Optimizing Machine Learning Techniques for Dark Matter Detection: Enhancing Position Resolution in SuperCDMS Detectors

Stephanie Cui, Department of Physics, University of Toronto

Summer 2024

1 Abstract

The Super Cryogenic Dark Matter Search (SuperCDMS) at SNOLAB employs detectors with phonon and charge sensors to record energy depositions, aiming to directly detect Weakly Interacting Massive Particles (WIMPs). This study implements and optimizes supervised machine learning (ML) techniques to enhance position reconstruction of recoils that is crucial for background discrimination. By examining the position dependence and correlation features in the reduced quantities (RQs), which serve as inputs for ML models, additional information are incorporated to improve the Z resolution of Geting Qin's ensemble model to approximately 3.29 mm. Analysis of maximum amplitude and minimum time delay channel pairs provides insights into the relationship between energy and timing characteristics as well as a measure of energy sharing. A comparison between features in simulated sample and experimental data also highlights critical discrepancies, indicating the need for future revisions to the Detector Monte Carlo (DMC).

2 Introduction

Abundant astrophysical and cosmological evidence points to the existence of dark matter (DM), implying that the current Standard Model of Particle Physics (SM) is incomplete. While the effects of DM are clearly observed on galactic scales, interactions between normal matter and DM are extremely rare, making the detection of DM on Earth a significant challenge. The hunt for dark matter consists of three main types: 1) direct detection experiments that attempt to observe the interaction between dark matter particles and SM targets in underground laboratories [1], 2) indirect detection experiments that search for the products of dark matter annihilation or decay in the form of surplus SM particles [2], and 3) collider search that looks for dark matter production from SM interactions.

The Super Cryogenic Dark Matter Search or SuperCDMS at SNOLAB is a direct detection experiment that aims to detect nuclear recoil of the most promising low-mass dark matter candidates, namely the Weakly Interacting Massive Particles (WIMPs) [1]. A continuation of the Cryogenic Dark Matter Search (CDMS), SuperCDMS specifically probes the mass region below 10 GeV/c^2 and sets out to exceed the state of the art sensitivities by at least an order of magnitude [1]. It utilizes detectors made of Ge and Si semiconductor crystals equipped with phonon and/or ionization sensors that complement each other in terms of phonon resolution and background discrimination—the limiting factors of their sensitivities [3].

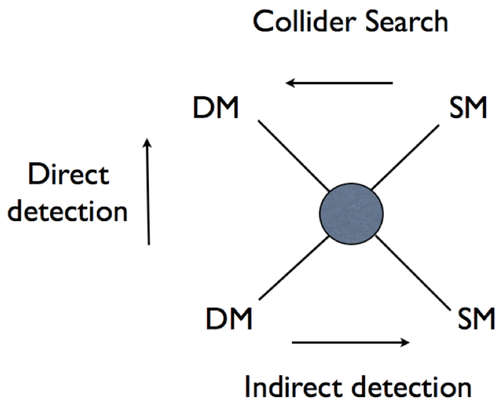

Figure 1: Diagram illustrating the three methods of dark matter detection [4].

The energy and position of energy depositions within the detectors (events) are essential to background rejection. On the one hand, the nuclear recoil energy of low-mass dark matter particles can be estimated [1], so an excellent energy resolution provides the tool to filter out electron recoil events. On the other hand, knowledge of the true position of events enables the placement of cuts that exclude those outside the defined fiducial volume from our analysis. The designs of SuperCDMS detectors leave open the possibility for both energy and position reconstruction [1]. However, this information is buried in the characteristics of raw signals. This unique problem calls for the employment of machine learning (ML) techniques, which are adept at discerning patterns in large statistics [5]. Indeed, members of this collaboration have leveraged the power of supervised ML to extract information from signals and reduced quantities (RQs) obtained by processing pulses through the reconstruction software CDMSBats.

Building on these previous efforts, the primary goal of this study is to optimize the performance of the ensemble neural network (ENN) developed by Geting Qin, especially with regard to its prediction of Z position. The ENN was trained using a HV100mm Full Detector Monte Carlo (DMC) sample, providing access to both the input reduced quantities (RQs) and the output position and energy truth data for analysis. To enhance our understanding of how the model learns position information, a comprehensive analysis of the position dependence and correlation features present in the data is necessary. We also endeavor to characterize the sharing of energy between sensors, which physically amplifies the position dependence of RQs. The insights gained from this project have been directly applied to the ensemble model to improve Z-resolution. Recognizing that the model's performance is contingent on the accuracy of our simulation, these findings are also compared with the first data collected by the SuperCDMS collaboration at the Cryogenic Underground Tower TEst (CUTE) facility.

3 Background

3.1 SuperCDMS Experimental Design

The SuperCDMS detectors are cylindrical in shape with a diameter of 100 mm, a thickness of 33.3 mm, and made of either high-purity germanium or silicon [1]. Apart from these commonalities, they can be classified into two different styles: the High Voltage (HV) detector and the interleaved Z-sensitive Ionization and Phonon (iZIP) detector. The former contains only athermal phonon sensors that record the amount of phonon energy absorbed per unit time, with a high bias voltage applied to amplify phonon production [1]. The latter contains additional charge sensors that enable the discrimination between nuclear and electron recoils. Since this study utilizes only simulated and experimental data from HV detectors, the focus will be exclusively on the physics relevant to the phonon sensors.

When a WIMP particle interacts with a nucleus in the detector, it transfers some of its energy to the nucleus, causing it to recoil. This disturbance generates athermal recoil phonons, which are excited vibrational states in the crystal lattice. Additionally, free charge carriers, consisting of electrons (e^-) and holes (h^+) (commonly referred to as electron-hole pairs), are produced. The energy deposition excites electrons into higher energy states, allowing them to move through the lattice, while the absence of electrons leaves behind net positive charges, or "holes". Although phonons and holes are quasiparticles, both travel through the crystal like particles. Under the influence of an applied voltage, the electron-hole pairs drift to opposite sides of the detector. Holes typically move straight down due to the electric field, whereas electrons travel along paths where their effective mass is minimized within the crystal, known as "valleys," which are determined by the crystal lattice structure [1]. These charges eventually recombine with the crystal lattice, releasing "recombination phonons", so that the total recoil energy is ultimately converted into the energy of both prompt and recombination phonons [6].

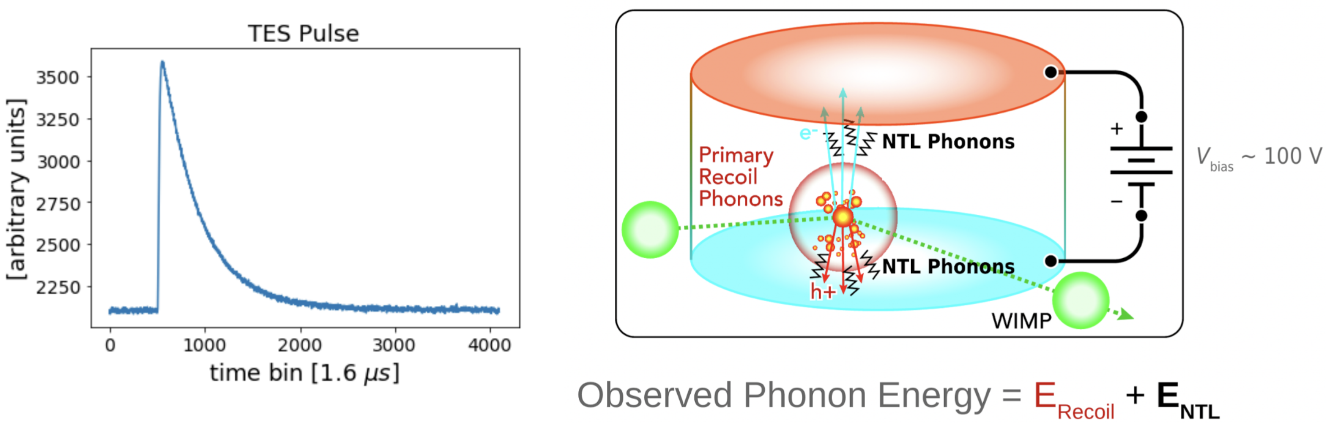

Figure 2: Left: TES pulse that shows the (inverted) decrease in current measured [6]. Right: Diagram of charge and phonon production and transport within a HV detector biased at 100V.

The primary phonons propagate quasi-isotropically, and their energy alone may not be enough to be picked up by the athermal sensors. This issue is mitigated by the Neganov-Trofimov-Luke (NTL) effect, which is another consequence of the electric field. The excess energy used to propel electron-hole pairs generates secondary phonons, which radiate outward in a cone-like pattern along the direction of the charge carriers [13]. The total phonon energy is therefore the sum of the recoil energy and the energy of NTL phonons where $E_{NTL} = n_{eh} e V_b$. To measure this energy, phonons that reach the top and bottom surfaces are absorbed by superconducting aluminum collection fins that act like antennas. The phonons with enough energy can break Cooper pairs to generate quasiparticles that act similarly to free electrons, which diffuse and heat up the tungsten Transition-Edge Sensors (TESs) attached to the fin. These sensors are held at the

[1]Electrons also occasionally jump between them in a phenomenon known as inter-valley scattering, which is not modelled in the current DMC [6].

critical temperature where their resistance transitions from superconducting to neutral, so a minimal change in temperature leads to a significant change in electrical resistance. This rise in resistance is detectable because it leads to a current flow proportional to the strength of the energy deposited. By negative electrothermal feedback, the TESs return to their normal resistance after some time [1]. This change in current is reflected in the shape of our signal, which is often inverted to resemble a pulse.

Six detectors are stacked in towers and mounted in the SNOBOX, a low-radioactivity copper cryostat further surrounded by six layers of shielding blocking various sources of background [7]. The SNOBOX is connected to a delusion refrigerator, which is a part of a cryogenic system capable of creating a 15 mK environment for the experiment [1]. The detector tower is connected to the trigger and readout system that produces data. The details of these data acquisition (DAQ) systems are not covered here.

3.1.1 HV Detector Layout and Energy Sharing

The HV detector contains six phonon channels on each side: an inner core, three wedge shaped channels, and two outer rings. The channels are named from A-F on each side, and the phonon channel A on side 1 is referred to as PAS1. The wedge channels on side 2 are each rotated from those on side 1. Note that PCS2 and PDS2 are rotated 60 degrees from PCS1 and PDS1, respectively, while PES2 and rotate 180 degrees (reflected) from PES1. The detector is designed in a way that all 12 channels cover regions with the same area and contain the same number of TESs [1].

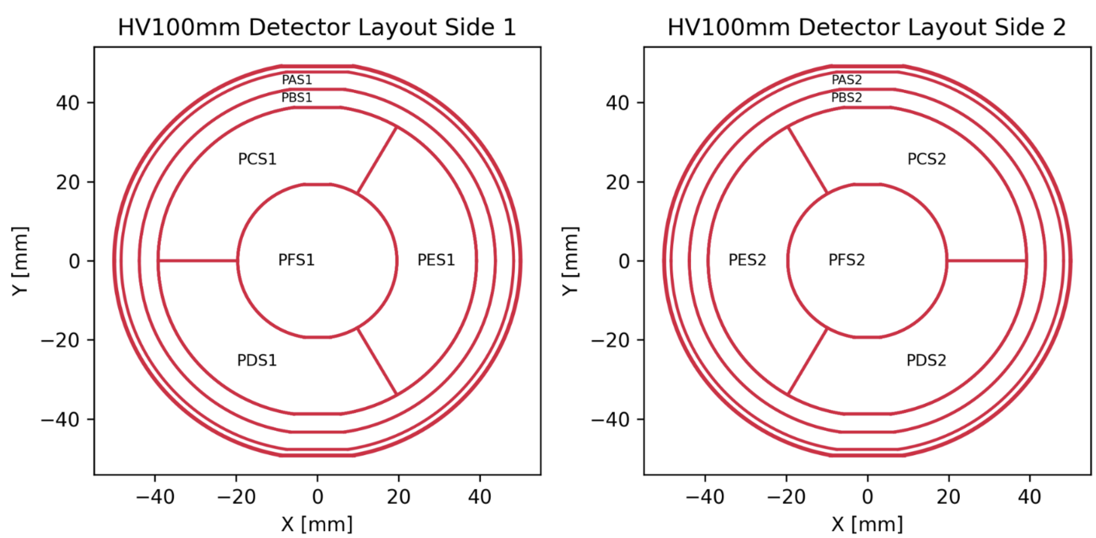

Figure 3: Two sides of the HV100mm detector with labelled channel layout, drawn using *detectorLayout.py* by courtesy of Warren Perry.

The sensitivity in radial position is from the signal-sharing between sensors. For each event, every channel measures the phonon energy deposited and produces a signal. If energy deposition occurs near the edge of a channel, the phonon energy is shared among neighboring channels. Without signal-sharing, the radial position sensitivity would be restricted by the channel size. Thus, it is crucial for us to gain a deeper understanding of energy sharing to aid our goal of position reconstruction.

3.2 Simulation Infrastructure

To better understand the physical processes described above, the SuperCDMS experiment has a series of simulations that model particle behavior and detector response using GEANT4. It begins with the simulation of incident particles, such as WIMPS, calibration sources, background sources etc., in a stage called "SourceSim". SourceSim uses configuration macros to generate data on interaction locations and energy deposition of various particles [6]. This data can be analyzed directly or passed to the next stage of the simulation pipeline. DetectorSim has three components (CrystalSim, TESSim, and FETSim) that model the detector response to interactions at different locations and outputs a smooth trace representing the change in current. They are grouped together under the name of "Detector Monte Carlo" or "DMC" [2] and are of the most importance to this project. CrystalSim simulates the propagation of phonons and charges in the crystal, TESSim simulates the behavior of the TESs, and FETSim simulates the charge readout. The output of DMC includes the hit positions of each event, i.e., where the energy deposition occurred, within the detector.

The final component of DetectorSim, DAQSim, adds noise to the trace using noise Power Spectral Densities (PSDs) of real noise recorded at previous iterations of the CDMS experiment [6]. The pulses generated are in the same format as real pulses from DAQ.

