Machine learning to denoise pulses from a p-type point contact germanium detector

CAP Congress 2022 | Hamilton, ON

Mark Anderson
anderson.mark@queensu.ca

June 8th, 2022

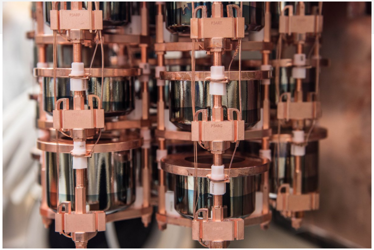
Overview

1. **Introduction**
   - Rare event searches using HPGe detectors
   - Problem statement and motivation
   - Autoencoders, denoising autoencoders
2. **Methodology**
   - Overview of our detector
   - Datasets (simulated and real)
   - Training procedures
3. **Results**
   - Denoising performance on simulations
   - Verification with real detector data
4. **Conclusions and future work**
   - Broadly applicable to particle astrophysics

Largely a summary of paper:

arXiv:**2204.06655**

Introduction



[10]

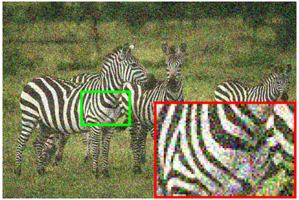
- High-purity germanium (HPGe) detectors widely used in rare event searches
    - Neutrinoless double-beta decay[2, 3, 4]
    - Dark matter searches[5, 6]
    - Other beyond Standard Model physics[7, 8, 9]

- Adaptive denoising techniques are not typically applied to data

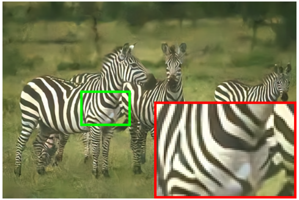
- Noise removal could help advance searches for certain rare event interactions
    - Identify low-energy signal events that would otherwise be dominated by electronic noise
        - Relevant for solar axions, Pauli Exclusion Principle violation, electron decay, etc.[11]
    - Improved background rejection based on pulse shapes
        - For example, slow energy-degraded pulses[12, 13]
    - More accurate measurements of pulse amplitudes $\rightarrow$ better energy resolution

Introduction

- Deep learning is frequently used to remove noise in other fields
  - Often outperforms traditional denoising methods (e.g., moving average, Savitzky-Golay filtering, wavelet thresholding)
- Most applications are to 2-dimensional images



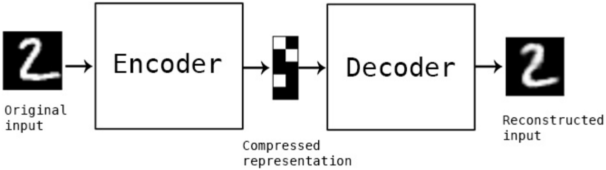
→ [14]



- Will show that deep learning is also very effective at denoising 1-dimensional electronic signals (demonstrated specifically on HPGe detector pulses)

Autoencoders

- An autoencoder is an algorithm used to learn a useful representation of data
  - Trained to map the inputs to the inputs (with some form of constraint)



[15]

- By definition, an autoencoder is *lossy*
  - The goal is to retain as much useful information as possible
- Typically a neural network
- Widely used for dimensionality reduction, anomaly detection, and generative modelling

Autoencoders

- Components include an encoder, f_θ, and decoder, $g_{\theta'}$

  Internal/latent representation, y, is obtained by applying encoder to input:

  $$f_\theta(x) = y$$

  Input reconstruction, z, is obtained by applying decoder to latent representation:

  $$g_{\theta'}(y) = g_{\theta'}(f_\theta(x)) = z$$

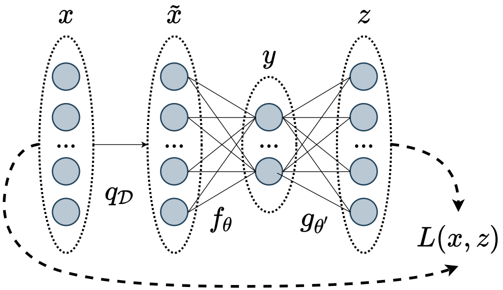
  Minimize some **loss function** quantifying the reconstruction of x, $L(x, z)$, to train the autoencoder

  - e.g., mean squared error

  $$L(x, z) = \frac{1}{N} \sum_{i}^{N} \|z_i - x_i\|_2^2$$

Denoising autoencoders

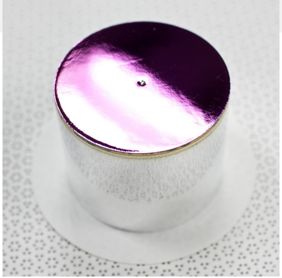
- Denoising autoencoders impose the constraint that reconstruction must also remove noise
  - Proposed as a method to extract robust features[16]
    - Purpose was not for denoising, but to learn a better representation for classification tasks
  - Input becomes a corrupted version of x, $\tilde{x}$, by some process $q_{\mathcal{D}}$



