Optimizing the Energy Resolution of the SuperCDMS Multi-channel, Multi-template Event Reconstruction Algorithm

by

Souren Salehi

H.B.Sc., University of Toronto, 2024

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Physics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2026

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Optimizing the Energy Resolution of the SuperCDMS Multi-channel, Multi-template Event Reconstruction Algorithm

submitted by Souren Salehi in partial fulfillment of the requirements for the degree of Master of Science in Physics

Examining Committee:

Scott Oser, Professor, Physics and Astronomy, UBC
Supervisor

Christopher Hearty, Professor, Physics and Astronomy, UBC
Supervisory Committee Member

Abstract

The SuperCDMS dark matter detection experiment uses a multi-channel, multi-template (N×M) fitting algorithm to reconstruct events in the detector. This algorithm fits each signal from the 12 detector channels with a weighted sum of linearly independent templates and optimizes a χ^2 function for a set of N×M amplitudes for each respective template and channel.

The standard energy reconstruction method uses a weighted sum of the N×M amplitudes along with a trained Boosted Decision Tree (BDT) to map each event to its energy and a correction factor respectively. The proposed improvement to this method consists of an improved template generation technique along with an optimized energy reconstruction algorithm.

N×M assumes linear scaling of the signals with the deposited energy inside the detector. At higher energies, however, the channels experience some saturation due to the temperature-resistance curve of the Transition Edge Sensors (TES). The revised template generation method fits for five templates. The first three templates are generated using events at lower energies where saturation is negligible, as in the original implementation of N×M. The final two templates are derived from the residual of a saturated signal and the appropriate linear combination of the first three non-saturated templates. This creates a set of five templates, three non-saturated and two saturated, that match the resolution of the five standard templates at lower energies but have a 2-5 eV improvement at energies above 10 keV.

For energy reconstruction, events are separated into center and edge events using a decision tree. For energy reconstruction of center events using N×M amplitudes, we propose the use of a quadratic energy estimator that includes linear and quadratic functions of the N×M amplitude, and appropriate weights trained on a set of simulated events. The range of energies is split into discrete bins and a distinct set of quadratic weights is calculated for each bin. A general linear estimator trained on the 0.5-12 keV range of events is used to place each event in the appropriate bin. A MultiLayer Perceptron (MLP) neural network is used for the energy reconstruction of the edge events. This method provides a resolution of 0.7 - 2.3% within our energy spectrum.

Lay Summary

SuperCDMS is a dark matter search experiment taking place in the SNO-LAB underground laboratory in Canada. The detectors aim to detect the energy that dark matter particles deposit when interacting with Standard Model particles. The goal of this thesis is to then reconstruct this energy from the recorded data.

We measure the energy deposition on the detectors as signals, and we can fit those signals with a set of templates. To improve the shape of our templates we create two different sets: one set of three templates created at lower energies and a set of two templates made from higher energy events to capture the range of possible shapes.

For each template and each detector channel we get an amplitude that tries to match the template to the original signal. Our algorithm then uses these amplitudes to create energy estimates for each event. A quadratic function of the amplitudes is created and then for different energy intervals a different weight is applied to the terms in this function in order to get an optimized energy estimator. Events in the edge of the detector are separately reconstructed using machine learning techniques due to their complex behaviour.

Preface

SuperCDMS is an international collaboration with more than two decades of history. Accordingly, this thesis substantially builds on the previous work of hundreds of collaborators. Chapters 1 and 2 provide a general overview of dark matter detection using SuperCDMS detectors. This text is my own, although the content described in it is not my original work. The cited references in these chapters provide additional background and details.

The work in Chapters 3 and 4 expands on the N$\times$M analysis algorithm devised by the SuperCDMS collaboration in order to fit the detector signals [1]. The foundation of the energy reconstruction shown in Chapter 5 is from work done by Aditi Pradeep in her PhD thesis on linear estimators and Boosted Decision Trees (BDT) as an energy estimator [2].

The simulated data were created by the SuperCDMS simulations working group and the data were thoroughly used for training and testing of the reconstruction algorithm.

In Chapter 4, I have extended previous work on template generation using Principal Component Analysis to the case of saturated events. This extension is my own original work, as are the results described in this chapter. In Chapter 5, I present several novel approaches for estimating energies from the amplitudes produced by the N$\times$M fitter. This chapter, and all of the results in it, are my original work unless explicitly noted in the text.