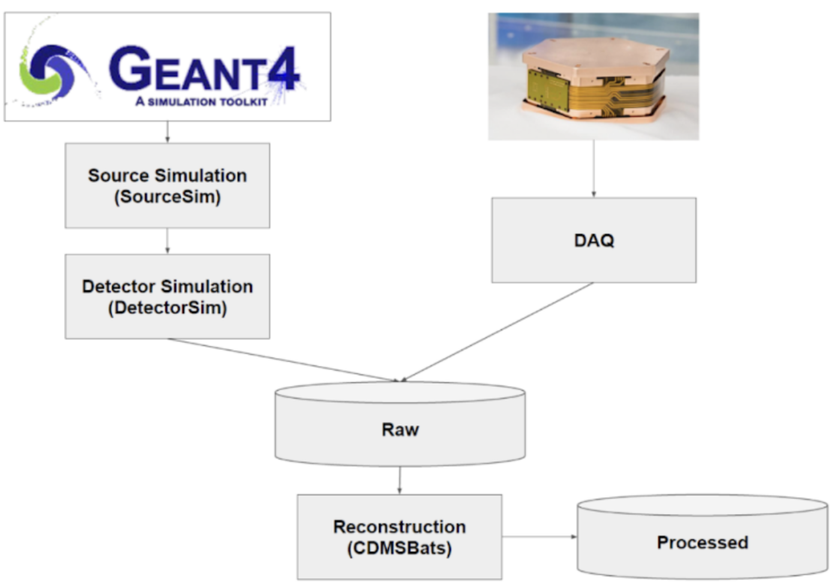

Figure 4: Flowchart of the simulation pipeline and experimental process that both result in raw data being processed by CDMSBats.

All "raw" pulses, whether from experiments or simulations, are processed using the reconstruction package CDMSBats. It takes in the pulse, a pulse template derived from real data, and a noise PSD function and runs them through a 1D Optimal Filter (OF). The filtering technique optimally aligns the signal with the template by shifting it both vertically and horizontally while minimizing noise. The outputs of CDMSBats contain Reduced Quantities (RQs) that describe these shifts, the goodness of fit, and other pulse characteristics.

[2] There are two versions of DMC: FullDMC and FastDMC. FullDMC generates and tracks the path of each individual phonon or charge through the crystal to the sensors, allowing us to make predictions about detector response. Each trace outputted by FullDMC will therefore look distinct. FastDMC, however, uses knowledge about the position dependence of phonon collection efficiency and energy sharing to estimate the detector response. These approximations are then used to output traces with the same shape but different amplitudes scaled vertically from templates. Thus, we can not gain new understanding of our detector from FastDMC, but its faster runtime is beneficial for generating high statistics.

The three most relevant RQs for this project are P*OFamps (pulse amplitude), P*OFdelay (time delay from the global trigger), and P*OFchisq (chi square of the optimal filter) [3]. Running simulation through CDMSBats is particularly productive, as it allows us to 1) trace back interesting output from CDMSBats to the SourceSim or DetectorSim to understand the physical processes that contribute to them [6], and 2) compare distilled features of simulated and experimental data processed through the same software.

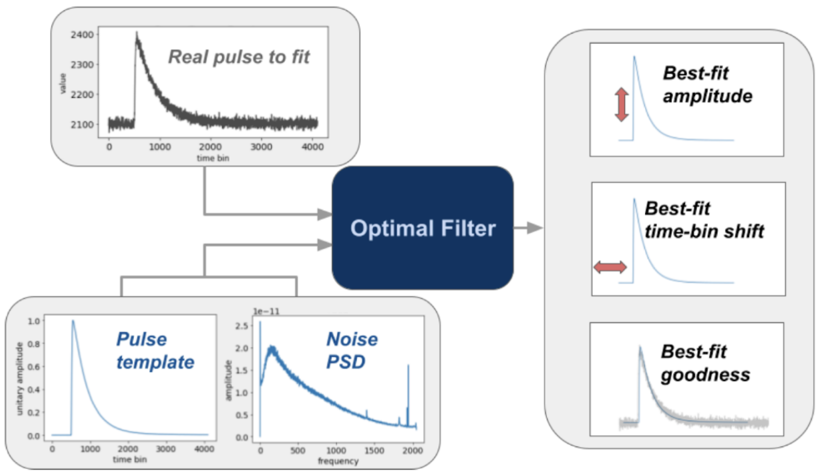

Figure 5: Flow chart showing the inputs to the optimal filter and three most relevant outputs to this study.

4 Previous Work & Motivations

4.1 Motivation for Using Machine Learning for Event Reconstruction

Energy and position information are crucial to the delineation of genuine DM signals from background. The sources of background in this experiment include radioactive contamination from ^{238}U, ^{232}Th, and ^{40}K within the detectors, material activation from exposure to cosmic rays, dust, and radon plate-out on the surfaces of the detector, and so on [6]. These sources can be categorized as either occurring at the surface or bulk of the detectors and as either resulting from nuclear or electron recoil. Background models are developed to gain a better understanding of the nature and location of these sources. Coupled with event reconstruction, we can implement energy and position cuts to eliminate events most affected by background.

Machine Learning (ML) has proven to be the most effective tool for this task and is implemented widely across different High Energy Physics (HEP) experiments for particle tagging, pulse denoising, and so on. It is capable of detecting patterns within a large volume of complex data in a speedy and automated way. The advance of Deep Learning (DL) has been described as a "qualitative shift" to the state of HEP data analysis with its performance surpassing traditional ML techniques [5]. Deep neural networks (DNN) are inspired by the structure of biological neural networks, where neurons form interconnected layers to fire signals within the brain. Thus, they contain units of information termed "neurons" that make up hidden layers. Beginning with the layer that takes in the dataset, each subsequent layer takes in the output of the previous layer.

[3]The * is a placeholder for channel names, e.g. "PAS1OFamps".

The connection between two neurons in different layers is assigned a weight, which measures the strength of their correlation. Consequently, each neuron in layer 1 can be expressed as the output of an activation function, where the input is the weighted sum of all the neurons in layer 0, plus a bias term. The most popular activation function for regression models is the ReLU (Rectified Linear Unit) function where $ReLU(x) = max(0, x)$, which has replaced the sigmoid activation due to its higher efficiency and computability [8]. The weights and biases are parameters within the model that are adjusted during training. The activation function, however, is an example of a hyperparameter, which contributes to the overall model architecture and is fixed during the training process.

ML methods can be supervised, where the target output values corresponding to each input feature is fed into the network (i.e., the training data is "labeled"), or unsupervised, where no labels are present. In this case, the true positions of each event from DMC are available and are used as labels to the RQs produced by CDMSBats, so our model begins with a baseline understanding of what the correct outputs are. Moreover, ML can be used to solve either classification or regression problems. Classification involves qualitative variables in the form of classes or categories, so a classification problem could be allocating particles into different categories based on their characteristics. Regression problems are interested in the prediction of quantitative variables, which in this case are the energy and position of an event.

The goal of training is to minimize (i.e., find the global minimum of) the cost function, which (for supervised learning) measures the error between the prediction made by the model and the actual labeled data. The most common cost function used in regression models is the Mean Squared Error (MSE) [8]:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 \tag{1}$$

where n is the number of data points, y_i is the actual value, and $f(x_i)$ is the predicted value.

4.2 NxM Optimal Filters

Aditi Pradeep from the University of British Columbia has been a pioneer in this area within the SuperCDMS collaboration. She applied Principal Component Analysis (PCA), an unsupervised machine learning technique that reduces the dimensionality of the feature space by creating new features, known as "principal components," which are combinations of the original variables. Using a FullDMC dataset of 10,000 monoenergetic events at 1.0 keV from a HV detector biased at 100V, she aimed to reduce the dimensionality of the raw phonon signals produced by the 12 channels per event. By constructing a covariance matrix that captures the relationships between the signals, she selected 3 principal components with the highest correlation to the signals and generated 12x3 templates [11]. These templates were then passed through the optimal filtering in CDMSBats to generate 12x3 phonon amplitudes [4].

Using these amplitudes as the input and the position truth information as the output, she used two supervised ML techniques—Feed-forward Neural Network (FNN) and a Regularized Gradient Boosted Machine (RGBM)—to reconstruct the position of FullDMC events. Both these models produce better results than linear fits, with the RGBM outperforming the neural network [9].

[4] It is important to note that this NxM optimal filtering technique can only produce the phonon amplitudes.

4.3 Comparison between Different Neural Network Input Data

Arsalan Khan from the University of Toronto builds on Pradeep's work by optimizing the Feed-forward Neural Network. He trained the network not only using NxM phonon amplitudes but also experimented with the standard CDMSBats RQs. The RQs used are the phonon amplitudes (P*OFamps) and phonon time delays (P*OFdelay). Khan found that the FNN performs better when trained with 12x2 phonon amplitudes (24 inputs per event) compared to using the standard amplitude RQs (12 inputs per event) [11].

Moreover, he discovered that training with the time delay RQs alone outperforms both models trained using amplitudes. In fact, training with both phonon amplitudes and time delays yield the best result [11]. The improvement is most apparent in the Z position as evident in Fig, where the predicted and actual values are normalized, so a more diagonal line represents less discrepancy. Notably, the Z predictions obtained by training with phonon amplitudes (upper right figure) are almost flat, meaning that the network is facing challenges to identity and/or learn the information embedded in the amplitude RQs.

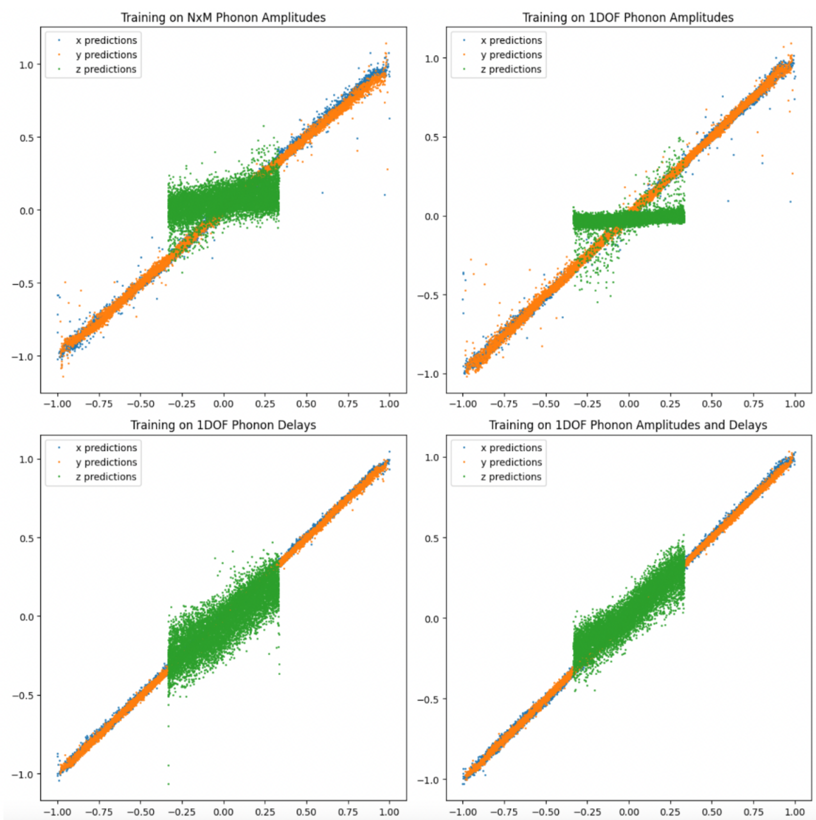

Figure 6: FNN training results using 12x2 phonon amplitudes, 12x1 phonon amplitudes, 12x1 phonon time delays, and 12x1 phonon amplitudes and time delays. The x-axes are the actual results, and the y-axes are the predicted values.

4.4 Ensemble Model Architecture & Results

Working parallel to Khan, Geting Qin built an ensemble neural network (ENN) that combines the outcomes of three baseline models [12]. She used the same FullDMC monoenergetic dataset and the amplitude and timeline RQs from the 12 channels as the inputs into each baseline model [5]. The outputs are the (X, Y, Z) positions measured with respect to the center of the detector in Cartesian coordinates as well as the energy

[5]These inputs are normalized by dividing each amplitude/time delay by the maximum of the absolute value of the phonon amplitude/time delay in each channel.

deposition of each event [6]. 80% of the 10,000 events are used as the training set, and the remaining 20% as the test set. This split is necessary to evaluate the performance of the network by testing it with data it has not seen before. The performance is quantified by calculating the standard deviation of the residual (Predicted − Actual), and this value is referred to as the resolution.

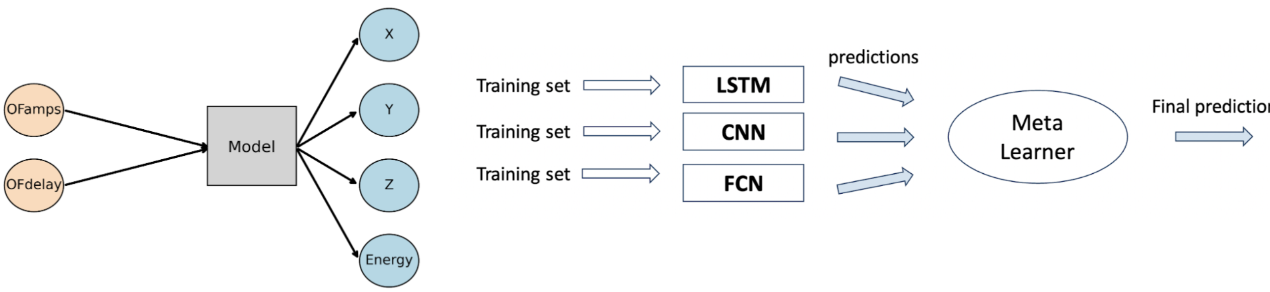

Figure 7: Left: Flowchart illustrating the RQ inputs and position and energy outputs of the model. Right: Flowchart representing the ensemble model structure [12].

Qin selected a Long Short-Term Memory (LSTM) network, a Convolutional Neural Network (CNN), and a Fully Connected Network (FCN) as her base-line models[7] due to their common application in experimental HEP [12]. Qin used the stacking method to combine the baseline models. After each model made predictions on the training set separately, a meta-learner compared the stacked predictions to the true positions and used linear regression to select the model that minimizes the residual for each event. Qin found that stacking all three models produced the best results, which suggests that each model picked up different features in the data. The ensemble model's prediction of event locations are summarized in Figure 8.

