Xenon Signal Denoising via Supervised, Semi-Supervised, and Unsupervised Models

(*arXiv:2603.27005*)

Neutrino Physics and Machine Learning 2026

June 19th, 2026

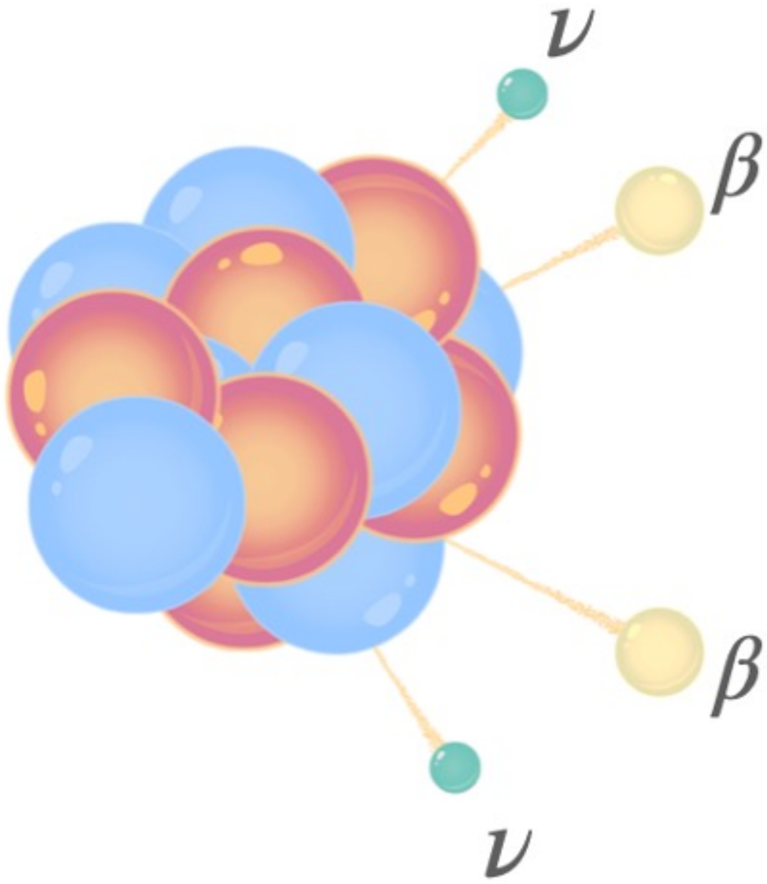
Dr. Grant Kendrick Parker, Postdoctoral Researcher

Nuclear & Chemical Sciences Division / Physical & Life Sciences Department

Prepared by LLNL under Contract DE-AC52-07NA27344.

Outline

- Introduction
  - Motivation: Measure Neutrinoless Double Beta Decay
  - Goal: Reduce Noise in Measurements
- Experiment Design & Data Properties
  - Decay Events & Simulation
  - Charge and Energy Resolution
- Solution: Machine Learning
  - Architecture: U-Net Autoencoder
  - Model Types
  - Results
- Discussion and Future Work



Introduction

Neutrinoless Double Beta Decay

- Nature of neutrino mass is unknown
  - May be Dirac (all other particles)
  - Or Majorana (it's own antiparticle)
- If neutrinos are Majorana, a new, neutrinoless double beta decay (0νββ) is available
  - Violates conservation of lepton number
  - Suggests mechanism for matter-antimatter asymmetry in the Universe
- Problem: 0νββ would be much more rare than double beta decay

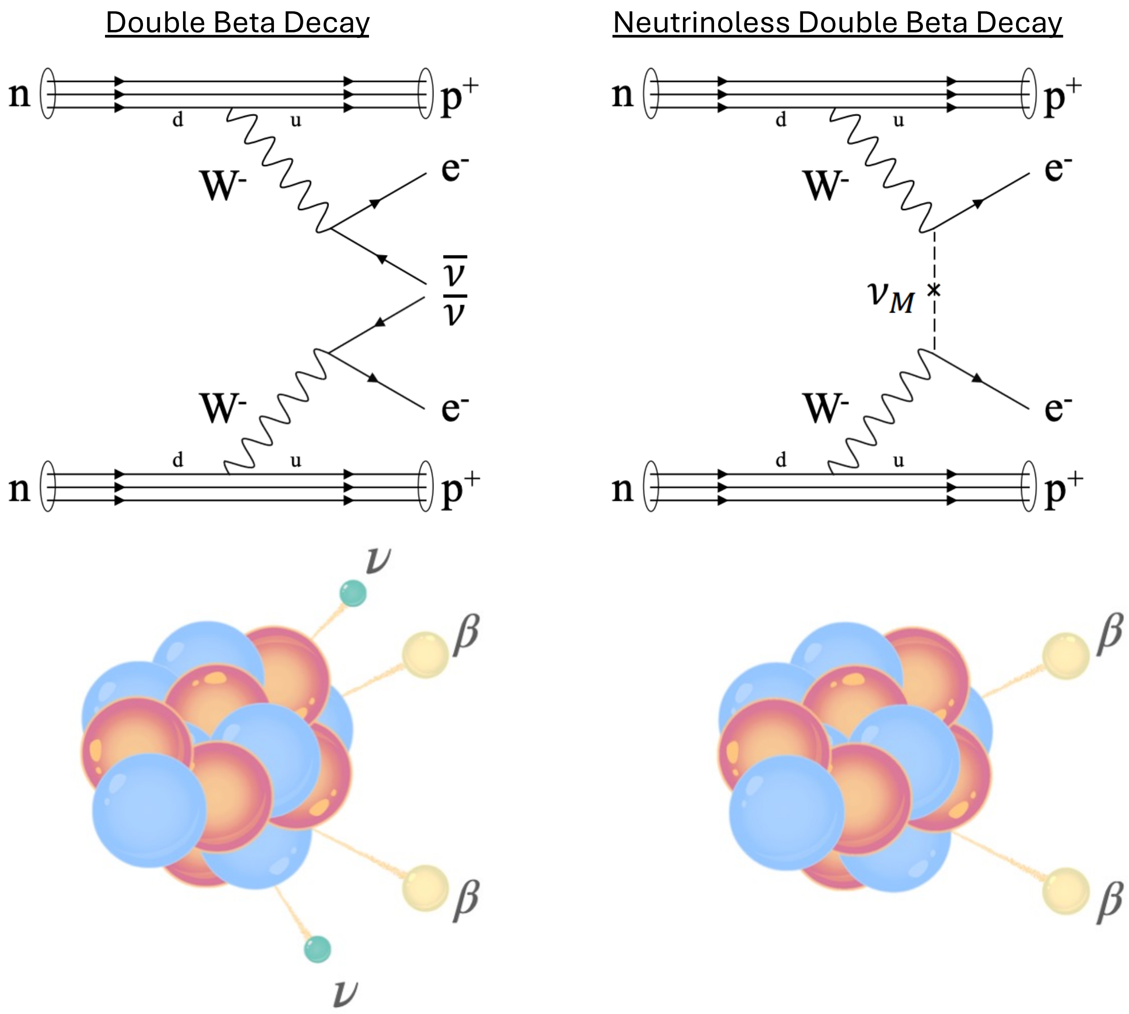


Figure: Feynman diagrams [Cao 2018] and illustrations [V. Palušová] of neutrinoless / double beta decay

