Towards Improved Pileup Rejection through Physics-Informed Denoising of Cryogenic Bolometer Pulses

Colin J. DuHamel
Department of Computer Science
California Polytechnic State University - San Luis Obispo
San Luis Obispo, CA 93407
cduhamel@calpoly.edu

Thomas D. Gutierrez
Department of Physics
California Polytechnic State University - San Luis Obispo
San Luis Obispo, CA 93407
tdgutier@calpoly.edu

Sumona Mukhopadhyay
Department of Computer Science
California Polytechnic State University - San Luis Obispo
San Luis Obispo, CA 93407
mukhopad@calpoly.edu

Abstract

Cryogenic bolometers are highly sensitive detectors that leverage extremely cold temperatures to detect nuclear decays, resulting in a high sensitivity to noise. This leads to complications during data processing, such as a reduction in pileup rejection accuracy. While current attempts convolve over the original pulse in 1 dimension, using the phase portrait representation of the pulse may better illuminate the differentiating features of pileup events. However, since the phase portrait utilizes the time derivative of the amplitude, it is highly sensitive to noise, requiring a powerful denoiser to remain viable. We explore a novel approach to denoising bolometer pulses using a Physics-Informed Neural Network (PINN) that utilizes a modified loss term based on current models of detector dynamics and response. By balancing data-driven pulse denoising with existing bolometer models, we aim to develop an improved denoiser that, in future work, may be integrated into a more effective pileup rejection algorithm.

1 Introduction and Background

In the search for rare events, it is essential to have highly sensitive detectors that can accurately identify and classify events of interest. However, the high sensitivity of these detectors often leads to a higher sensitivity to noise. This is especially true for cryogenic bolometers, which operate at extremely low temperatures (on the order of millikelvin) to detect nuclear decays, making them particularly susceptible to noise. This affects the accuracy of pileup rejection.

Pileup, an irreducible source of background in rare event searches, occurs when multiple events occur near simultaneously in a single detector, resulting in a detector response that resembles a single pulse. To reject these events, we can utilize the phase portrait representation of the pulse, which plots the time derivative of the amplitude against the pulse amplitude. This representation "stretches out" the rising edge of the pulse, which is where the differentiating features of pileup events are most prominent. However, due to noise amplification from the time derivative, the phase portrait becomes unintelligible, motivating the need for a powerful denoiser to remain viable. While other methods exist for denoising signals [1][2], we explore the integration of bolometer dynamics through a Physics-Informed Neural Network (PINN) approach to help constrain the network during training to physically accurate reconstructions.

Conference on Physics and AI at Stanford University (PAI 2026)

1.1 Cryogenic Bolometers

Cryogenic bolometers are particle detectors that measure incident particle energy through the resulting temperature change in the detector material. Since this temperature change is proportional to deposited energy, it can be used for particle identification and event analysis. To resolve the extremely small energy deposits of interest, cryogenic bolometers operate at temperatures of roughly 5–20 mK, minimizing thermal noise and enabling sensitivity to temperature changes of order $10\ \mu$K. This allows precise measurements of keV-scale energy deposits, where 1 keV $\approx 1.6 \times 10^{-16}$ J.

CUORE [3] and its upgrade, CUPID [4], are cryogenic-bolometer experiments searching for neutrinoless double beta decay $(0\nu\beta\beta)$, a lepton-number-violating process that would indicate neutrinos are Majorana particles. CUPID reduces backgrounds, including alpha backgrounds, using scintillating bolometers, which enable particle discrimination through simultaneous heat and light detection. However, CUPID is expected to face increased pileup due to the higher $2\nu\beta\beta$ rate in $\mathrm{Li_2MoO_4}$ compared with $\mathrm{TeO_2}$ in CUORE.

1.2 Physics-Informed Neural Networks

Physics-Informed Neural Networks (PINNs) are neural networks that encode physical laws and constraints into the model architecture or training process [5]. By incorporating known physical principles, typically expressed as partial differential equations (PDEs) or ordinary differential equations (ODEs), PINNs encourage predictions that adhere to established models of the system being studied.