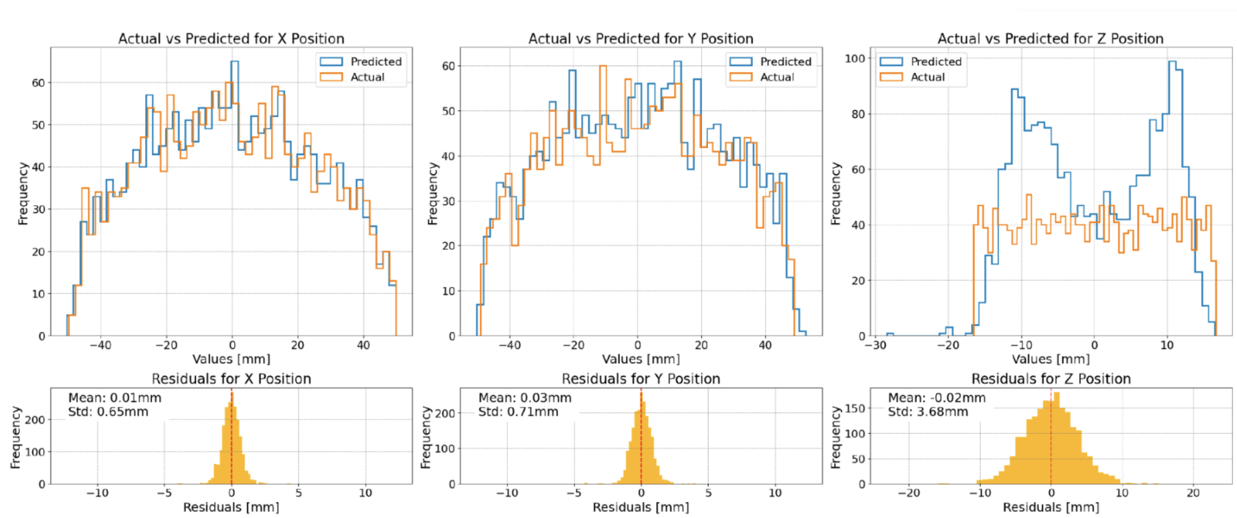

Figure 8: Top: Histograms of the actual (in orange) and predicted (in blue) values for X, Y, and Z positions for Qin's ensemble model. Bottom: Histograms of the residuals for three positions with the mean and standard deviation labelled [12].

[6]The position values are normalized to range from 0 to 1 [13].

[7]A brief overview of each model and their performance in position reconstruction is provided in Appendix section ??, but one should refer to Qin's report for a detailed description of their architecture as well as access to her code on Github.

The ensemble model can achieve a resolution of 0.65 mm for the X position, 0.71 mm for Y, and 3.68 mm for Z [8]. The distributions of X and Y predictions closely resemble the distributions of true positions, while the distribution of Z prediction has two apparent peaks close to the surfaces of the detector. This problem will be discussed more in Section 6.

It is important to understand what information the network is extracting from the RQs, both as a process of validation and optimization. One avenue to investigate is the correlation between Z position and the time delay RQ, which significantly improves the model's performance as an input, suggesting that it introduces additional information. This allows us to verify whether the network's predictions are based on sensible features in our data. If additional features are found, we can include them as inputs into our model to improve its performance. Furthermore, characteristics of the RQs can inform us on how the DMC is modelling detector response, and any unexpected features may point to shortcomings in the simulation. These features can then be directly compared with those of the experimental data taken at the (CUTE) facility during Tower 3 testing.

4.5 Radial Position Dependence of RQs

To discern what our network is learning from the input RQs, we must first understand their physical meaning. P*OFamps is the best-fit amplitude of a raw pulse when fitted to a template in CDMSBats, measured in units of amperes (A). This amplitude is proportional to the phonon energy of an event, as higher energy causes the quasiparticles generated from phonon-induced Cooper pair breaking to increase the temperature of the TES (Transition Edge Sensor) more significantly (see Experimental Design section). P*OFdelay is the horizontal delay of the raw pulse from the global trigger, measured in units of seconds (s). Since the x-axis of a raw pulse corresponds to time bins, this RQ provides the timing information of the an event. The trigger point marks the moment a signal is detected, as recognized by electronics in real-life. The time delay is therefore a measure of when phonons arrive at a channel [9].

Warren Perry from the University of Toronto conducted a validation of the position dependence of the RQs by plotting a 2D histogram of the X-Y true positions on the detector layout, with the color scale set to the amplitude or time delay measured by a channel [13]. He hypothesized that when setting the color scale to PBS1OFamps (amplitude recorded by PBS1), that particular channel should "light up", i.e., the events with position corresponding to the PBS1 channel should record the highest amplitude. The opposite behavior is expected for time delay, where the channel should "darken" as events with corresponding position should record the lowest time delay. This is precisely what Perry observed for all 12 channels (see Fig), concluding that the RQs contain dependence on radial position.

[8] All three baseline models attained excellent energy resolutions since all events are simulated to have an energy of 1.0 keV.

[9] Triggering is modelled differently in simulations, and there is ongoing work within the collaboration to investigate its behavior. Specifically, the trigger offset is not accurately modelled, causing simulated data to produce time delay RQs that lack physical meaning. Thus, instead of examining individual time delays, this study deals with relative timing information exclusively.

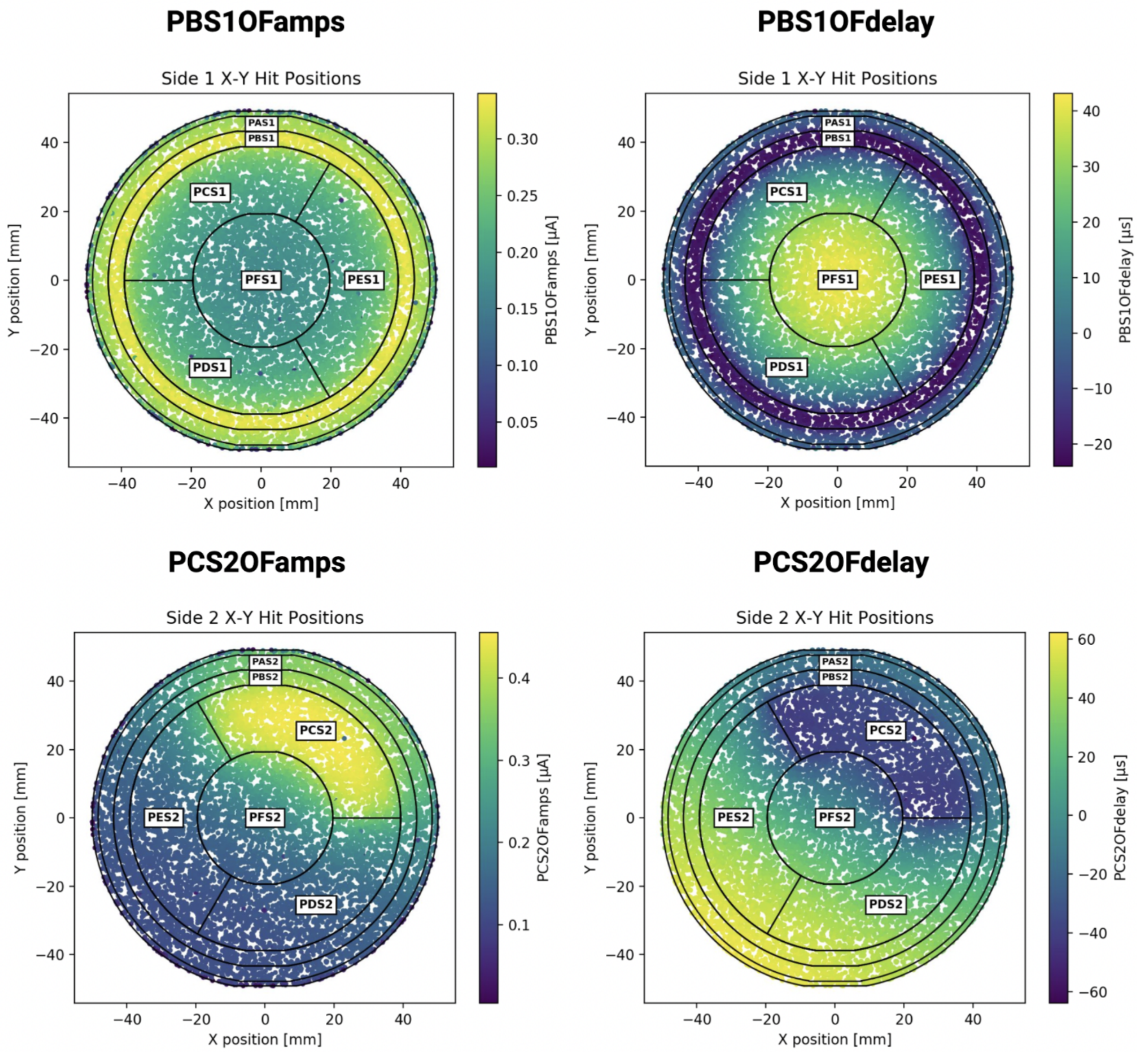

Figure 9: 2D histograms of radial event positions with the color scale set to amplitude or time delay measured by a channel (Top: PBS1, Bottom: PCS2). The channel layout on both sides are drawn and labelled [14].

5 Channel Pair Analysis

5.1 Breakdown of Max Amplitude & Min Time Delay Channel Pairs

To investigate the relationship between the RQs and Z position, it is more useful to examine top and bottom channel pairs instead of individual channels for a single event. If an event occurs near the top of the detector, it takes less time for the emitted phonons to reach the top surface, so the top channels should measure lower time delay than the bottom channels. Moreover, since the phonons must travel a greater distance to reach the bottom surface, they are more likely to collide with the bare side of the detector where there are no TESs. These phonons are absorbed by the crystal surface and experience anharmonic downconversion, where each phonon decays into two lower energy phonons that reflect off of the surface [16]. Thus, the top channels should also measure higher amplitude values than the bottom channels.

In order to find the most relevant channels, I identified the channels that measure the maximum amplitude across the top and bottom sides, respectively, and called them the “maximum amplitude channel pair” for each event. Then, I identified the channels that measure the minimum time delay across each side and called them the “minimum time delay channel pair”. Due to the clear radial position dependence of amplitude and time delay, these channels should correspond to where the energy deposition happened on the X-Y plane.

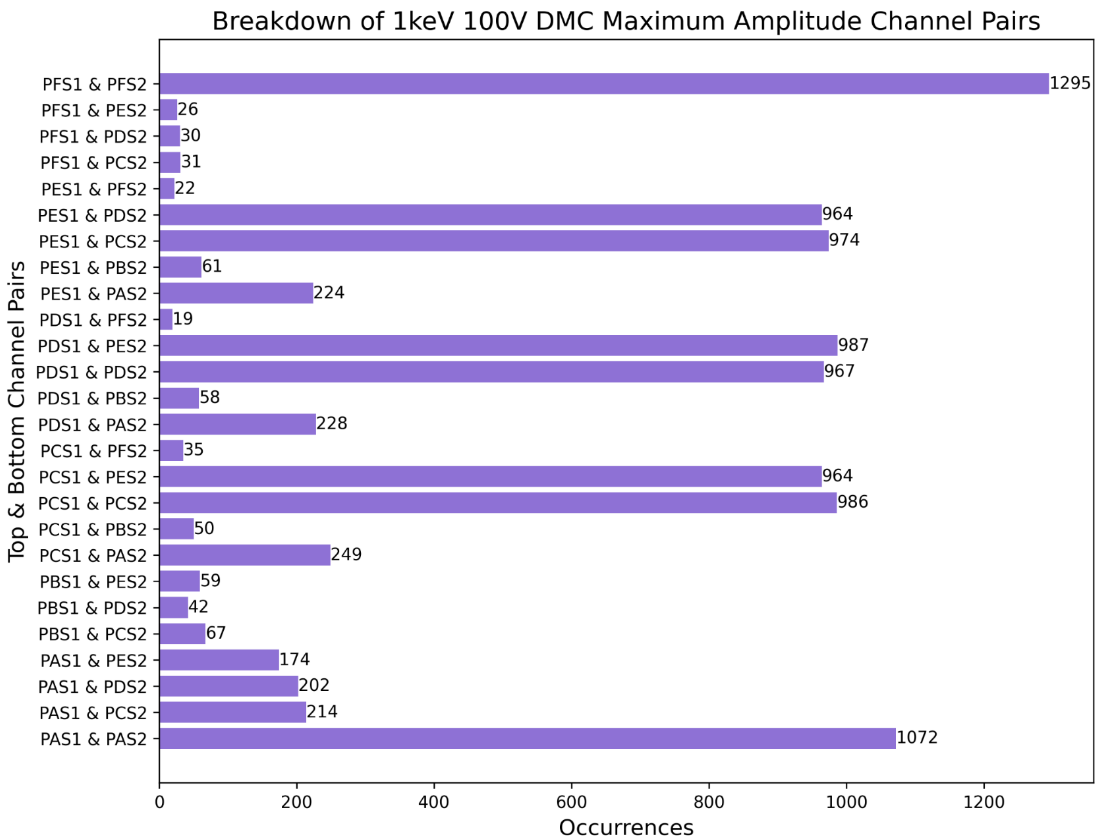

Figure 10: Bar plot showing the number of events each channel pair record maximum amplitude across their respective sides in the 1keV 100V DMC sample.

Figure 10 shows a bar graph of the breakdown of maximum amplitude channel pairs that confirms this hypothesis. The two channel pairs with the most occurrences are the outer ring channels (PAS1 & PAS2) and the inner core (PFS1 & PFS2), which are directly on top of each other. The other six channel pairs with more than 900 occurrences are combinations of the wedge shaped channels (PCS1 & PCS2, PCS1 & PES2, PDS1 & PDS2, PDS1 & PES2, PES1 & PCS2, PES1 & PDS2). These channel pairs are consistent with the offset between top and bottom wedge channels, where for instance, PCS1 lies above PCS2 and PES2. Channel pairs such as PCS1 & PAS2 may appear odd at first glance, since channel C and channel A do not directly neighbor each other and are separated by channel B. However, since electrons drift along the crystal valleys rather than traveling straight up, the Luke phonons emitted can reach the outer ring channel although the energy deposition may have happened at a radial position that corresponds to channel C.

Notably, the second outer ring channel pair (PBS1 & PBS2) is absent, and there is a general lack of channel pairs involving the B channels. This is likely due to the detector’s channel layout and geometry.

Channel B is adjacent to the maximum number of channels—one outer channel and three wedge channels. However, unlike channel F, which also neighbors three wedge channels, channel B's ring shape results in a larger region near its edge. Consequently, channel B tends to share the maximum amount of energy with neighboring channels. It is unlikely to record the maximum amplitude across all channels on one side for a given event, as one of its neighboring channels is more likely to do so. Nonetheless, the overall energy measured by channel B should be relatively high (see Figure 23), and it is expected to register the second-highest amplitude in a significant number of events, confirmed by Fig 11. This phenomenon can also be observed in the plots Perry obtained for the preliminary position dependence study. In Figure 9, PBS1OFamps range from $\sim$ 0.05 to $\sim$0.30 μA, whereas PCS2OFamps range from $\sim$ 0.1 to $\sim$0.4 μA. In the top left plot, although events with positions corresponding to PBS1 has the brightest color, locations far away from the B channel also sees PBS1OFamps $> 0.15 \mu A$. In the bottom left plot, however, locations furthest away from the C channel are measured to have PCS2OFamps $< 0.10 \mu A$.