The Measurement Problem

- $2\nu\beta\beta$: combined β energy is a spectrum (neutrinos carry away energy)
- $0\nu\beta\beta$: combined β exit with all of the product energy; signature is a mono-energetic line
- Experiment: energy resolution is major $0\nu\beta\beta$-$2\nu\beta\beta$-background discriminator
- Problem: Very good energy resolution is extremely challenging

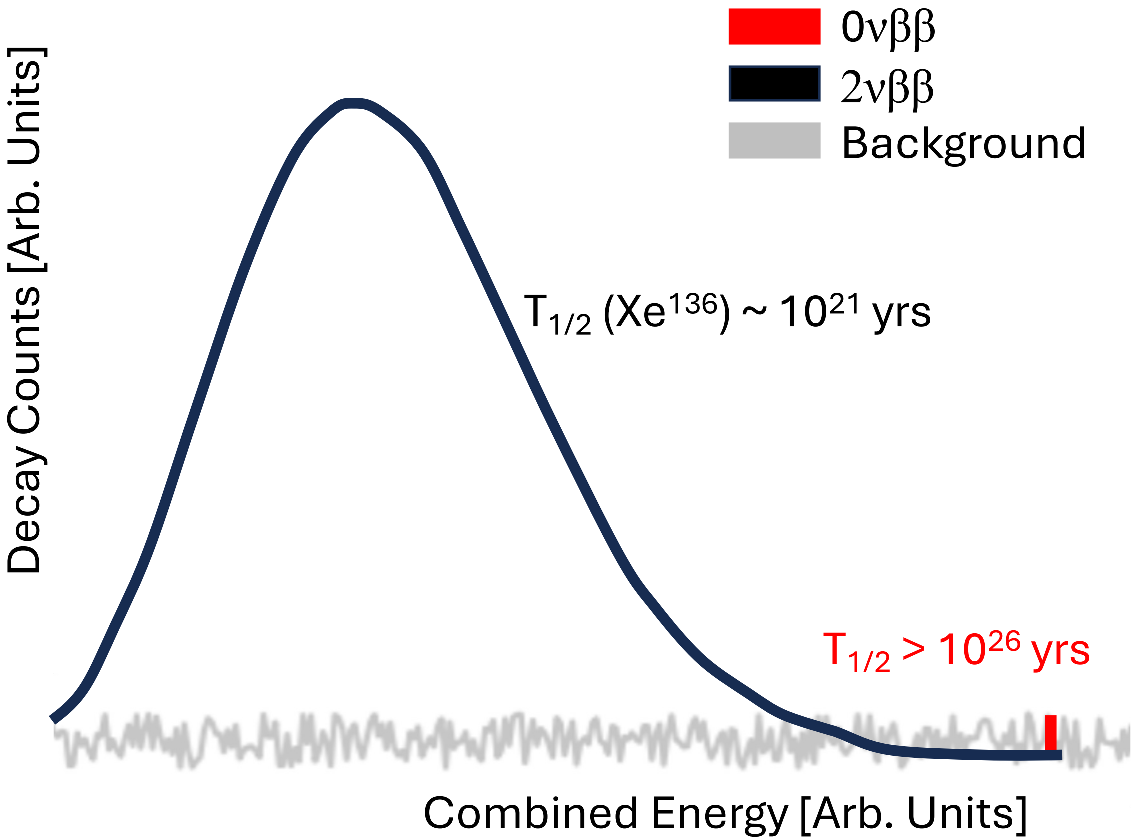
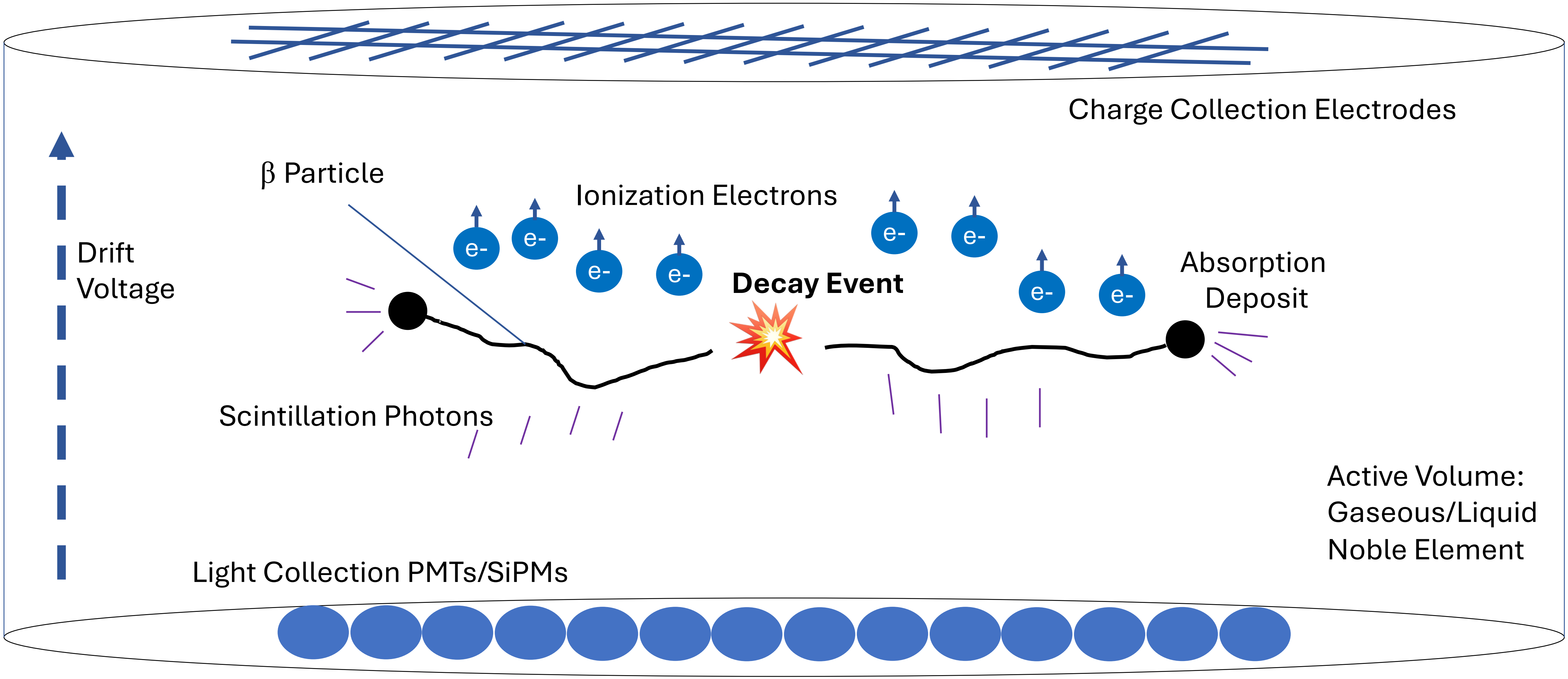