Generative AI, specifically Google Gemini was used for code debugging and as an auto-complete tool within the coding environment. No generative AI tool was used for editing or writing of this thesis.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADC Analog-to-Digital
BDT Boosted Decision Tree
CMB Cosmic Microwave Background
CP Charge-Parity
DAQ Data AcQuisition
DCRC Detector Control and Readout Cards
DMC Detector Monte Carlo
ER Electron Recoil
FIR Finite Impulse Response
GBDT Gradient Boosted Decision Tree
HEMT High Electron Mobility Transistors
HV High Voltage
iZIP interleaved Z-dependent Ionization and Phonon
L1 Level 1
L2 Level 2
LAT Large Area Telescope
LHC Large Hadron Collider
MLP MulitLayer Perceptron
OF Optimal Filter
NR Nuclear Recoil
NTL Neganov-Trofimov-Luke
PCA Principal Component Analysis
QET Quasi-particle-assisted Electrothermal feedback Transition edge sensors
RQ Reduced Quantity
SuperCDMS Super Cryogenic Dark Matter Search
SQUID Superconducting QUantum Interference Devices
TES Transision-Edge Sensors
WIMP Weakly Interacting Massive Particle

Acknowledgments

I would like to express my gratitude to my supervisor, Professor Scott Oser. During this time, he provided unconditional support and guidance and made this journey an enjoyable and productive experience. I always felt welcomed and felt reassured that I was following the correct path in my academic journey. I look forward to the work that we can produce in the future.

I would also like to thank the SuperCDMS collaboration; they welcomed me as an undergraduate student and have been a pivotal part of my journey as a graduate student. Their support in providing feedback and guidance through this journey was instrumental in this work and every line in this paper is an extension of the brilliant work done by my peers in the collaboration.

Dedication

To my parents
Babak Salehi and Haleh Hooshmandi
and to Idil Yaktubay

Chapter 1

Dark Matter

The visible world around us, consisting of baryonic matter, only makes up about 5% of the mass/energy content in the universe. The rest of the universe seems invisible to our eyes and the nature and origin of the other 95% remains an elusive and central question to modern physics. Dark matter and dark energy fill 27% and 68% [3] of the energy content in the universe, respectively. While they interact with ordinary matter through gravity, they do not couple to the electromagnetic force and are undetectable with current instrumentation. Their interaction with other matter through gravity has given the first clues of dark matter's existence. Current research efforts aim to uncover the nature and interaction of dark matter with Standard Model particles. One noteworthy effort is from SuperCDMS, a cryogenic direct dark matter detection experiment set in SNOLAB in Canada. This chapter focuses on the astrophysical and gravitational evidence for dark matter, along with its role in the Standard Model of particle physics and some of the current efforts in the detection of dark matter.

1.1 Evidence for Dark Matter

1.1.1 Rotation Curves of Galaxies

One of the early pieces of evidence for the existence of mass outside of ordinary matter came from Ford and Rubin's observations of the rotation curves of galaxies [4]. Equating the gravitational force on a star orbiting in its galaxy with the central force, we have

$$m\frac{v^2}{r} = \frac{GM(r)m}{r^2}, \tag{1.1}$$

where v is the orbital velocity of the star, r is its distance from the galactic center, m is its mass, $M(r)$ is the total mass of the galaxy within radius r and G the gravitational constant. Thus we obtain the velocity vs. radius relationship

$$v = \sqrt{\frac{GM(r)}{r}}. \tag{1.2}$$

The majority of the visible mass for galaxies is concentrated in the central disk and bulge. Thus, we can assume a constant M at large radii. We would then expect the velocity of the orbiting stars or gas to decrease as we move closer to the edge of the galaxy. However, Vera Rubin measured a constant velocity for the stars and gas in galaxies until the edge, as shown in Figure 1.1.

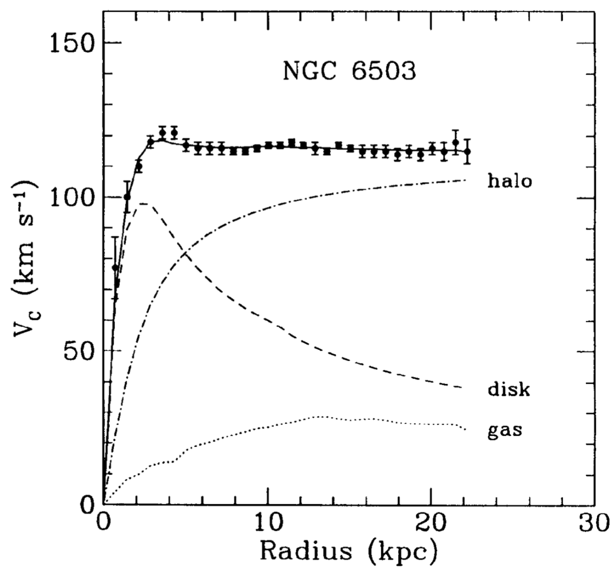


Figure 1.1: Galactic rotation curve measurement by Vera Rubin for NGC 6503, with the dark matter halo mass required to match the rotational velocity. The mass distribution of the visible disk and gases are also shown [5].

This unexpected result points to the need for additional mass, invisible to our instruments, inside galaxies in large enough quantities to have a significant influence on galactic rotation curves.