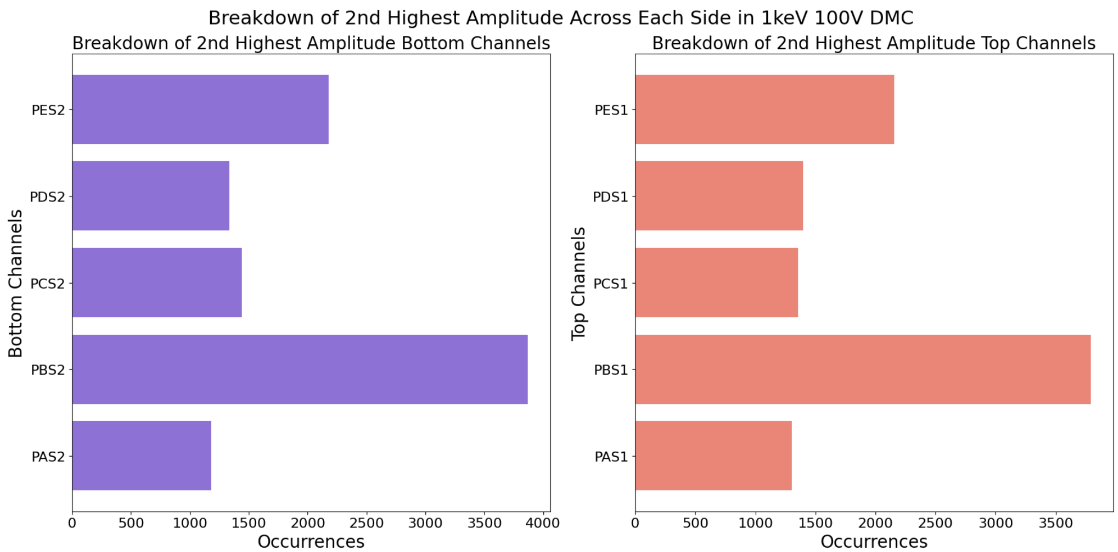

Figure 11: Two bar plots showing the number of occurrences a channel records the 2nd highest amplitude across its respective side in the 1keV 100V DMC sample.

The bar graph for minimum time delay channel pairs displays a similar trend, with PAS1 & PAS2 and PFS1 & PFS2 as the two most commonly occurring pairs. A few differences can be observed. First, the number of times PAS1 & PAS2 record the minimum time delay overwhelmingly exceeds that of the other channel pairs. Second, all the channel pairs of channels that do not align radially have much fewer occurrences, with the most frequent being PCS1 & PAS2 that record minimum time delay for 20 events. The absence of PBS1 & PBS2 here makes less sense than the maximum amplitude case as energy sharing does not affect time delay. Events that have PBS1 & PBS2 measure minimum time delay may have been misattributed to PAS1 & PAS2, but it is unclear why.

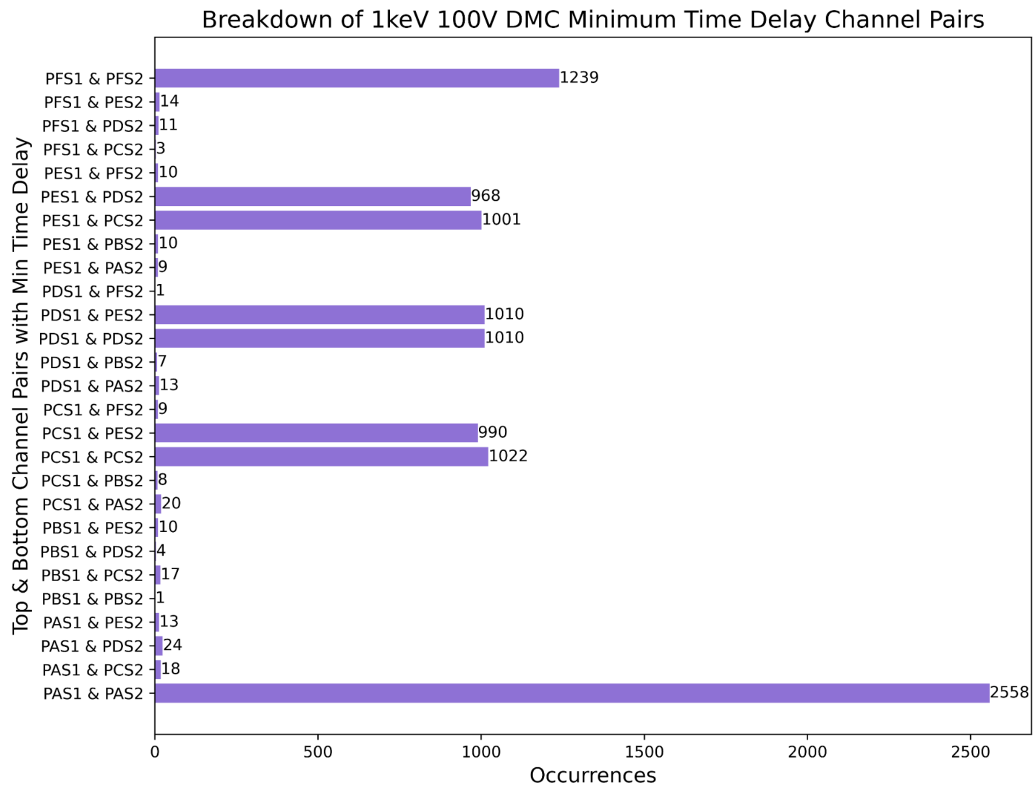

Figure 12: Bar plot showing the number of events each channel pair record minimum time delay across their respective sides in the 1keV 100V DMC sample.

A more direct comparison of the two bar graphs involves counting the number of events with matching top and bottom channels when measuring both maximum amplitude and minimum time delay on their respective sides. We found that 7,980 events have ”matching” channel pairs, representing approximately 80% of the 10,000 events. This result reinforces the relationship between amplitude and time delay RQs in interpreting the physics of phonon propagation. Our hypothesis that higher energy corresponds to shorter travel time is supported by the observed correlation between maximum amplitude and minimum time delay.

5.2 Position Dependence of Matched Events Using DMC True Position

Since DMC contains position information of the energy depositions, we can inspect the position dependence of events with matching channel pairs. The two top graphs in Figure 16 shows the locations of “matched” and “unmatched” events, and the bottom graphs show the same positions with the detector channel layout overlaid on top. Recall that there are 7980 events with the same channel pair recording maximum amplitude and minimum time delay, and from the left graphs, they occur uniformly within the inner core channel and the wedge channels. From the right graphs, it is clear that at the channel boundaries and certain locations in the two outer ring channels, a mismatch occurs.

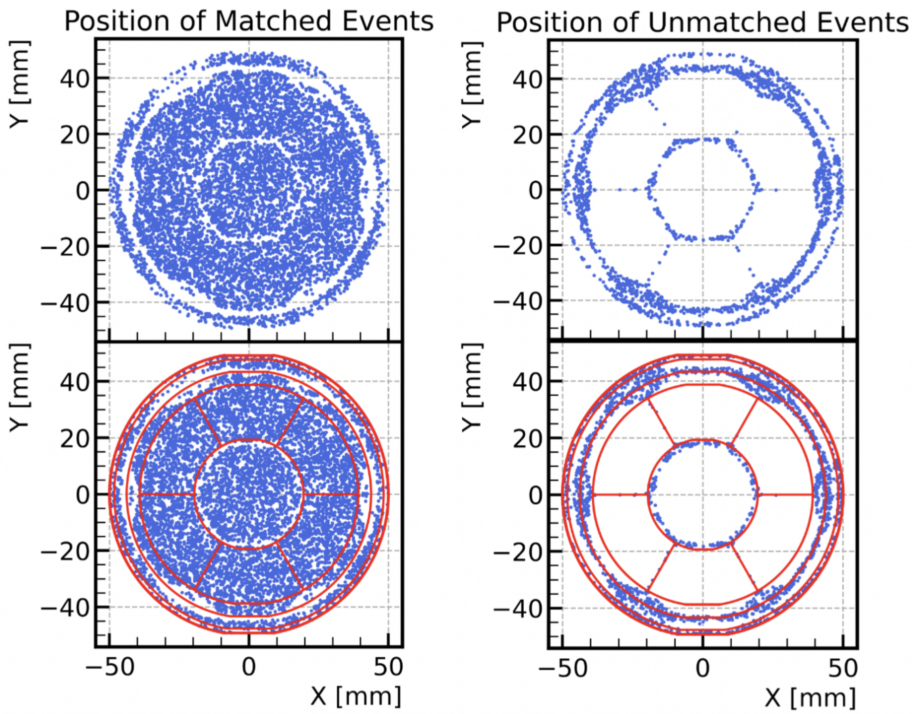

Figure 13: Top: position of events with/without matching channel pairs. Bottom: the same plots with channel boundaries on both sides drawn in red.

The mismatch happens where the number of neighboring channels differ between the two sides of the detector. For instance, the six inner wedge channel edges and a hexagonal shape are traced out. These are where a channel boundary occurs on side 1 but not on side 2 due to the rotation of wedge channels. For instance, on side 1 the majority of energy may be shared between two wedge channels, but on side 2 most of the energy is absorbed by one wedge channel. These events may have the same minimum time delay channel pair but different maximum amplitude pair.

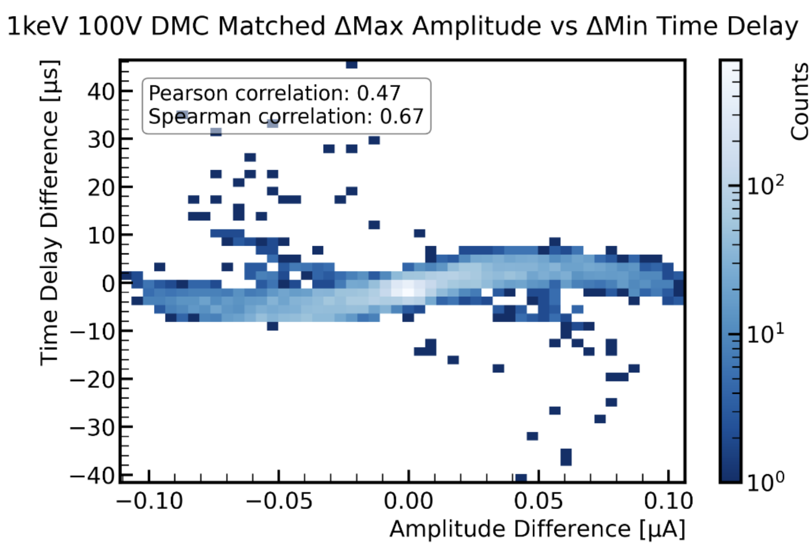

Figure 14: histogram of Δmax amplitude vs. Δmin time delay for events with matching maximum amplitude and minimum time delay channel pairs, with the Pearson and Spearman correlations labelled.

From Figure 14, it is evident that the central region contains the highest amount of counts as well as the highest correlation between the parameters. In fact, if we perform a cut at $|\Delta\text{time delay}| <= 8\mu s$ and $|\Delta\text{amplitude}| <= 0.050\mu A$, the correlation coefficients of these events are Pearson = 0.72 and Spearman = 0.68. We can also identify the two tailing ends in the graph with $|\Delta\text{time delay}| >= 10\mu s$.

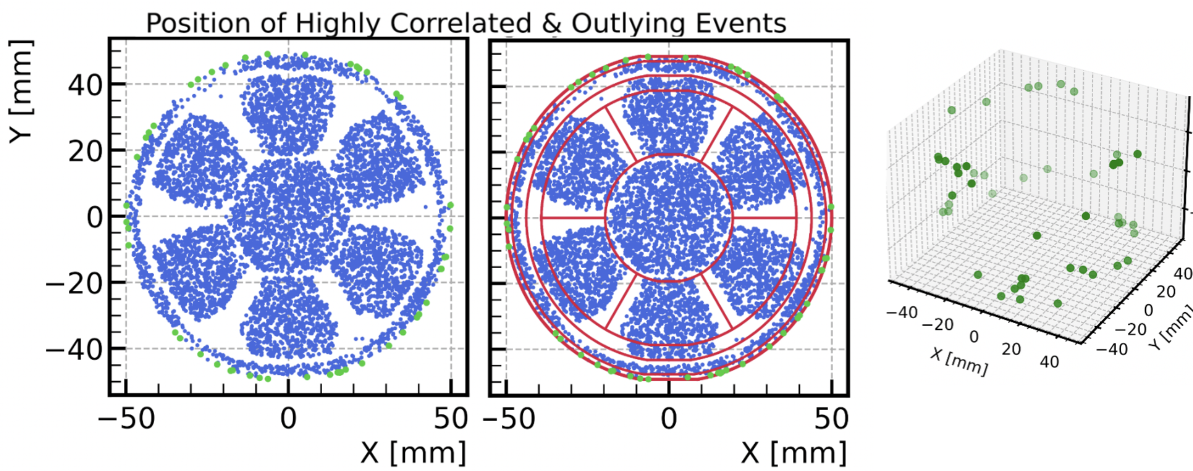

Figure 15: Left: Radial position of events with $|\Delta\text{time delay}| <= 8\mu s$ and $|\Delta\text{amplitude}| <= 0.050\mu A$ (in blue) and event with $|\Delta\text{amplitude}| <= 0.050\mu A$ (in green) as well as the channel boundaries of both sides overlaid on top in red. Right: 3D position of events with $|\Delta\text{amplitude}| <= 0.050\mu A$.

Figure 15 shows the position of the highly correlated events in the center and the outlying events at the top left and bottom right of Figure 14. The outlying events (in green) all occur at the radial edge and towards the top and bottom of the detector. These are places where phonons get absorbed by the sensors on one side of the detector very quickly compared to the other side, causing the energy depositions to have large $|\Delta\text{time delay}|$ magnitudes.

The highly correlated events (in blue) form a pattern with empty band gaps at channel boundaries that are especially apparent at the six wedge channel boundaries. The widths of these band gaps can be a measure of energy sharing, which causes the mismatch of maximum amplitude and minimum time delay. This method can only be applied to simulated samples where the true position is available, but it may serve as a validation for the reconstructed positions using ML techniques.

5.3 Correlations Between Amplitude/Time Delay Difference and Z

Having identified two channel pairs for each event, we can compute the difference between the amplitude and time delay values of these channels (top - bottom) and find their correlations with the Z position of the event. Consider an event occurring near the top of the detector. If its maximum amplitude channel pair is PFS1 and PFS2, indicating that it is near the radial center of the detector, then PFS1 should measure a larger amplitude than PFS2, making the difference between these amplitudes positive. This difference, denoted as Δmax amplitude, measures the distance from the energy deposition to the top and bottom sides, thereby revealing the Z position of the event. Similarly, we can find Δmin time delay for each event.

