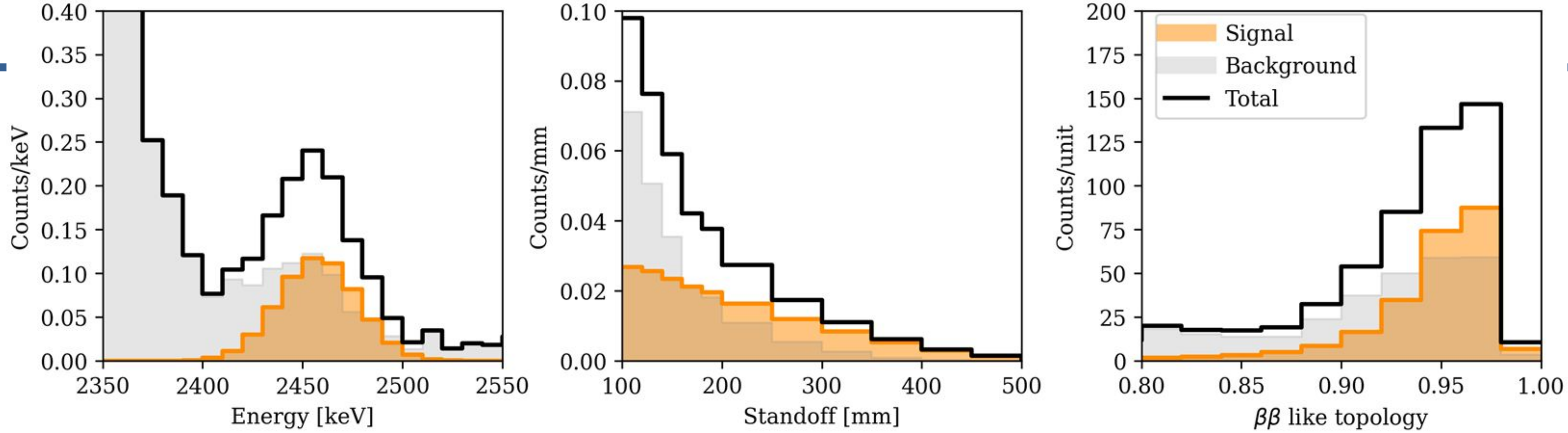
Signals are monoenergetic, single-site events uniform in the detector



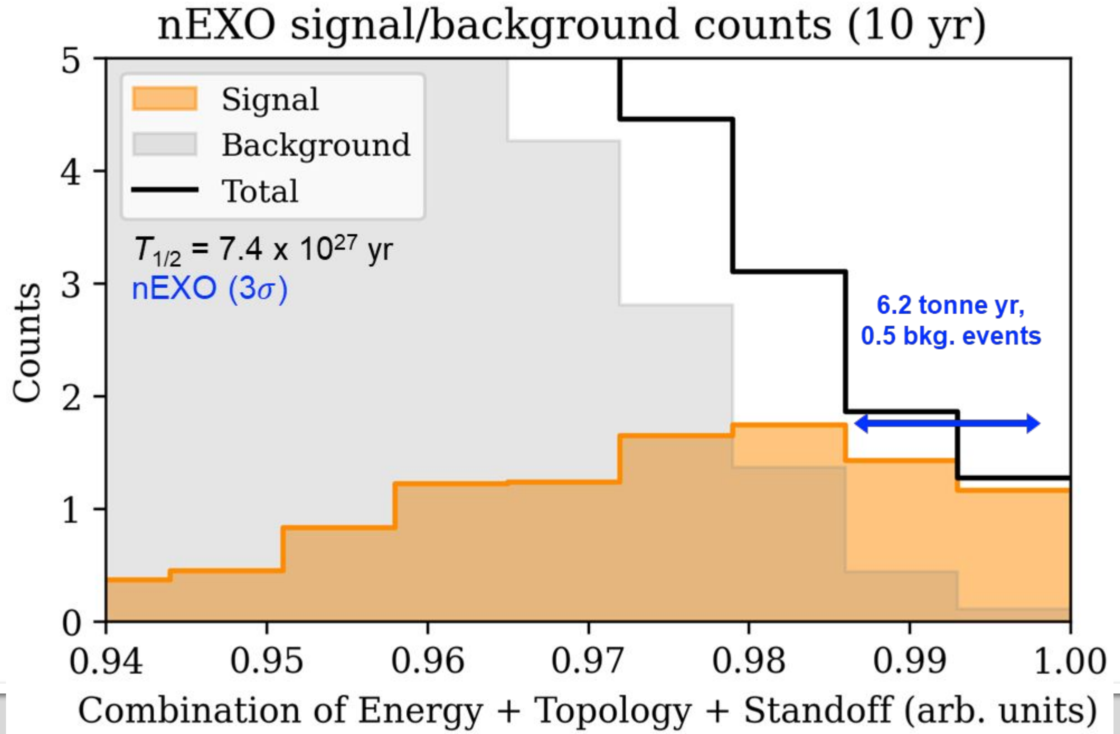
1D projections



Multi-axis discrimination is powerful



Combine energy, topology, and standoff (preserving correlations)