1.1.2 Bullet Cluster

Due to dark matter's abundance and mass, evidence for its existence can be observed in many astrophysical phenomena. Gravitational lensing for example can look at the distortion of light from far away stars and galaxies and deduce concentrations of mass in the light's pathway. An example of the application of gravitational lensing for dark matter evidence is the Bullet Cluster, itself consisting of two massive clusters of galaxies colliding. The majority of ordinary mass lies in hot gases from these clusters, and in a collision these gases are slowed due to drag forces. A non-interacting particle like dark matter, however, would go through the collision without slowing down. Gravitational lensing can then be used to deduce the distribution of total mass before and after the collision and deduce if there is a large mass density unaffected by the drag forces [5].

Figure 1.2: Region of the Bullet Cluster made up of two colliding galaxy clusters with the hot gases located by X-ray emission shown in pink. The mass distribution of dark matter required to match the observed gravitational lensing is shown in blue [6].

Analysis of NASA's James Webb Space Telescope's images of the Bullet Cluster then provides evidence for this missing mass. Figure 1.2 uses gravitational lensing observations to deduce the dark matter distribution required for the observed light distortion after the collision of the homogeneous hot gases.

1.1.3 Other Evidence for Dark Matter

Studying the Cosmic Microwave Background (CMB) can provide further evidence concerning the nature and characteristics of dark matter. CMB gives a glimpse into the universe as it was 400,000 years after the Big Bang. The high energy state of the universe allowed for a dense plasma of free electrons and protons that could continuously interact with photons, creating an equilibrium. The expansion of the universe, however, caused the free particles to combine and form stable atoms in an event called Recombination. Atoms like hydrogen and helium have a much lower intersection cross section with photons compared to free particles, thus the photons became decoupled from the particles and travelled freely, red-shifting due to the universe's expansion and finally reaching us as CMB [7].

The photon-baryon coupling created a phenomenon called acoustic oscillation, due to the restorative gravitational force pulling matter inwards and radiation pressure pushing outwards. These acoustic oscillations cause small fluctuations in the mostly isotropic CMB temperature, and the power spectra of these oscillations can give an insight into the form and interactions of matter. Dark matter is unique as it experiences the gravitational pull in these acoustic oscillations but is decoupled from the radiative force that only influences baryonic matter. The CMB power spectra then can be used as evidence for dark matter. The three largest peaks of the spectra can give insight into the nature of the universe and its matter content [7].

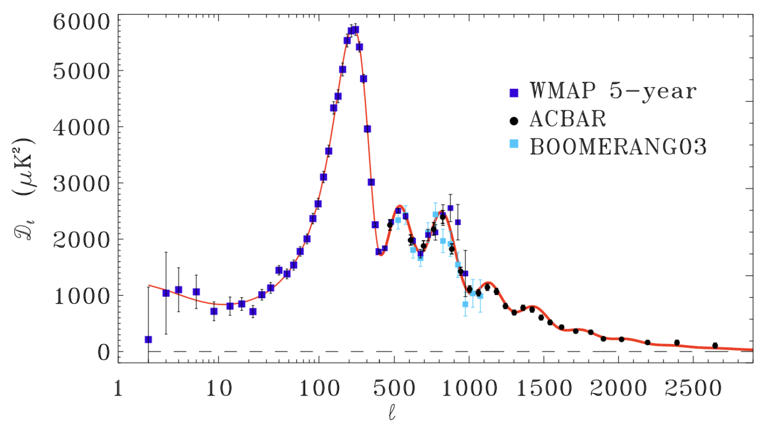


Figure 1.3: Temperature anisotropy power spectrum of CMB as measured by WMAP, ACBAR and BOOMERANG03 given as a function of multipole moments [7].

As shown in Figure 1.3m there are three large peaks in the CMB power spectra. The first peak accounts for a single inward pull of matter towards a gravitational well. The scale of this movement is known and it can be used to trace the trajectory of the photons from the CMB. Its location provides evidence of a flat universe. The heights of the second and third peaks are interesting as they account for inwards and outwards oscillations of matter. The second peak captures the scale of matter that was pulled by a gravitational well and then pushed outwards by radiative pressure, giving insight into the baryonic contents of the universe. The third peak then captures oscillations in and out and then back into the gravitational wells. The third peak however is larger than expected for a purely baryonic universe and is thus a clue for the existence of matter with gravitational contributions that is impervious to the radiative pressure. This indicates the presence of dark matter.

Big Bang Nucleosynthesis also provides further independent evidence for the existence of non-baryonic dark matter. Minutes after the Big Bang, the universe had cooled enough to allow protons and neutrons to form stable nuclei and light elements like helium. This light element formation was dependent on the density of baryonic matter and thus, if dark matter were baryonic, it would contribute to this light element formation, and a different abundance of light isotopes would be observed in the universe [8] compared to what we see. These cosmological and astrophysical findings motivate the search for non-baryonic matter as a candidates for dark matter.