Figure 16 shows the 2D histograms of Δmax amplitude vs. Z and Δmin time delay vs. Z, along with two correlation coefficients. These coefficients range between -1 and 1, describing the strength and direction of the correlation (with ± 1 being the strongest, 0 being the weakest; >0 indicating a positive correlation, <0 indicating a negative correlation). The Pearson coefficient captures linear relationships, whereas the Spearman coefficient depicts monotonic relationships, which can be either linear or nonlinear.

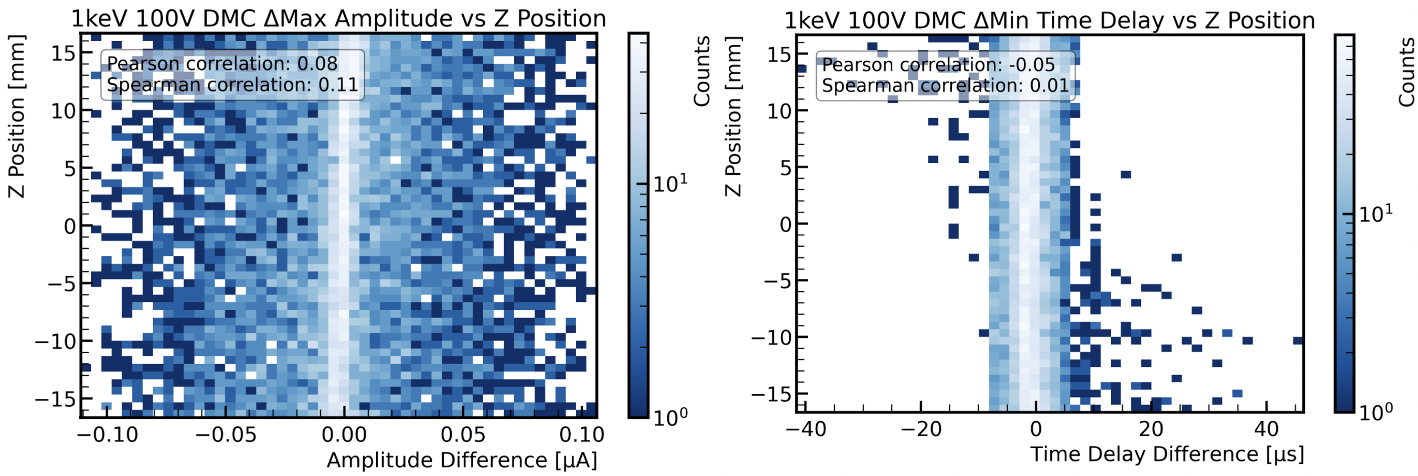

Figure 16: Left: 2D Histogram of Z position vs Δmax amplitude with the Pearson and Spearman correlation coefficients labelled. Right: 2D Histogram of Z position vs Δmin time delay with the Pearson and Spearman correlation coefficients labelled.

We observe that Δmax amplitude values range from -0.10 to 0.10 microamps, while Δmin time delay values are mostly between -10 and 10 microseconds, with some outlying events at higher and lower Z positions. To assess if our detector has sufficient timing resolution to capture the Z position, we can estimate the travel time of phonons. Given that phonons propagate in the crystal at approximately 8000 m/s, the travel time for a phonon from one surface of the detector to the other is calculated as follows:

$$\Delta t = \frac{3.33 \times 10^{-2}\ \text{m}}{8 \times 10^{3}\ \text{m/s}} = 4 \times 10^{-6}\ \text{s} = 4 \text{ microseconds (assuming phonons travel vertically without any deviation)}$$

which is the same order of magnitude as Δmin time delay. We expect the travel time to be longer in reality due to phonons not traveling in a straight line and bouncing off the sidewalls of the detector. Thus, information about Z should be encoded in Δmin time delay. However, these histograms show no observable trend in Δmax amplitude/Δmin time delay as Z varies. The correlation coefficients are also both minimal for Δmax amplitude/Δmin time delay.

Two more parameters can be found as each maximum amplitude channel pair also measures time delay, and each minimum time delay channel pair corresponds also to amplitude values. Thus, we can find the differences between the time delay values recorded by the maximum amplitude channel pairs (Δtime delay at max amplitude) as well as the differences between amplitudes recorded by the minimum time delay channel pairs (Δ amplitude at min time delay). The correlations between Δamplitude at min time delay and Z are found to be Pearson = 0.08, Spearman = 0.14, and the correlations between Δtime delay at max amplitude and Z are Pearson = -0.09, Spearman = -0.03. These results are not much stronger than the previous correlations, and they do not explain why inputting time delay into our model improves its performance.

One possibility is that the overall correlation with Z is obscured because some channel pairs have stronger correlations with Z while others do not. To address this, I focused on the eight channel pairs with more than 900 occurrences, as they contribute the most to the overall correlation. I then calculated the correlations between amplitude and time delay differences and Z for events involving these channel pairs.

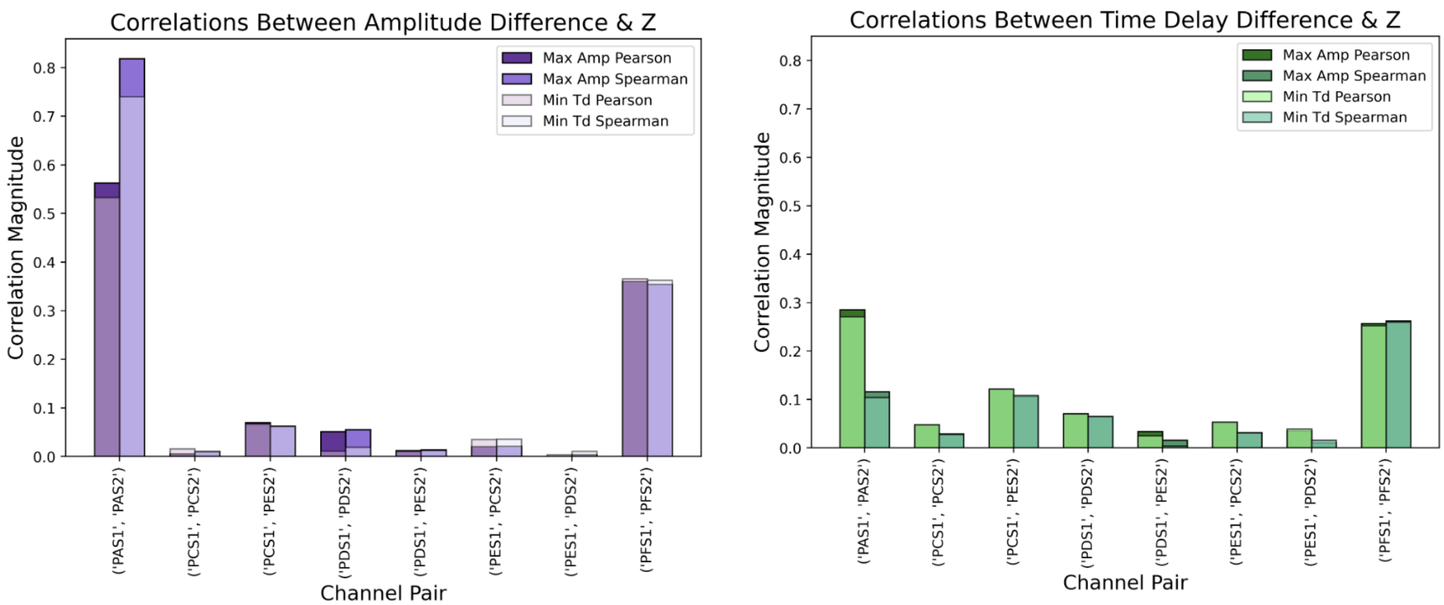

Figure 17: Left: Bar plot of the absolute value of Pearson and Spearman coefficient correlations between Δmax amplitude and Z position (dark purple) and Δamplitude at min time delay and Z position (light purple) of the 8 maximum amplitude channel pairs with more than 900 occurrences. Right: Δmin time delay and Z position (dark green) and Δtime delay at max amplitude and Z position (light green) of the 8 minimum time delay channel pairs with more than 900 occurrences.

Figure 17 summarizes the correlations found for two sets of amplitude differences (Δmax amplitude and Δamplitude at min time delay) and two sets of time delay differences (Δmin time delay and Δtime delay at max amplitude) with Z. Note that the bar heights represent the magnitude (absolute value) of the correlation coefficients. It's evident that different channel pairs yield varying levels of correlation, with the PAS1 & PAS2 and PFS1 & PFS2 pairs showing the strongest correlation with Z. The six wedge channel pairs have much weaker correlations in comparison. The correlations between time delay differences and Z are weaker for the outer ring and inner core channel pairs but more significant for the wedge channel pairs compared to the correlations between amplitude differences and Z. This may explain why training with time delays alone is more effective than training with amplitudes alone.

6 Optimization of Ensemble Model

6.1 Training with Cylindrical Coordinates

The first trial to optimize the ensemble model involved transforming the output position data from Cartesian coordinates to cylindrical coordinates. Since the SuperCDMS detectors are shaped like cylinders, representing positions using cylindrical coordinates is more intuitive. As both Cartesian and cylindrical systems share the Z coordinate, the model's performance in predicting Z can be directly compared.

Qin and Khan previously trained models using the coordinates (R, Φ, Z) but did not achieve favorable results. Here, R is given by $\sqrt{X^2+Y^2}$ and Φ is given by $\arctan\left(\frac{Y}{X}\right)$ [10]. A key issue with this approach is that the Φ angle becomes indeterminate as R approaches zero. To address this, we can train the model using $(R, R\cdot\Phi, Z)$ instead. Unlike X and Y positions, the distributions of R and $R\cdot\Phi$ are non-uniform: the number of events increases proportionally with the radius, while $R\cdot\Phi$ peaks around zero.[11]

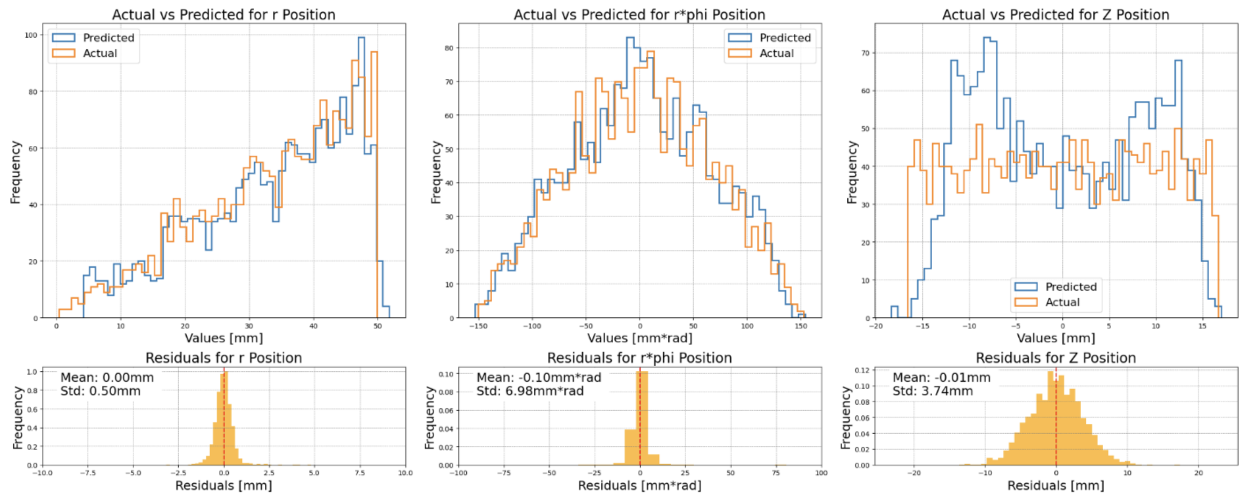

Figure 18: Top: Histograms of the actual (in orange) and predicted (in blue) values for R, $R\cdot\Phi$, and Z positions for Qin's ensemble model. Bottom: Histograms of the residuals for three positions with the mean and standard deviation labelled. [12]

The resolution in Z is around 3.6-3.8 mm, which does not improve over the original ensemble model. The X and Y position predictions reconstructed from R and R*Phi have resolutions that fluctuate around 3 mm, which are significantly worse than the results obtained before. This shows that training with cylindrical coordinates does not have any advantage over Cartesian coordinates.

6.2 Training with Weighted Amplitude and Time Delay Differences

Next, I incorporated amplitude and time delay differences into the network, in the hope that additional information about Z would improve the model's Z predictions. Two sets of correlations are found for amplitude differences and time delay differences with Z for each event (see Appendix section 10.2). Thus, one of Δmax amplitude and Δamplitude at min time delay as well as one of Δmin time delay and Δtime delay at max amplitude are added to the input data. I selected the two differences based on which has greater correlation with Z and multiplied each difference by the corresponding correlation. Then, I input the normalized difference at an index after the 24 RQs [12]. There are a total of 52 indices that represent the amplitude and time delay differences of each of the 26 channel pairs. For instance, if PAS1 & PAS2 records maximum

[10]The inverse tangent here is the four quadrant version, i.e., $arctan2$ in Python, where the output ranges from $-\pi$ to π. This is important because both X and Y values can be either positive or negative.

[11]Qin trained models with $(\Phi, R\cdot\Phi, Z)$ that achieved similar resolutions.

[12]These inputs are normalized by max-min normalization [9].

amplitude on their respective sides for an event and Δmax amplitude has greater correlation with Z than Δamplitude at min time delay, then Δmax amplitude is inputted in the first index, and the seven subsequent indices are set to zero. A time delay difference is selected by the same process and inputted into one of the last 26 indices. The total input for each event has $24 + 52 = 76$ indices.

In order to accommodate the larger training set, adjustments to the baseline model hyperparameters are also necessary. Recall that Qin's ensemble model has two peaks in its Z predictions. They originate from the FCN model, meaning that the FCN's predictions in Z on the training data most closely resemble the corresponding true positions (see Appendix section 10.1). Increasing the number of epochs and the batch size to 64 cause two peaks to appear in the LSTM Z prediction. Figure 19 shows the emergence of the peaks that become apparent at around 400 epochs [13].