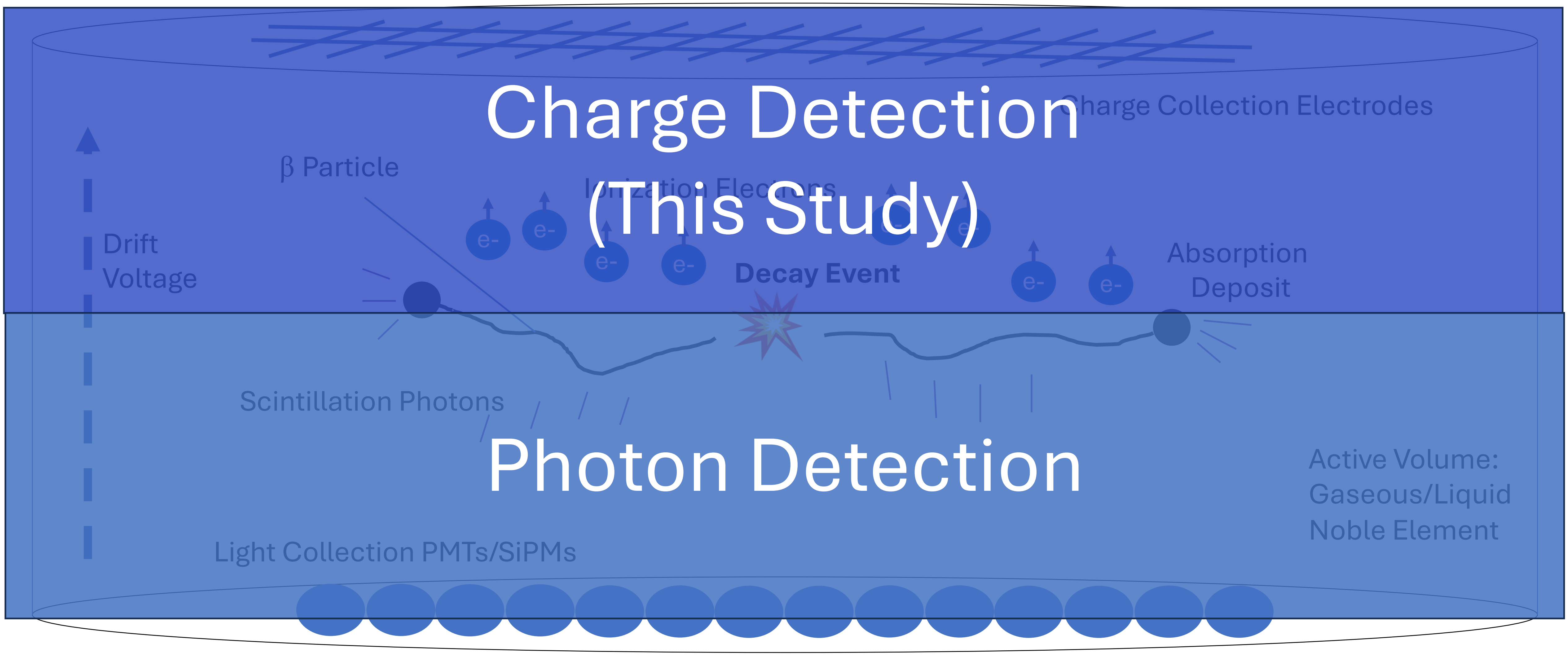
Figure: Illustration of $2\nu\beta\beta$ spectrum with hypothetical $0\nu\beta\beta$ spectrum and background.

Experiment & Data

Single-Phase Time Projection Chamber (TPC)



Single-Phase Time Projection Chamber (TPC)



nEXO Charge Collector Conceptual Design

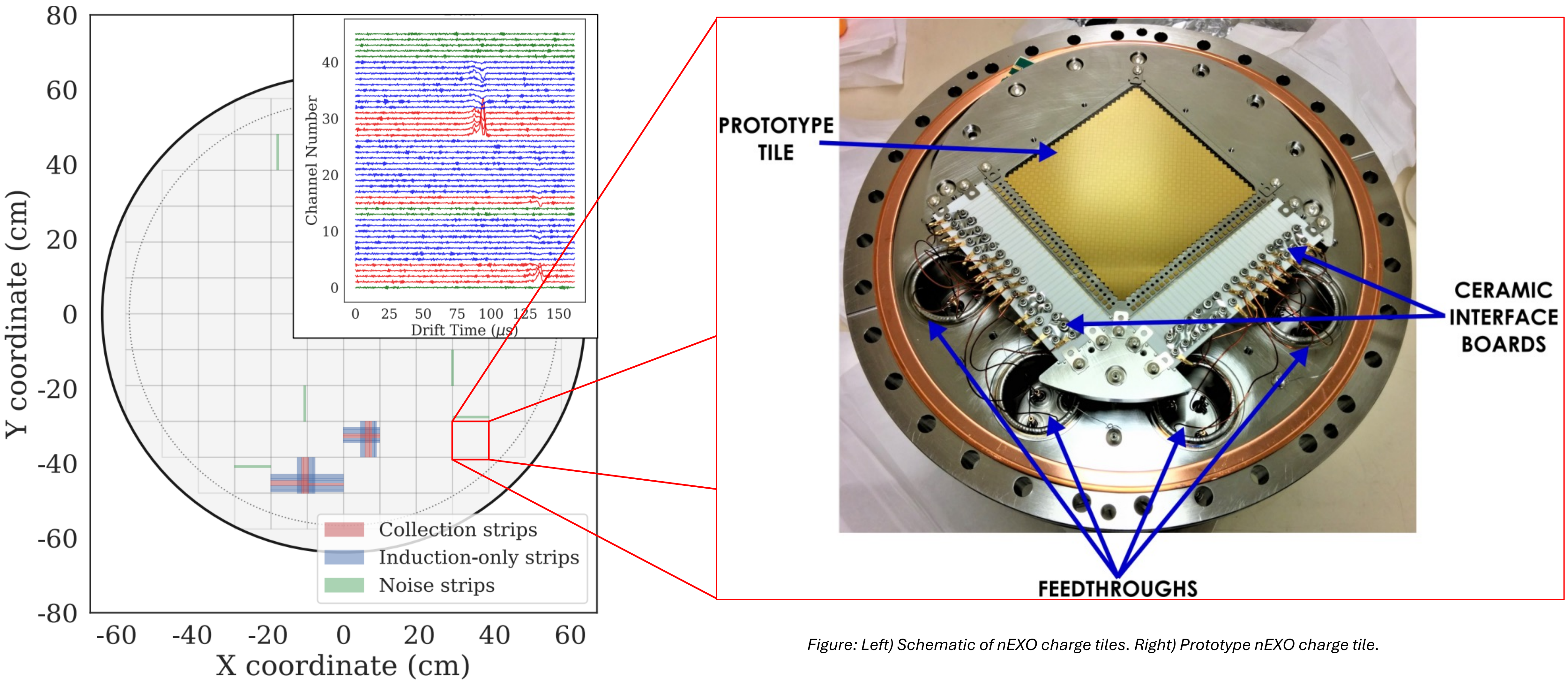


Figure: Left) Schematic of nEXO charge tiles. Right) Prototype nEXO charge tile.