1.2 Dark Matter and the Standard Model

To date, all the empirical evidence for dark matter is through astrophysical and cosmological findings. However, the question of the nature of dark matter extends to particle physics as well. The Standard Model, despite being the best description of the subatomic world, fails to provide answers to some of physics' most fundamental questions, like a quantum description of gravity, what gives mass to neutrinos, and the Hierarchy Problem [9]. The nature of dark matter and its interaction with ordinary matter fall in this beyond the Standard Model physics. By the 1980s, most physicists had accepted the existence of non-baryonic matter and a search had started for unknown species of elementary particles. Some of these candidates included sterile neutrinos, supersymmetric particles and axions [10].

1.2.1 WIMPs

In the early universe, the Standard Model particles went through a process of thermal equilibration due to the high energies that allowed for constant pair production and annihilation. Due to the expansion of the universe, the photons cooled until they lacked the energy to go through pair production. Thus, these particles eventually froze out. This left the Standard Model particles with a relic abundance, which is the matter density that roughly exists to this day. This idea can be expanded to dark matter, where the relic dark matter density $\Omega_\chi h^2 \sim 0.12$ (h is the dimensionless Hubble constant derived from H_0, the Hubble constant: $h = H_0/$ (100 Mpc·km/s)) can be attained through cosmological data. The relic dark matter density is related to the thermally averaged interaction cross section of dark matter times the relative velocity between the annihilating particles $\langle \sigma v \rangle$ by

$$\Omega_\chi h^2 \propto \frac{1}{\langle \sigma v \rangle}. \tag{1.3}$$

The cross section times velocity is then calculated to be $\langle \sigma v \rangle \sim 10^{-26} cm^3/s$, remarkably similar to the weak force cross section [10]. This motivates the hypothesis that dark matter could consist of massive particles with weak scale cross sections. Along with the requirement of a massive particle required for the freeze-out process, Weakly Interacting Massive Particles (WIMPs) have been one of the most studied candidates for dark matter.

The weak interaction also enables scattering with ordinary matter, annihilation of the dark matter and direct detections, which make WIMPs an attractive area of study and makes experimental searches viable. There are also multiple possible WIMP candidates: heavy neutral particles of GeV masses could match the relic abundance that has been measured for dark matter, but more generally any stable particle in the MeV-TeV mass range can be a viable candidate for WIMPs. Cosmological evolution, such as possible decay of dark matter particles, can change the relic abundance and push the possible mass of dark matter particles to over 100 TeV, however most current WIMP experiments operate in the 1 GeV - 10 TeV ranges[11].

1.2.2 Axions

In contrast to the study of heavy dark matter candidates, axions are the leading candidate in the ultralight dark matter region. The original proposition for the axion was as a Beyond the Standard Model solution to the strong Charge-Parity (CP) problem. The Standard Model does not require CP to be a conserved quantity but experimental data fail to show evidence of CP violation by the strong force. An example is the electric dipole moment of the neutron. If CP was not conserved, a neutron would be free to take an electric dipole moment and its proportional relationship with spin would be broken under symmetry transformation, but no such electric dipole moment has been measured so far. A solution to this problem proposed by Roberto Peccei and Helen Quinn consists of a field that would be created in the early universe whose ripple can give rise to light particles [12]. This provided a candidate for dark matter called axions, constrained by experimental evidence to have masses in the $\sim 1 - 100$ μeV range or even smaller. This range is required to ensure that axions produced in the early universe reach a relic abundance consistent with the observed dark matter density of the universe [10]. Axions are hypothesized to couple with photons and nucleons and these are the basis of current axion detection research [12].

1.3 Dark Matter Detection

There are three current avenues of searching for dark matter, namely, direct detection, indirect detection, and production.

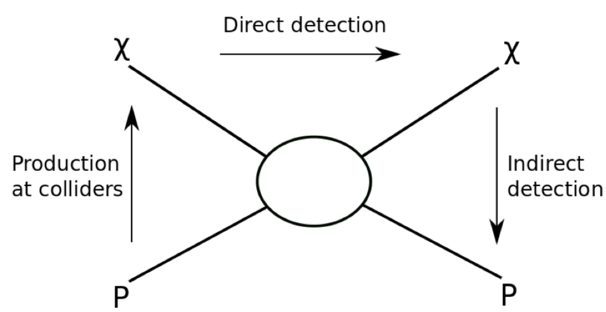


Figure 1.4: Figure showing the possible avenues of dark matter detection using production and annihilation of dark matter and its scattering with Standard Model particles. [13]

Production takes place at particle accelerators, requiring high energy annihilation of Standard Model particles to produce a pair of dark matter particles. Indirect detection involves the observation of the byproducts of dark matter annihilation, and the direct search consists of scattering with Standard Model particles, all shown in Figure 1.4.