In practice, these physical laws are often incorporated through the loss function. This allows the standard data-driven loss, such as mean squared error for regression, to be balanced against a physics-based loss that measures how well the network predictions satisfy the governing equations. The relative weighting of these terms balances fitting the data and enforcing physical consistency.

2 Related Work

2.1 Signal Denoising with Neural Networks

Current methods for denoising signals with neural networks often involve using autoencoders [6][7], such as convolutional neural network encoder-decoder architectures [8][9]. These architectures are designed to learn a mapping from noisy input signals to clean output signals by training on pairs of noisy and clean data.

Traditional denoising autoencoders consist of an encoder network that compresses the input signal into a lower-dimensional space, and a decoder network that reconstructs the clean signal from this lower dimensional representation. The networks are trained to minimize the difference between the reconstruction and the clean signal, most often using Mean Squared Error (MSE) as the loss function.

2.2 Machine Learning for Pileup Rejection

The CUPID collaboration has published a paper [10] exploring the usage of one-dimensional Convolutional Neural Networks for pileup rejection. The CNN was trained on synthetic "heater pulses" calibrated to mimic real $2\nu\beta\beta$ (and subsequent pileup) events. Their approach involved 10 equally sized convolutional blocks (convolutional, pooling, ReLU activation) followed by fully connected layers and was determined to have similar performance to optimal filtering, with a $>$90% pileup discrimination efficiency down to Δt $>$ 1 ms [11].

3 Methods

3.1 Pole-Zero Bolometer Model and Dataset Generation

To generate synthetic bolometer pulses for training our denoising network, we utilize a Pole-Zero model of the bolometer's response detailed in Nutini et al. [12]. While this model is a simplification of the true dynamics of the bolometer, it captures the key features of the pulse shape and serves as a basis for deriving the physics-informed loss term for our network.

The poles and zeros of the system also come from Nutini et al. [12], which are sampled uniformly from a range of values to simulate the natural variation in pulse shapes seen in real detectors. The ranges for the parameters used are:

$$z_1 \in [-3.21, -2.93], p_1 \in [-2.07, -1.91], p_2 \in [-14.6, -14.5], p_3 \in [-62, -60]$$

An example of a pulse generated using this model is shown in Figure 1.

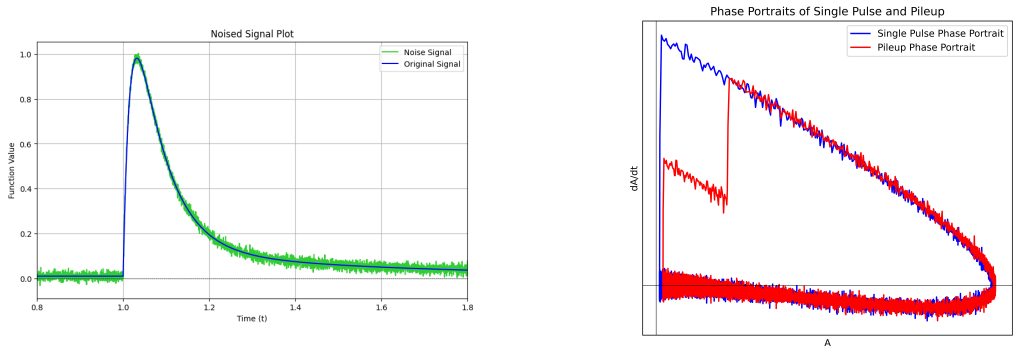

Figure 1: **Left:** An example synthetic bolometer pulse generated using the Pole-Zero model. **Right:** The phase portrait of a normal pulse (blue) and a pileup pulse (red) after a significant noise reduction.