Simulated Decay Event (nEXO-Type TPC)

- Energy resolution is related to charge resolution by the following:

$$\frac{\sigma_E}{E}\bigg|_{0\nu\beta\beta} \approx \frac{\sqrt{\sigma_P^2 + \sigma_Q^2}}{\langle n \rangle}$$

where σ_E/E is the fractional energy resolution, σ_P is the photon resolution, σ_Q is the charge resolution, n is the total quanta

- Target is $\sigma_E/E < 1\%$

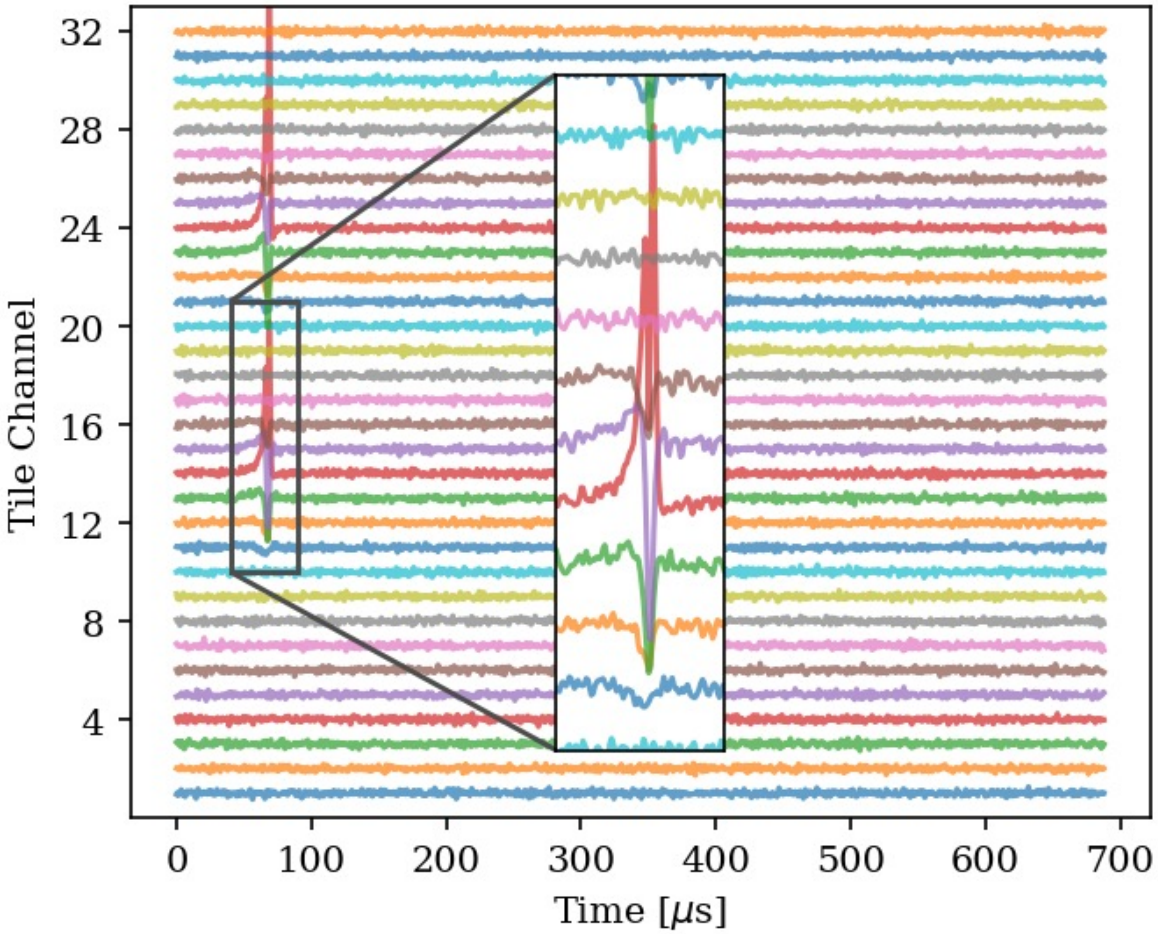


Figure: Simulated charge collection event for decay.

Denoising with ML

Optimal Denoising Limit: Wiener Filter

- Wiener filter* assumes certain noise characteristics
  - Noise has time-independent statistics, is additive
  - Noise has known power spectrum
- Noisy event given by $\hat{y}(t) = Ay(t) + n(t)$, where A is a linear estimator, n is noise, y is clean true signal
- Optimal filter output is given by $\hat{y}_{\text{optimal}}(t) = \hat{A}y(t)$, where:

$$\hat{A} = \frac{\sum_k \hat{Y}_k' Y_k'^*}{\sum_k |Y_k'|^2}$$, such that

$$\boxed{Y_k = \text{rFFT}\{y\}_k \Rightarrow Y_k' = Y_k/p_k}$$

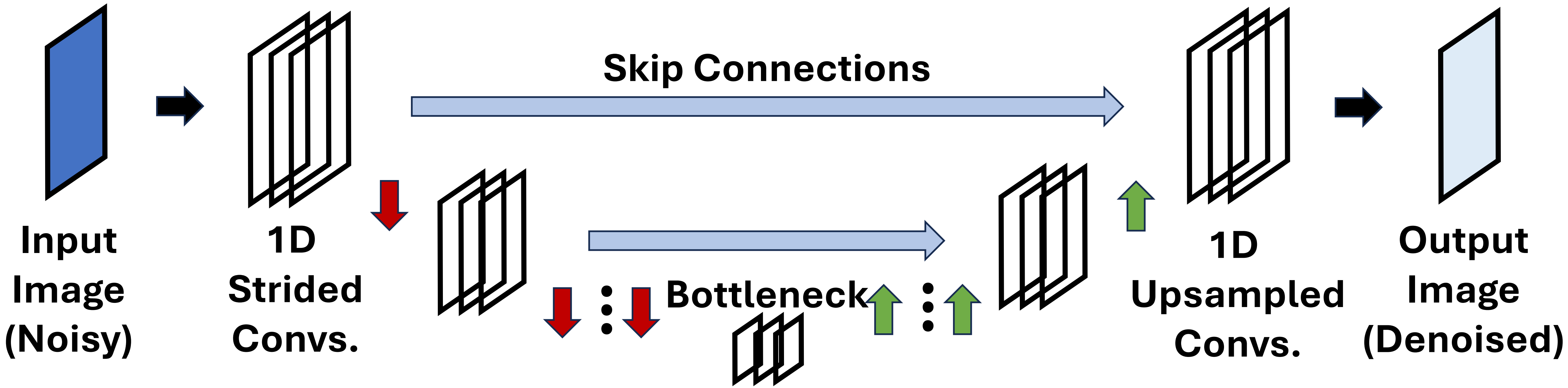
Inaccessible in a real experiment

$$\underbrace{\hat{Y}_k = \text{rFFT}\{\hat{y}\}_k}_{\text{Real-valued Fourier coefficients}} \Rightarrow \hat{Y}_k' = \hat{Y}_k/p_k$$

Noise amplitude spectral density

* N. Wiener, Extrapolation, interpolation, and smoothing of Stationary time series: With engineering applications (MIT Press, 1949).