1.3.1 Production

Particle accelerators like the Large Hadron Collider (LHC) aim to search for new particles produced in collisions of beams of Standard Model particles either with each other or with fixed targets. The detection of dark matter in particle accelerators thus requires evaluating events with missing momentum and energy, looking for interactions like

$$pp \to \chi\bar{\chi} + x, \tag{1.4}$$

where x are the other possible products like hadronic jets or photons [13]. Sub-GeV dark matter searches become difficult for direct detection as the nuclear scattering signature becomes too small. Accelerator experiments can fill this lower mass void, specifically fixed target experiments. An example of such an experiment is MiniBooNE, which is a fixed target beam dump experiment that has an incident relativistic beam incident on a zero velocity target. It aims to detect dark matter particles produced in the target in a downstream detector [14].

1.3.2 Indirect Detection

Indirect dark matter detection mainly consists of searching for signatures of dark matter annihilation into Standard Model particles such as emission of gamma-rays or neutrinos coming from sources like the galactic center. Using knowledge of astrophysical sources and ordinary matter, the aim is to measure excess radiation produced by dark matter annihilation. Dark matter is the densest at the center of galaxies and their inner halos, however those are also the most complex astrophysical systems and thus it is difficult to model their exact signatures. Thus, most searches focus on regions with higher dark matter to ordinary matter ratio in order to maximize the signal [15].

Some expected products to measure are gamma-rays, X-rays and radio waves, neutrinos, or positrons and electrons along with protons and antiprotons, each with their own limitation in experiment. gamma-rays for example are a popular signature, as they conserve both energy and momentum information, however the neutrality of dark matter means gamma-rays can only be produced in secondary interactions thus there will be a suppression of the flux. The Fermi Large Area Telescope (LAT) is one of the gamma-ray experiments. This satellite has taken the largest set of gamma-ray data from galactic centers and it has allowed for precise measurements of possible dark matter interactions [16].

1.3.3 Direct Detection

Direct dark matter detection looks for signs of ambient dark matter scattering from Standard Model particles, mainly nuclei, and it has been the largest research effort among the different avenues of dark matter detection. Direct searches look for scattering of dark matter particles of masses within the ∼1-100 GeV range that can produce nuclear recoils leaving energy depositions in the range of a few keV to ∼100 keV. Due to the low probability of dark matter scattering, the signature can be reduced to the interaction of a single dark matter particle with a nucleus, with a differential recoil spectrum that can be written as a function of energy E and time t:

$$\frac{dR}{dE}(E,t) = \frac{\rho_0}{m_\chi m_A} \int v f(v,t) \frac{d\sigma}{dE}(E,v) d^3v, \tag{1.5}$$

where m_χ and m_A represent the dark matter and nucleus masses, respectively, $d\sigma/dE$ is the differential cross section, v is the velocity of the dark matter in the lab frame and ρ_0 and $f(v,t)$ are the astrophysical parameters for dark matter density and WIMP velocity distribution. Two common approaches used by direct detection experiments are to observe the recoil energy spectrum or its time dependence. Energy dependence is the more common approach that aims to quantify the relationship between the energy of dark matter particles and the rates and energies of the resulting nuclear recoils. Time dependence can look at annual modulation caused by the variation in the relative velocity of dark matter particles with respect to Earth over different months of the year [13].

One type of direct detection dark matter experiment uses scintillator crystals that can emit light due to radiation passing through the scintillating crystal. The DAMA experiment [13] is one such example that uses Nal(TI) crystals at room temperature and searches for annual modulation. Liquid noble gases can also be used as an experiment medium due to their higher ionization yield and scintillation. DEAP [13] is an example of a noble gas experiment taking place at SNOLAB in Canada using liquid argon. Cryogenic bolometers are another research pathway looking at thermal and athermal phonons to study the deposited energy of dark matter particles inside a crystal lattice. The bolometers can be coupled with charge measurements to discriminate between electron and nuclear recoils. One such example is the SuperCDMS experiment, which is the focus of the rest of this thesis [17].

Chapter 2

The SuperCDMS Experiment

Super Cryogenic Dark Matter Search (SuperCDMS) is a direct detection experiment situated at SNOLAB in Sudbury, Ontario. The SNOLAB facilities are located approximately 2km underground, and the rocky environment provides the required shielding against cosmic rays [18]. The basic schematic of the experiment setup is shown in Figure 2.1. The SuperCDMS detectors are made of semiconducting crystals in a cryogenic system, utilizing ionization and phonon production to investigate scattering of dark matter particles. The towers of detectors are made of interleaved Z-dependent Ionization and Phonon (iZIP) detectors, which measure both ionization and phonon signals, and High Voltage (HV) detectors that take advantage of the Neganov-Trofimov-Luke (NTL) effect to amplify the signal of the detectors and improve the experiment's sensitivity [18].

2.1 Dark Matter Detection Principles

SuperCDMS operates semiconducting detectors at cryogenic temperatures ($\sim$ 20 mK) for charge and heat signal detection. The experiment utilizes a mix of silicon (Si) and germanium (Ge) cylindrical crystals with a diameter of 100 mm and thickness of 33.3 mm. The detectors are split into two separate designs, HV and iZIP shown in Figure 2.2. This provides a wider range of sensitivity to the mass of dark matter with HV having improved sensitivity at masses below 5 GeV/c^2 and iZIP having better sensitivity at around 5 GeV/c^2 [20].