Background-free experiment *plus*

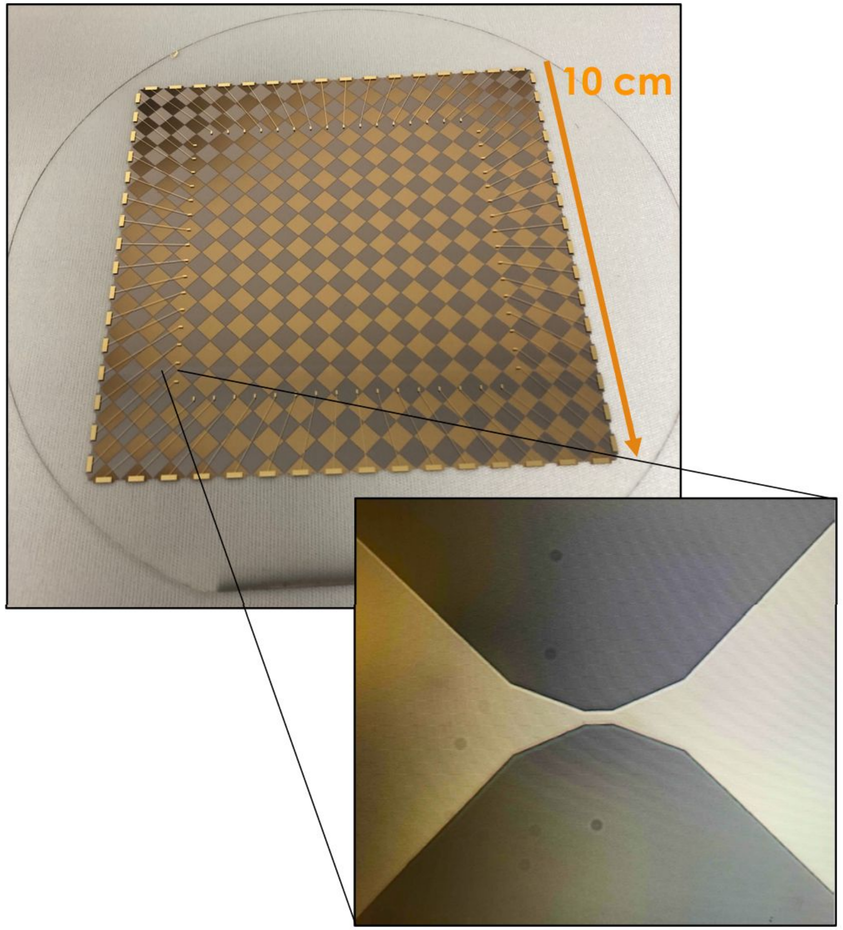
Charge Sensors

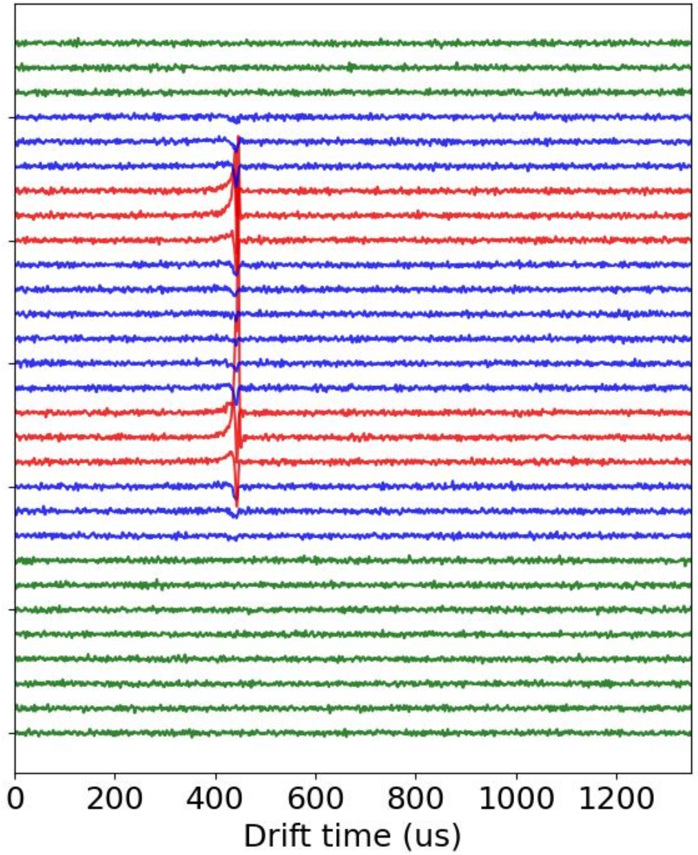
Interwoven pattern of strips on multiple 10cm tiles.

In some ways similar to crossed wire planes, in many ways different.

Main focus of ML analysis, due to lots of suitable features:

- High-dimensional signals
- Signal correlations
- Noise that poses a meaningful challenge to conventional analysis





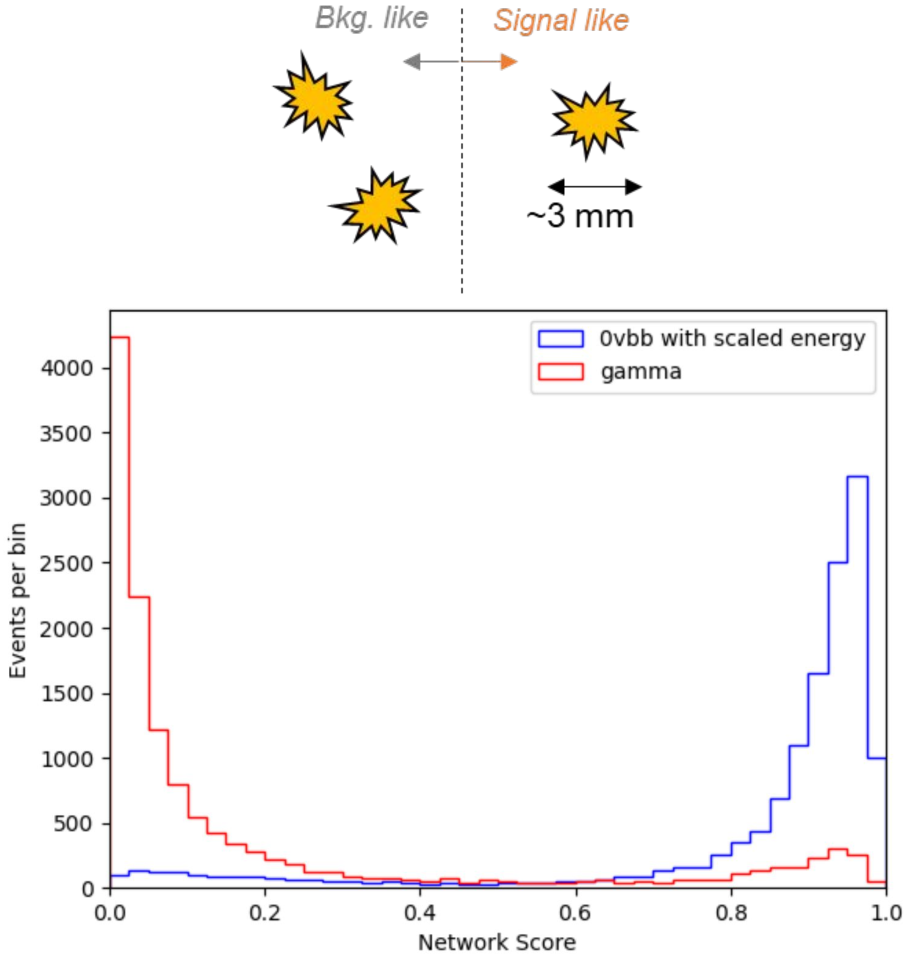
Approach 1: Topology Discriminator

Neural network classifier

Looking at charge signals to spot gamma-like multiple sites

Trained on simulated gammas and double beta decays with identical energy & standoff distributions

Topology:

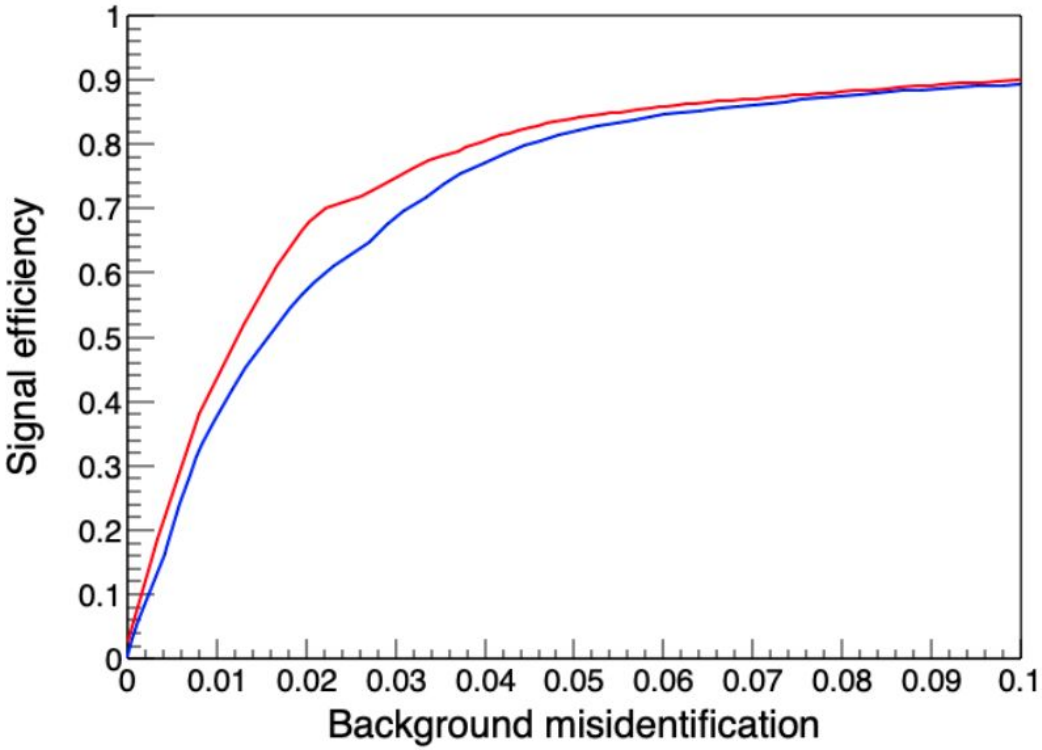


The importance of the ML discriminator

- Performance: substantially improved ROC curve compared to conventional algorithm
- Statistical power: fits with nEXO's strategy of multiparameter fits of signal vs background.
- Adaptability: nEXO is in conceptual design phase, where engineering alternatives must be evaluated, and ML adapts quickly to test alternative designs.

Discriminator Performance

Neural network on signals performs better than feature-based discriminators such as a boosted decision tree

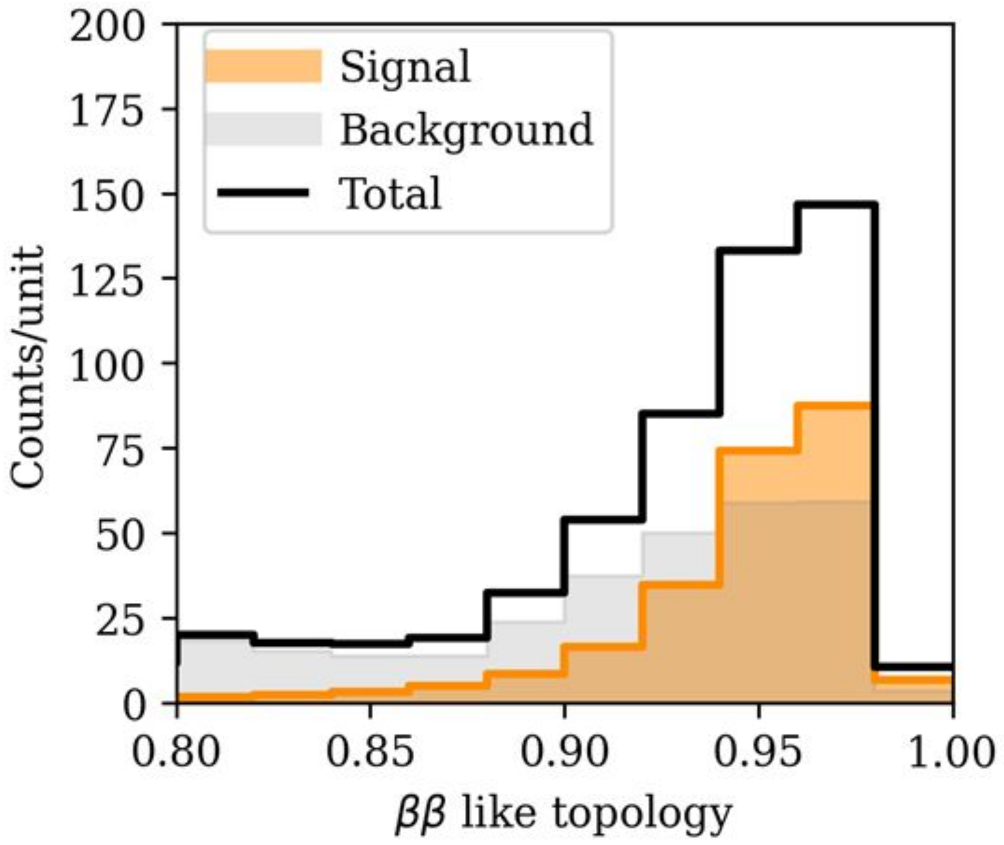


Older comparison between neural network (red) and boosted decision tree (blue)

Statistical power

Not all neutrinoless double beta decays look the same, and some gamma backgrounds are almost (but not entirely) perfect mimics.

We can harness finely-binned statistically fitting of the score for a 20% improvement in scientific reach compared to a binary discriminator.



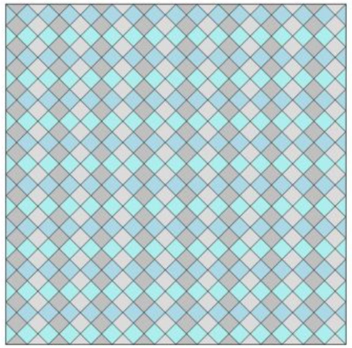
Adaptability

During the current conceptual design phase, nEXO engineering & construction project requires quantitative evaluation of alternative designs against baseline

Earlier discriminators that required more manual design performed (unfairly) better in the baseline case they were designed for

ML adapts immediately to the training data.