Shared Architecture: U-Net Autoencoder



We Train Models of Each Scenario:

- Supervised Learning
  - Clean (true physics) signals and noisy counterparts, basic loss function
  - Comes with uncertainty; deployed detector measurements will never exactly match simulation, so denoiser performance will be suboptimal
- Unsupervised Learning
  - Only noisy measurements fed to model, all instructions and knowledge encoded in loss function
  - Very difficult to train, loss space fluctuates greatly
- Semi-Supervised Learning (Two-Stage Learning)
  - Simulated clean-noisy pairs, then just noisy (containing true physics) measurements

Supervised Learning

- Easiest (least-realistic) learning method:
  - Provide noisy-clean simulated image pairs
  - Loss function is Smooth L1

$$\mathcal{L}_{\mathrm{SL1}}(\hat{x}, x) = \begin{cases} \sum_{i=0}^{N} \frac{(x_i - \hat{x}_i)^2}{2N\beta}, & \text{if } |x_i - \hat{x}_i| < \beta \\ |x_i - \hat{x}_i| - 0.5\beta, & \text{otherwise} \end{cases}$$

- Training:
  - 83M simulated event pairs, 200K reserved for validation, 69K for final evaluation
  - 300K training steps at batch size 256

Unsupervised Learning

- Hardest learning method:
  - Provide only noisy simulated events, rely entirely on loss function
- MC-SURE Loss Function*
  - SURE** loss assumes noise is Gaussian and white with variance σ^2:

$$\mathcal{L}_{\text{SURE}} = \text{MSE}(\hat{x}, y) + 2\sigma^2 \text{div}[f(y)]$$

  - Noise in our events is colored, with covariance matrix $\mathbf{\Sigma}$:

$$\mathcal{L}_{\text{UNSURE}} = \text{MSE}(\hat{x}, y) - n\text{Tr}[\Sigma] + 2\text{Tr}\left[\Sigma \frac{\partial f}{\partial y}(y)\right]$$

  - Last term is computationally expensive

<u>Approximation via Monte Carlo</u>

$$2\text{Tr}\left[\Sigma \frac{\partial f}{\partial y}(y)\right] \approx \frac{2}{\epsilon k} \sum_{i=1}^{k} z_i^T \underbrace{[f(y + \epsilon z_i) - f(y)]}_{\text{Finite-Difference Derivative Approximation}}$$

Scaling Factor

Independent Draw From Noise Distribution

<u>Note:</u> Quantization is the performance floor for MC-SURE models, as quantized noise breaks SURE conditions

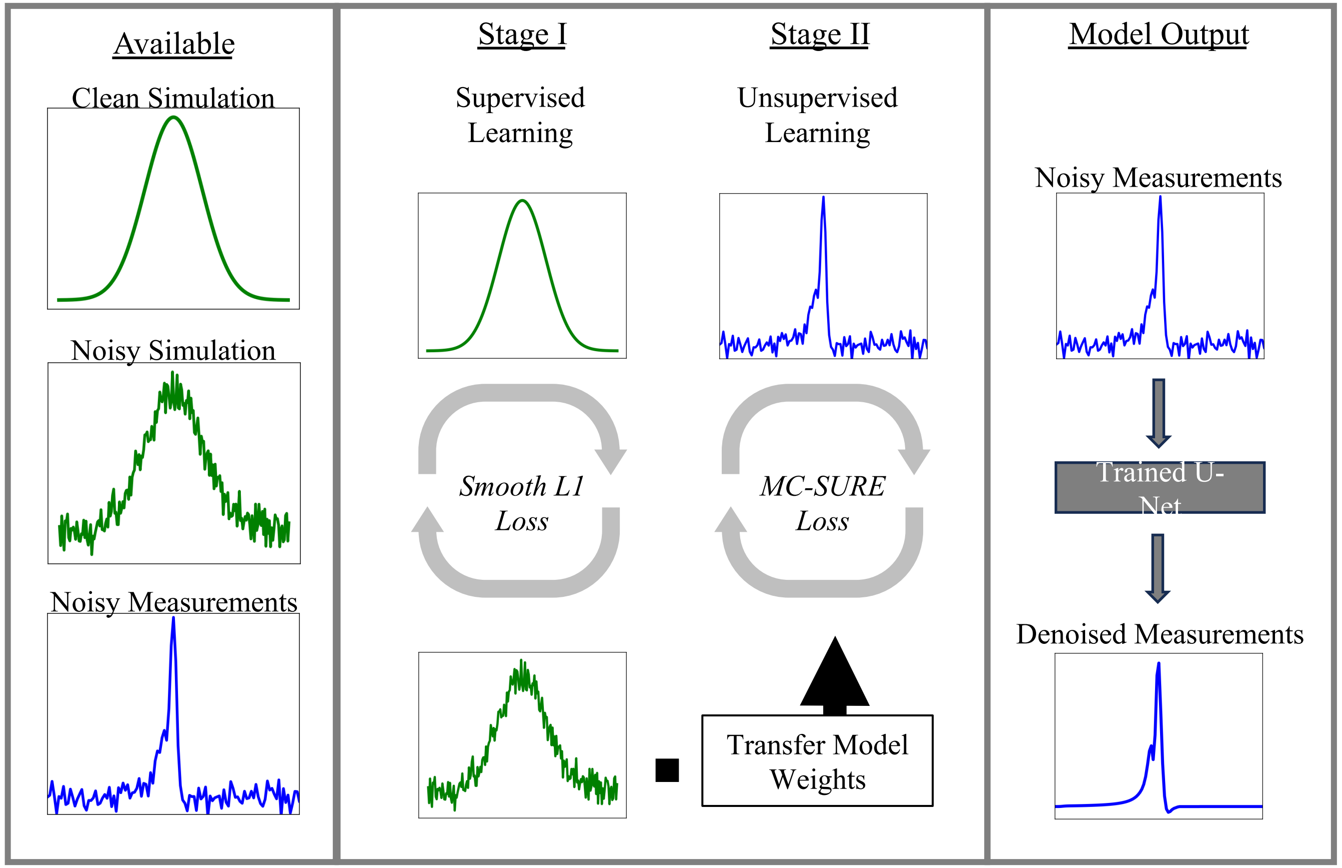
*S. Ramani, T. Blu, and M. Unser, Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms, Trans. Img. Proc. 17, 1540–1554 (2008).
**C. A. Metzler, A. Mousavi, R. Heckel, and R. G. Baraniuk, Unsupervised learning with stein's unbiased risk estimator (2020), arXiv:1805.10531 [stat.ML].

