



Performance of a convolutional autoencoder designed to remove electronic noise from p-type point contact germanium detector signals

Mark R. Anderson^{1,a}, Vasundhara Basu¹, Ryan D. Martin¹, Charlotte Z. Reed¹, Noah J. Rowe¹, Mehdi Shafiee^{1,2,3}, Tianai Ye¹

¹ Department of Physics, Engineering Physics and Astronomy, Queen's University, Kingston, ON, Canada

² Present Address: Department of Electrical and Computer Engineering, Nazarbayev University, Nur-Sultan, Kazakhstan

³ Present Address: Energetic Cosmos Laboratory, Nazarbayev University, Nur-Sultan, Kazakhstan

Received: 14 April 2022 / Accepted: 4 November 2022 / Published online: 1 December 2022
© The Author(s) 2022

Abstract We present a convolutional autoencoder to denoise pulses from a p-type point contact high-purity germanium detector similar to those used in several rare event searches. While we focus on training procedures that rely on detailed detector physics simulations, we also present implementations requiring only noisy detector pulses to train the model. We validate our autoencoder on both simulated data and calibration data from an ^{241}Am source, the latter of which is used to show that the denoised pulses are statistically compatible with data pulses. We demonstrate that our denoising method is able to preserve the underlying shapes of the pulses well, offering improvement over traditional denoising methods. We also show that the shaping time used to calculate energy with a trapezoidal filter can be significantly reduced while maintaining a comparable energy resolution. Under certain circumstances, our denoising method can improve the overall energy resolution. The methods we developed to remove electronic noise are straightforward to extend to other detector technologies. Furthermore, the latent representation from the encoder is also of use in quantifying shape-based characteristics of the signals. Our work has great potential to be used in particle physics experiments and beyond.

1 Introduction

High-purity germanium (HPGe) detectors are used in the search for rare events such as neutrinoless double-beta decay ($0\nu\beta\beta$) [1–3], dark matter [4,5], and other beyond Standard Model physics [6–8]. HPGe detectors are a common choice due to their intrinsic purity, good energy resolutions,

and ability to be fabricated from material enriched in the $0\nu\beta\beta$ candidate isotope ^{76}Ge . Some of the strongest limits on the $0\nu\beta\beta$ half-life have been set using HPGe detectors and are in the range of $10^{25}\text{ yr} - 10^{26}\text{ yr}$ [1,2].

Due to the infrequent occurrence of signal events, extraordinary measures are taken to reduce backgrounds associated with particles that can deposit energy in the detectors, including locating the experiments in ultra-clean laboratories deep underground [9]. Further analytical background reduction and discrimination techniques are still required as experiments probe larger regions of the parameter space associated with the processes that they are trying to detect.

An efficient denoising algorithm can help advance searches for rare event interactions. Noise reduction techniques can allow one to identify low-energy signal events that would otherwise be dominated by electronic noise. This is of direct relevance to experiments using germanium detectors to search for rare events at low energies, such as solar axions, violation of the Pauli Exclusion Principle, and electron decay [10]. A reduction in noise would also allow for better background rejection techniques that are based on pulse shapes, such as the rejection of slow energy-degraded pulses in germanium detectors searching for signals at low energies [11,12].

Denoising could also provide more accurate measurements of pulse amplitudes, leading to a better energy resolution. Many pulse height estimation algorithms [13,14] use effective averaging windows with a given shaping time. While an overall reduction in the energy resolution is difficult to achieve compared to these highly efficient algorithms, denoising the pulses beforehand can reduce the shaping time required to obtain a comparable energy resolution. This can allow for shorter traces to be collected (more efficient data

^a e-mail: anderson.mark@queensu.ca (corresponding author)

storage), a smaller sampling period to be used (more detailed pulses), and/or a higher data collection rate (lower energy thresholds).

In this paper, we demonstrate the effectiveness of deep learning to strongly reduce electronic noise in the charge pulses from a p-type point contact (PPC) HPGe detector. We conduct several studies on both real and simulated data to verify the performance of our deep learning-based model and compare it to traditional noise reduction techniques. We also show that an effective noise reduction model can be trained using *only noisy detector data*. Since this method does not require underlying clean pulses for training, our approach can be implemented *without* detailed modelling and simulations of the detector. Furthermore, our results are not limited to PPC HPGe detectors; this efficient and flexible denoiser which requires only detector data is widely applicable to the particle astrophysics community and beyond.

Section 2 provides an overview of both traditional denoising methods and autoencoders. In Sect. 3, we describe the design of our neural network architecture. In Sect. 4, we outline the experimental setup and datasets used for training and evaluation, while in Sect. 5 we describe our methodology. The results are presented in Sect. 6, first on simulations to demonstrate both pulse shape preservation and improvements over traditional methods, and then on real detector data to further validate pulse shape preservation and to show the impact of denoising on the energy resolution. We conclude our results and emphasize the potential of our work for other experiments in Sect. 7.

2 Autoencoders for denoising

2.1 Traditional denoising methods

We briefly discuss non deep learning-based approaches to denoise signals before introducing autoencoders. We focus on a subset of techniques that were evaluated for comparison with the autoencoder in Sect. 6.1.2. The first one is a moving average filter over w samples. The selection of w requires a trade-off between the level of noise reduction and the preservation of details such as edge sharpness in the pulse shape. An extension to the moving average is the exponential filter [15], where the smoothed output of a given sample is the weighted sum of the current sample and the previous prediction. The algorithm is recursive and is defined by the smoothing constant $\alpha \in (0, 1)$.

We also consider the more advanced Savitzky–Golay filter [16], which removes noise by fitting a degree- p polynomial to w adjacent samples centred about a given point of the signal and evaluating the fitted polynomial at this point. The Savitzky–Golay filter can be implemented as a weighted

Table 1 Wavelet functions used for denoising in this work, including the mother wavelet and order (where applicable)

Mother wavelet	Order
Haar	N/A
Daubechies	2–38
Coiflet	1–17
Symlet	2–19
Biorthogonal	1.1, 1.3, 1.5, 2.2, 2.6, 2.8, 3.1, 3.3, 3.5, 3.7, 3.9, 4.4, 5.9, 6.8
Reverse Biorthogonal	1.1, 1.3, 1.5, 2.2, 2.6, 2.8, 3.1, 3.3, 3.5, 3.7, 3.9, 4.4, 5.9, 6.8
Meyer	N/A (finite impulse response approximation)

moving average with coefficients that depend on w and p [16].

Another method we investigated applies thresholding rules to the wavelet decomposition of the noisy signal [17]. Wavelet-based denoising requires a choice of mother wavelet and order as well as thresholds for the wavelet coefficients [18]. Table 1 lists the set of mother wavelets that were explored. Additionally, we tested two methods of thresholding: VisuShrink [19], which applies a global threshold to the wavelet coefficients, and BayesShrink [20], which determines thresholds at each subband of the wavelet by minimizing Bayesian squared error risk. For each method, we considered hard thresholding (keeping the coefficient if greater than the threshold) and soft thresholding (shrinking the coefficient toward zero by the threshold).

The last technique examined in this work is the Kalman filter [21], which relies on knowledge of an underlying model of the data. Kalman filtering provides an estimate of the true value based on a model of the event itself and the measured samples. By knowing the accuracy of the event modeling technique as well as the measurement error, an interpolation can be made between the two.

2.2 Autoencoders

An autoencoder is an unsupervised machine learning algorithm used to encode data by learning from the data. The idea of the autoencoder was introduced in 1986 to learn an efficient coding for orthogonal inputs [22]. The concept was further explored in the mid to late 1980s; for example, coupled hierarchical autoencoders were demonstrated to converge much faster on encoding a representation than a typical multi-hidden layer feedforward network at the time [23]. Today, autoencoders are widely used for dimensionality reduction, anomaly detection [24], and generative modelling [25]. Autoencoders are also still frequently used for obtaining useful features or representations for other tasks. In [26], for

instance, an autoencoder was trained to learn a compressed representation of pulses from germanium detectors, and the encoded input was then used to train another network for pulse shape discrimination.

An autoencoder has two main components: the encoder is a function, f , that produces a latent representation of the input, while the decoder is another function, g , which takes the latent representation and generates a reconstruction of the original input. Using x as the input, y as the internal representation, and z as the reconstructed output, the encoder transforms x to y by $f_\theta(x) = y$ and the decoder transforms y to z by $g_{\theta'}(y) = z$. The optimal parameters θ and θ' are selected by minimizing the reconstruction error L between x and z . Typically, the encoder and decoder are connected and trained as one model. In almost all references to autoencoders in the literature, both f_θ and $g_{\theta'}$ are neural networks, and further discussion herein will assume this. For a comprehensive review of neural networks and deep learning, including the core concepts of advanced architectures such as convolutional networks, see [27] and [28].