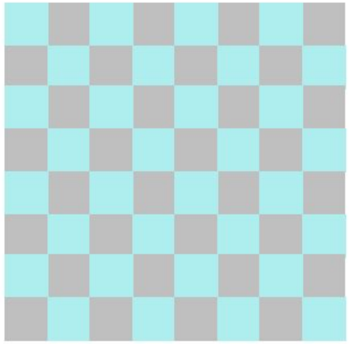
Strips



- Interwoven braids - 96 mm length & 6 mm pitch
- ~288 mm^2 area per channel
- 32 channel per tile

VS

Pads

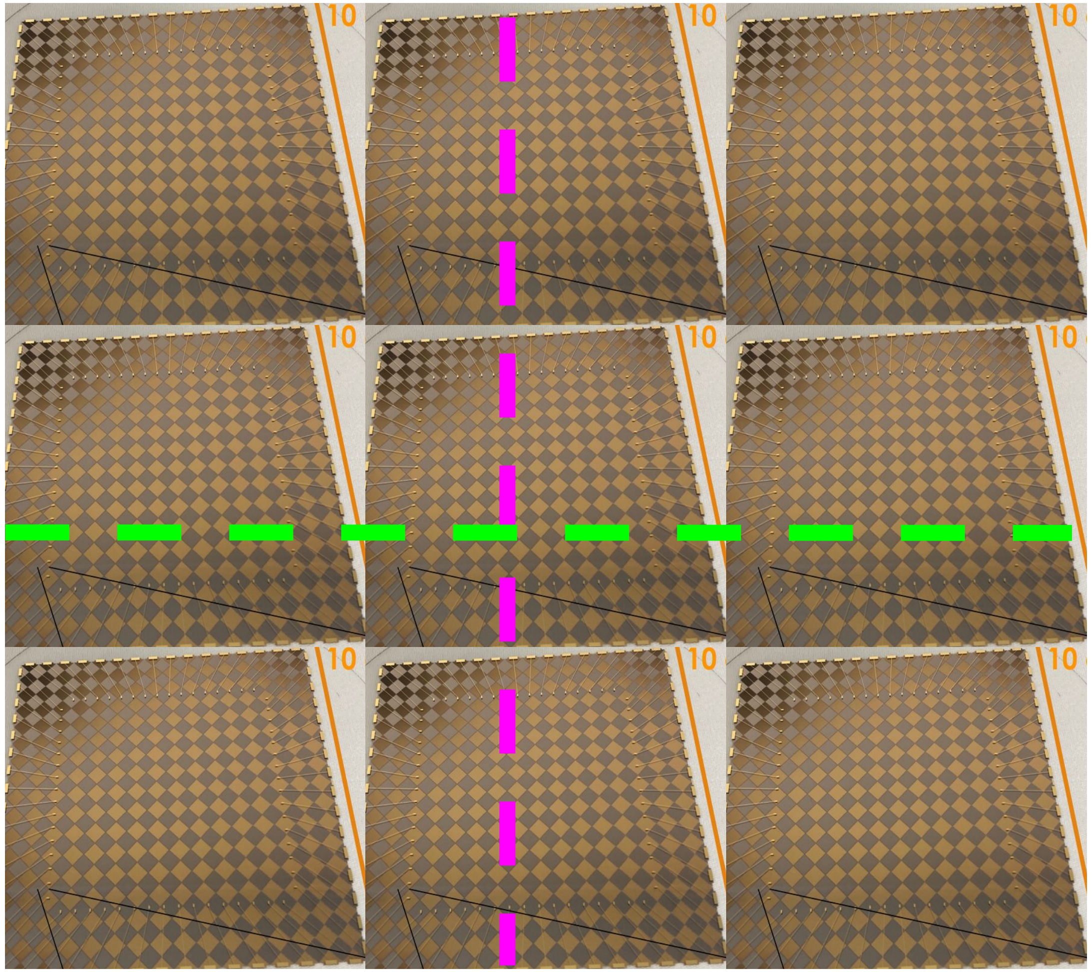


- Square design - length & width of 12 mm
- 144 mm^2 area per channel
- 64 channel per tile

Dimensionality in the Discriminator

nEXO has ~4000 charge sensors and records about 1500 time points, starting an entire event at 6e6 dimensions.

Baseline dimensionality reduction sums “in line” strips, comparable to having wire planes spanning full diameter. 4000 -> 400 channels.