Compromise: Semi-Supervised Learning

- Scenario: Experimentalist has approximate simulation of true signal, followed by noisy measurements

- Stage I - Supervised:
  - Train in supervised mode: pairs of events with estimated ("distorted") signal coupled with noisy counterpart

- Stage II - Unsupervised:
  - <u>Resume</u> training in unsupervised mode with noisy measurements that contain *correct* signal



A Simulated Experimentalist: Distorting True Signals

- To simulate experimentalist's knowledge, we distort true signals
- Distortion: Gaussian filter
  - Distortion quantized by information loss:

$$D = \frac{\mathbb{E}[(y_{\text{recovered}} - y_{\text{original}})^2]}{\mathbb{E}[y^2_{\text{original}}]}$$

- Inverting the filter does not return the true signal – information is irrecoverably lost, and the ML <u>cannot</u> just learn the filter inversion

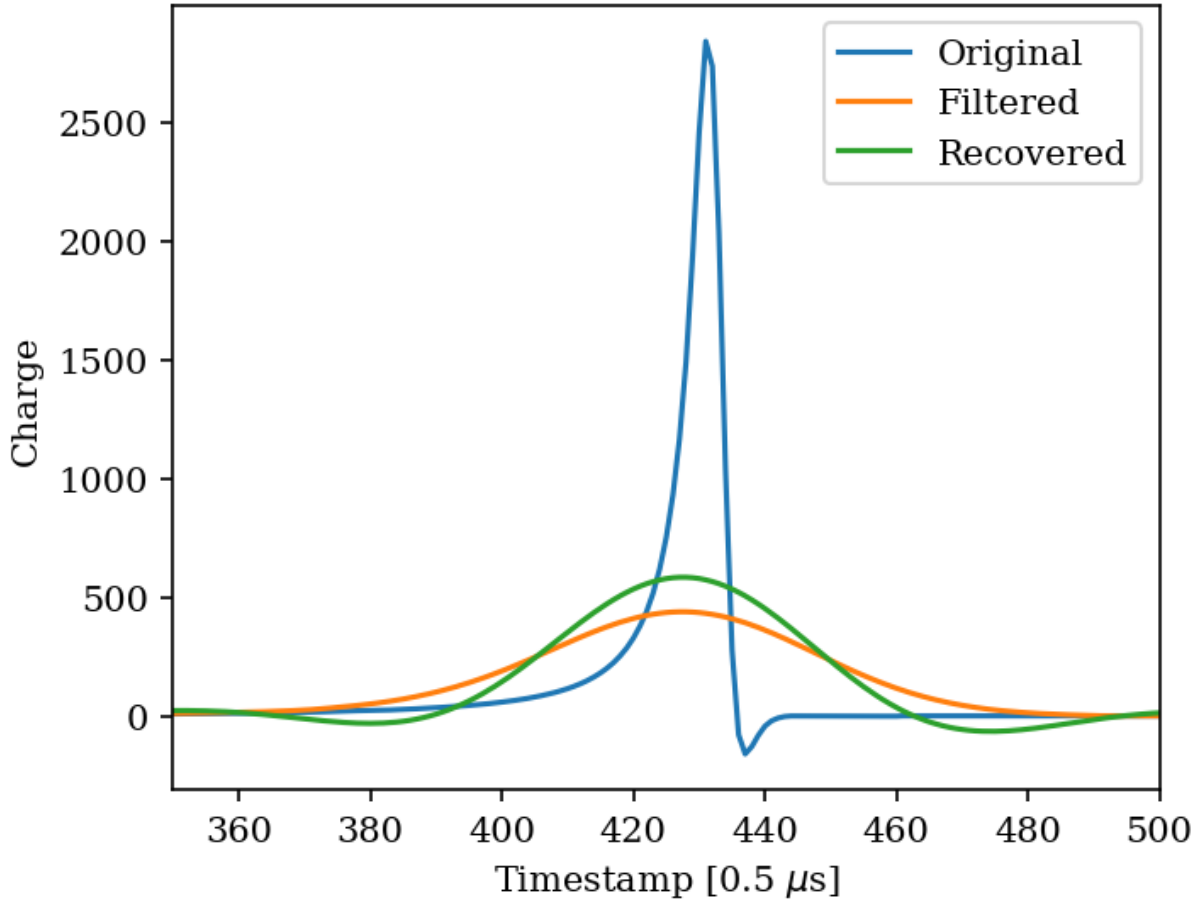


Figure: Example of applying Gaussian filter, output orange, to true signal (blue). Attempting to invert the filter returns green profile.

Results

Example of Model Output

- All models successfully eliminate noise in regions distant from signal
- Models successfully capture bulk signal features
- Primary error sources are sharp, high frequency signal regions