2.3 Denoising autoencoders

Standard autoencoders can remove some of the input noise if the latent representation is highly compressed, as the network will be forced to extract only the most useful features. For example, in [26], it was observed that the autoencoder removed some of the noise from the inputs, despite not being trained to do so. Denoising autoencoders take this a step further and make the objective to remove noise explicit. With denoising autoencoders, the input is instead an artificially corrupted version of the clean signal, which we denote by \tilde{x} . The encoder is thus a mapping $f_\theta(\tilde{x})$ of the noisy signal to the latent representation, and the decoder a mapping $g_{\theta'}(f_\theta(\tilde{x}))$ of the latent representation to a reconstruction of the clean signal. As the output is the same, the reconstruction error between x and z , $L(x, z)$, is still minimized to train the network. The basic concept of the denoising autoencoder is illustrated in Fig. 1.

Denoising autoencoders were originally proposed to extract robust features from the inputs [29,30]. The primary goal was not to remove noise. Rather, denoising was used as a criterion to produce encoded representations which performed better on a variety of classification tasks.

Autoencoders specifically for denoising purposes have been successful in a variety of domains, although much of the existing research focuses on denoising images rather than one-dimensional signals. In [31], convolutional networks are used to denoise images with both known and unknown Gaussian noise levels, achieving comparable or superior performance to state-of-the-art methods at the time. In [32], autoencoders are used for both denoising and removal of superim-

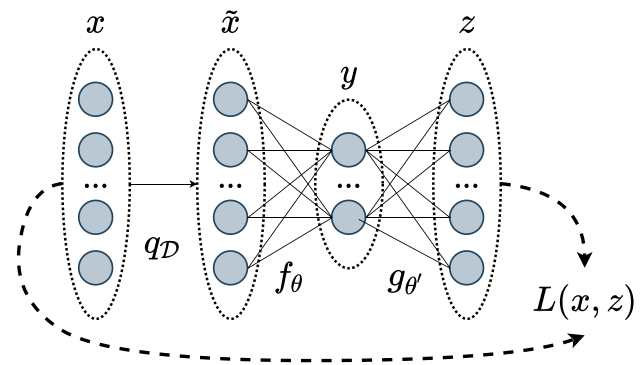


Fig. 1 Basic concept of denoising autoencoders. An input x is artificially corrupted by some noise process q_D to become \tilde{x} . The encoding portion of the autoencoder, f_θ , produces a new representation y . The decoder portion of the network, $g_{\theta'}$, attempts to reconstruct the clean input. Its estimate is given by z , and a loss function, $L(x, z)$, quantifies the reconstruction. Notation is largely based off of figure from [29]

posed text on images, although a separate network is trained for each noise level.

More recent works take advantage of modern advances in deep learning to train convolutional networks to remove Gaussian noise, as well as other tasks [33]. Convolutional networks with noise estimation subnetworks have also been applied to more complex noise models [34]. Furthermore, denoising convolutional autoencoders have now been demonstrated to perform well on several practical applications in various fields, such as denoising medical images [35], “repairing” images of corrupted printed circuit boards [36], and speech enhancement [37].

2.4 Noise2Noise

The Noise2Noise procedure is an approach used to train a denoising model without the need for clean target examples [38], which, in our case, require detailed physics simulations. Although we outline the method here, we refer the reader to the original paper for details. The central idea is that the mean of the target distribution minimizes the sum of squared errors, or L_2 loss, between the target and the prediction. In the simple case of point estimation, the L_2 loss for a set of measurements $\{x_i\}$ and prediction z is given by

$$L_2 = \sum_i \|z - x_i\|_2^2. \tag{1}$$

The best estimate z that minimizes the L_2 loss is simply the mean, $z = \sum_i x_i$. One can observe that z is unchanged as long as the mean of the measurements is unchanged. Given infinite data, the addition of zero-mean noise would produce the same estimate z . The same logic can be applied to the denoising task of minimizing the L_2 loss between the clean signals, x_i , and the output of the autoencoder,

$$z_i = g_{\theta'}(f_{\theta}(\tilde{x}_i)):$$

$$L_2 = \sum_i \|g_{\theta'}(f_{\theta}(\tilde{x}_i)) - x_i\|_2^2. \quad (2)$$

If instead the target x_i is replaced with corrupted versions of itself, \hat{x}_i , with samples drawn from a distribution or distributions with a mean of x_i , then, the optimal estimate z_i remains unchanged. Put another way, the corruption of x_i with zero-mean noise will produce the same estimate, given enough data. The loss function becomes

$$L_2 = \sum_i \|g_{\theta'}(f_{\theta}(\tilde{x}_i)) - \hat{x}_i\|_2^2, \quad (3)$$

where \hat{x}_i is a different noisy realization of the same underlying signal. By minimizing Eq. (3), the autoencoder should learn to predict the mean of the distribution of noisy pulses. Intuitively, the task of mapping one version of a noisy pulse to another is impossible if the model is not over-parameterized, and the best it can do is predict the mean. With finite data, the above results are only approximately true. However, we show in this paper that the Noise2Noise approach, with some additions, can produce excellent results in practice.

3 Model

3.1 Inputs and outputs

Signals are collected from a PPC HPGe detector described in Sect. 4.1. Each signal is a pulse that consists of M voltages sampled at 8 ns intervals. M can technically be any number, but is typically either 4096 or 8192 samples in our work. Thus, the inputs and outputs of the models are M -long vectors of voltage samples.

All data pulses are preprocessed to have unit amplitude using a trapezoidal filter [13], as described in Sect. 5.2. An example of three preprocessed detector pulses with different rise times – indicating different event positions in the detector – and different signal-to-noise ratios is shown in Fig. 2. The full length pulses are plotted and the rise region between approximately 16 μ s and 17 μ s is highlighted in the inset.

3.2 Network architecture

3.2.1 Motivation

Our architecture is fully convolutional and thus benefits from superior computational and memory efficiency over fully connected networks [28]. As well, the convolutional nature of our model allows for a variable length input pulse, regardless of the size of the pulses it was trained on.

The parameter sharing aspect of convolutional networks is particularly important to our work as it forces the network

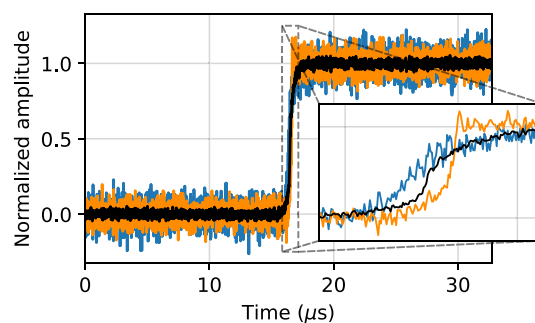


Fig. 2 Three example pulses from a germanium detector, each with a substantially different rise time. All pulses are preprocessed such that they have a baseline of zero and an amplitude of one. The black pulse has apparently less noise because it was originally a higher amplitude pulse, so that the electronic noise is smaller relative to its amplitude

to learn to remove noise in a consistent manner across the entire pulse, emphasizing feature locality. It also ensures that the network is equivariant to shifts on the time axis and thus independent of the relative position of the pulse in the trigger window. Finally, weight sharing greatly reduces the number of parameters to train, effectively acting as a regularizer to prevent overfitting.

3.2.2 Design

For a general one-dimensional convolution, the size of the output O is given by

$$O = \frac{W - K + 2P}{S} + 1, \quad (4)$$

where W is the input size, K is the filter or window size, S is the stride length, and P is the padding. For a stride length or window size greater than one, the size of the output of the convolution may be different from the input, which has consequences on the selection of layer parameters. While the output size can be forced by padding the input, it is not apparent how to pad the pulses. Prepending or appending a constant value does not account for misalignment due to the imperfect normalization. Padding also ignores the noise.

An alternative approach is to use transposed convolution layers in the decoder to “undo” the size changes caused by non-padded convolutions. The transposed convolution operation uses the transpose of the convolution matrix – a sparse matrix containing elements of the filter for computing the convolution using matrix multiplication – to switch the forward pass with the backward pass [39]. This effectively acts as a form of upsampling.

The architecture of our autoencoder is described in Table 2, with each layer consisting of its stride, window, and output size. The first element of the output is the temporal length after the operation is applied, while the second element is the number of filters. The batch size is not included in the output

Table 2 Summary of the convolutional autoencoder architecture used in this paper for a fixed pulse size of 4096. Included in this table is the type, strides length S , window size K , and output shape O of each layer

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
Trans. convolution	1	33	491, 32
Upsampling	2	2	982, 64
Trans. convolution	1	33	1014, 64
Upsampling	2	2	2028, 64
Trans. convolution	1	17	2044, 32
Upsampling	2	2	4088, 32
Trans. convolution	1	9	4096, 16
Convolution (output)	1	1	4096, 1
Total number of parameters: 286,145			

shape. For illustration, the table uses a fixed input length of 4096 samples. A different input shape will only change the temporal dimension of the output shape as the number of filters does not depend on the input.