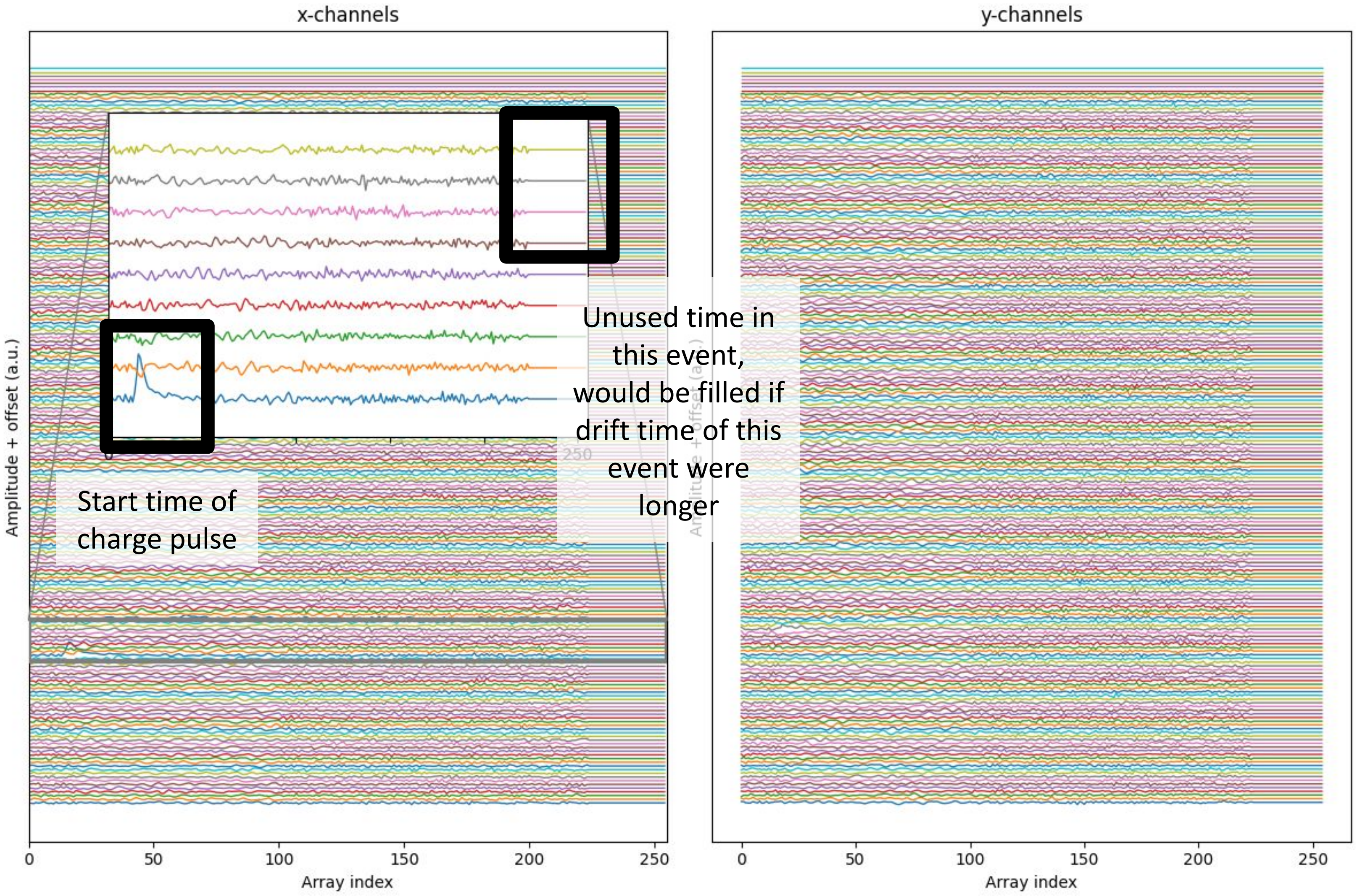


Dimensionality in the Discriminator

Reduce time points from 1500 -> 255, by removing majority of points far from pulse time.

Reverse time order so pulse time appears in consistent position.

400x255 = 1e5, instead of 6e6

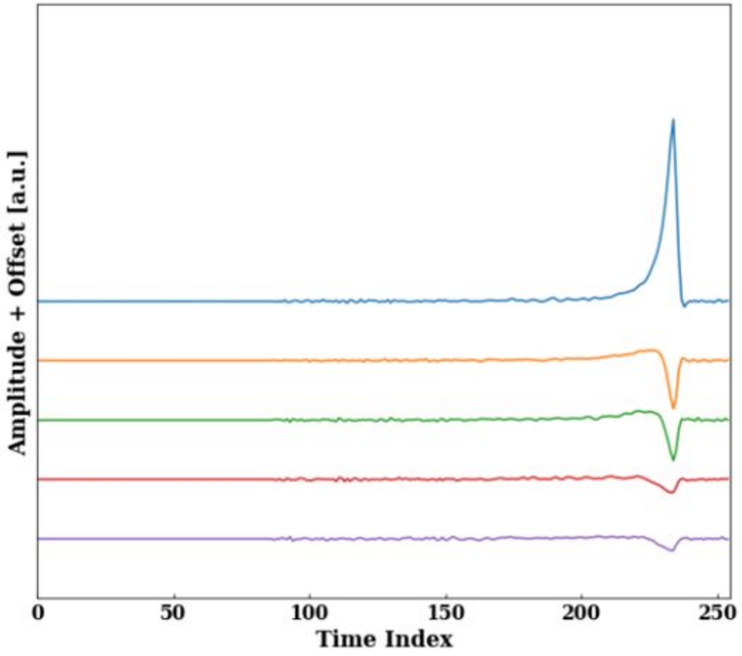


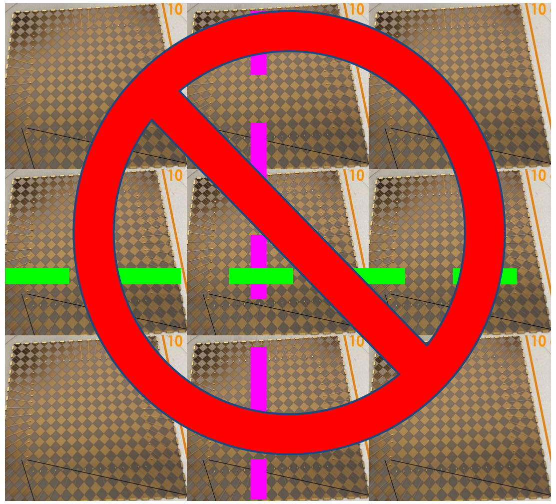
Signals can be sparse

Zero suppression can reduce signals down to a few channels and a few time points of interest

Perhaps not ideal in all situations, but input dimension reduction is necessary for some circumstances

E.g., in-line strip summing does not apply to pads, so input dimension jumps from 400x255 -> 8000 x 255





SparseConvNet

https://github.com/facebookresearch/SparseConvNet

Scales with number of **active** elements of the input, not total number of inputs

Successfully used for pads evaluation

Active	Type	C	SC	SSC
Yes	FLOPs	$3^d mn$	amn	$\boxed{a}mn$
	Memory	n	n	n
No, $a > 0$	FLOPs	$3^d mn$	amn	0
	Memory	n	n	0
No, $a = 0$	FLOPs	$3^d mn$	0	0
	Memory	n	0	0

Table 2: Computational and memory requirements of three convolutional operations: regular convolution (C), sparse convolution (SC), and submanifold sparse convolution (SSC). We consider convolutions with size $f = 3$ and padding $s = 1$ at a single location in d dimensions. Notation: a is the number of active inputs to the spatial location, m the number of input feature planes, and n the number of output feature planes.

Approach 2: Denoising ML

Take noisy signals and make best estimate of original, noiseless signal

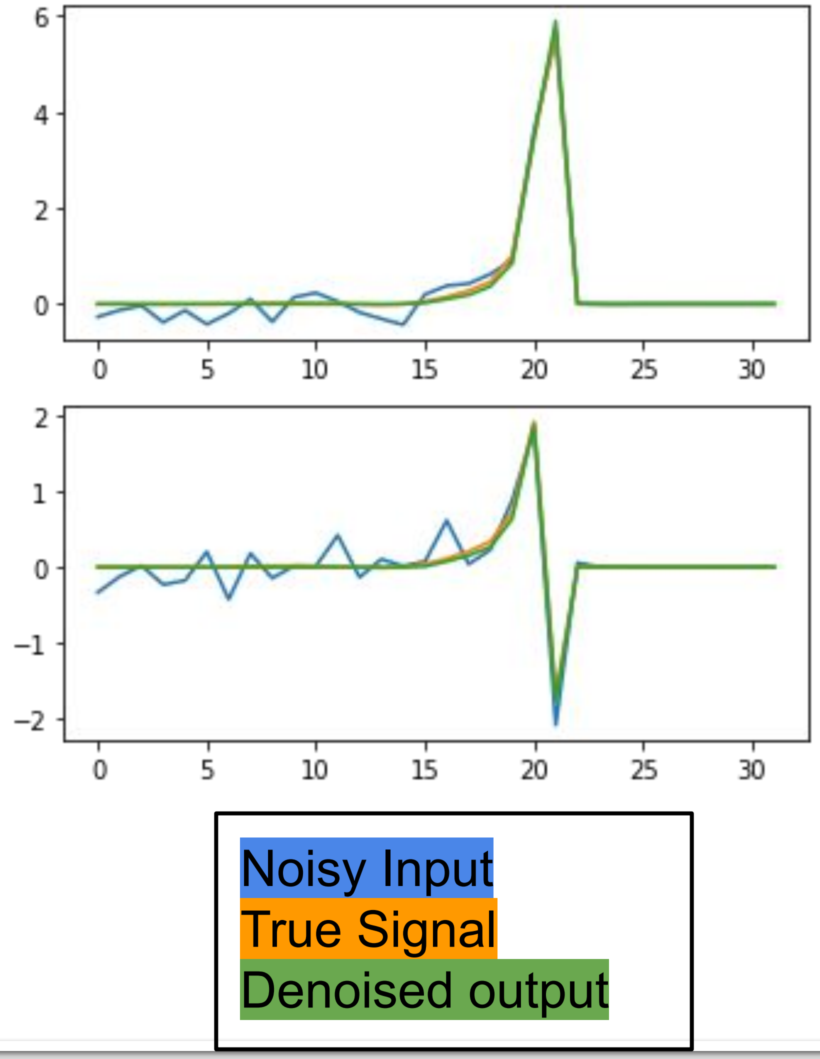
Not magic: noise is always lossy

But:

Improves energy resolution

Convenient

Credible



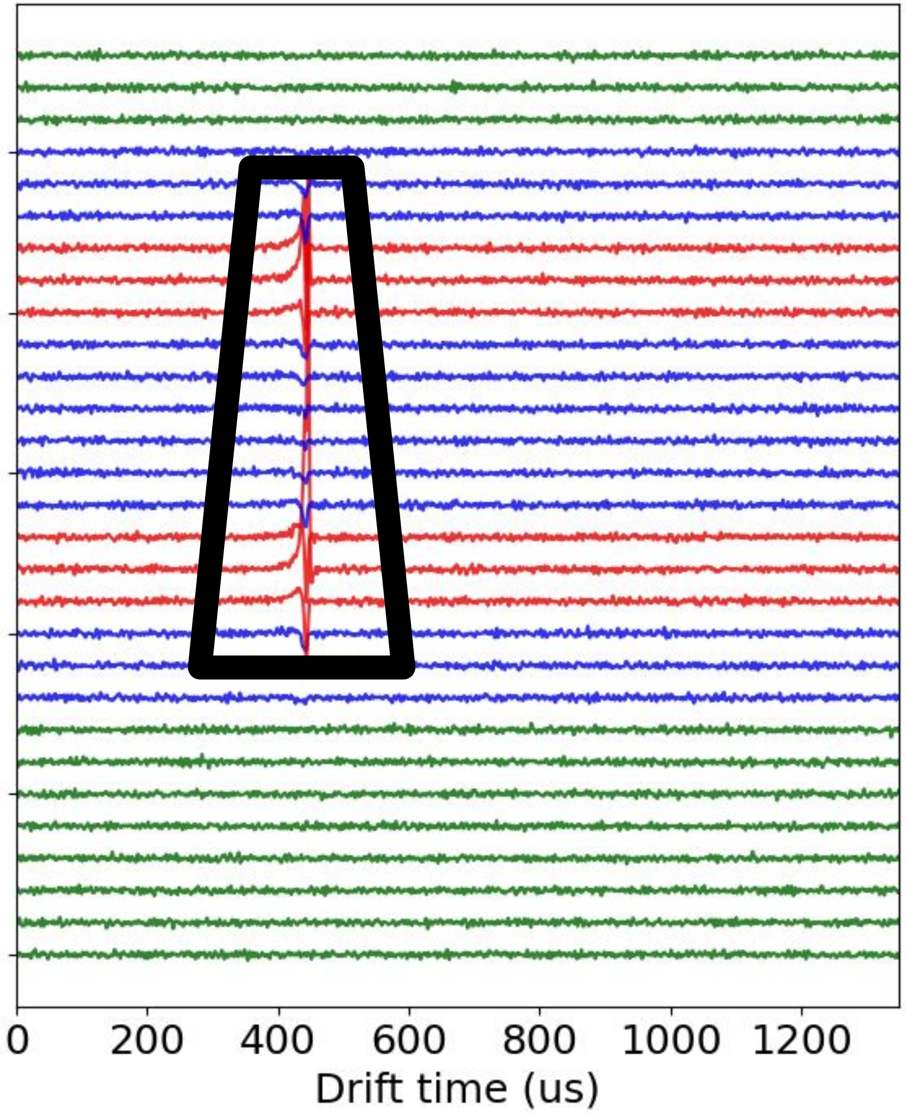
Charge channel noise

Charge channels have a meaningful amount of electronic noise.

Energy $\propto$ drift charge $\propto \sum_{channels} \sum_{time}$ current

But actually summing current over *all* time and channels introduces lots of noise

Basic approach: trapezoidal(-ish) filter, including only times and channels with good signal:noise.