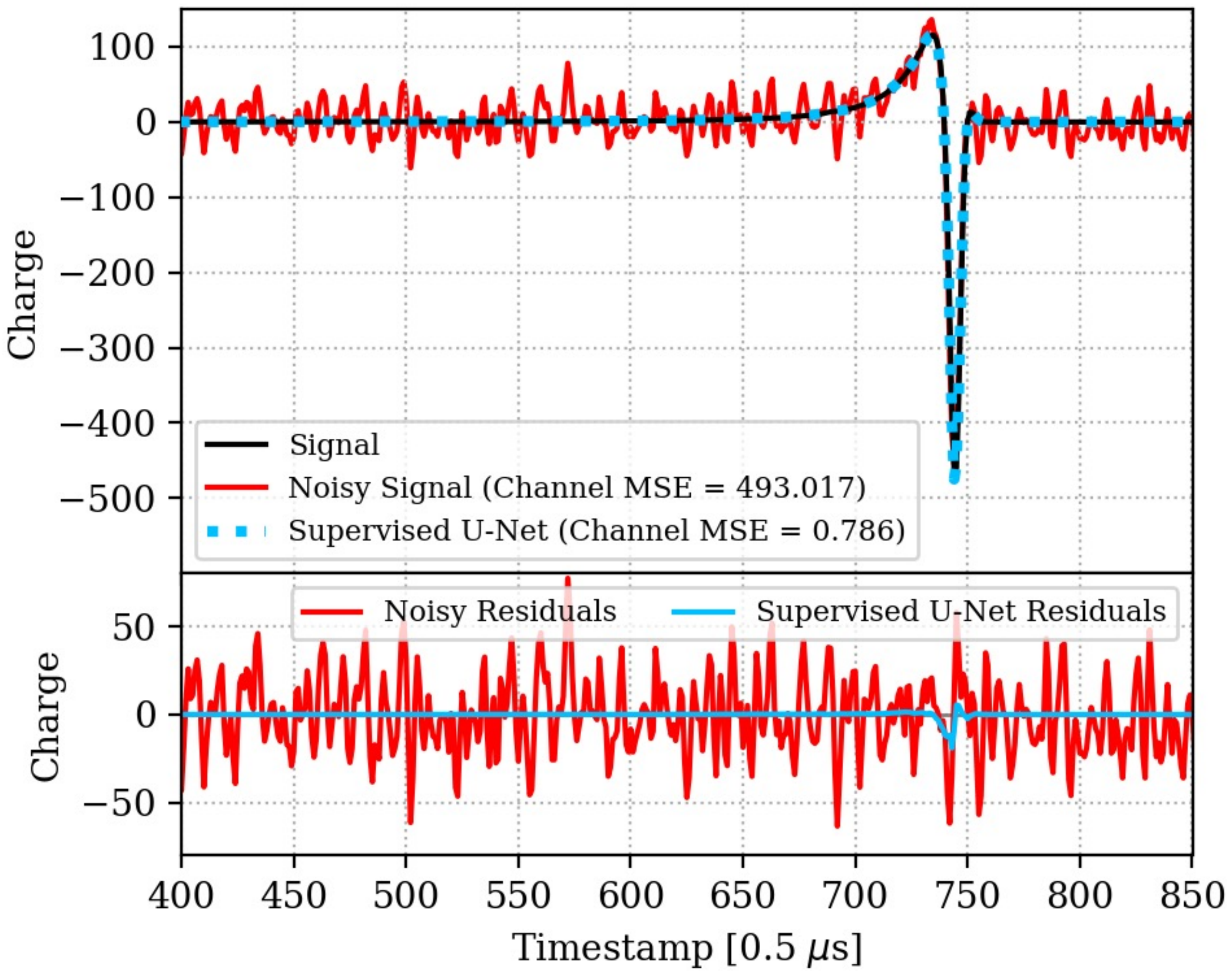


Figure: Top) Single-channel example of supervised model denoising output (blue) of noisy measurement (red), compared to true signal (black). Below) Corresponding residuals.

Results For All Models

- Supervised model produces near-optimal denoising result

- Unsupervised model performs moderately

- Semi-supervised models in Stage I perform as a function of signal distortion

- **Stage II performance approaches optimal limit** despite high amounts of signal distortion

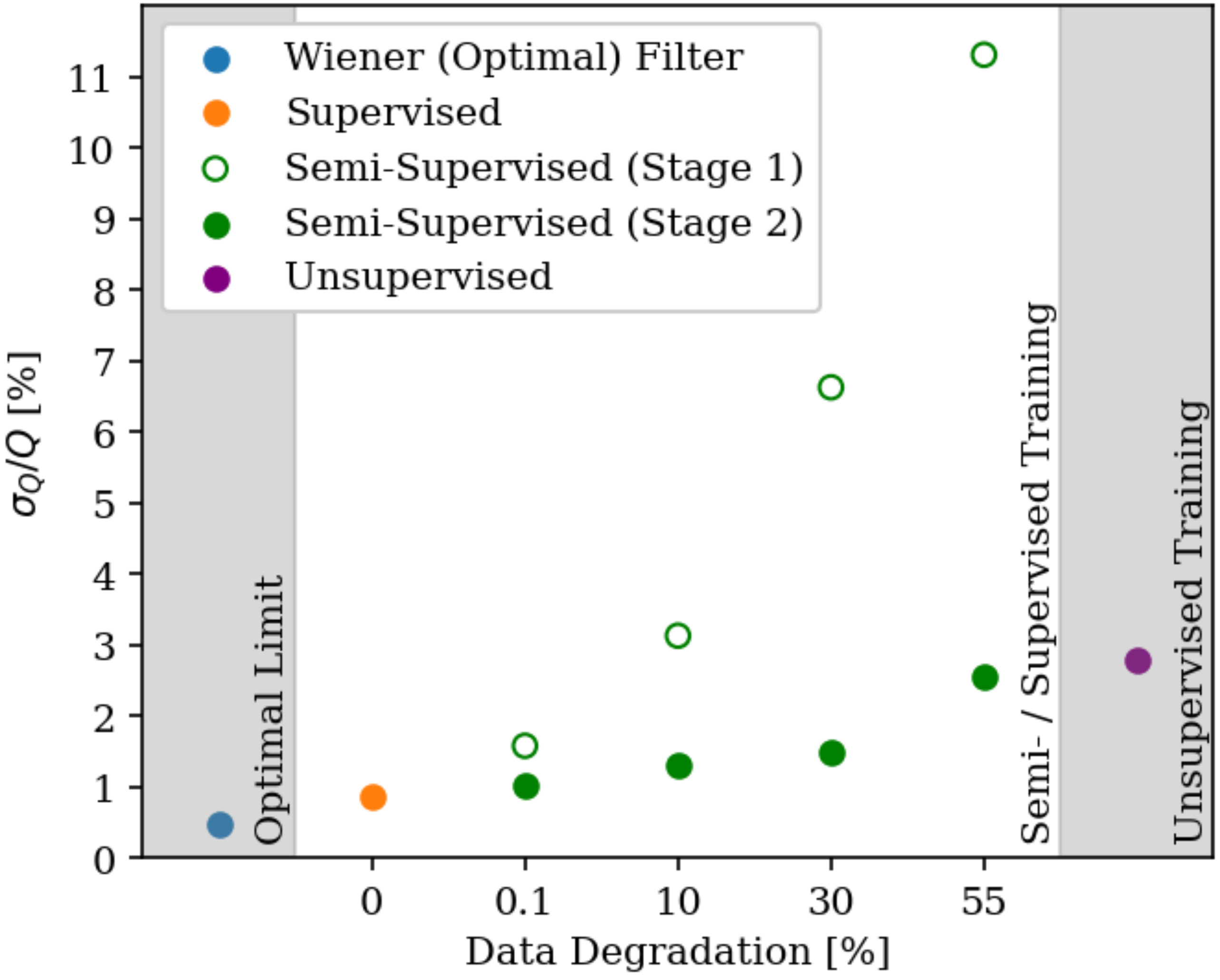
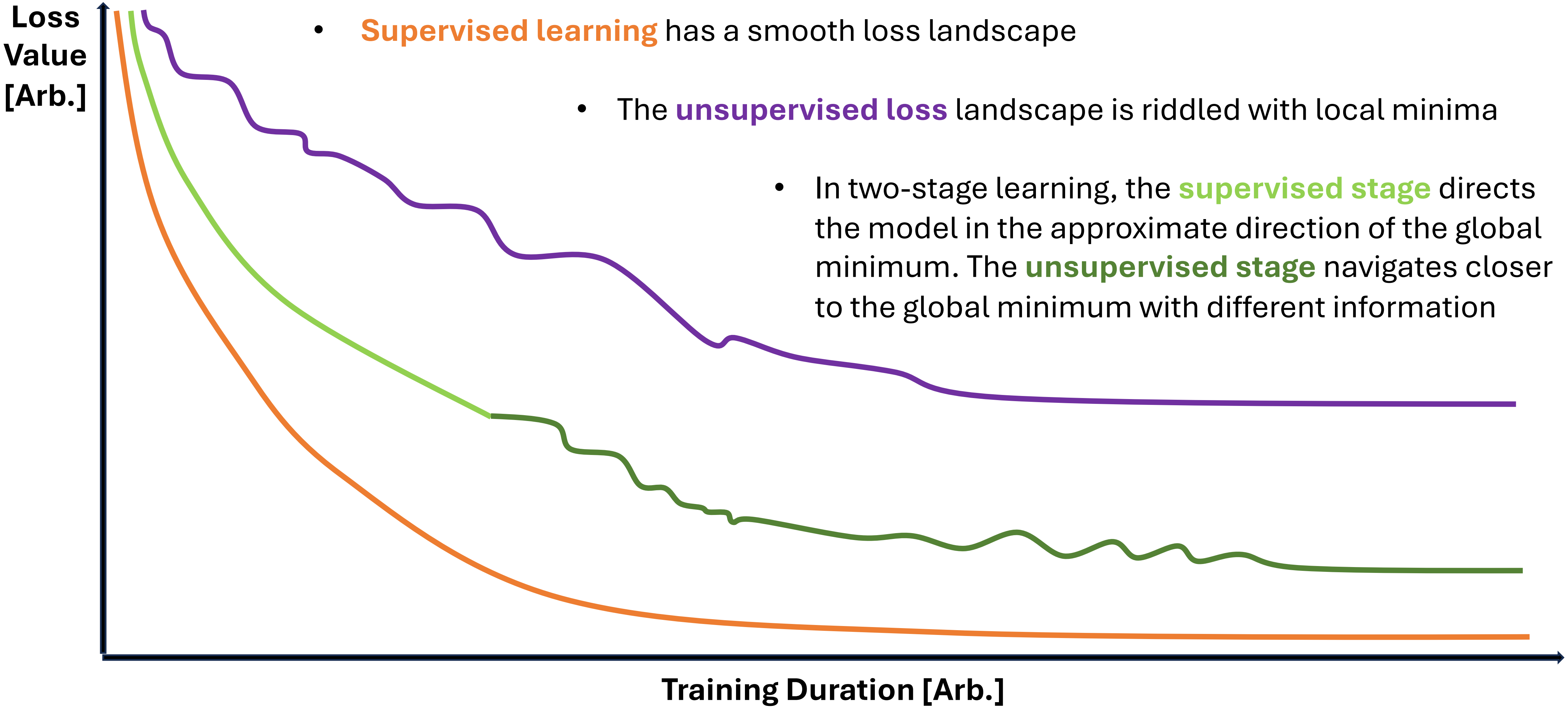