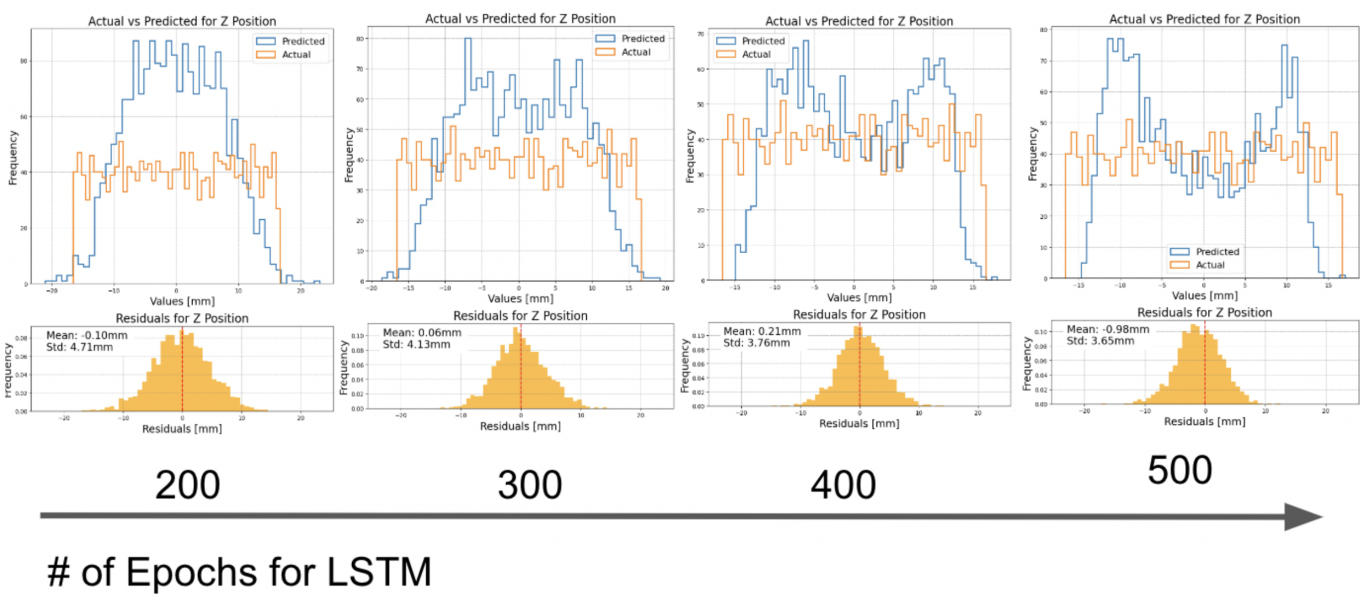

Figure 19: Histograms of the actual (in orange) and predicted (in blue) values for Z position of the LSTM model when training with 200, 300, 400 and 500 epochs.

The two peaks near the centers of the detector's top and bottom sides suggest that the network can distinguish between the two sides but lacks the sensitivity to further differentiate between surface and bulk events on each side [14].

Figure 20 shows the result of training the ensemble model with additional amplitude and time delay differences. The LSTM trained for 300 epochs with batch size 64, CNN trained for 200 epochs with batch size 64, and FCN trained for 500 epochs with batch size 128, all without early stopping. The resolutions in X and Y are around the same as those obtained from the original ensemble model, but the Z resolution decreased from 3.68 mm to 3.29 mm. Indeed, training with additional differences improves the network's ability to

[13] Plotting the training and validation loss curves, there is no sign of over-training at 500 epochs (see Appendix section 10.3).

[14] To test this hypothesis, we have developed classification models that categorize events into different Z slices of the detector. The best performing model can reach over 90 % accuracy when classifying events near the center of the detector vs. at the top and bottom surfaces of the detector.

learn Z position. However, the overall minimal correlations between amplitude/time delay differences and Z suggest that the RQs from this DMC sample may be fundamentally limited in terms of Z position sensitivity.

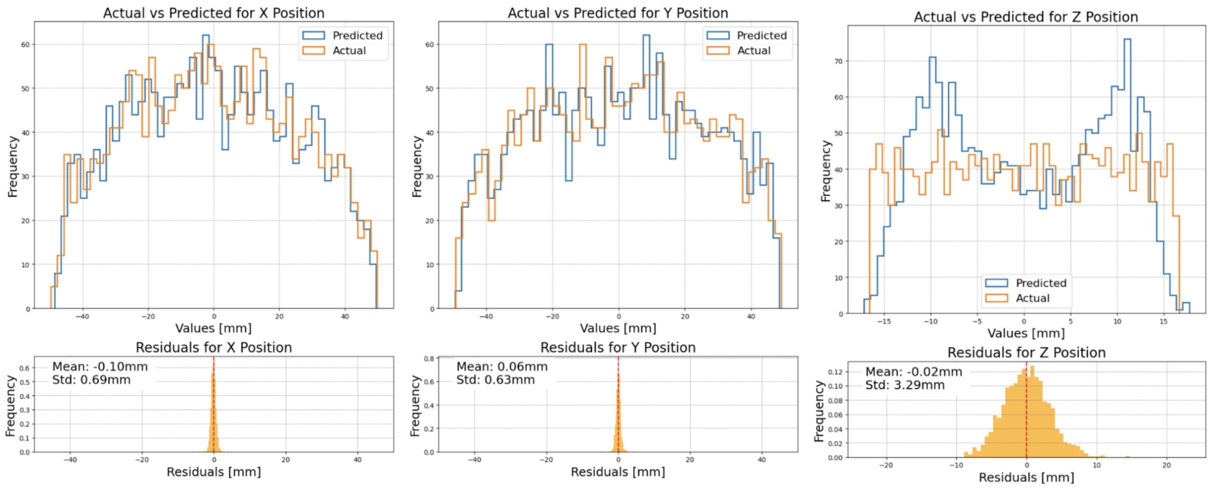

Figure 20: Top: Histograms of the actual (in orange) and predicted (in blue) values for X, Y, and Z positions for the optimized ensemble model trained with RQs and amplitude/time delay differences. Bottom: Histograms of the residuals for three positions with the mean and standard deviation labelled.

7 Comparison of RQ Features Between DMC and CUTE Data

7.1 Comparison of Channel Pair Characteristics

Run 37 during the third tower testing at the CUTE facility has yielded comparable data to the high voltage DMC sample analyzed in this study during germanium activation. Ge activation is a process that produces distinct peaks in the energy spectrum, which are used for energy calibration because they correspond to well-known and precise energy values associated with the decay of specific radioactive isotopes. In SuperCDMS Ge detectors, Ge-70 gets excited by neutron capture and becomes Ge-71, an unstable isotope that decays into Ga-71 via electron capture, releasing phonons and Auger electrons [15]. Three activation lines are expected at 160 eV, 1.3 keV, and 10 keV and correspond to peaks in the total phonon amplitude RQ (PTOFamps) spectrum. 20 "Golden series" contain data with all three Ge activation peaks on detectors Z1 and Z3, both biased at 50V. This analysis focuses solely on data from Z1, as channel A on side 1 of Z3 is non-functional. Given that the DMC sample consists of monoenergetic events at 1 keV, the Ge calibration events around the 1.3 keV peak are the most relevant for comparison.

Before identifying the activation peak, preliminary cuts are applied to filter out undesirable events. During data taking, LEDs are flashed every 10 mins for 5s at 2 Hz to monitor stability, and these events must be removed when performing analysis. In order to look at physics events, Trigger Type must be set to 1. Lastly, PTOFamps below 0 are filtered out because some events have meaningless RQ values of -999999. From the left plot of Figure 21, three distinct peaks are observable, corresponding to the three Ge activation peaks. Applying a chi square cut of PTOFchisqLF < 3000 to remove high energy events, the right plot displays the two low energy peaks, with the 1.3 keV peak at $\sim 3.5\ \mu A$. 391 events from Z1 can be identified around the peak with PTOFamps between 2.5 μA and 4.0 μA.

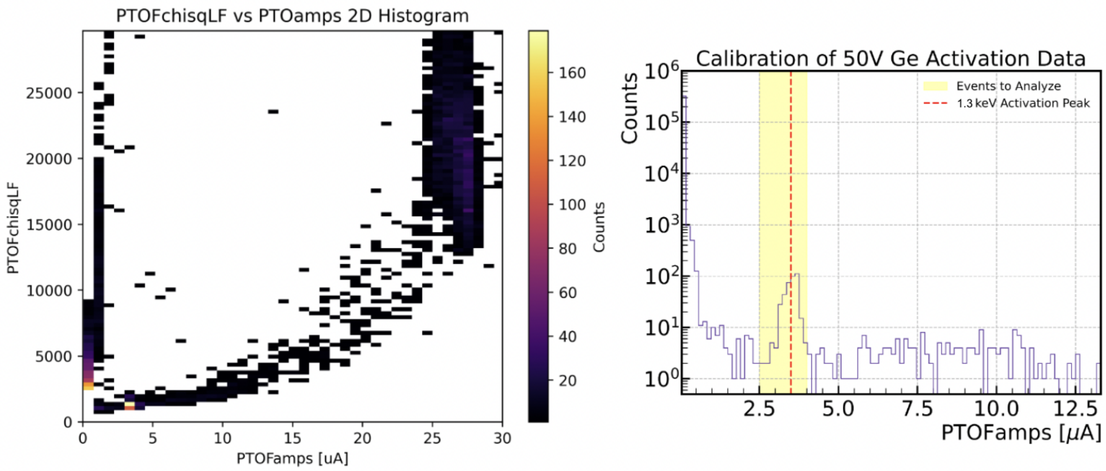

Figure 21: Left: 2D histogram of PTOFchisqLF vs PTOFamps after applying preliminary cuts. The color scale indicates the density of data points. Right: PTOFamps histogram after the chi square cut. The 1.3 keV activation peak is indicated by a red dotted line, and the events to analyze analysis are highlighted.

Using the same method as section 5, the maximum amplitude and minimum time delay channel pairs are identified for each event. Figure 22 shows the break down of maximum amplitude channel pairs.

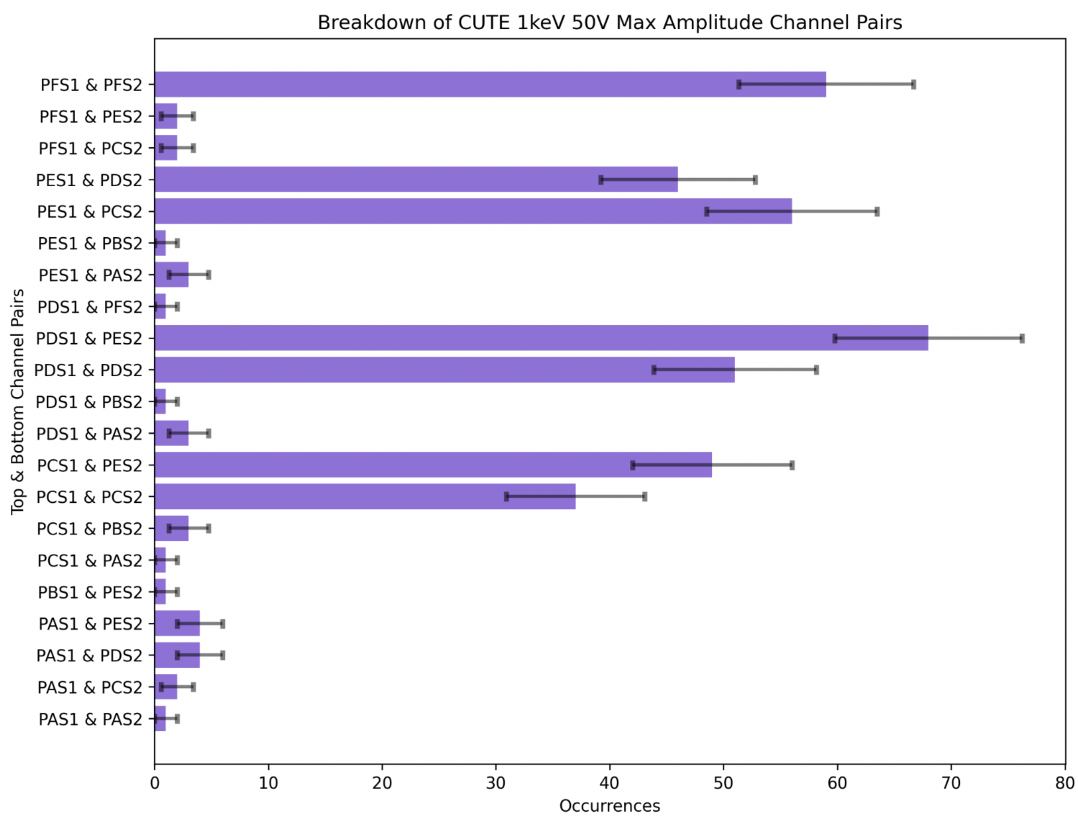

Figure 22: Bar plot showing the number of events each channel pair record maximum amplitude across their respective sides in the CUTE 1keV 50V Ge Activation data. Error bars are plotted instead of labeled occurrences due to the low amount of statistics.

In contrast to Figure 10, PAS1 & PAS2 is not a maximum amplitude channel pair for a high number of events. The other common channel pairs are all present. This discrepancy is concerning as it reveals that DMC does not model the energy absorption of the outer ring channel pair accurately. Figure 23 shows that the percentage of energy absorbed by the A channels in DMC have the highest discrepancy among all channels compared to the energy distribution of CUTE data.

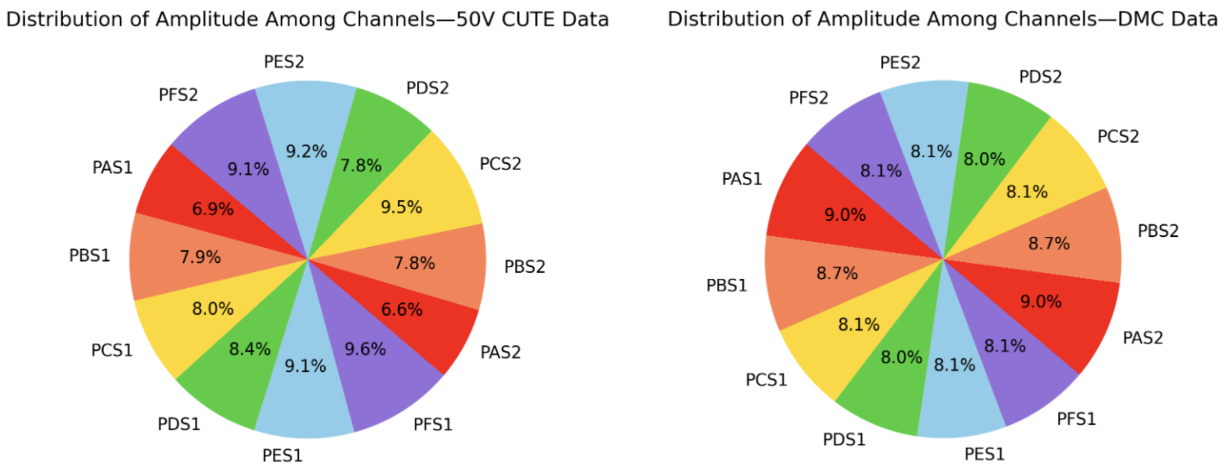

Figure 23: Pie charts showing the percentage of total amplitude each channel measures among all channels for the 50V CUTE data (left) and the 100V DMC data (right).