The iZIP detectors are designed for effective electron recoil (ER) and nuclear recoil (NR) discrimination. Electron and nuclear recoil manifest from the interactions of particles with the electrons and the nuclei of our semiconductors respectively. Energy deposits in the crystals with energies above the bandgap cause electrons to leave the valence band into the conduction band. This is improved by the small bandgap of the crystals which is 0.74eV

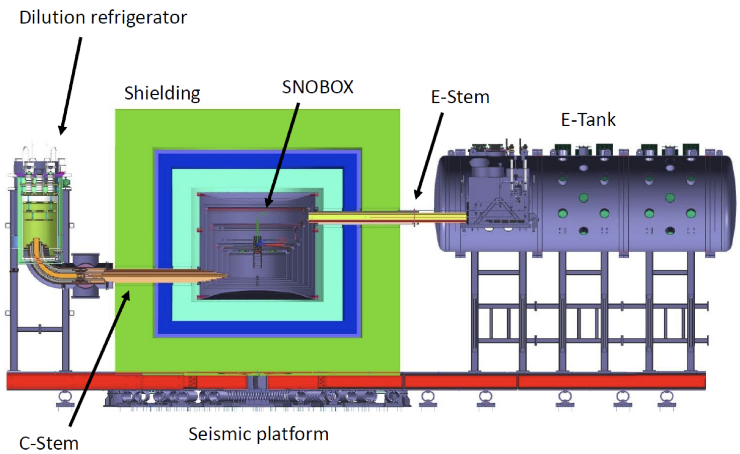


Figure 2.1: Cross section of the design of the SuperCDMS SNOLAB experiment. The detectors are placed within a cryostat surrounded by high-density shielding. The stems are used for connection of the detectors to outside systems to minimize the exposure points of the shielding. The SNOBOX cryostat and shielding are placed on a seismic platform in order to protect and isolate the experiment from vibrations[19].

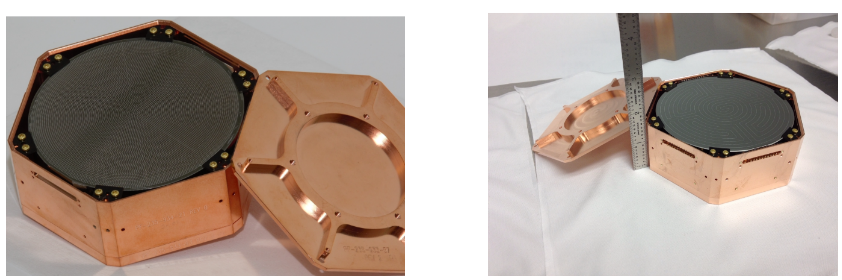


Figure 2.2: SuperCDMS SNOLAB detectors showing an HV detector (Left) and iZIP detector (Right) inside copper housing [2].

and 1.18eV for Ge and Si, respectively [21]. The freed electrons leave holes behind that act as positive quasi-particles and these electron-hole pairs are able to move through the bulk in the presence of an electric field and produce an ionization signal. The average energy required to create these hole-pairs in Ge and Si is 3 eV and 3.8 eV, respectively [20].

Some fraction of the recoil energy will be dissipated as phonons. Phonons are quantized units of vibration within the lattice of the crystals. These vibrations can travel through the lattice and be read as heat signatures. Both ER and NR can produce electron-hole pairs and phonons and the information regarding the type of interaction is contained in the ratio of the charge to heat released. Thus, iZIP detectors have the ability to capture both phonon and charge signals and use these ratios to suppress background events which are expected to be mostly ER events [21].

The HV detectors are unable to directly measure ionization and thus specialize in maximizing the phonon signal of the particle interaction. The experiment uses the NTL effect by applying a 100V bias across the detectors that increases the final readout energy in our detector by an extra NTL term E_{NTL}. When an electron-hole pair is created, it can be accelerated by applying a voltage across the crystal. The accelerated pair can then release more electron-hole pairs and more phonons while traveling through the crystal. This allows us to reach lower energy thresholds and thus sensitivity to smaller energy deposits and hence lower mass WIMPs [20]. The final energy that reaches the measurement channels is then

$$E_{final} = E_R + E_{NTL} = E_R + \frac{Y E_R}{\epsilon} e \Delta V, \tag{2.1}$$

where E_R is the recoil energy, Y is the ionization yield (ratio of energy deposited in the form of ionization to the energy that an electron of the same recoil energy would produce as ionization), ϵ is the average energy to create an electron-hole pair, e is the absolute magnitude of the charge of an electron and ΔV is the voltage applied. The NTL term can be simplified to $Ne\Delta V$ with N being the number of free electrons released and accelerated to the sensors [21]. The HV data will likely be dominated by ER due to the inability to discriminate between ER and NR, however the NTL effect will be able to somewhat reduce the ER background due to ER's ionization yield of $Y = 1$. NRs have an energy-dependent ionization yield of <0.3 so they are less amplified by the NTL effect. Thus, ER events will be mapped to higher energies compared to NR, creating a method to partially separate the ER and NR.