The network begins with an eight-filter convolution operation with a stride length and window size of one. This does not change the temporal dimension of the input, but rather acts as a sort of “preprocessing” layer that increases the overall dimensionality of the signal. The decoder also ends with a convolution layer where the stride length, window size, and number of filters are set to unity to recover the original shape of the pulse.

Both the encoder and decoder have three blocks of layers. A “block” consists of a convolution followed by a two-fold average pooling operation or a two-fold upsampling operation followed by a transpose convolution for the encoder and decoder respectively. The encoder compresses the original input by a factor of approximately eight, due to both the downsampling operation and Eq. (4). Every convolution and transpose convolution layer *except* for the last is followed by a rectified linear unit (ReLU) activation function [40]. No activation is applied to the final convolution layer as pulses can have values outside of the range (0, 1) or (−1, 1), and bounding the output was found not to be necessary.

The window sizes are chosen to *increase* for each layer in the encoder and *decrease* for each layer in the decoder. The

window sizes were all chosen to be smaller or approximately the same size as a typical rise time. However, they are also subject to some restrictions. This can be seen from Eq. (4), which simplifies to

$$O = W - K + 1 \quad (5)$$

for a stride length of one and no padding. If W is even (odd), K must be odd (even) in order for O to always be even. O must be even to ensure that the size of the input to the subsequent downsampling layer is divisible by two. The window size in each convolution layer, K , is selected such that this condition is true for each layer in the network starting with an input of 4096 samples. Conversely, pulses of variable length are subject to this requirement for the chosen window sizes.

4 Experimental setup, data, and simulations

4.1 Detector

Detector pulses were collected from a 1 kg PPC HPGe detector manufactured by ORTEC/AMTEK [41]. The detector is a cylinder with a radius of approximately 3 cm and 5 cm in height. The detector was operated in a PopTop cryostat, and preamplifier signals were recorded using a 16-bit 125 MHz SIS3316 digitizer fabricated by Struck Innovative Systems [42]. The detector depletion voltage is 2750 V and was operated at a bias of 3700 V, as recommended by the manufacturer. Data were stored and preprocessed with a custom C++ analysis suite that is based on CERN’s ROOT software [43].

4.2 Real data

4.2.1 ^{241}Am source

Data were collected from a $10\mu\text{Ci } ^{241}\text{Am}$ encapsulated source, which produces 60 keV gamma rays. At this energy, electronic noise is a significant component of the pulse. By using a low energy collimated gamma source, the location of the collimator can be used to infer the location of energy depositions since Compton scattering is unlikely and the gamma rays do not typically penetrate more than 1 mm in the detector. Furthermore, the interactions of these gamma rays in the detector are almost entirely single-site. Each trace is 8192 samples in length. Data were collected in December 2021.

4.2.2 ^{60}Co source

Data were also collected from a ^{60}Co source that produces gamma rays with energies of 1173 keV and 1332 keV. In practice, data were collected over energies ranging from approximately 500 keV to 3 MeV, including events from

room backgrounds and the ^{60}Co source. These data include many multi-site events from gamma rays Compton scattering in the detector. Each trace is 4096 samples in length. Data were collected in June 2020 and October 2020.

4.2.3 Noise

A large number of noise traces were collected from the detector in order to train the models with realistic electronic noise. These were obtained by randomly triggering the SIS3316 digitizer to read out signals from the PPC detector. Noise data were collected at three different times over a period of two and a half years: July 2019, January 2021, and December 2021. At each time point, the detector was under different operating conditions, resulting in a diverse noise set. Traces collected in December 2021 are 8192 samples in length, while the remaining sets are 4096 samples in length.

Signals in the noise data were filtered to remove any events that occasionally occur in the same trigger window. This was done by rejecting outliers based on the baseline and root mean square (RMS) of the baseline, calculated over the first 1000 samples of each trace. Similarly, pulses with outliers in the minimal or maximal values from a trapezoidal filter with a gap window of $1.8\ \mu\text{s}$ and a rise window of $6\ \mu\text{s}$ were removed from this dataset.

4.3 Simulated data

4.3.1 Library pulses

A detailed physics-based simulation was used to create a set of 1724 “library” pulses on a $1\ \text{mm} \times 1\ \text{mm}$ grid in radius and height to represent pulses uniformly in the azimuthally symmetric PPC detector [44]. This position-dependent basis set was created using the *siggen* simulation software which models the propagation of charges in germanium detectors [45]. The library pulses can be used to infer realistic clean signals underlying actual events in the detector using a χ^2 minimization between a normalized data pulse and every library pulse in the basis set [44],

$$\chi^2 = \sum_j^M \frac{(z_{k,j} - x_{i,j})^2}{\sigma_i^2}, \quad (6)$$

where $x_{i,j}$ is the j th sample of the i th data pulse, $z_{k,j}$ is the j th sample of the k th library pulse in the basis set, and the quantity is evaluated over M samples. σ_i represents the noise level in the data pulse, which we take as the RMS of the noise in the pre-trigger baseline.

Multiple sets of library pulses were generated with different preamplifier time constants to further expand this dataset. For a given position, this affects the rise time and the curvature of the pulse. Using preamplifier time constants of 0–80 ns

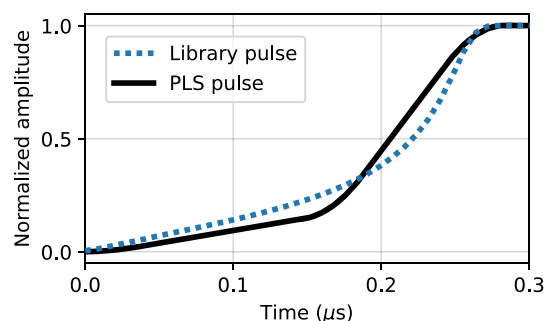


Fig. 3 An example of a simulated PLS pulse (solid line) and library pulse (dotted line), each with a rise time of 280 ns

in increments of 10 ns, the library pulse dataset was expanded to 15,516 unique traces.

4.3.2 Piecewise linear smoothed pulses

A set of pulses that look similar to the library pulses was generated by using piecewise mathematical functions. These pulses mimic the shape of the library pulses without requiring any complex physics simulation, and do not depend on the details of a specific detector. Each pulse was composed of two linear pieces connected at a varying fixed point to mimic the slow and fast portions of the rise. All pulses have an amplitude of unity and a rise time between 25 samples and 125 samples. The pulses were smoothed using a moving average with a window size of 5% of the rise time. A total of 20,070 unique traces were created using this procedure. We refer to these signals as piecewise linear smoothed (PLS) pulses.

Figure 3 shows an example library pulse and PLS pulse, each with a short rise time of 280 ns. The PLS pulses tend to be noticeably sharper than the library pulses and only roughly approximate their general shape.

5 Methodology

5.1 Datasets

The library and PLS simulated pulses described in Sect. 4.3 were used as distinct datasets to train and evaluate the performance of the model. Pulses in these sets were padded to have 4096 samples, which was found to be sufficient for the network to denoise the flat regions of the pulses well without distorting the important components of the pulses.

Data from the ^{241}Am source were used to validate the denoising algorithm by comparing denoised pulses to their known shapes from simulations. Since data from the ^{60}Co source include a variety of high-energy, multi-site events, they provided an ideal set of diverse pulses for training the autoencoder with the Noise2Noise method.

5.2 Data preprocessing

The baseline, defined as the mean over the first 1000 samples of a pulse, was calculated and subtracted from each signal. The pulse amplitudes were calculated by applying a trapezoidal filter to each trace [13], and the baseline-removed pulses were then scaled by their amplitudes. Additionally, a pole-zero correction was applied to each pulse to remove the main component of the exponential decay from the resistive feedback preamplifier that was used to read out the detector. We refer to this entire preprocessing procedure as “amplitude normalization”, noting that it includes the pole zero correction for data pulses (the simulated pulses do not have an exponential decay). Because of the noise, the amplitude normalized pulses roughly, but not exactly, range in height from zero to one.

We also explored standardizing each pulse to have a mean of zero and standard deviation of 0.5 after the pole-zero correction. A value of 0.5 was used to ensure that horizontally centred pulses have an amplitude of roughly one. We refer to this method of preprocessing as “standardization”, noting again that it includes the pole zero correction for data pulses. However, models trained with standardized pulses were found to perform similarly or even slightly worse in most circumstances. Additionally, these models were found to depend heavily on the absolute position in the pulse where the rise region begins. Due to the lack of robustness to horizontal shifts, all models described in the results section can be assumed to have been trained with amplitude normalized pulses unless mentioned otherwise.