The breakdown of second highest amplitudes across each side closely resembles that of the DMC data, which further confirms that DMC does not reflect the reality of energy absorption in the A channels.

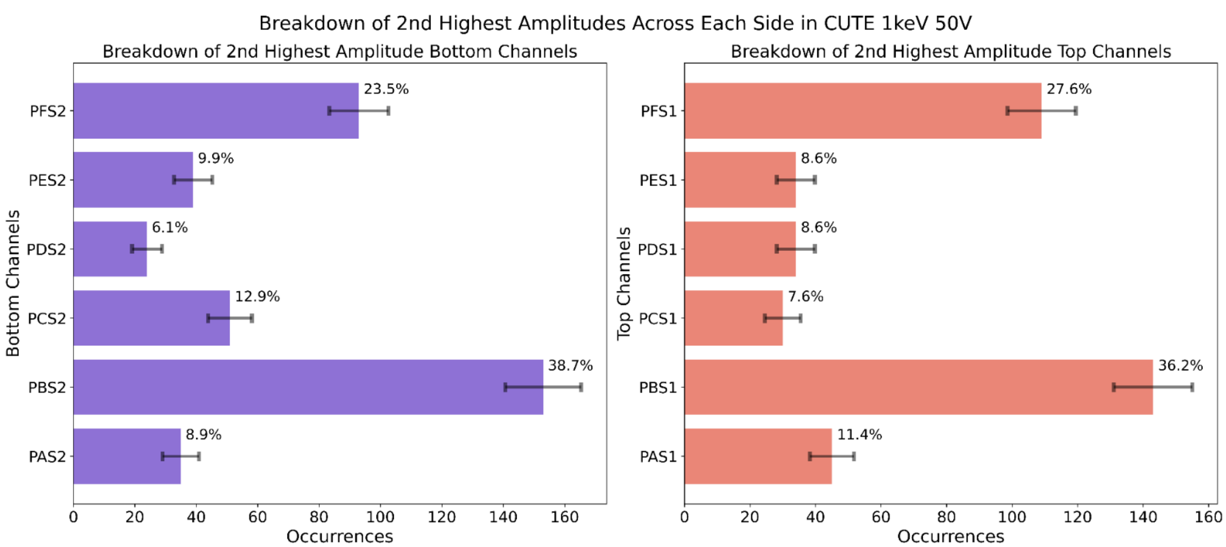

Figure 24: Two bar plots showing the number of occurrences a channel records the 2nd highest amplitude across its respective side in the 1keV 100V DMC sample. Error bars along with percentage labels are shown due to the low amount of statistics.

In opposition to channel pairs that measure maximum amplitude, PAS1 & PAS2 are among the most frequent minimum time delay channel pairs. This aligns with what we observed from the DMC sample (Figure 12). However, PBS1 & PBS2 is present in CUTE data but not in DMC, which verifies the speculation that the DMC sample is incorrectly modeling the behavior of B channels in terms of time delay.

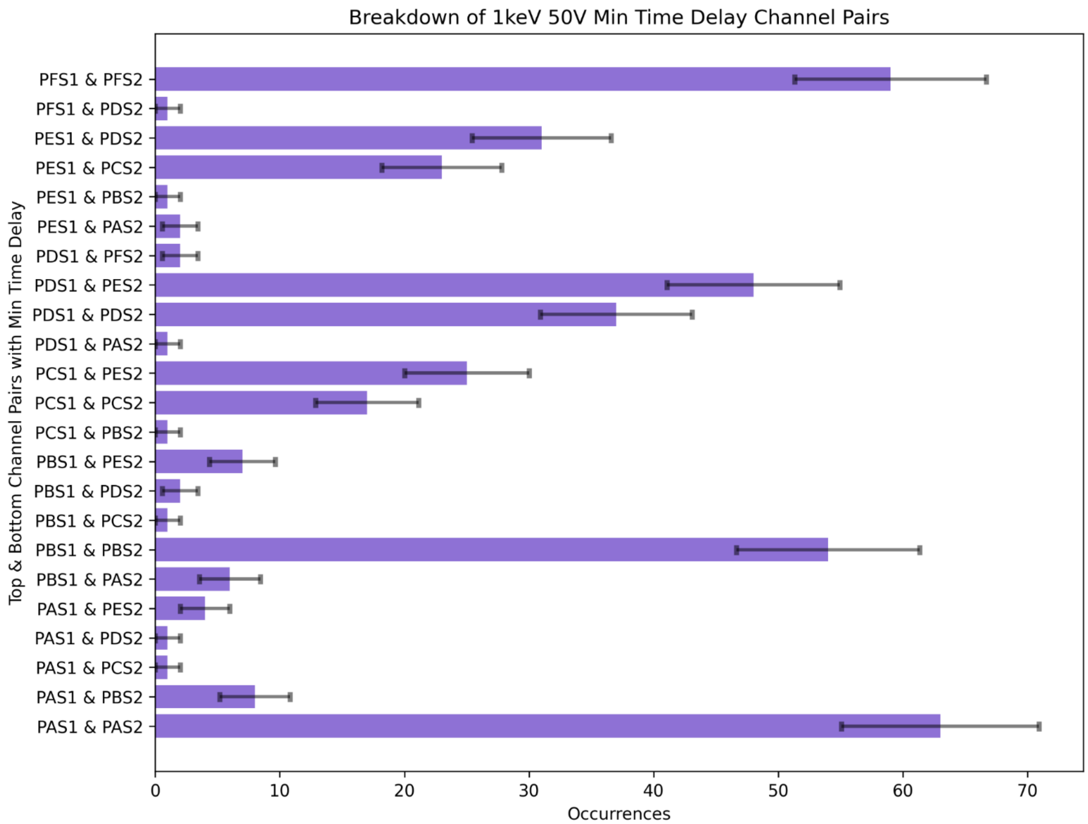

Figure 25: Bar plot showing the number of events each channel pair record minimum time delay across their respective sides in the CUTE 1keV 50V Ge Activation data. Error bars are plotted instead of labeled occurrences due to the low amount of statistics.

Evidently, there is a greater difference between the maximum amplitude and minimum time delay channel pair breakdowns for CUTE compared to the DMC sample. Thus, only 237 out of 391 events ($\sim 60\%$) have the same channel pair recording maximum amplitude and minimum time delay on their respective sides, which is 20% less than what was found for DMC.

7.2 Discussion of Discrepancies Between DMC and CUTE Data

Some discrepancy between the channel pairs of CUTE and DMC data is expected since the simulated detector is biased at 100V and the CUTE detector is biased at 50V. This voltage difference impacts the energy sharing among channels. At a lower voltage, the electric field strength is weaker, so the charge carriers drift more slowly and have more time to diffuse.

Moreover, there are many aspects of DMC that do not accurately reflect real-life physics that could explain its failure to model the behavior near the detector radial edge. For instance, the electric field is assumed to be uniform in the simulated detector, while in reality, it bends near the detector sidewalls [13]. To address this, members of the collaboration have developed Epot files over the summer that more accurately model the electric fields, enabling the generation of more realistic samples.

Similarly, the current DMC sample does not account for downconversion, which may result in a significant number of phonons failing to break Cooper pairs into quasiparticles, causing the overall energy spectrum to appear higher than in reality. However, the DMC does have the capability to simulate downconversion by adjusting the surface anharmonic decay coefficient, allowing for tests on how downconversion impacts phonon energy absorption near the radial edge [18]. Additionally, the simulation assumes phonon reflections off the sidewalls are perfectly specular, while in reality, these reflections are diffusive. This discrepancy alters phonon trajectories and affects the amount of energy recorded by the TESs. Indeed, the phonon collection efficiency (the ratio between measured phonon energy and the expected phonon energy) of the DMC sample is 100% for a majority of the events (see Appendix section 10.4).

Another factor contributing to this discrepancy is the modeling of TESs. Given the large number of TESs on the detector surfaces, simulating the response of each individual TES would extend the run time to impractical lengths. To address this, X TESs are randomly grouped into a single pseudo-TES, which has X times the true volume of tungsten, heat capacity, and other relevant properties [17]. The result is 49 pseudo-TESs per channel, and the location of which affect how accurately signal sharing is modeled, with the two outer ring channels being particularly impacted since the actual TES locations are more spread out. Thus, a more reasonable grouping of TESs may need to be developed to account for detector geometry.

8 Conclusion

Motivated by the goal of optimizing Qin's ensemble model performance, particularly for Z position accuracy, this study developed a method to identify channel pairs with maximum amplitude and minimum time delay for each event. The occurrences of channels pairs were shown to be consistent with the radial position dependence of amplitude and time delay RQs. The differences in these values, as recorded by each channel pair, revealed a Z position dependence when Pearson and Spearman correlation coefficients were calculated with respect to Z. By adjusting the baseline architecture and incorporating the amplitude and time delay differences into the input data, the Z resolution of the original ensemble model was improved to 3.29 mm. Notably, events where the same channel pair recorded both maximum amplitude and minimum time delay exhibited high correlations between amplitude and time delay differences, with the position of these events indicating the extent of energy sharing among channels. The overall minimal correlations found between RQ features and Z position as well as the two peaks in the distribution of Z predictions point to the inherent lack of Z sensitivity in the 1 keV 100V DMC sample.

It is important to recognize that the success of event reconstruction using ML is contingent on how closely our simulations resemble reality. Analyzing channel pairs allows for direct comparison between features in the simulated sample and experimental data. When applying this analysis to data from the CUTE facility, significant discrepancies were observed, particularly in the energy absorption by the outer ring channels. These discrepancies may be attributed to assumptions regarding the electric field, surface anharmonic downconversion, and TES grouping, although this list is not exhaustive. The next step is to generate DMC samples under conditions more comparable to the CUTE Ge activation data, allowing for a clearer identification of the factors causing our simulations to diverge from real data. We can then initiate an iterative process: training machine learning models on increasingly realistic samples, applying these models to analogous experimental data, and identifying further opportunities to enhance the DMC.

9 Acknowledgement

I am deeply grateful to my supervisor, Prof. Pekka Sinervo, for giving me the opportunity to join the SuperCDMS collaboration and for his unwavering guidance and support in both group meetings and one-on-one discussions. I am also indebted to Dr. Madeleine Zurowski and Warren Perry, who closely collaborated with me on this project. Their kindness and approachability were instrumental in helping me become familiar with the experiment and data analysis tools. I am fortunate to have worked alongside Geting (Janice) Qin, whose patience in answering my many questions was matched only by her friendship, and Matthew Penner, who generously shared his expertise in ML. I also extend my gratitude to the many other members of the collaboration, such as Stefan Zatschler, Mehar Sahato, and Jason Kahn (to name a few), who provided invaluable feedback and made my research experience truly special.

References

[1] SuperCDMS Collaboration, SuperCDMS SNOLAB Technical Design Report. https://confluence.slac.stanford.edu/display/CDMS/SuperCDMS+SNOLAB+Project?preview=/180554723/445874563/supercdms-snolab-fdr.pdf

[2] Leane, R. (2020). Indirect Detection of Dark Matter in the Galaxy (MIT-CTP/5199). [arXiv:2006.00513]

[3] Pyle, M. (2012). Optimizing the Design and Analysis of Cryogenic Semiconductor Dark Matter Detectors for Maximum Sensitivity [PhD Thesis, Stanford University]. https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/mpyle.pdf

[4] Zatschler, S. (2024). Status and prospects of the SuperCDMS Dark Matter Experiment [Presenter]. GUINEAPIG, Canada. https://indico.triumf.ca/event/521/contributions/5991/

[5] Karagiorgi, G., Kasieczka, G., Kravitz, S. et al. Machine learning in the search for new fundamental physics. Nat Rev Phys 4, 399–412 (2022). https://doi.org/10.1038/s42254-022-00455-1

[6] Winchell, J. (2023). Characterizing photon and neutron responses in CDMS detectors using real and simulated CF-252 data [PhD Thesis, Texas A&M University]. https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/winchell.pdf

[7] Agnese, R. *et al.* (2017). Projected sensitivity of the SuperCDMS SNOLAB experiment. Physical Review. D/Physical Review. D., 95(8). https://doi.org/10.1103/physrevd.95.082002

[8] Araz, J. Y. & Spannowsky, M. (2021). Combine and conquer: Event reconstruction with bayesian ensemble neural networks. J. High Energy Phys. [10.1007/jhep04(2021)296]

[9] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. In Springer texts in statistics. https://doi.org/10.1007/978-1-4614-7138-7

[10] Pradeep, A. (2023) Ml applications in the nxm filter. SLAC Conflu.

[11] Liu, Y., & Oser, S. (2022). The NxM Optimal Filter. SLAC Conflu.

[12] Khan, A. (2024). Neural Networks for Improved Dark Matter Detection: Advances in SNOLAB SuperCDMS Detector Resolution. SLAC Conflu.

[13] Qin, G. (2024). Ensemble Learning Model for Improving Position and Energy Resolution for SuperCDMS. SLAC Conflu.

[14] Perry, W. HV100mm Per-Channel Energy Collection in CDMSBats. SLAC Conflu.

[15] Nelson, J. (2024). Lowering backgrounds and thresholds in the search for light dark matter with SuperCDMS [PhD Thesis, University of Minnesota]. https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/nelson.pdf

[16] Kozorezov, A. G., Volkov, A. F., Wigmore, J. K., Peacock, A., Poelaert, A., & Hartog, R. D. (2000). Quasiparticle-phonon downconversion in nonequilibrium superconductors. Physical Review. B, Condensed Matter, 61(17), 11807–11819. https://doi.org/10.1103/physrevb.61.11807

[17] Kelsey, M., & Toback, D. (2022). Improving TES Simulation [Slide show; SLAC Conflu.]. SuperCDMS DMC Sub-Working Group Meeting.

[18] Paudel, S., Gaven, R., (2023) Surface Anharmonic Down Conversion Study. SLAC Conflu.

10 Appendix

10.1 Overview of ENN Baseline Model Architecture & Results

Model	STD in X [mm]	STD in Y [mm]	STD in Z [mm]	STD in E [keV]
LSTM	0.90	1.02	4.27	1.07×10^{-6}
CNN	2.93	2.30	5.66	1.15×10^{-3}
FCN	1.57	2.21	3.87	1.20×10^{-3}

Figure 26: Resolution in X, Y, Z, and Energy of the three baseline models of the ENN.