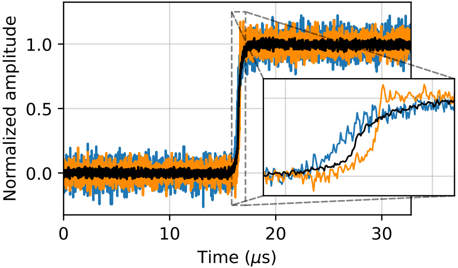
[1]

The detector

- Signals are from a 1 kg p-type point contact detector located at Queen's University
  - Cylindrical with a radius of 3 cm and height of 5 cm
  - Manufactured by ORTEC/AMTEK
  - Operated in a PopTop cryostat



[17]



[1]

- Each signal is a sequence of voltages sampled at a fixed interval
  - Recorded with 16-bit digitizer
  - Observed noise levels after preprocessing reflect energy of pulse

Datasets: real detector data

Americium-241 source

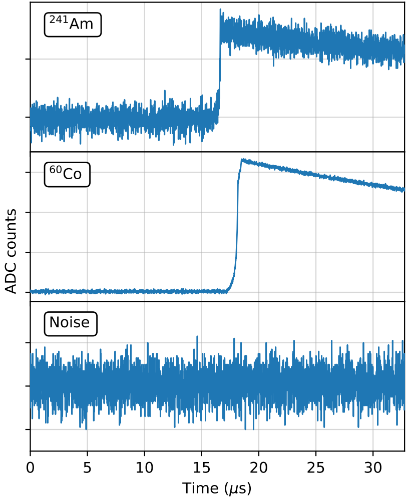
- Produces 60 keV γs
- Almost entirely single-site events

Cobalt-60 source

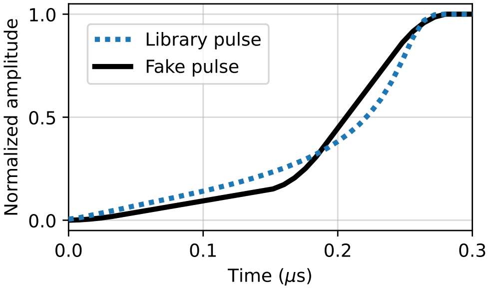
- Produces 1173 keV and 1332 keV γs
- Numerous multi-site events from Compton scatters

Detector noise

- Collected by randomly triggering the detector
- Signals filtered to remove actual events that occur in the same trigger window



Datasets: simulated data



[1]

Library pulses

- 1724 simulated "library" pulses[18]
- Each pulse corresponds to point on $1\,\text{mm} \times 1\,\text{mm}$ azimuthally symmetric grid
- Created using `siggen` simulation software[19]
- Used to infer position of real events

Fake pulses

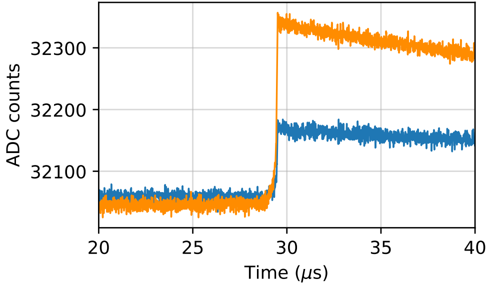
- Generated with piecewise functions
- Mimic the general shape of the library pulses without the requirement of complex physics simulations

Preprocessing

- Data pulses preprocessed to remove baseline
- Data pulses have exponential decay removed with pole zero correction
- Data pulses scaled by amplitude (calculated with a trapezoidal filter)

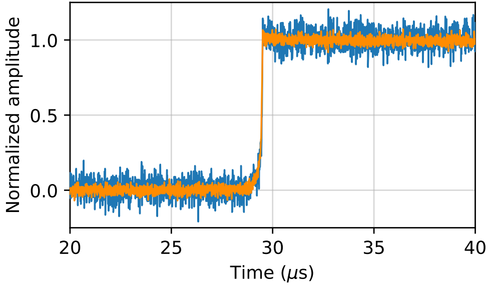
} **Amplitude normalization**

- *Simulated pulses do not require this preprocessing* (already amplitude normalized)