3.2 Derivation of Bolometer Dynamics

Since the bolometer model used to generate the pulses is a Pole-Zero model defined in the Laplace domain, we first need to derive the corresponding ordinary differential equation (ODE) that governs the time domain dynamics of the bolometer. To do this, we can start with our transfer function, and set it equal to the ratio of the Laplace transforms of the output and input signals, $Y(s)$ and $U(s)$, respectively:

$$\frac{Y(s)}{U(s)} = \frac{(s - z_1)}{(s - p_1)(s - p_2)(s - p_3)} \tag{1}$$

We can then cross-multiply, rearrange, and expand the equation, yielding:

$$U(s)(s - z_1) = Y(s)(s^3 - s^2(p_1 + p_2 + p_3) + s(p_1p_2 + p_2p_3 + p_3p_1) - p_1p_2p_3)$$

Finally, we can take the inverse Laplace transform of both sides to obtain the time domain ODE. Assuming all zero initial conditions, leading to the correspondence $s^nY(s) \leftrightarrow \frac{d^ny}{dt^n}$, we can express the ODE as follows:

$$\frac{du}{dt} - z_1u = \frac{d^3y}{dt^3} - (p_1 + p_2 + p_3)\frac{d^2y}{dt^2} + (p_1p_2 + p_2p_3 + p_3p_1)\frac{dy}{dt} - p_1p_2p_3y \tag{2}$$

3.3 Physics-Informed Loss Function

To incorporate the physics of the bolometer into the training of our denoising network, we craft a specialized physics-informed loss term into our loss function. Since we do not know the impulse input $u(t)$, we can carefully mask out the small trigger region of the pulse for the calculation of the physics loss. This allows us to enforce a homogenous version of the ODE in Equation 2 on the rest of the pulse, if we assume the impulse input can be approximated as a delta function. We then utilize a root mean squared error (RMSE) formulation instead of mean squared error (MSE) to prevent the loss from becoming too large and dominating the data loss term during training.

We, however, must make a few adjustments for this loss term to be viable. First, we must address the blowup in error in the calculation of higher order derivatives. To do this, we can apply a scaling term to each derivative term in the loss function, which can be expressed as follows:

$$D_i = \sigma^i \cdot \frac{d^i}{dt^i}\hat{y}_i(t) \tag{3}$$

where σ is a hyperparameter that can be tuned to balance the error contributions from each derivative term. This allows us to prevent the higher order derivatives from dominating the loss and causing instability during training. This makes our total physics loss term as follows:

$$\mathcal{L}_{physics} = \left(\frac{1}{N} \sum_{i=1}^{N} \left(D_3 - (p_1 + p_2 + p_3) D_2 + (p_1 p_2 + p_2 p_3 + p_3 p_1) D_1 - p_1 p_2 p_3 \hat{y}_i(t) \right)^2 \right)^{\frac{1}{2}}$$

Since we have chosen to mask out the trigger region of the pulse, we must accommodate the fact that our ODE holds for many solutions of different amplitudes. To do this, we implement a physics loss scheduler that gradually increases the weight of the physics loss term as the data loss decreases during training. The motivation behind this is that we can use the data loss to guide the network towards a pulse with the correct shape, and then constrain the shape of the pulse later in the training process with the physics loss. The weighting term, $\lambda_{physics}$, can be expressed as a sigmoid centered around a tuned center point, with a tuned width k. This allows us to gradually increase the influence of the physics loss as the data loss decreases during training.

3.4 Network Architecture

To test the viability of our physics-informed loss function, we trained a simple feedforward neural network (three hidden layers of sizes 256, 128, and 64 with tanh activation) implemented in PyTorch [13] on a small dataset of synthetic bolometer pulses generated using the Pole-Zero model. We can then compare the performance of the network trained with the physics-informed loss function to a baseline network trained with only the data loss term.

Our synthetic dataset consists of 500 noisy–clean bolometer pulse pairs, each with length n_{samples}. To compute the time derivatives required by the physics loss, each pulse is replicated n_{samples} times, with each copy assigned a different time value corresponding to one sample in the pulse. The network therefore receives both the full noisy input pulse $\mathbf{x}$ and a scalar time coordinate t, and learns a function of the form $\hat{y}(t) = f(\mathbf{x}, t)$, where $\hat{y}(t)$ is the denoised pulse amplitude at time t.