5.3 Data augmentation

5.3.1 Multi-site event generation

The set of library pulses form a basis of position-dependent signals in the detector and thus contain only single-site events. Similarly, the PLS pulses are simple mathematical functions constrained by two points and are analogous to single-site events in shape. To augment the training data, artificial multi-site events are created by adding randomized combinations of simulated pulses together, without mixing between the library and PLS sets.

Events are generated with up to five sites. This upper bound is chosen because for a Poisson process with an expected rate of two, the probability of a number greater than five is less than 2%. Furthermore, an equal number of events are generated for each number of sites, rather than being based on a physical distribution for the number of Compton scatters. The number of multi-site events is thus four times larger than the number of single-site events. We found this to be optimal for ensuring that the network does

not smooth over multi-site events while also preserving the shape of single-site events.

In order to generate an n -site event, n pulses are drawn from a set of simulated pulses. A random amplitude is drawn from a uniform distribution, $\sim U(0, 1)$, for each pulse. Each pulse is also horizontally shifted by a value drawn from a discrete uniform distribution, $\sim U(0, 100)$, to account for the possible drift times from different points in the detector. The scaled and shifted pulses are then added together to create an artificial multi-site event, and the result is rescaled to have an amplitude of one.

5.3.2 Shifting and scaling

The process of amplitude normalization on detector signals is imperfect due to noise. Simulated pulses, in contrast, are noiseless and have perfect normalization. Preliminary results showed that the network would overfit the beginning and the end of the predicted pulses if the clean pulses started at exactly zero and ended at exactly one. To combat this issue, random shifting and scaling is applied to all simulated pulses during training. This ensures that the network sees a variety of imperfect normalizations and is able to better generalize to real detector data while avoiding overfitting. This type of artificial data augmentation can also help improve the overall performance of the network by effectively providing more data than is available. In order of application, the shifting and scaling procedure consists of:

1. **Amplitude scaling:** Each pulse is rescaled to have an amplitude drawn from the uniform distribution $\sim U(0.9, 1.1)$.
2. **Vertical shifts:** A random vertical shift, drawn from a uniform distribution $\sim U(-0.1, 0.1)$, is applied to each pulse.
3. **Horizontal shifts:** For each pulse of length 4096 samples, a number is drawn from a discrete uniform distribution $\sim U(1000, 3000)$, and the pulse is shifted such that the rise begins at this randomly chosen sample.

While vertical scaling and shifting of $\pm 10\%$ of the pulse amplitude is proportionally much larger than observed with typical data pulses, we use such a wide range of random variations to further improve generalization of the autoencoder, particularly to high noise pulses and outliers such as pile-up events. As well, horizontal shifts only have a major effect near the edges of an event due to the receptive field of the convolution layers. Thus, the unrealistic range of values used in this procedure is primarily for data augmentation purposes. We also verified that shifting and scaling parameters can be selected differently to favour performance on outliers.

5.3.3 Noise addition

Noise collected from the detector is added to the clean pulses after the previous data augmentation steps are applied. The detector noise dataset is sufficiently large such that each clean pulse has a unique, randomly chosen noise trace. Thus, noise seen in training is not seen in the validation and test phases, and noise pulses are not shared between datasets. All detector noise pulses are standardized to have a mean of zero and a standard deviation of σ .

In order to understand the effect of real detector noise, zero-mean normally distributed noise with no covariance over time, $\sim \mathcal{N}(0, \sigma)$, is also used, separately, to create independent datasets for comparison. For both noise types, σ is drawn randomly for each trace from a uniform distribution, $\sim U(0, 0.2)$, to simulate the effect of varying the signal-to-noise ratio.

5.4 Training procedures

Two training procedures are explored in this paper: a regular training procedure which requires a noisy input and clean output, and a procedure which does not require the clean version of the noisy pulse. The latter procedure uses and extends the Noise2Noise approach developed in [38] and summarized in Sect. 2.4.

The networks are implemented using the Keras API from TensorFlow [46,47]. Training, validation, and testing were done using NVIDIA GeForce GTX Titan X and GeForce RTX 3090 GPUs.

5.4.1 Regular training

The regular training procedure consists of applying the augmentation recipe above to a set of simulated data d times, resulting in $N = d \times 5N_0$ pulses from an initial set of N_0 clean, single-site pulses. This is done separately with library and PLS pulses to understand whether detailed physics simulations are required for training the algorithm. The augmentation procedure is also applied twice per simulated set using Gaussian noise and real noise from the detector, for a total of four distinct datasets. We set $d = 45$ for the library pulses and $d = 35$ for the PLS pulses to ensure that the sizes of each training set are comparable. A copy of the pulses after the shifts and scales, but before the noise addition, is created and used for the clean targets.

The network is then trained to map the noisy pulses to the corresponding clean pulses. We use the Adam algorithm – a stochastic gradient-based adaptive optimization procedure [48] – with a learning rate of $3 \cdot 10^{-4}$ to minimize the mean squared error between the true clean pulse and the denoised pulse. We use a batch size of 128, which is large enough to ensure that the network trains reasonably quickly while not

being so large as to affect the convergence of the network. Training is conducted over 100 epochs.

For each of the augmented simulated datasets, 10% of the data are withheld for testing. Of the remaining 90%, 10% of that are withheld for validation and the remainder are used for training. At every epoch of training, the network is run on the validation set and the model at the epoch with the lowest loss is saved. The validation set is also used to select the best hyperparameters, including the batch size, learning rate, and architecture details. The best model – as determined by the validation set – is then run on the reserved test set. All results shown in this paper are calculated only on the test set.

5.4.2 Noise2Noise training

Unlike the regular training procedure, the Noise2Noise approach can be applied to *both* simulated and raw detector data. The procedure follows the regular training procedure with the amendment of adding noise to the target pulse as well. For real detector data, where no clean underlying pulse exists, we add noise to the already noisy signals. Of course, the optimal solution to the minimization problem is then the underlying noisy pulse rather than the true clean pulse. To alleviate this issue, we add a simple penalty on the total variation of the denoised signal to the loss function L ,

$$L = L_0 + \frac{\lambda}{N} \sum_i^N \sum_j^{M-1} |z_{i,j+1} - z_{i,j}|, \quad (7)$$

where L_0 is the original loss function (which in our case is the L_2 loss), $z_{i,j}$ is the j th sample of the i th denoised output pulse, M is the number of samples in the pulse, N is the size of the training set, and λ is a scaling factor. Minimization of the total variation, with some criterion of similarity to the original signal, was introduced as a method to denoise signals in 1992 [49]. Total variation is particularly useful in retaining the important components of a signal, including sharp edges and discontinuities, while avoiding the over-compensated smoothing present in many traditional denoising techniques [50]. Since the true pulses are monotonically increasing functions, the regularizer in Eq. (7) will evaluate to the amplitude in the case of perfect denoising. If the denoised pulse still contains components of the noise and is very jagged, the penalty will become large.

The augmentation procedure for the simulated datasets is the same, except that a different noise pulse is also added to the target pulse. In the augmentation procedure for detector data, the artificial multi-site event generation step is skipped as the ^{60}Co dataset used in training already contains multi-site events. As well, pulses are not horizontally shifted because the network is already equivariant to such shifts and because it is difficult to handle the bounds. However, the remainder of the procedure is still applied because

the ^{60}Co dataset consists of primarily high energy events, meaning that the shifts, scales, and noise are unrepresentative of lower energy pulses and outliers. For the noise addition, σ is instead drawn from the uniform distribution $\sim U(0.025, 0.25)$. These numbers are chosen so that at least some noise is added to the data pulses, and to minimize correlated noise between the inputs and the targets.

The network is then trained to map one version of a noisy pulse to another. We again use the Adam algorithm to minimize the loss in Eq. (7). We use the validation set of the augmented library pulses with detector noise to select the best model as the data pulses do not have a target pulse. We specifically choose the model which minimizes the mean squared error, not the total loss, as the total variation penalty tends to dominate the validation loss. A batch size of 128 is still used, although we set the learning rate to 10^{-5} as the learning rate in the regular training procedure was found to be too large.

6 Results

6.1 Evaluation on simulated data

6.1.1 Comparison between data augmentation and training procedures

We evaluate the performance of the denoising convolutional autoencoder trained with several different datasets for the two training procedures. For each model, we test its generalization on the four available simulated datasets. The results are shown in Table 3. The first three columns are related to the training method and show the procedure, the dataset, and the type of additional noise used in training, respectively. The remaining columns show the mean squared error evaluated by denoising the simulated test datasets, containing either Gaussian or real detector noise on single- and multi-site events generated using either the library or the PLS pulses, for which the clean target is known. All pulses in the test sets were amplitude normalized and horizontally centred before denoising and comparison.