The LSTM network improves upon the traditional Recurrent Neural Network (RNN)'s difficulty in learning dependencies between data as inputs grow. In order to capture these connections, each input is converted into a sequence with shape (1, 24). Qin used grid search to settle on two LSTM layers with 150 LSTM cells and a batch size of 32. She set the model to train for 200 epochs with early stopping. This mechanism monitors the validation loss and stops the training if it has not decreased for a certain number ("patience") of consecutive epochs. The LSTM achieved resolutions of less than 1 mm in the X and Y positions, but the Z predictions appeared biased towards the center of the detector. Thus, she trained an additional Support Vector Machine (SVM) that trained on the residual of the LSTM's predictions in Z, which "flattened" out the predicted Z distribution to better resemble the quasi-uniform distribution of truth position.

The CNN model is typically used for image recognition by identifying low-level features in the input image and combining them to form high-level features. It contains convolutional layers made of filters that determine the presence of local features. Qin's model contains two 1D convolutional layers, some intermediate layers that enable data reduction and dropout (random clamping of 50% of hidden units to prevent overfitting), and a final output layer using a linear activation function [13]. She also used L2 regularization in the convolutional and intermediate layers, which strikes a balance between the accuracy of the fit and the smoothness of the output function. She found the best batch size to be 64 using grid search and set the model to train for 100 epochs with early stopping. The performance of CNN is worse than the LSTM with a resolution of 2-3 mm in X and Y.

Lastly, the FCN is a simple model that does not require a long training time. It consists of three dense layers where all the neurons are connected to every neuron in the previous layer and final output layer with a linear activation function. The best batch size was determined via grid search to be 32, and the model trained for a total of 100 epochs with early stopping. The FCN achieved better X and Y resolutions than the CNN and outperformed both the CNN and the LSTM in predicting Z. However, the predicted distribution in Z contains two peaks near the top and bottom detector surfaces (see Figure 27).

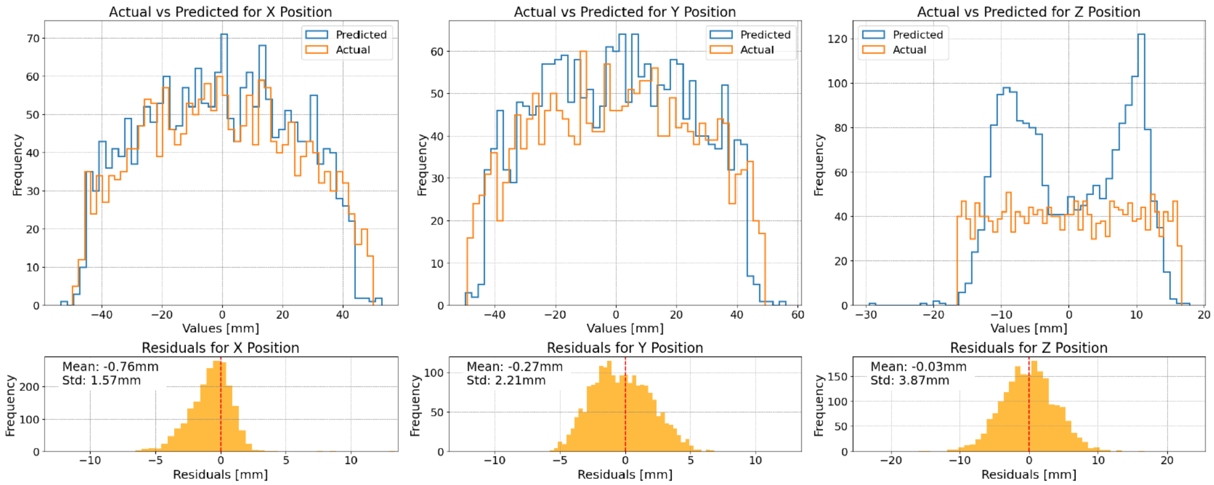

Figure 27: Top: Histograms of the actual (in orange) and predicted (in blue) values for X, Y, and Z positions for the FCN model. Bottom: Histograms of the residuals for three positions with the mean and standard deviation labelled [13].

10.2 Correlation Features Summary

Table 1: Correlations between Δmax amplitude and Z

Channel Pair	Pearson Correlation	Spearman Correlation
(PAS1, PAS2)	0.5630	0.8183
(PAS1, PCS2)	0.4732	0.4920
(PAS1, PDS2)	0.5153	0.5215
(PAS1, PES2)	0.3978	0.3610
(PBS1, PCS2)	0.2439	0.4035
(PBS1, PDS2)	-0.1206	0.0544
(PBS1, PES2)	0.2232	0.3108
(PCS1, PAS2)	0.4309	0.4640
(PCS1, PBS2)	0.1494	0.2057
(PCS1, PCS2)	0.0058	0.0101
(PCS1, PES2)	0.0693	0.0625
(PCS1, PFS2)	-0.2432	-0.2787
(PDS1, PAS2)	0.4954	0.5242
(PDS1, PBS2)	0.2153	0.3764
(PDS1, PDS2)	0.0511	0.0550
(PDS1, PES2)	-0.0120	-0.0139
(PDS1, PFS2)	0.1447	0.0474
(PES1, PAS2)	0.4793	0.4899
(PES1, PBS2)	0.1082	0.1750
(PES1, PCS2)	-0.0203	-0.0213
(PES1, PDS2)	-0.00003	-0.0039
(PES1, PFS2)	-0.0719	0.0898
(PFS1, PCS2)	0.1017	0.0931
(PFS1, PDS2)	0.2801	0.2957
(PFS1, PES2)	0.1287	0.2232
(PFS1, PFS2)	0.3603	0.3547

Table 2: Correlations between Δamplitude at min time delay and Z

Channel Pair	**Pearson Correlation**	**Spearman Correlation**
(PAS1, PAS2)	0.5330	0.7403
(PAS1, PCS2)	0.2917	0.1538
(PAS1, PDS2)	0.0395	-0.0052
(PAS1, PES2)	0.5304	0.4121
(PBS1, PBS2)	0	0
(PBS1, PCS2)	0.4640	0.6029
(PBS1, PDS2)	0.7657	0.4000
(PBS1, PES2)	0.7747	0.7455
(PCS1, PAS2)	0.4179	0.4015
(PCS1, PBS2)	-0.1507	0.1429
(PCS1, PCS2)	-0.0161	-0.0109
(PCS1, PES2)	0.0669	0.0621
(PCS1, PFS2)	0.1946	0.0833
(PDS1, PAS2)	0.3464	0.4890
(PDS1, PBS2)	0.6164	0.6786
(PDS1, PDS2)	0.0113	0.0190
(PDS1, PES2)	-0.0104	-0.0125
(PDS1, PFS2)	0	0
(PES1, PAS2)	0.6066	0.6000
(PES1, PBS2)	0.3294	0.2242
(PES1, PCS2)	-0.0353	-0.0362
(PES1, PDS2)	-0.0041	-0.0103
(PES1, PFS2)	0.5628	0.4909
(PFS1, PCS2)	-0.8962	-1.0000
(PFS1, PDS2)	-0.8083	-0.6273
(PFS1, PES2)	0.3797	0.4945
(PFS1, PFS2)	0.3663	0.3623

Table 3: Correlations between Δmin time delay and Z

Channel Pair	**Pearson Correlation**	**Spearman Correlation**
(PAS1, PAS2)	-0.2708	-0.1039
(PAS1, PCS2)	-0.3010	-0.0340
(PAS1, PDS2)	-0.0195	0.0276
(PAS1, PES2)	-0.6423	-0.5913
(PBS1, PBS2)	0	0
(PBS1, PCS2)	-0.3694	-0.5285
(PBS1, PDS2)	-0.3112	0.2108
(PBS1, PES2)	-0.3023	-0.4799
(PCS1, PAS2)	-0.5225	-0.3801
(PCS1, PBS2)	0.5217	0.6508
(PCS1, PCS2)	0.0484	0.0276
(PCS1, PES2)	0.1219	0.1082
(PCS1, PFS2)	-0.1009	-0.1419
(PDS1, PAS2)	-0.4338	-0.2782
(PDS1, PBS2)	0.6949	0.5777
(PDS1, PDS2)	0.0701	0.0649
(PDS1, PES2)	0.0245	0.0040
(PDS1, PFS2)	0	0
(PES1, PAS2)	-0.7030	-0.6109
(PES1, PBS2)	0.4280	0.3737
(PES1, PCS2)	0.0535	0.0314
(PES1, PDS2)	0.0389	0.0157
(PES1, PFS2)	0.1317	0.0824
(PFS1, PCS2)	0.9085	1.0000
(PFS1, PDS2)	0.2948	0.4374
(PFS1, PES2)	-0.1013	-0.0207
(PFS1, PFS2)	0.2521	0.2601

Table 4: Correlations between Δtime delay at max amplitude and Z

Channel Pair	Pearson Correlation	Spearman Correlation
(PAS1, PAS2)	-0.2853	-0.1163
(PAS1, PCS2)	-0.3871	-0.3906
(PAS1, PDS2)	-0.3790	-0.4003
(PAS1, PES2)	-0.2969	-0.3188
(PBS1, PCS2)	-0.0117	-0.1552
(PBS1, PDS2)	0.1033	-0.0928
(PBS1, PES2)	-0.1537	-0.2935
(PCS1, PAS2)	-0.2910	-0.2231
(PCS1, PBS2)	0.1537	0.2803
(PCS1, PCS2)	0.0471	0.0290
(PCS1, PES2)	0.1211	0.1063
(PCS1, PFS2)	0.2313	0.2863
(PDS1, PAS2)	-0.3207	-0.3503
(PDS1, PBS2)	0.2146	0.2385
(PDS1, PDS2)	0.0706	0.0651
(PDS1, PES2)	0.0338	0.0157
(PDS1, PFS2)	-0.0040	-0.0628
(PES1, PAS2)	-0.3609	-0.3844
(PES1, PBS2)	0.1678	0.1521
(PES1, PCS2)	0.0521	0.0312
(PES1, PDS2)	0.0352	0.0106
(PES1, PFS2)	0.2153	0.2062
(PFS1, PCS2)	-0.1008	-0.0923
(PFS1, PDS2)	-0.1073	-0.0830
(PFS1, PES2)	-0.1768	-0.2894
(PFS1, PFS2)	0.2563	0.2626

10.3 Training and Validation Loss for Optimized ENN

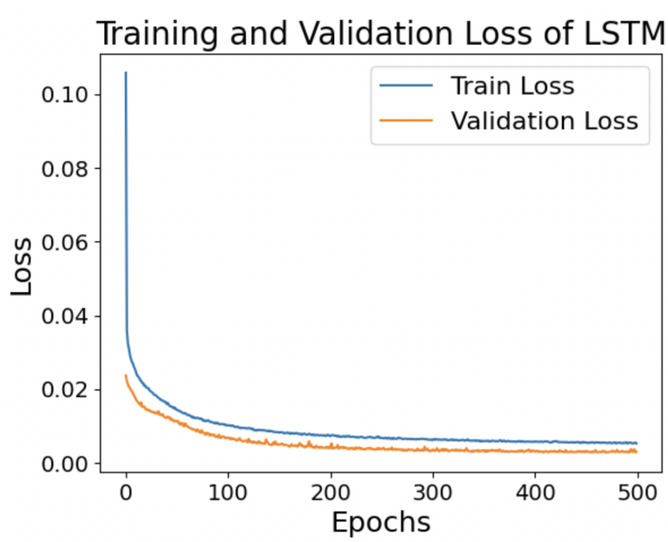

Figure 28: Training (in blue) and validation (in orange) loss vs number of epochs for the LSTM trained using 24 RQs and weighted amplitude/time delay differences. No significant overfitting is evident since both the training loss and validation loss decrease over the epochs and converge to relatively low values, and no significant gap is present between the two curves.

10.4 Phonon Collection Efficiency

Phonon collection efficiency is the ratio between measured phonon energy and the expected phonon energy. The expected phonon energy is given by

$$E_{\text{Ph}} = E_{\text{recoil}} \left(1 + \frac{e\Delta V}{\epsilon}\right) \tag{2}$$

where E_{recoil} is the recoil energy, $e\Delta V$ is the elementary charge times the voltage bias, and ϵ is the energy required to ionize an electron in Germanium (2.96 eV) or Silicon (3.81 eV).

For the Ge 100V HV detector,

$$E_{\text{Ph}} = E_{\text{recoil}} \left(1 + \frac{e100V}{2.96eV}\right) \tag{3}$$

where E_{recoil} is the true energy given by DMC. The total phonon energy collected is also given. The phonon collection efficiency can therefore be calculated for each event. From Figure **??**, most events exhibit a collection efficiency close to 100%, though there is a trailing end with lower collection efficiencies. The 2D histogram on the right reveals that events occurring at the radial edge of the detector, regardless of their Z positions, have low phonon collection efficiencies.

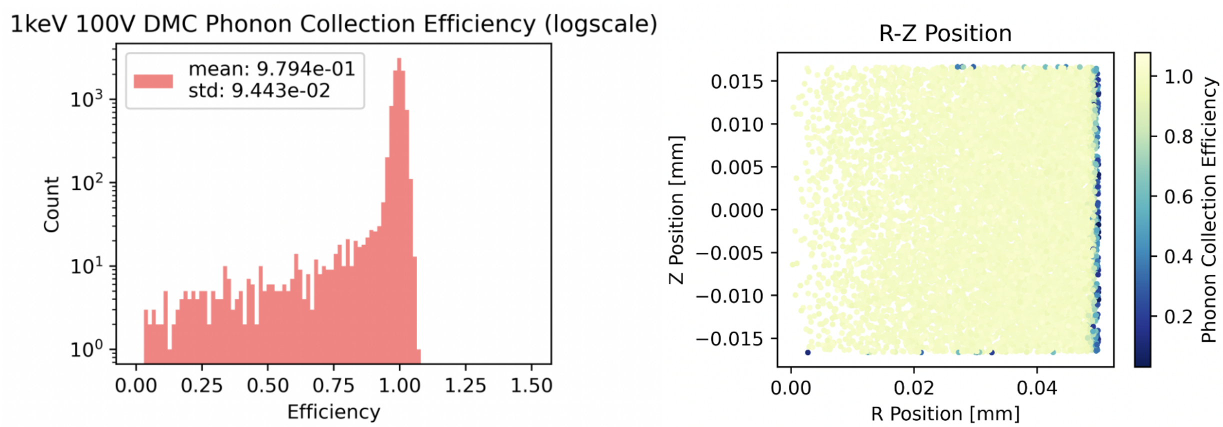

Figure 29: Left: Histogram of phonon collection efficiency of each event in the 1keV 100V DMC sample using Equation 3 with the y-axis in log scale. Right: 2D Histogram of Z position vs R position where the colorscale is set to the phonon collection efficiency.