Before preprocessing

→

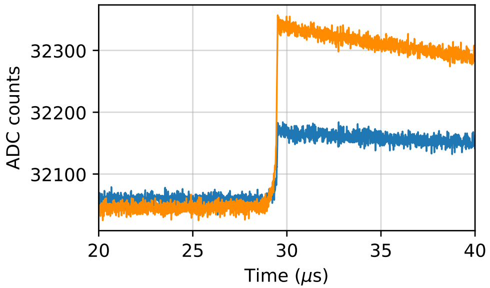


After preprocessing

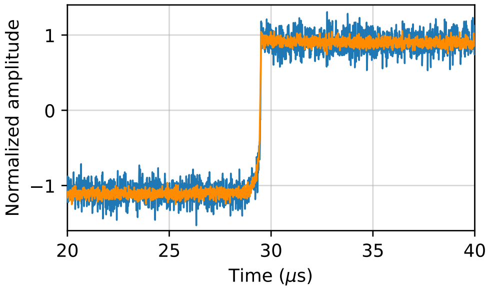
Preprocessing

- Data pulses preprocessed to remove baseline
- Data pulses have exponential decay removed with pole zero correction
- Data pulses are mean-subtracted and scaled to have standard deviation of 0.5
- *Simulated pulses only require last step*

} **Standardization**



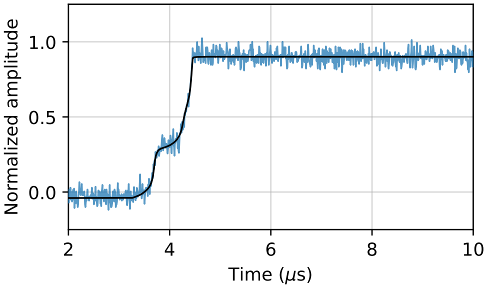
Before preprocessing

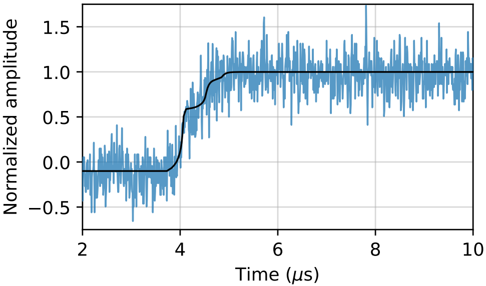


After preprocessing

Data augmentation

- From the simulated single-site event pulses, can create a diverse training set
  - Combine single-site simulated pulses to create *artificial multi-site events*
  - Apply random *horizontal shifts, vertical shifts, and amplitude scales* to each pulse
  - Add *detector noise* to each pulse with a random standard deviation





Training procedures

Regular

- Trained to map the noisy pulse to the corresponding clean underlying pulse
- Must know the true pulse – *only works on simulated data*

Noise2Noise[20]

- Trained to map noisy pulse to *another* noisy pulse (different noisy realizations of same underlying pulse)
  - An impossible task in practice
  - Model will instead learn to predict the mean, given infinite different noisy realizations
- Can be used with either simulations or real data
  - For detector data, add even more noise to the already noisy pulse
  - Include a *total variation* penalty[21] to original loss function L_0 to account for the noisy true mean
  - Penalize the absolute difference between given sample (j) and subsequent sample $(j+1)$ in pulse
  - Apply scaling factor λ to control weighting

$$L = L_0 + \frac{\lambda}{N} \sum_{i}^{N} \sum_{j}^{M-1} |z_{i,j+1} - z_{i,j}|$$

Training procedures

Regular

- Trained to map the noisy pulse to the corresponding clean underlying pulse
- Must know the true pulse – *only works on simulated data*

Noise2Noise[20]

- Trained to map noisy pulse to *another* noisy pulse (different noisy realizations of same underlying pulse)
  - An impossible task in practice
  - Model will instead learn to predict the mean, given infinite different noisy realizations
- Can be used with either simulations or real data
  - For detector data, add even more noise to the already noisy pulse
  - Include a *total variation* penalty[21] to original loss function L_0 to account for the noisy true mean
  - Penalize the absolute difference between given sample (j) and subsequent sample ($j+1$) in pulse
  - Apply scaling factor λ to control weighting

$$L = L_0 + \frac{\lambda}{N} \sum_{i}^{N} \sum_{j}^{M-1} |z_{i,j+1} - z_{i,j}|$$

Training procedures

Regular

- Trained to map the noisy pulse to the corresponding clean underlying pulse
- Must know the true pulse – *only works on simulated data*

Noise2Noise[20]

- Trained to map noisy pulse to *another* noisy pulse (different noisy realizations of same underlying pulse)
  - An impossible task in practice
  - Model will instead learn to predict the mean, given infinite different noisy realizations
- Can be used with either simulations or real data
  - For detector data, add even more noise to the already noisy pulse
  - Include a *total variation* penalty[21] to original loss function L_0 to account for the noisy true mean
  - Penalize the absolute difference between given sample (j) and subsequent sample ($j+1$) in pulse
  - Apply scaling factor λ to control weighting

$$L = L_0 + \frac{\lambda}{N} \sum_{i}^{N} \sum_{j}^{M-1} |z_{i,j+1} - z_{i,j}|$$