Each model trained on simulated data tends to perform best on the same class of pulses that it was trained on. Models trained with Gaussian noise tend to generalize slightly worse to detector noise than the other way around. As well, models trained using library and PLS simulated pulses have similar performance. However, models trained using library pulses tend to generalize better to PLS pulses than the other way around. This is evidenced by the difference between the performance on the test set of library and PLS pulses for a given noise type; the gap is larger for the model trained on PLS pulses.

The Noise2Noise models trained on simulated pulses perform nearly identically on the test set as the corresponding regular models in most cases. Differences in the mean squared error are typically on the order of $10^{-7} - 10^{-6}$, which is expected due to statistical fluctuations in training. No total variation penalty was used as it was not found to improve the performance in the case of simulations. On simulated pulses with detector noise, neither Noise2Noise model trained on pulses from the ^{60}Co source outperforms any of the models trained with simulated data, regardless of the procedure. However, the Noise2Noise model is still effective at denoising and requires only noisy detector data. A larger set of detector data, including from other high-energy sources which produce multi-site events, could improve the performance of this model.

Furthermore, we evaluated the performance of the Noise2Noise model with and without a total variation penalty and note that unlike the models trained on simulated data, the total variation penalty has a non-negligible impact on the denoising performance. We selected a penalty of $\lambda = 10^{-2}$ for Table 3 as it was found to have the best mean squared error on the validation set, although using any penalty in the range $10^{-4} - 10^{-1}$ produced similar results. The mean squared error is approximately 30% lower by setting λ in this range.

Figure 4 shows an example library pulse multi-site event that has been denoised by two different versions of the autoencoder. The top panel shows the pulse denoised with the regular library pulse model while the bottom panel shows the denoised pulse using the Noise2Noise model trained with ^{60}Co data with a total variation penalty.

Qualitatively, the regular library pulse model appears to fit the true underlying pulse the best, in line with the results in Table 3. However, the differences are subtle, and the Noise2Noise model appears to remove most of the noise without much additional distortion, and can be suitable for many physics applications that do not require highly detailed shapes to be preserved. Furthermore, although not shown here, visualization of the denoised pulses illustrates the impact of including a total variation penalty for the Noise2Noise method, as it removes much of the noise still present in the pulse denoised without it.

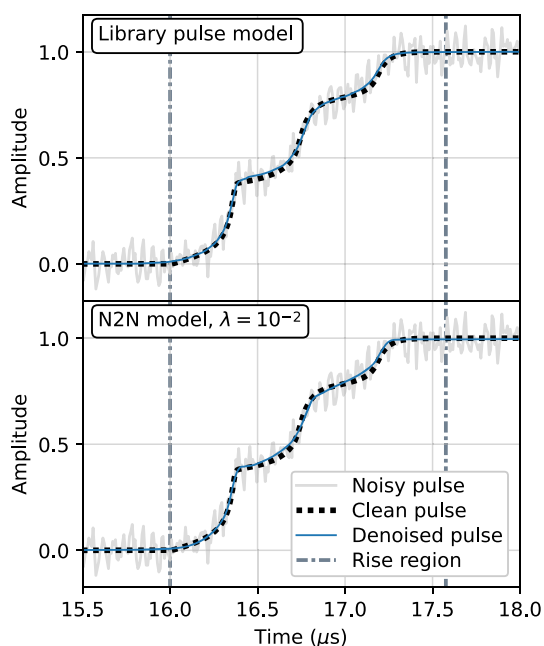
We emphasize that while all methods perform well, only the regular training procedure with library pulses required careful simulations of the detector.

6.1.2 Comparison to traditional denoising methods

The denoising performance of the autoencoder was compared to the traditional noise removal methods described in Sect. 2.1. The mean squared error between the clean target and denoised output was used as the metric and evaluated on the simulated library pulse test set. The performance was

Table 3 Summary of the mean squared error on the test sets for different models. The procedure, dataset, and noise type used in the training are given

Training procedure and data			Mean squared error ($\times 10^{-5}$)			
Procedure	Data	Noise	Gaussian noise		Detector noise	
			Library	PLS	Library	PLS
Regular	Library	Detector	4.12	4.72	3.76	4.21
Regular	Library	Gaussian	3.40	3.82	4.50	4.77
Regular	PLS	Detector	5.10	4.48	4.15	3.57
Regular	PLS	Gaussian	3.93	3.36	5.02	4.31
N2N ($\lambda = 0$)	Library	Detector	3.90	4.37	3.86	4.20
N2N ($\lambda = 0$)	Library	Gaussian	3.46	3.87	4.57	4.82
N2N ($\lambda = 0$)	PLS	Detector	5.11	4.48	4.14	3.55
N2N ($\lambda = 0$)	PLS	Gaussian	3.85	3.46	4.97	4.43
N2N ($\lambda = 0$)	Detector	Detector	6.54	6.30	7.78	7.40
N2N ($\lambda = 10^{-2}$)	Detector	Detector	4.17	4.54	5.04	5.26

**Fig. 4** An example multi-site event from the library dataset. Included in each plot is the simulated pulse with artificial noise (solid light line), the clean underlying pulse (dotted line), and the corresponding denoised pulse (solid dark line) from the regular library pulse model (top) and Noise2Noise model (bottom)

evaluated over the entire pulse as well as over two distinct sections of the pulse: the rise region and flat region. The rise region begins where the simulated pulse deviates from the flat baseline and ends where the pulse reaches its maximum amplitude. The flat region is defined as the area outside of the rise region. The example pulse in Fig. 4 includes an illustration of the region boundaries.

The mean squared error comparison between the traditional methods in Sect. 2.1 is shown in Fig. 5. The performance of two autoencoder models trained with the regular procedure (one using the library set and one using the PLS set) and one model trained with the Noise2Noise procedure (using the ^{60}Co data) are included. Figure 5 contains two variations of each traditional method: one using the optimal parameter(s) for the rise region and one using the optimal parameter(s) for the entire pulse, both optimized on the validation set by minimizing the mean squared error. For wavelet-based methods, we used VisuShrink to determine a global threshold and soft thresholding to shrink the coefficients towards zero.

In addition to the four methods in Fig. 5, a Kalman filter was implemented using the denoised output from the autoencoder as the underlying model and the baseline RMS of the noisy pulse as the measurement error. The filter was optimized by selecting the level of extrapolation between the noisy data and the underlying model that resulted in the least error. This process resulted in an optimized model with zero extrapolation, representing a copy of the autoencoder model, and so the Kalman filter was omitted from Fig. 5.

The autoencoder outperformed all traditional methods in both the rise region and flat region of the pulses. While the method requires training to denoise a specific type of data, it does offer improvements over traditional denoising, as evaluated using the mean squared error. The structural similarity index measure (SSIM) [51] was also used to compare the performance of the traditional denoising methods as it is designed to quantify image degradation with reference to human perception. However, the relative performance of each method using SSIM is nearly identical to that using the mean squared error, and so the corresponding SSIM comparison figure is not shown.

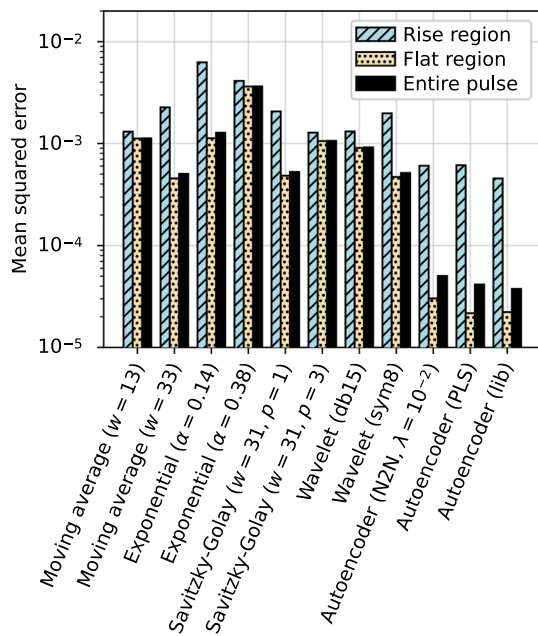


Fig. 5 Mean squared error comparison of different noise removal methods inside and outside the rise region of the pulse (defined in the text and shown in Fig. 4). The mean squared error inside the rise region is indicated by the slanted line hatch while outside the rise region is indicated by the dotted hatch. Solid fill corresponds to the mean squared error over the entire pulse

6.1.3 Energy resolution comparison

In this section, we use simulations to calculate the energy resolution before and after denoising using the amplitude normalized library pulse model. For a given pulse, the energy is calculated from the amplitude of a trapezoidal filter with a given gap and shaping time. The energy resolution is then defined as the full width at half maximum (FWHM) of the energy peak.