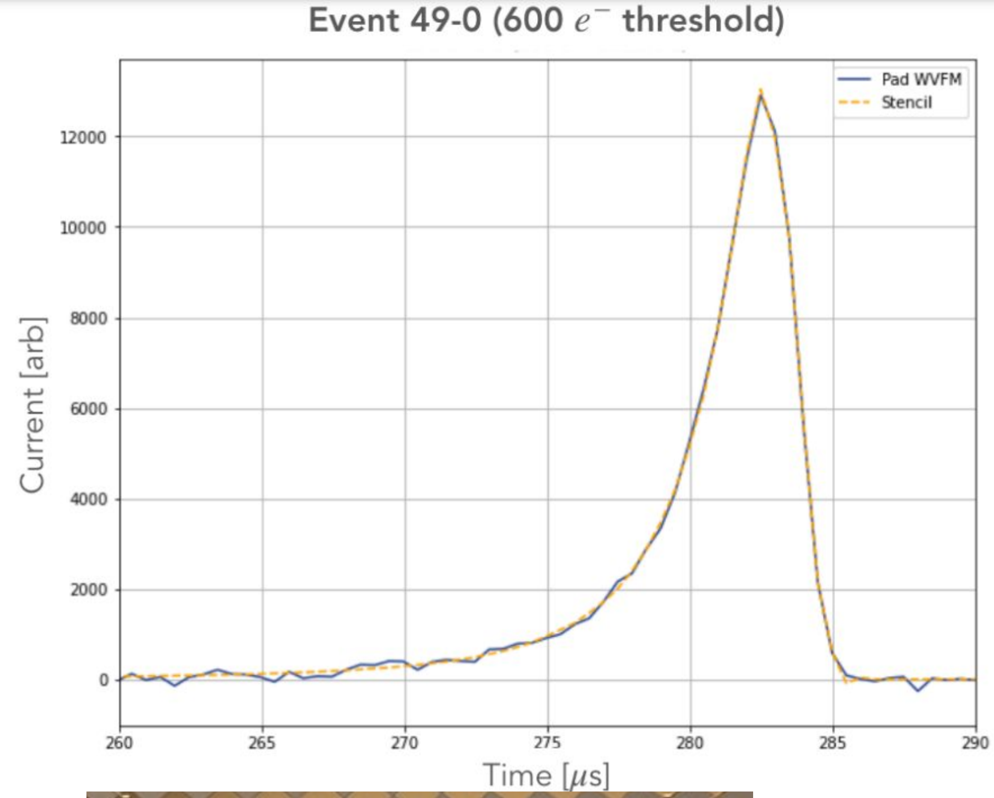
Energy estimation using a stencil

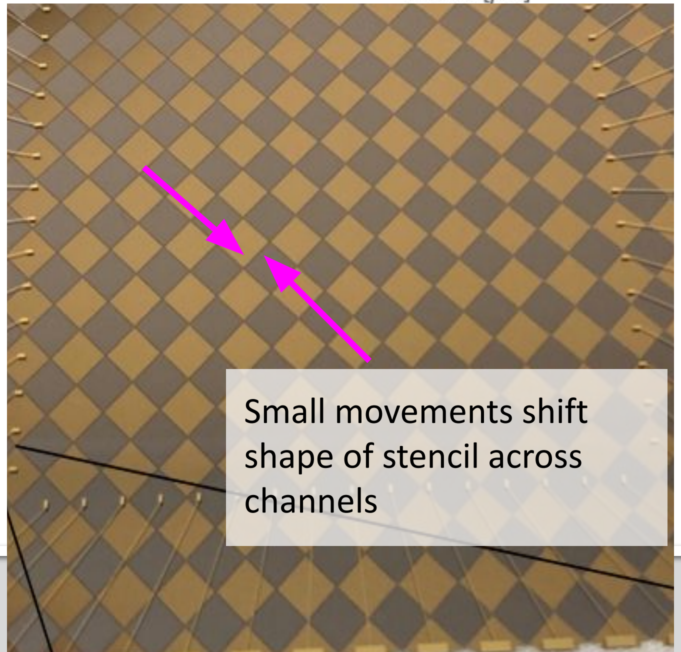
Estimating the amplitude of a signal with a known shape & known noise power spectrum has a mathematically optimal, analytic solution

The shape of the stencil changes from event to event, as charge distributes differently between adjacent channels

Reconstructing correct stencil is tricky using conventional techniques

Denoising translates into an easy application of this mathematically optimal estimator





Lever 1: True signals are low-dimensional, noise is high

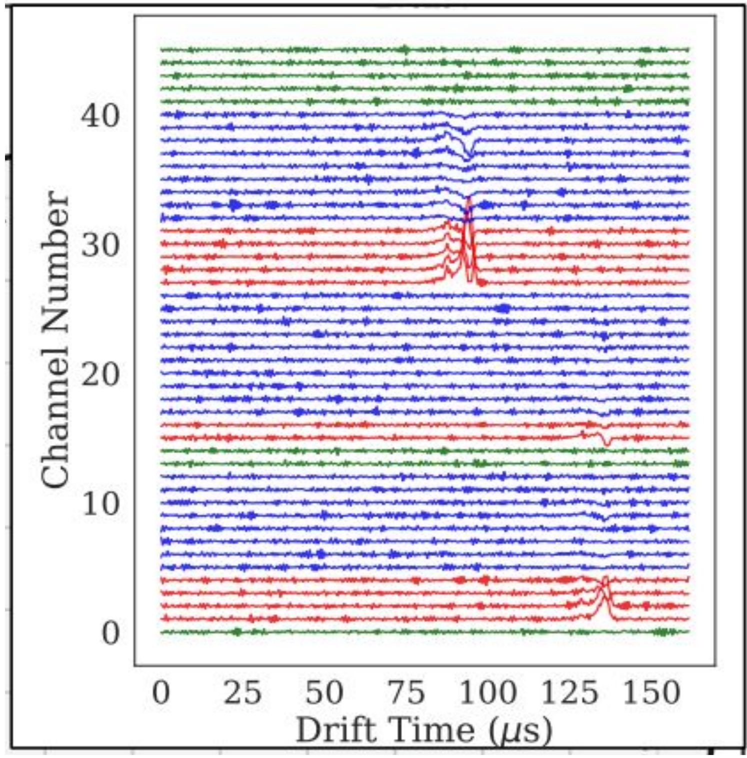
How do we simulate charge signals? Start with just a few parameters:
x, y, z, amount of charge

f(x,y,z,c) = [..................../\.......]

You can describe a true signal with the few parameters needed to generate it. (Perhaps a few extra for the nuances…)

Meanwhile, the noise generation function may be simple but the noise in a given event requires millions of parameters to describe.

"Every signal is alike (up to a few parameters), but every noise is noisy in its own way" – Anna KarenEXO



Lever 2: Easy access to noise-only samples

Lots of literature on ML denoising, but needs adaption for the physics case.

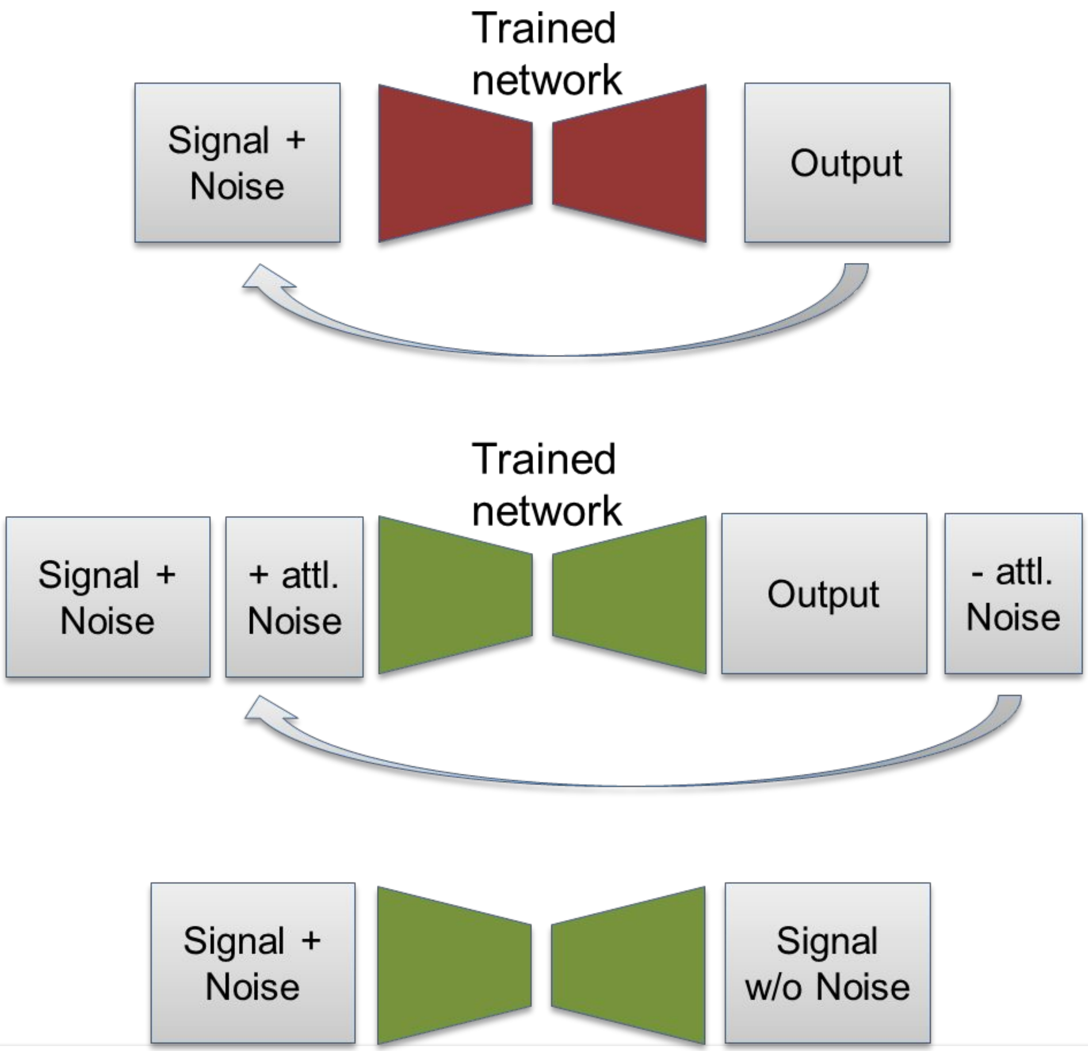
Recorrupted2Recorrupted (illustrated here)