Training procedures

Regular

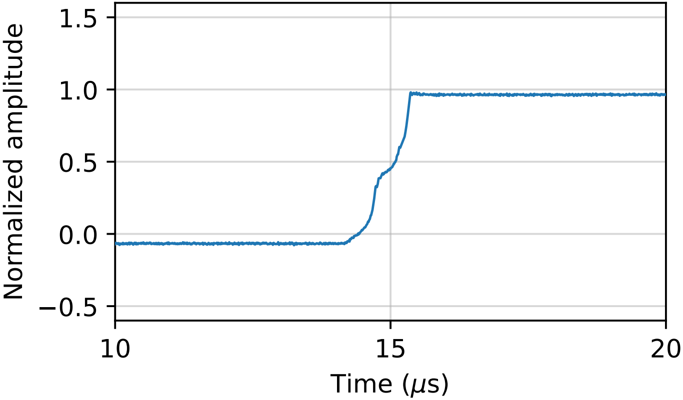
- Trained to map the noisy pulse to the corresponding clean underlying pulse
- Must know the true pulse – *only works on simulated data*

Noise2Noise[20]

- Trained to map noisy pulse to *another* noisy pulse (different noisy realizations of same underlying pulse)
  - An impossible task in practice
  - Model will instead learn to predict the mean, given infinite different noisy realizations
- Can be used with either simulations or real data
  - For detector data, add even more noise to the already noisy pulse
  - Include a *total variation* penalty[21] to original loss function L_0 to account for the noisy true mean
  - Penalize the absolute difference between given sample (j) and subsequent sample ($j+1$) in pulse
  - Apply scaling factor λ to control weighting

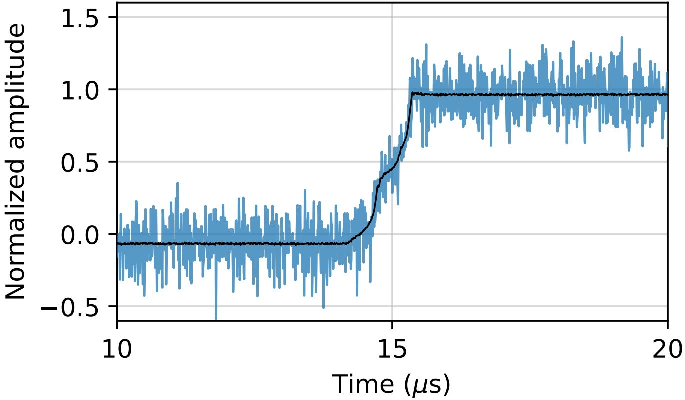
$$L = L_0 + \frac{\lambda}{N} \sum_{i}^{N} \sum_{j}^{M-1} |z_{i,j+1} - z_{i,j}|$$

Original data pulse

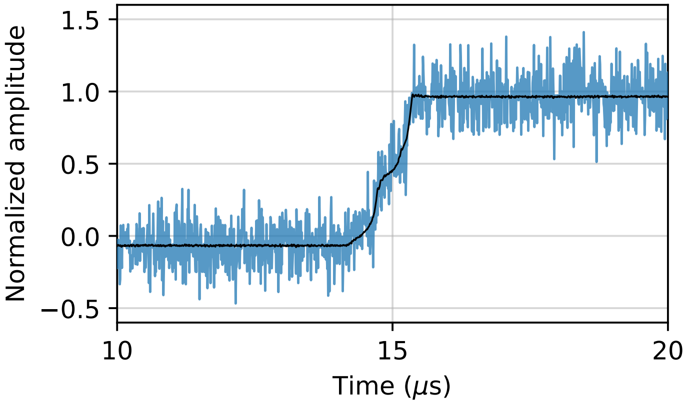


Noise2Noise

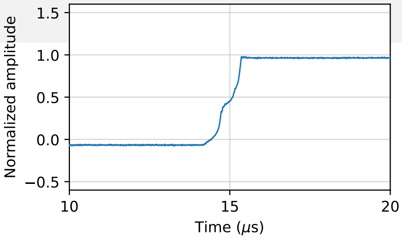
Original data pulse with a random noise pulse