We use the set of library pulses and real detector noise, distinct from the training set, to create new sets of 172,400 noisy single-site event pulses for evaluation. A dataset is created for noise levels ranging from 0.02 to 0.2 in increments of 0.02. The energy resolution as a function of the trapezoidal filter shaping time on one such simulated dataset with a noise level of 0.1 is shown in Fig. 6 with and without denoising. This noise level is roughly the same as we observe with ²⁴¹Am gamma rays. For all datasets, the gap time in the trapezoidal filter is fixed at 1.8 μs, which was found to be sufficiently large for even the slowest rise times. Clean library pulses all have an amplitude of unity before noise addition, and so the “true energy” is one, meaning that the noise level and the resolution in Fig. 6 can both be interpreted as a fraction of the amplitude.

Figure 7 shows the relative improvement in the energy resolution after denoising as a function of the noise level. Each curve corresponds to a given shaping time, and each

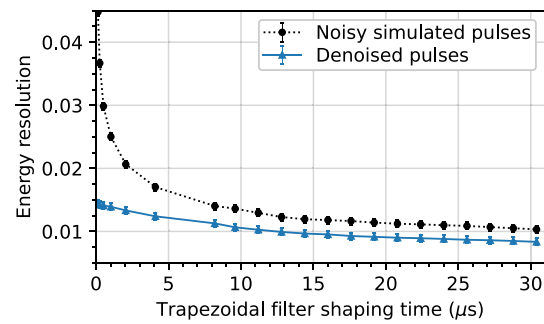


Fig. 6 Energy resolution, defined as the FWHM of the energy distribution, as a function of trapezoidal filter shaping time. Calculated on library pulses with real detector noise with a baseline RMS of 0.1 before denoising (dotted line, circle markers) and after denoising with the amplitude normalized model (solid line, triangle markers)

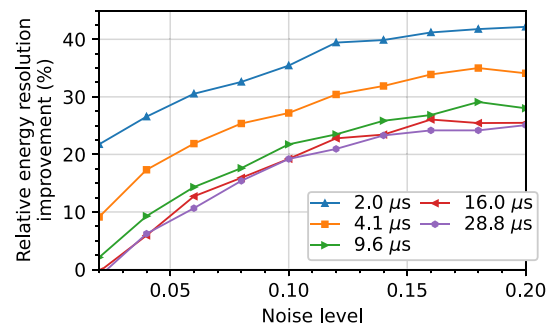


Fig. 7 Relative improvement in the energy resolution from denoising as a function of noise level on simulated pulses. Each curve corresponds to a single trapezoidal filter shaping time

point on a given curve corresponds to a point from a plot such as Fig. 6.

At every shaping time, the energy resolution on simulated data is lower with denoising. This is particularly prominent at shaping times less than 5 μs, where the improvement in energy resolution exceeds 20% for all but the lowest noise levels. At higher shaping times, the averaging window of the trapezoidal filter is large enough to smooth out the noise, and so the improvement is generally smaller. Overall, the best improvement in energy resolution is obtained when the pulses have a high level of noise and when the shaping time used to calculate the energy is small. However, we observe improvements in the energy resolution resolution at every noise level and shaping time evaluated.

6.2 Evaluation on detector data

In this section, we evaluate the denoising performance using data collected with the ²⁴¹Am source, as it provides monoenergetic pulses with a reasonable amount of noise to evaluate the autoencoder. We select events within ±2 keV of the 60 keV peak associated with the source gamma rays, as seen in Fig. 8. The event energies are estimated using a trapezoidal

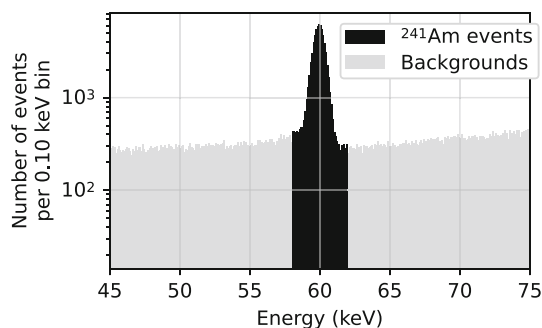


Fig. 8 Amplitude distribution (converted to energy) of the ^{241}Am data, as calculated from the trapezoidal filter. Events in the dark region are within ± 2 keV of the peak and taken to be signal events from the ^{241}Am source, while events in the light region are taken to be backgrounds. Note that the y-axis is logarithmic

filter. Pulses with outliers in the slopes on either the baseline or end of the trace are removed from the dataset.

Furthermore, the 60 keV events from the source are essentially all single site events for which the underlying true shape is close to the ones simulated in the library dataset. This allows us to infer a reasonable guess of the clean target by selecting the library pulse that minimizes the χ^2 value defined in Eq. (6). Note that the basis set used here was generated with a preamplifier time constant of 20 ns as that was found to best match our detector.

6.2.1 Qualitative evaluation

When evaluating the performance on real data, one does not have a true clean signal with which to compare, as was the case with simulated pulses. Instead, we conduct several studies to understand how the autoencoder affects the shape of the denoised pulses. We start with a qualitative evaluation on a data pulse. Figure 9 shows a single-site event from the ^{241}Am source data that has been denoised by two versions of our model. The top panel shows the pulse denoised with the regular library pulse model while the bottom panel shows the denoised pulse using the Noise2Noise model trained with ^{60}Co data with a total variation penalty. Each plot also includes the library pulse with the lowest chi-squared value as a best estimate of the true underlying pulse.

Qualitatively, we observe that the performance of all models on the single-site event are promising and tend to better capture the shape of the pulse than the best fit library pulse. Furthermore, while not shown here, the denoising performance of all models on the multi-site events present in the set is also very good. As was observed with the results on simulated data, the library pulse model tends to perform the best. For the Noise2Noise models, the total variation penalty is important as the model trained without it retains some of the noise.

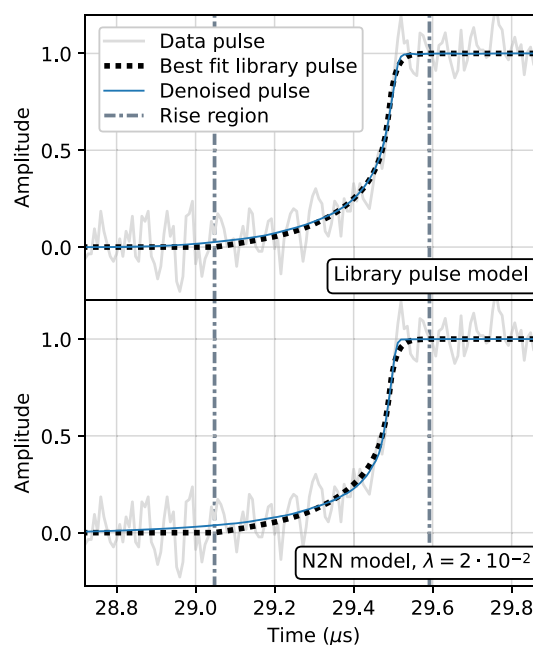


Fig. 9 An example single-site event from the ^{241}Am dataset. Included in each plot is the noisy data pulse (solid light line), the best fit library pulse (dotted line), and the corresponding denoised pulse (solid dark line) from the regular library pulse model (top) and Noise2Noise model (bottom)

6.2.2 Chi-squared comparison

We compare the χ^2 value between the data pulse and denoised pulse for events in the ^{241}Am dataset. Although we do not know the true shape of the data pulses without noise, we can use the χ^2 distribution to determine if, statistically, the denoised pulses are consistent with their noisy progenitors. It is important to use the χ^2 instead of the mean or sum of squared errors to account for the signal-to-noise ratio of the pulses.

Figure 10 shows the distribution of χ^2 values between the data pulses and corresponding denoised pulses using the regular library noise removal model. It also includes the distribution of χ^2 values between the data pulses and corresponding best matching library pulses as determined by the χ^2 -minimization across all library pulses in the basis set.

The χ^2 value for each pulse is computed from sample 3600–3800, meaning that the number of degrees of freedom (NDF) is 200. This range is chosen because the pulses are horizontally centred within this window and it is long enough to capture the important components of the pulse for the largest rise times. A modified χ^2 distribution for NDF = 200 expected from our detector noise is also overlaid on Fig. 10 for comparison. This distribution is calculated directly on the detector noise, independent of any signal. Note that since the detector noise is not Gaussian, the modified χ^2 distribution is shorter and wider than the true χ^2 probability

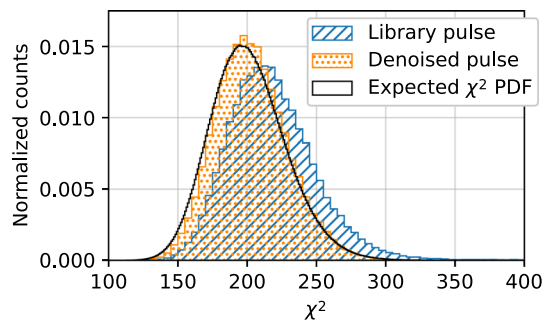


Fig. 10 χ^2 distribution computed between the data and both the denoised pulse (slanted line hatch) and best fit library pulse (dotted hatch)

density function for $NDF = 200$ which assumes normally distributed noise.