Noisier2Noise

Autoencoders learn (or approximate) identity function, reproducing signal + noise

Theorems show modified autoencoder learns denoising instead of identity

Trainable on experimental data with no ground truth signals



Proof of Concept

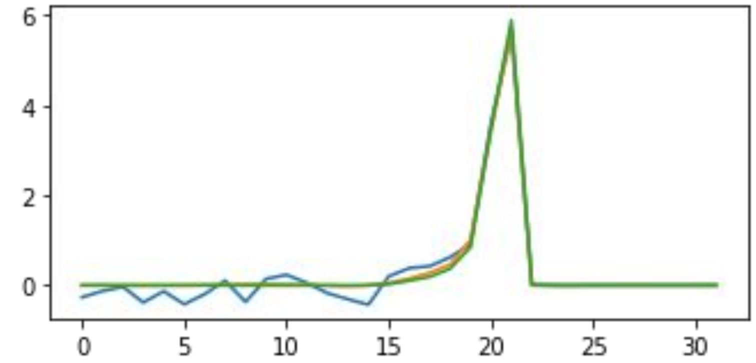
nEXO baseline charge signals: 4000 channels x 1500 time samples (full drift)

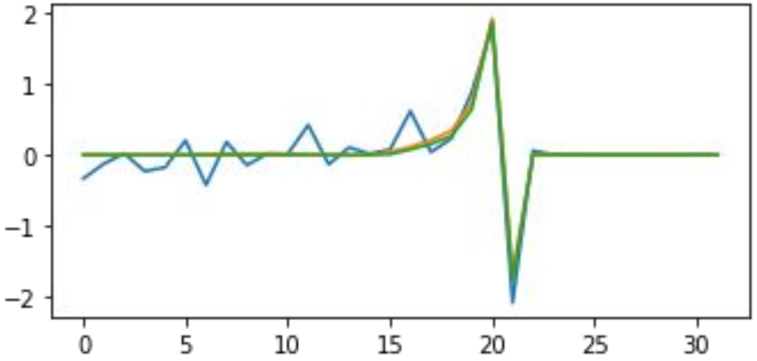
Make the proof of concept feasible by shrinking (crudely) to many few dimensions.

Heavily compress original data into fewer channels and time points:

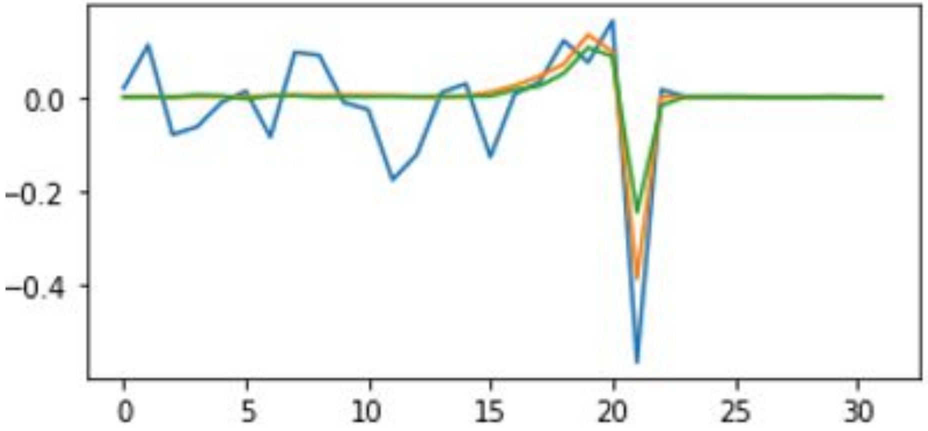
64 channels (32 X & 32 Y), 32 time points.
Sum “in line” strips, and sum adjacent channels.
(Of course, this is a lossy compression that makes the noise worse)

Inheriting some old features of the nEXO chargesim–noise is already zero suppressed in empty channels and after the pulse.





Noisy Input
True Signal
Denoised output



Given noisy inputs, algorithm outputs denoised waveforms.

The denoised output is a good match to the true signal, despite the algorithm having never seen a true signal during unsupervised training

Sometimes under/overshoot peaks.
Scale of inaccuracy similar to noise.
Is this an expression of the noise limit?

Other future approaches

- ML inference: learn dependence of likelihood function on more parameters
  - In nEXO's case, principally useful for detector nuisance parameters
- ML simulations debiasing
  - Measure discrepancies between simulations and data and determine most reliable corrections
- Differentiable simulation for calibration and inverse solver
  (thanks for a great talk, Yifan)

Probably not (in my opinion):

- ML surrogate for G4-based background simulations: rare simulation outcomes are extremely important for nEXO's physics

Summary

nEXO's ML focuses on "close to the ground" instrumental signals analysis

Lots of interesting high-dimensional information in the instruments, even for relatively simple events

Discriminator working well, denoiser about to launch

Lots of potential for un/semi-supervised training incorporating experimental data

LLNL is hiring!

Upcoming work needs a talented postdoc

Email CVs to brodsky3@llnl.gov

Official posting going up soon

Target start date: now–spring 2024

…but don't hesitate to drop us a line if you're looking further out