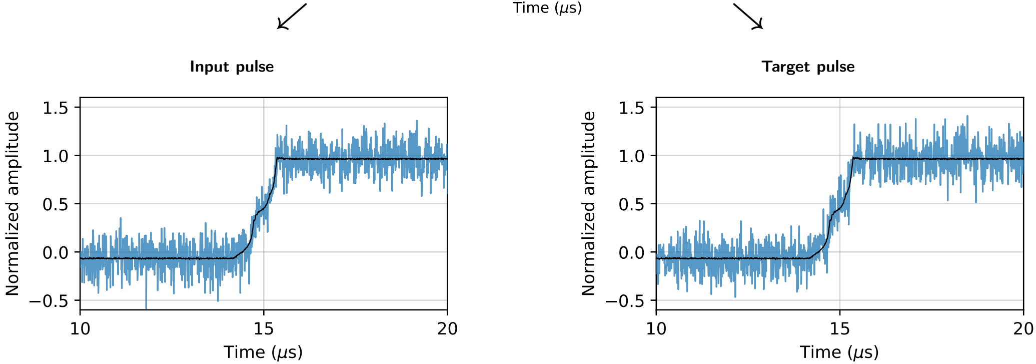


Original data pulse with *another* random noise pulse



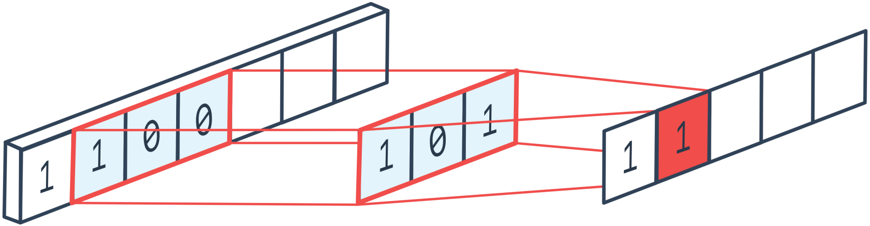
Noise2Noise





Network architecture

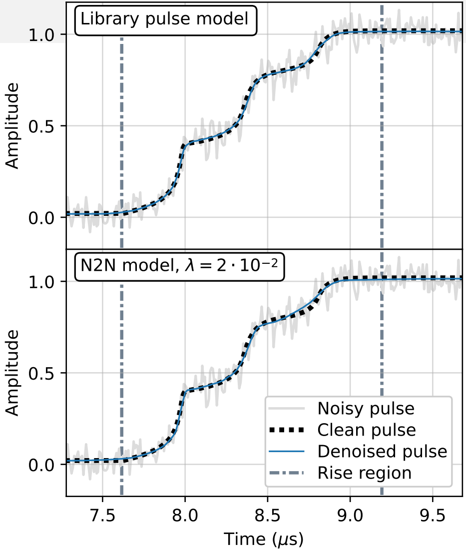
- Fully convolutional autoencoder
  - Weight sharing provides consistent noise removal across pulse
  - Feature locality and shift equivariance
  - Allows for a variable input shape (subject to some restrictions)
  - Significant reduction in the number of trainable parameters



[22]

Results: simulations

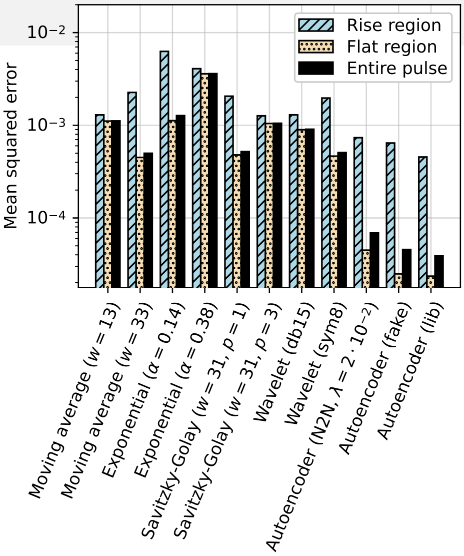
- Qualitatively, denoising with deep learning performs very well on simulations
- Autoencoder is superior to all traditional denoising methods investigated
  - Compared mean squared error on test set containing simulated single- and multi-site events
  - Each method optimized on a separate validation set to select hyperparameters
- Regular training procedure (with simulations) outperforms Noise2Noise procedure (with ^{60}Co data)
  - Still very good performance with Noise2Noise



[1]

Results: simulations

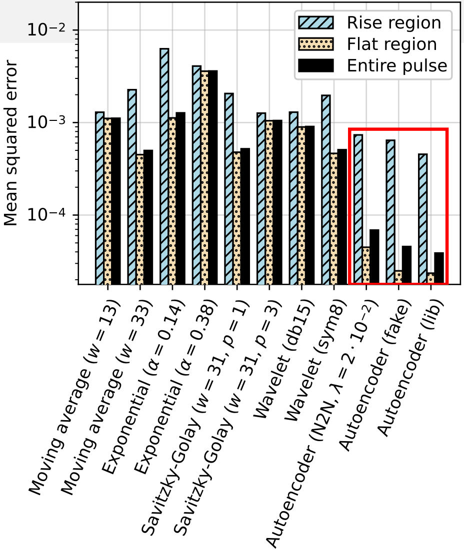
- Qualitatively, denoising with deep learning performs very well on simulations
- Autoencoder is superior to all traditional denoising methods investigated
  - Compared mean squared error on test set containing simulated single- and multi-site events
  - Each method optimized on a separate validation set to select hyperparameters
- Regular training procedure (with simulations) outperforms Noise2Noise procedure (with ^{60}Co data)
  - Still very good performance with Noise2Noise