The results show that the denoised pulses fit the data better than the best matching library pulses, as indicated by the lower mean and the better fit of the distribution to the χ^2 distribution expected from the detector noise. This could indicate that there are not enough library pulses in the basis set, or that some other parameter of the detector is being modelled improperly. Additionally, while not shown here, the fit is better on *all* events, not just those in the energy range within the 60 keV peak, indicating that our model is properly denoising multi-site events and low-energy events for which the pulse fitter fails to converge.

6.2.3 Energy resolution comparison

In this section, we compare the energy resolution of the 60 keV peak before and after denoising, similar to what was done in Sect. 6.1.3 on simulated data. The results are shown in Fig. 11, which illustrates the energy resolution on both the original noisy and denoised pulses as a function of the trapezoidal filter shaping time. The gap time in the trapezoidal filter is again fixed at $1.8 \mu\text{s}$. The results for two denoising models (both trained with the regular procedure and library pulses) are included: one using amplitude normalized pulses and the other using standardized pulses.

While denoising does not achieve a lower energy resolution overall, it does allow us to obtain a comparable energy resolution with a lower shaping time. This is especially true for the model trained with standardized pulses, where the optimal energy resolution is achieved even at shaping times under $2 \mu\text{s}$. The discrepancy between the two denoised models at lower shaping times is surprising as the only difference is the preprocessing of the pulses. However, all data pulses are horizontally centred, meaning that the lack of robustness to horizontal shifts for standardized pulse models is not an issue here. We note that in general, models trained with standardized pulses performed similarly in terms of both

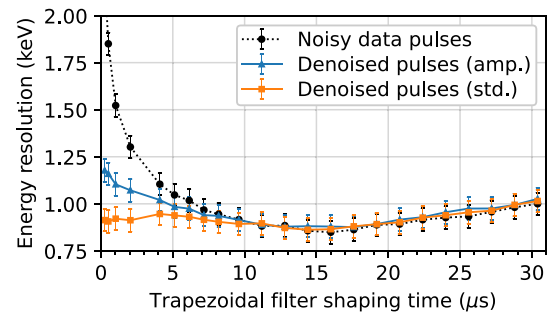


Fig. 11 Energy resolution, defined as the FWHM of the energy distribution, as a function of trapezoidal filter shaping time. Calculated on data from the ^{241}Am source before denoising (dotted line, circle markers) and after denoising with the amplitude normalized model (solid line, triangle markers) and standardized model (solid line, square markers)

mean squared error (on simulations) and χ^2 (on data), and the energy resolution is the only metric where a significant performance deviation is observed.

These results are important, as the optimal shaping time required to obtain a reasonable energy resolution directly affects data collection. Specifically, a longer shaping time requires longer traces, which in turn occupies more disk space and increases the chances of event pile-up. A lower shaping time thus has many practical implications with regards to easier/more efficient data storage and analysis. This is also of interest to higher rate applications of HPGc detectors, rather than just in rare-event searches.

As shown in Sect. 6.1.3, we expect an overall improvement in the energy resolution from simulations under ideal conditions. However, we do not observe this on the real ^{241}Am data. As well, the shape of the energy resolution curve is different; specifically, it decreases as a function of the shaping time up until the largest shaping time, which is limited by the length of the noise traces. This is in contrast to Fig. 11, where there is a clear minimum at around $15 \mu\text{s}$. A prominent imperfection with data is that there are multiple sources of exponential decay, while the pole zero correction only accounts for one “effective” source. This results in visually imperceptible changes to the ends of the pulses that broaden the energy resolution, particularly at high shaping times. As the models are not trained to remove this effect, it is still present to some extent even after denoising. This was confirmed by adding multiple exponential decays and then applying only one correction to a simulated test set. However, as we do not know the properties of the exponential decays present in our setup, we have not accounted for this in the training data or removed it from the real data.

6.2.4 Frequency spectra comparison

The frequency power spectrum of denoised pulses was analyzed for a subrun of the ^{241}Am dataset described in Sect. 4.2

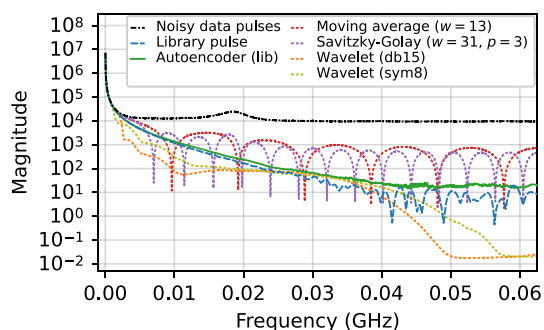


Fig. 12 Comparison of the frequency spectra for simulated library pulses, noisy pulses, pulses denoised with the autoencoder trained using the regular procedure and library pulses, and pulses denoised with four different traditional denoising methods

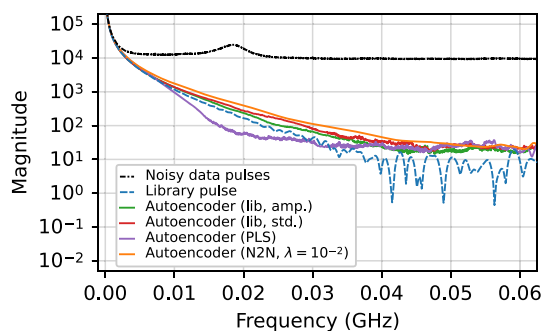


Fig. 13 Comparison of the frequency spectra for simulated library pulses, noisy pulses, and pulses denoised with four different autoencoder models

where the source was fixed at roughly half the height of the detector. The data were denoised with some of the traditional denoising methods evaluated in Sect. 6.1.2, as well as with some of our autoencoder models. The discrete Fourier transform was computed on these denoised data in addition to the original noisy pulses in the subrun. In computing the discrete Fourier transform, a Hanning window [52] was used in order to remove discontinuities resulting from the step-like shape of the pulses and to lessen spectral leakage.

The resulting frequency spectra, obtained by summing the individual spectrum of each pulse for a given set of data, are shown in Fig. 12 for the moving average, Savitzky–Golay, and wavelet denoising methods (all using the optimized parameters as in Fig. 5). Figure 13 shows the frequency spectra comparisons for regular autoencoder models trained with library and PLS simulated data and the Noise2Noise model trained with real data. It also includes the regular library pulse autoencoder trained with standardized pulses. In each figure, the frequency spectra for the noisy data and a clean simulated pulse at the position in the detector corresponding to the location of the ^{241}Am source are also shown for comparison.

Of the methods evaluated, the pulses denoised by the autoencoders have the closest frequency spectra to that of the

clean simulated pulse. In Fig. 12, both the moving average and Savitzky–Golay method spectra have periodic artifacts from their windows. Both wavelet denoising method spectra tend to be overly aggressive in removing noise, specifically at lower frequencies (up to approximately 0.02 GHz) and the highest frequencies (greater than 0.05 GHz).

In Fig. 13, both regular autoencoder models trained with library pulses and the Noise2Noise model have similar frequency spectra. The Noise2Noise model has the smoothest spectrum, but is less aggressive in removing noise at virtually all frequencies, particularly at frequencies larger than 0.04 GHz and around 0.06 GHz, indicating that it fails to remove some of the higher frequency noise. This is consistent with visual observations of the denoised pulses. The autoencoder trained with PLS pulses has the most divergent spectrum and is removing some portion of the signal at the mid-range of frequencies. It does, however, perform similarly to the library pulse models at higher frequencies. This could be due to the fact that the PLS pulse model is more likely to distort and “smooth out” the pulse in and around the rise region, while in principle it should denoise the flat regions just as well as the library pulse models. These results are in agreement with those from Fig. 5, which shows that while the Noise2Noise model is comparable or slightly better at preserving features in the rise region, it is inferior at denoising the overall pulse.

7 Conclusions

We have applied deep convolutional autoencoders to strongly suppress electronic noise from one-dimensional signals collected with a p-type point contact high purity germanium detector, while demonstrating that the underlying pulse shape is preserved to a high degree of accuracy. Comparisons with simulated data, which allow for the underlying clean signal to be known, show that the autoencoder outperforms various traditional denoising methods in this task. We also showed the excellent performance of the autoencoder using real detector data. We found that the denoised pulses are statistically consistent with the original noisy pulses collected from an ^{241}Am calibration source. Notably, the distribution of chi-squared values made from the denoised pulses is in better agreement with the expected distribution than that made with the best fitting simulated pulses obtained from detailed physics simulations. We also showed that denoising allows one to reach the optimal energy resolution with a trapezoidal filter at significantly reduced shaping times for low energy pulses. This has practical implications on data collection, storage, and analysis as it can allow for shorter traces to be collected without loss in energy resolution. We expect that, under certain circumstances, the energy resolution at low energies can even be reduced with denoising, in particular

if any residual exponential decays in the pulse shape can be removed.