2.2 Charge and Phonon Detection

The detectors use Transition-Edge Sensors (TES) for phonon signal readout. The TES leverage the steep resistance-temperature gradient of a superconducting film to operate as a highly sensitive bolometer for phonon detection. Superconductors possess zero resistance below their critical temperature, but they experience a sharp resistance increase in the presence of thermal energy and thus are suitable measurement devices for cryogenic dark matter detectors as shown in Figure 2.3.

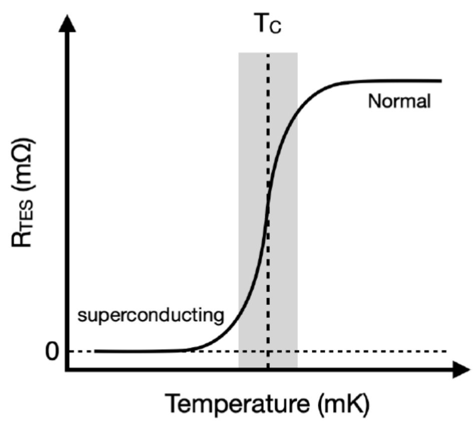


Figure 2.3: Diagram of a TES resistance-temperature curve [2]. TES can be operated in the transition region, shaded in grey in the figure. Upon energy absorption, such as phonon absorption, they can move to the normal resistance regime.

The specific SuperCDMS sensors are Quasi-particle-assisted Electrothermal feedback Transition-edge sensors (QETs). For the sensor, a superconducting aluminum collection fin is added to the TES which allows for better surface coverage. The phonons created inside the crystal will travel to the aluminum fins. They can then break Cooper pairs inside the superconducting aluminum creating quasi-particles that will diffuse into the tungsten TES. This process is demonstrated in Figure 2.4. The temperature-resistance curve of the TES can be approximated to be linear during the superconducting transition, thus for a range of energies the response of the phonon channel is nearly proportional to the deposited energy. A limitation of the TES is the response to higher energies than the ones within the transition temperature region, and this saturation effect will be discussed further.

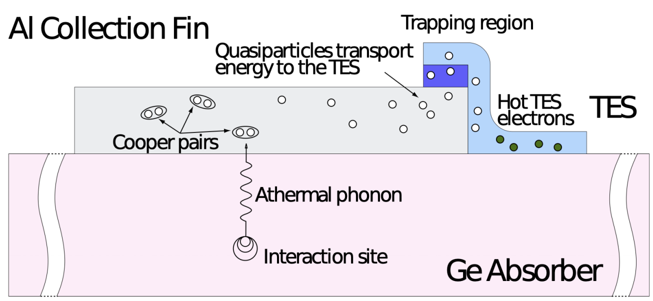


Figure 2.4: Diagram of QET with aluminum fins that aid in the phonon collection on the surface of the SuperCDMS detectors. The phonons break the Cooper pairs allowing for the transport of the quasi-particles into the tungsten TES [21].

The TES is placed in series with an inductor in order to translate the energy release into the phonon sensor as a resistance change. A bias current is run through the TES in order to keep the resistance in the middle of the resistance-temperature curve. During a sudden increase in the resistance of the TES, the current in the circuit decreases, changing the magnetic field of the inductor. Superconducting QUantum Interference Devices (SQUIDs) are then used to read and amplify the magnetic field change in the inductor aiding in the sensitivity of the detector. The process of the phonon readout channel is shown in the diagram below.

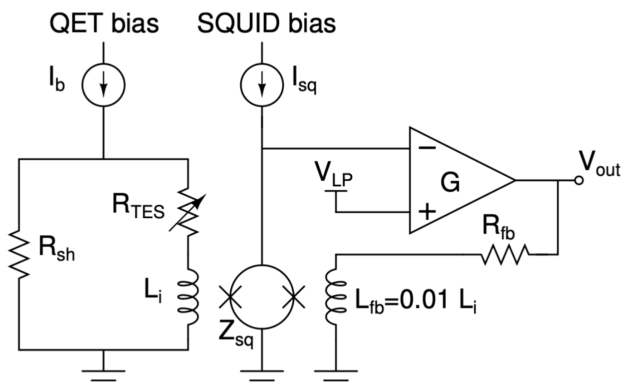


Figure 2.5: Diagram of the SuperCDMS phonon readout with the TES R_{TES} and SQUID Z_{sq} circuits [21].