[1]

Results: simulations

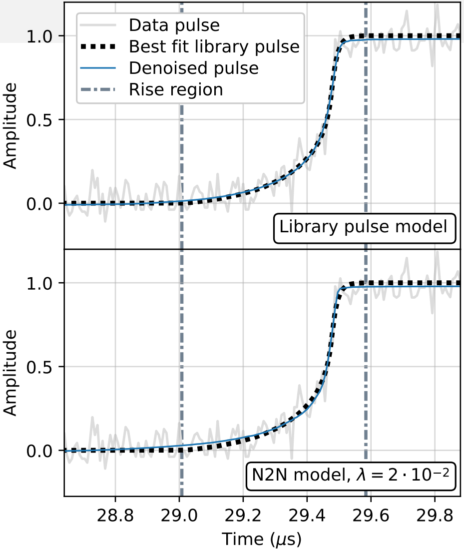
- Qualitatively, denoising with deep learning performs very well on simulations
- Autoencoder is superior to all traditional denoising methods investigated
    - Compared mean squared error on test set containing simulated single- and multi-site events
    - Each method optimized on a separate validation set to select hyperparameters
- Regular training procedure (with simulations) outperforms Noise2Noise procedure (with ^{60}Co data)
    - Still very good performance with Noise2Noise



[1]

Results: data

- Qualitatively, denoising with deep learning performs very well on data
- More difficult to quantify denoising
  - No true underlying pulse to compare to
- However, can make a statistical comparison to evaluate the performance



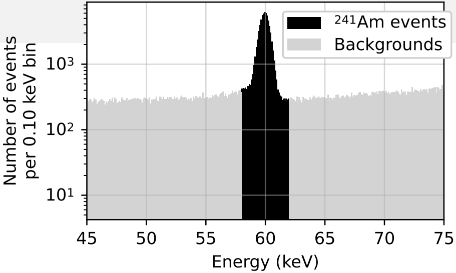
[1]

Results: data (χ^2 comparison)

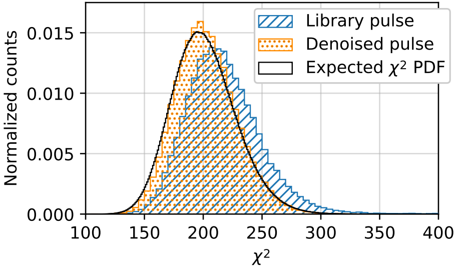
- ^{241}Am dataset contains mostly single-site events from 60 keV γs
- Use a χ^2 comparison between the original pulse and denoised pulse, best-fit library pulse

$$\chi^2 = \sum_{i=1}^{M} \frac{(z_i - x_i)^2}{\sigma_i^2}$$

- χ^2 distribution between noisy and denoised pulse is consistent with expected χ^2 distribution of our detector noise
  - Taken over 200 samples, contains rise region



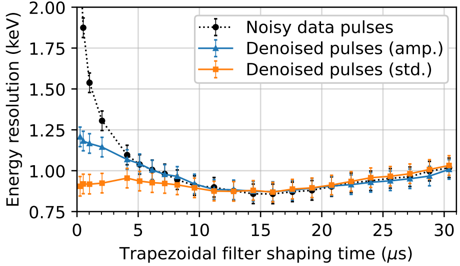
[1]



[1]

Results: data (energy resolution)

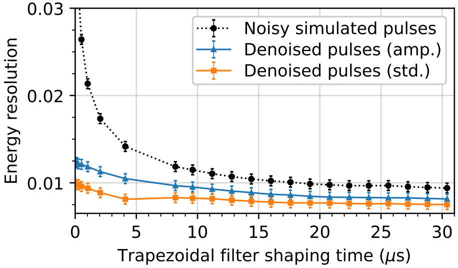
- Can also evaluate the effect of denoising on the energy resolution
    - Energy calculated from the amplitude of a trapezoidal filter
    - FWHM of 60 keV peak is the energy resolution
- Optimal energy resolution is comparable before and after denoising
- Much lower shaping time required to achieve good energy resolution
    - Especially true for standardized preprocessing
    - Important for data storage, analysis, etc.