In addition to the excellent performance of the models trained on library pulses, we have presented two methods to train models that can perform well without the need for detailed detector simulations. We created a set of pulses using piecewise linear functions with smoothing that are qualitatively similar to the library pulses. Models trained with these pulses performed almost as well as models trained with the library dataset. Additionally, we used a modification to an existing procedure – the Noise2Noise method – to train an autoencoder with no underlying clean pulse as a basis of truth. Instead, the model only required data and noise collected from the detector. While it did not outperform any of the standard models trained with simulated pulses, it still outperformed traditional denoising techniques such as wavelet-based methods. The Noise2Noise method is most limited by the amount and quality of data used to train it. More data, particularly from different sources over a wider energy range, would allow the network to see more variation in pulses and presumably improve the model.

Although we demonstrate the performance of the convolutional autoencoder on germanium detector data, there are many applications that can benefit from the removal of electronic noise in one-dimensional signals. We are beginning to apply our methods to other detector technologies such as gaseous proportional counters and bubble chambers with promising results. Furthermore, denoising autoencoders can be used to obtain a more robust latent representation of the original input [29]. We are exploring the use of the encoder portion of the network for other tasks, including clipping restoration, peak finding, and single-site/multi-site event discrimination (similar to the work in [26]). Our work has great potential to be expanded upon, and the methods demonstrated here are broadly applicable to not only the particle astrophysics community, but any field dealing with noisy one-dimensional signals.

Acknowledgements This work was primarily supported by the Natural Sciences and Engineering Research Council of Canada, funding reference numbers SAPIN-2017-00023 and CGSD3-546735-2020. We acknowledge the support of the Arthur B. McDonald Canadian Astroparticle Physics Research Institute through the Highly Qualified Personnel pooled funding opportunity, the Canada Foundation for Innovation John R. Evans Leaders Fund, and the Walter C. Sumner Memorial Fellowship. We also thank the NVIDIA corporation for their support through their academic hardware grant program.

Data Availability Statement The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP³. SCOAP³ supports the goals of the International Year of Basic Sciences for Sustainable Development.

References

1. S.I. Alvis et al., Search for neutrinoless double- β decay in ^{76}Ge with 26 kg yr of exposure from the MAJORANA DEMONSTRATOR. *Phys. Rev. C* **100**, 025501 (2019)
2. M. Agostini et al., Final results of GERDA on the search for neutrinoless double- β decay. *Phys. Rev. Lett.* **125**, 252502 (2020)
3. N. Abgrall, et al. LEGEND-1000 preconceptual design report. arXiv preprint [arXiv:2107.11462](https://arxiv.org/abs/2107.11462) (2021)
4. C.E. Aalseth et al., CoGeNT: A search for low-mass dark matter using p-type point contact germanium detectors. *Phys. Rev. D* **88**(1), 012002 (2013)
5. M. Agostini et al., First search for bosonic superweakly interacting massive particles with masses up to $1 \text{ MeV}/c^2$ with GERDA. *Phys. Rev. Lett.* **125**(1), 011801 (2020)
6. 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* **76**(11), 1–5 (2016)
7. 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.* **120**(21), 211804 (2018)
8. S.I. Alvis et al., Search for trinucleon decay in the MAJORANA DEMONSTRATOR. *Phys. Rev. D* **99**(7), 072004 (2019)
9. N. Abgrall et al., The MAJORANA DEMONSTRATOR neutrinoless double-beta decay experiment. *Adv. High Energy Phys.* **2014**, (2014)
10. 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.* **118**(16), 161801 (2017)
11. 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* **701**, 176–185 (2013)
12. C. Wiseman, A low energy rare event search with the MAJORANA DEMONSTRATOR. *J. Phys. Conf. Ser.* **1468**, 012040 (2020). IOP Publishing
13. V.T. Jordanov, G.F. Knoll, Digital synthesis of pulse shapes in real time for high resolution radiation spectroscopy. *Nucl. Instrum. Meth. A* **345**(2), 337–345 (1994)
14. M. Agostini et al., Improvement of the energy resolution via an optimized digital signal processing in GERDA Phase I. *Eur. Phys. J. C* **75**(6), 1–11 (2015)

15. R.G. Brown, *Exponential Smoothing for Predicting Demand* (Arthur D. Little Inc., Cambridge, Massachusetts, United States, 1956)
16. A. Savitzky, M.J. Golay, Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**(8), 1627–1639 (1964)
17. C. Taswell, The what, how, and why of wavelet shrinkage denoising. *Comput. Sci. Eng.* **2**(3), 12–19 (2000)
18. M. Shafiee, S.A.H. Feghhi, J. Rahighi, Analysis of de-noising methods to improve the precision of the ILSF BPM electronic read-out system. *J. Instrum.* **11**(11), 12020 (2016)
19. D.L. Donoho, J.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**(3), 425–455 (1994)
20. S.G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression. *IEEE T. Image Process.* **9**(9), 1532–1546 (2000)
21. R.E. Kalman, A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**(1), 35–45 (1960)
22. G. Hinton, D. Rumelhart, R. Williams, Learning internal representations by error propagation. *Parallel Distrib. Process.* **1**, 318–362 (1986)
23. Ballard, D.H.: Modular learning in neural networks. In: Proceedings of the Sixth National Conference on Artificial Intelligence, vol. 1, pp. 279–284. AAAI Press, Palo Alto, California, US (1987)
24. Sakurada, M., Yairi, T.: Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, pp. 4–11 (2014)
25. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: Proceedings of the Second International Conference on Learning Representations (2014)
26. P. Holl, L. Hauertmann, B. Majorovits, O. Schulz, M. Schuster, A. Zsigmond, Deep learning based pulse shape discrimination for germanium detectors. *Eur. Phys. J. C* **79**(6), 1–9 (2019)
27. Nielsen, M.A.: *Neural Networks and Deep Learning*. Determination Press, San Francisco, California, United States (2015). <http://neuralnetworksanddeeplearning.com>
28. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press, Cambridge, Massachusetts, United States (2016). <https://www.deeplearningbook.org>
29. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11**(Dec), 3371–3408 (2010)
30. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.-A.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning, pp. 1096–1103 (2008)
31. Jain, V., Seung, S.: Natural image denoising with convolutional networks. In: Advances in Neural Information Processing Systems, pp. 769–776 (2009)
32. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: Advances in Neural Information Processing Systems, pp. 341–349 (2012)
33. K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE T. Image Process.* **26**(7), 3142–3155 (2017)
34. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1712–1722 (2019)
35. Gondara, L.: Medical image denoising using convolutional denoising autoencoders. In: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pp. 241–246 (2016). IEEE
36. Khalilian, S., Hallaj, Y., Balouchestani, A., Karshenas, H., Mohammadi, A.: PCB defect detection using denoising convolutional autoencoders. In: 2020 International Conference on Machine Vision and Image Processing (MVIP), pp. 1–5 (2020). IEEE
37. Zezario, R.E., Huang, J.-W., Lu, X., Tsao, Y., Hwang, H.-T., Wang, H.-M.: Deep denoising autoencoder based post filtering for speech enhancement. In: 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 373–377 (2018). IEEE
38. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2Noise: Learning image restoration without clean data. In: Proceedings of the 35th International Conference on Machine Learning, vol. 80, pp. 2965–2974 (2018)
39. V. Dumoulin, F. Visin, A guide to convolution arithmetic for deep learning. arXiv preprint [arXiv:1603.07285](https://arxiv.org/abs/1603.07285) (2016)
40. X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 315–323 (2011)
41. ORTEC/AMETEK (2020). <https://www.ortec-online.com/>
42. Struck Innovative Systems (2022). <https://www.struck.de/>
43. R. Brun, F. Rademakers, ROOT - an object oriented data analysis framework. *Nucl. Instrum. Meth. A* **389**(1–2), 81–86 (1997)
44. Vasundhara: Pulse fitting for event localization in high purity germanium point contact detectors. Master's thesis, Queen's University (2020)
45. D.C. Radford, siggen. GitHub (2017). https://github.com/radforddc/icpc_siggen
46. F. Chollet, et al. Keras (2015). <https://keras.io>
47. M. Abadi, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems (2015). <https://www.tensorflow.org/>
48. D.P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
49. L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms. *Physica D* **60**(1–4), 259–268 (1992)
50. D. Strong, T. Chan, Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Probl.* **19**(6), 165 (2003)
51. Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity. *IEEE T. Image Process.* **13**(4), 600–612 (2004)
52. R.B. Blackman, J.W. Tukey, The measurement of power spectra from the point of view of communications engineering - Part I. *Bell Syst. Tech. J.* **37**(1), 185–282 (1958)