This formulation is enabled by using a feedforward architecture rather than a one-dimensional convolutional network. Since t is provided as an explicit differentiable input to the network, automatic differentiation can be used to compute derivatives of the output with respect to time, such as $\partial^i \hat{y}(t)/\partial t^i$, which are then used in the physics loss. The final dataset has shape $(500, n_{\text{samples}}, n_{\text{samples}} + 1)$, where the last dimension contains the full input pulse together with the time value at which the output is evaluated.

4 Results and Discussion

The standalone feedforward network trained without the physics-informed loss achieved a **test MSE of 0.003** and a **weighted physics loss of 0.027** after 1000 epochs, where the physics loss was computed only for evaluation. In comparison, the network trained with the physics-informed loss achieved a **test MSE of 0.057** and a **weighted physics loss of 0.016**. This indicates that the physics-informed loss is successfully constraining the network toward solutions that better satisfy the governing equations, but may also be pulling the reconstruction away from solutions supported by the data.

Several factors may explain the poor performance of the PINN loss term. The most likely issue is instability from higher-order derivatives, particularly the third derivative term, which may distort the loss landscape and provide little useful gradient information, effectively acting as noise during training. This may occur even after scaling the physics loss to be comparable to the data loss. Additionally, because the trigger region of the pulse is masked out, the trivial ODE solution, a flat line at 0, may act as an attractor and prevent the network from learning useful reconstructions.

Although physics-informed denoising remains promising, further modifications to the loss term are needed. Future work will explicitly include the input term $u(t)$, allowing enforcement of a nonhomogeneous ODE and helping prevent convergence to the trivial solution. It will also be necessary to address sensitivity to initial conditions, which can cause training to break or become unrecoverable, potentially through gradient clipping or a data-loss threshold in the physics-loss scheduler.

References

[1] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, March 1960. ISSN 0021-9223. doi: 10.1115/1.3662552.

[2] D. L. Donoho and I. M. Johnstone. Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, 81(3):425–455, 1994. doi: 10.1093/biomet/81.3.425.

[3] D. Q. Adams, C. Alduino, K. Alfonso, et al. Search for Majorana neutrinos exploiting millikelvin cryogenics with CUORE. *Nature*, 604(7904):53–58, 2022. doi: 10.1038/s41586-022-04497-4.

[4] K. Alfonso, A. Armatol, C. Augier, et al. CUPID, the CUORE upgrade with particle identification. *European Physical Journal C*, 85(7):737, 2025. doi: 10.1140/epjc/s10052-025-14352-1.

[5] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

[6] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning - ICML '08*, pages 1096–1103, Helsinki, Finland, 2008. ACM Press. ISBN 978-1-60558-205-4. doi: 10.1145/1390156.1390294.

[7] A. Creswell and A. A. Bharath. Denoising Adversarial Autoencoders, January 2018. arXiv:1703.01220 [cs].

[8] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing*, 26 (7):3142–3155, July 2017. ISSN 1057-7149, 1941-0042. doi: 10.1109/TIP.2017.2662206. arXiv:1608.03981 [cs].

[9] V. Jain and S. Seung. Natural Image Denoising with Convolutional Networks. In *Advances in Neural Information Processing Systems*, volume 21, pages 769–776. Curran Associates, Inc., 2008.

[10] G. Fantini, A. Armatol, E. Armengaud, et al. Machine Learning Techniques for Pile-Up Rejection in Cryogenic Calorimeters. *Journal of Low Temperature Physics*, 209(5):1024–1031, 2022. doi: 10.1007/s10909-022-02741-9.

[11] A. Armatol, E. Armengaud, W. Armstrong, et al. Novel technique for the study of pileup events in cryogenic bolometers. *Physical Review C*, 104(1):015501, 2021. doi: 10.1103/PhysRevC.104.015501.

[12] I. Nutini, C. Bucci, and O. Cremonesi. Modelling the shape of thermal pulses from low temperature detectors, 2021. arXiv:2101.05029.

[13] A. Paszke, S. Gross, F. Massa, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

Acknowledgements

We gratefully acknowledge the support of the National Science Foundation (NSF) under Grant No. NSF-PHY-1913374 and NSF-PHY-2412377, along with the CUORE and CUPID collaborations.