[1]

Results: data (energy resolution)

- Create simulated dataset to mimic ^{241}Am γs
  - Single-site library events with detector noise
  - Same noise level as ^{241}Am dataset
- Observe same effect as with real data
  - Comparable energy resolution with lower shaping time
  - Best results with standardized preprocessing
- Observe different patterns
  - Optimal energy resolution improves
  - Energy resolution continues decreasing up to the limit of the shaping time
- Under ideal conditions, denoising allows for an improvement in energy resolution



[1]

Conclusions

- Deep convolutional autoencoders are effective at removing noise from HPGe detector pulses
  - Outperforms various traditional denoising methods
  - Denoised pulses are statistically consistent with data pulses
  - Can reach optimal energy resolution with a lower shaping time
    - Can improve the optimal energy resolution under some circumstances

- Models can be trained without the need for detailed detector simulations
  - “Fake” pulses are a very rough approximation to library pulses
  - Noise2Noise method requires *only* noisy detector data
    - Results could likely be improved with more data

Conclusions

- Results presented here are focused on HPGe detector data
  - Noise removal is beneficial in many contexts
  - Our group is applying these methods to signals from other detector technologies
    - Gaseous proportional counters, bubble chambers
- The encoder portion of the network can also be used for tasks
  - The denoising autoencoder forces the encoder to learn a robust representation of the data
  - Exploring the use of the encoder output for clipping restoration, peak finding, and single-/multi-site event discrimination
- Work is broadly applicable to the particle astrophysics community
  - Great potential to be expanded on
  - See poster from Tianai Ye, talk from Noah Rowe

Thank You!

More details contained in the paper. Check it out!

arXiv:**2204.06655**

Acknowledgements

- Thanks to all other authors of the paper
  - V. Basu, R. D. Martin, C. Z. Reed, N. J. Rowe, M. Shafiee, T. Ye
- This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC)
- The author held a Walter C. Sumner Memorial Fellowship

References I

[1] 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," *arXiv preprint arXiv:2204.06655*, 2022.

[2] S. I. Alvis *et al.*, "Search for neutrinoless double-β decay in $^{76}\mathrm{Ge}$ with 26 kg yr of exposure from the Majorana Demonstrator," *Phys. Rev. C*, vol. 100, p. 025501, 2019.

[3] M. Agostini *et al.*, "Final results of GERDA on the search for neutrinoless double-β decay," *Phys. Rev. Lett.*, vol. 125, p. 252502, 2020.

[4] N. Abgrall *et al.*, "LEGEND-1000 preconceptual design report," *arXiv preprint arXiv:2107.11462*, 2021.

[5] C. E. Aalseth *et al.*, "CoGeNT: A search for low-mass dark matter using p-type point contact germanium detectors," *Phys. Rev. D*, vol. 88, no. 1, p. 012002, 2013.

[6] M. Agostini *et al.*, "First search for bosonic superweakly interacting massive particles with masses up to 1 MeV/c^2 with GERDA," *Phys. Rev. Lett.*, vol. 125, no. 1, p. 011801, 2020.

[7] N. Abgrall *et al.*, "Search for Pauli Exclusion Principle violating atomic transitions and electron decay with a p-type point contact germanium detector," *Eur. Phys. J. C.*, vol. 76, no. 11, pp. 1–5, 2016.

References II

[8] S. I. Alvis *et al.*, "First limit on the direct detection of lightly ionizing particles for electric charge as low as $e/1000$ with the Majorana Demonstrator," *Phys. Rev. Lett.*, vol. 120, no. 21, p. 211804, 2018.

[9] S. I. Alvis *et al.*, "Search for trinucleon decay in the Majorana Demonstrator," *Phys. Rev. D*, vol. 99, no. 7, p. 072004, 2019.

[10] "Underground neutrino experiment could provide greater clarity on matter-antimatter imbalance." `https://newscenter.lbl.gov/2018/03/26/neutrino-experiment-probes-matter-antimatter-imbalance/`. Accessed: 2022-05-18.

[11] N. Abgrall *et al.*, "New limits on bosonic dark matter, solar axions, Pauli Exclusion Principle violation, and electron decay from the Majorana Demonstrator," *Phys. Rev. Lett.*, vol. 118, no. 16, p. 161801, 2017.

[12] E. Aguayo *et al.*, "Characteristics of signals originating near the lithium-diffused n+ contact of high purity germanium p-type point contact detectors," *Nucl. Instrum. Meth. A*, vol. 701, pp. 176–185, 2013.

[13] C. Wiseman, "A low energy rare event search with the Majorana Demonstrator," in *J. Phys. Conf. Ser.*, vol. 1468, p. 012040, IOP Publishing, 2020.

[14] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE T. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

References III

[15] "Building autoencoders in keras." https://blog.keras.io/building-autoencoders-in-keras.html. Accessed: 2022-05-18.

[16] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[17] "P-type point-contact (PPC) germanium detectors." https://www.npl.washington.edu/majorana/design-technologies. Accessed: 2022-05-24.

[18] Vasundhara, "Pulse fitting for event localization in high purity germanium point contact detectors," Master's thesis, Queen's University, 2020.

[19] D. C. Radford, "siggen." https://github.com/radforddc/icpc_siggen, 2017.

[20] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," in *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, pp. 2965–2974, 2018.

References IV

[21] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[22] "1D convolution." https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/blocks/1d-convolution.
Accessed: 2022-05-13.