Figure: Charge resolution as a result of denoising for all models.

Aside: Hypothesizing Why Semi-Supervised Works

- **Supervised learning** has a smooth loss landscape
- The **unsupervised loss** landscape is riddled with local minima
- In two-stage learning, the **supervised stage** directs the model in the approximate direction of the global minimum. The **unsupervised stage** navigates closer to the global minimum with different information



Improvement in Energy Resolution

- Energy resolution (proportional to half-life sensitivity) calculated at nEXO photon resolution across all models*
- Supervised model returns sensitivity greater than nEXO projection
- Semi-supervised models approach nEXO sensitivity
- Performing denoising on photon component of decay events could push energy resolution even lower

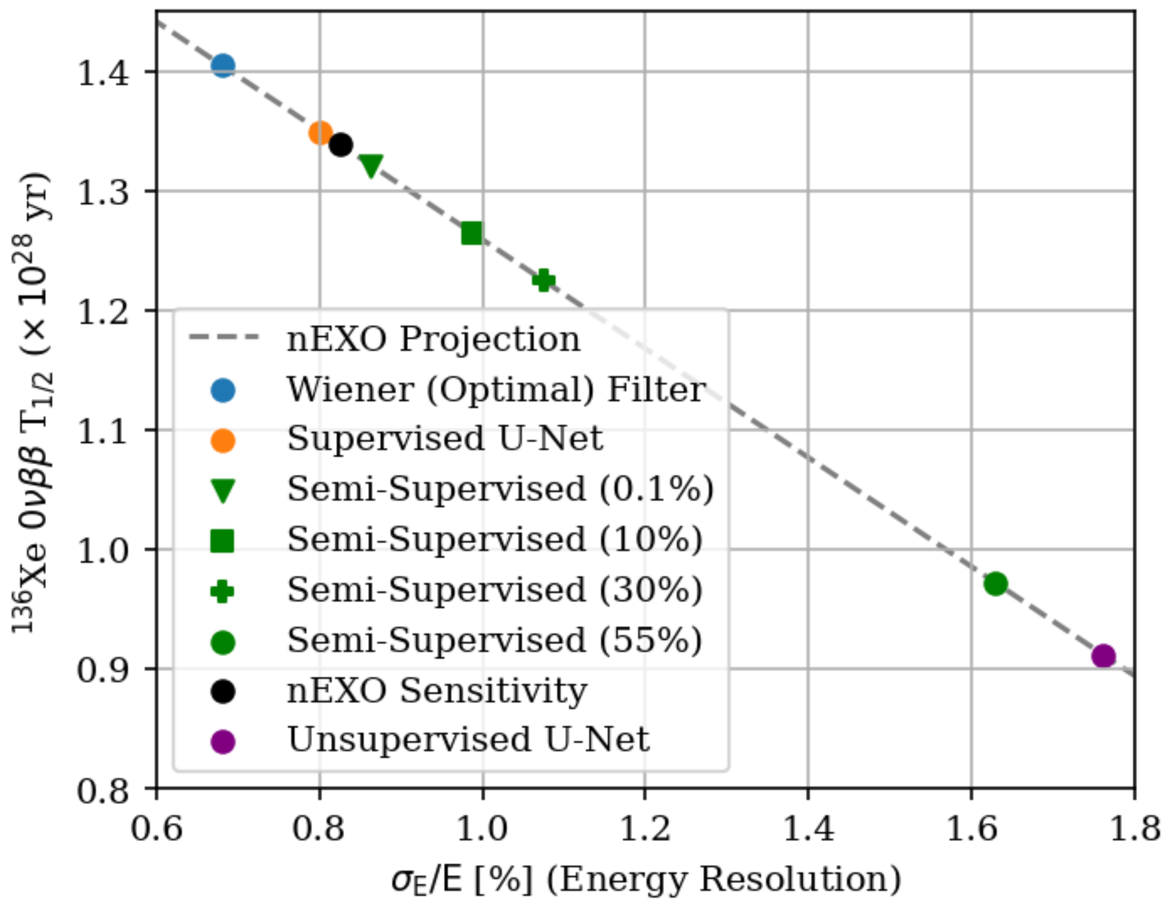


Figure: Decay half-life sensitivity as a function of energy resolution for all models.

*M. R. Anderson, V. Basu, R. D. Martin, C. Z. Reed, N. J. Rowe, M. Shafiee, and T. Ye, Performance of a convolutional autoencoder designed to remove electronic noise from p-type point contact germanium detector signals, Eur. Phys. J. C 82, 1084 (2022), arXiv:2204.06655 [nucl-ex].

Discussion and Future Work

Denoising for $0\nu\beta\beta$ and Beyond

- This study demonstrates
  - Supervised denoisers can reach near-optimal performance limits
  - **Even with limited signal knowledge, a semi-supervised denoiser can reach near-optimal performance**
- With the addition of photon signal denoising, $0\nu\beta\beta$ energy resolution of << 1% can be within reach
- Applications not limited to $0\nu\beta\beta$, **all detectors have noise**
- Potential semi-supervised benefit: post-deployment detector simulation generator
  - Latent space captures real physics + detector characteristics; available for fast, accurate simulation production

Thank You

Grant Parker
parker79@llnl.gov