[23] "Med Phys 4RA3, 4RB3/6R03 - Chapter 6: Pulse Processing." https://www.science.mcmaster.ca/radgrad/images/6R06CourseResources/4RA34RB3_Lecture_Note_6_Pulse-Processing.pdf.
Accessed: 2022-05-19.

Additional Slides

Model architecture

Layer	Stride	Window	Output
Input			4096, 1
Convolution	1	1	4096, 8
Convolution	1	9	4088, 16
Average Pooling	2	2	2044, 16
Convolution	1	17	2028, 32
Average Pooling	2	2	1014, 32
Convolution	1	33	982, 64
Average Pooling	2	2	491, 64
Convolution	1	33	459, 32
Transpose Convolution	1	33	491, 32
Upsampling	2	2	982, 64
Transpose Convolution	1	33	1014, 64
Upsampling	2	2	2028, 64
Transpose Convolution	1	17	2044, 32
Upsampling	2	2	4088, 32
Transpose Convolution	1	9	4096, 16
Convolution (output)	1	1	4096, 1

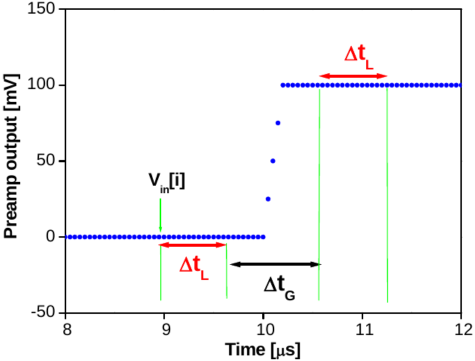
[1]

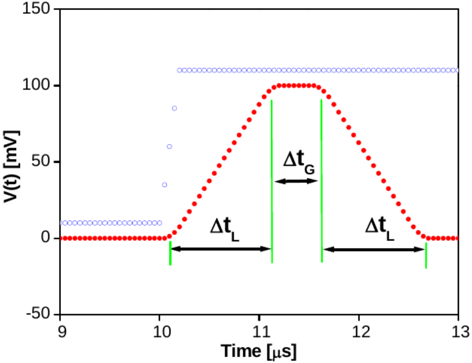
Results on simulations

Training procedure and data			Mean squared error ($\times 10^{-5}$)			
			Gaussian noise		Detector noise	
Procedure	**Data**	**Noise**	**Lib**	**Fake**	**Lib**	**Fake**
Regular	Library	Detector	4.37	4.94	3.88	4.26
Regular	Library	Gaussian	3.41	3.85	4.49	4.80
Regular	Fake	Detector	5.43	4.55	4.57	3.83
Regular	Fake	Gaussian	3.85	3.35	4.92	4.29
N2N ($\lambda = 0$)	Library	Detector	4.13	4.65	4.03	4.41
N2N ($\lambda = 0$)	Library	Gaussian	3.45	3.85	4.52	4.79
N2N ($\lambda = 0$)	Fake	Detector	4.87	4.08	4.62	3.87
N2N ($\lambda = 0$)	Fake	Gaussian	3.84	3.39	4.92	4.33
N2N ($\lambda = 0$)	Detector	Detector	9.28	9.10	10.02	9.97
N2N ($\lambda = 2 \cdot 10^{-2}$)	Detector	Detector	5.81	6.05	6.88	6.99

[1]

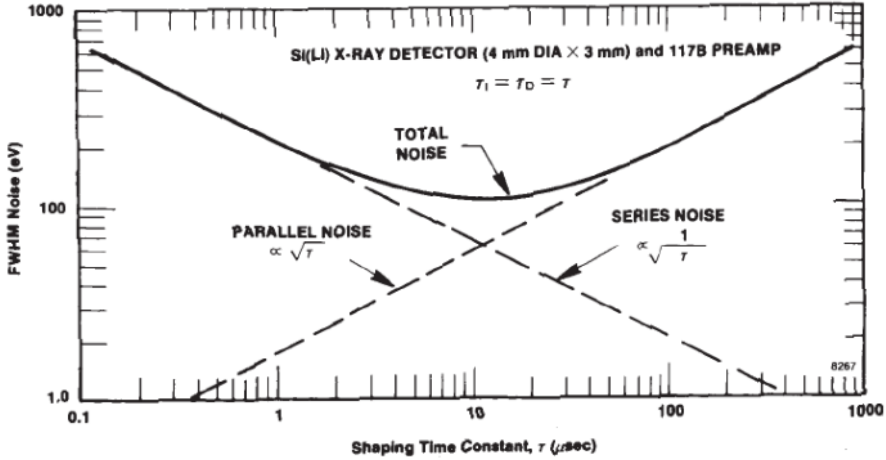
Trapezoidal filter





[23]

Noise curve



[23]