The surfaces of the iZIP detectors are covered with alternating charge and phonon sensors. The charge is measured using High Electron Mobility Transistors (HEMTs) which is a type of field-effect transistor. The HEMTs are kept at a constant and opposing voltage on either side of the detector and the TES are kept grounded. This design helps guide the created electron-hole pairs towards the electrodes. However, the grounded TES does create a surface trapping field that can attract either the electrons or holes created near the surface, whilst the partner electrons or holes will be read by the HEMTs on the other side [22]. The asymmetry in the charge collection thus allows the iZIP detectors to distinguish between events near the surface and bulk events. Figure 2.6 shows the electric field and potential lines near the surface of the iZIP detector.

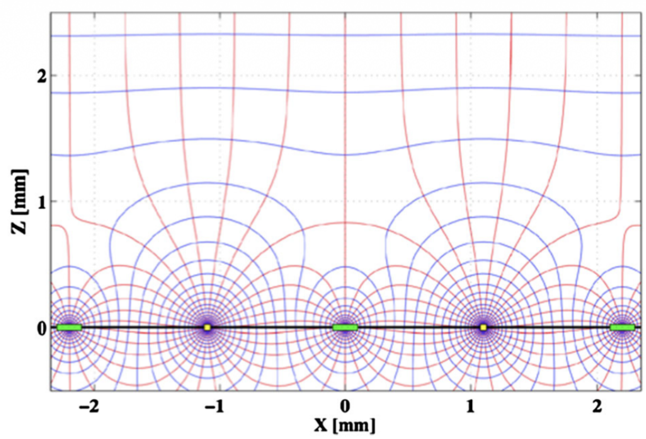


Figure 2.6: The electric field (red) and potential (blue) of an iZIP detector near the surface. Electron hole pairs are attracted to the biased electrodes (yellow) but near the surface the potential difference of the electrodes and the TES (green) produces a trapping field that can attract electron-hole pairs to the TES creating a charge asymmetry in the measurement [22].

Both HV and iZIP detectors have six phonon channels on each side of the detector, and iZIP detectors have an additional two ionization channels on each side. This thesis will focus on the phonon channels only, and their geometry layout will be given in Chapter 5.

2.3 Data Acquisition

The next level in the data pipeline is the SuperCDMS Data AcQuisition (DAQ) system that is responsible for the readout, digitization and early-stage processing and triggering. SuperCDMS-SNOLAB consists of two iZIP towers and two HV towers with six detectors in each. The Detector Control and Readout Cards (DCRCs) are responsible for the channel readouts and the level 1 (L1) triggering system. HV detectors use two DCRCs (one for each side of the detector) to read out the phonon channels and mediate the applied voltage across either side of the detector. The iZIP detectors use only a single DCRC to read out both the charge and phonon channels [2].

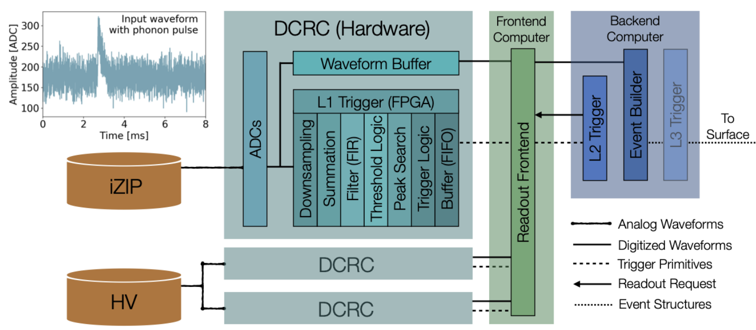


Figure 2.7: Diagram of the overview of the SuperCDMS data acquisition system. The input signals are detected by the one or two DCRCs for iZIP and HV respectively. The L1 trigger system is used to decide on the triggering of a single detector. The L2 system is then used to decide on sets of detectors to read out [23].

There is a two-level trigger system for the SuperCDMS readout. The original signal is digitized and stored in a circular buffer and the L1 trigger is responsible for the triggering of a single detector. The signal is first down-sampled and run through a Finite Impulse Response (FIR) filter to produce the L1 waveforms. A weighted sum of the waveforms from the different channels is used to make the trigger decision. A set of trigger primitives is also created and used later for the level 2 (L2) decision. The L2 trigger analyses the primitives for all the detectors and decides the set of detectors that should be read out [24]. The resulting waveforms chosen by the L2 trigger are then collected and assembled alongside other detector information using the Event Builder and are the basis of event reconstruction. This entire process is summarized in Figure 2.7.

Chapter 3

SuperCDMS Event Reconstruction

CDMSBats is the primary processing software responsible for converting the raw pulse traces from the detectors into Reduced Quantities (RQs). These RQs contain information like pulse amplitudes, time of the pulse, the χ^2 value for the fit of a template shape to the pulse, and other quantities that are later used for precise energy reconstruction. The event reconstruction algorithm inside CDMSBats is a multi-channel, multi-template method used for fitting of the raw pulses, called N×M. This algorithm is designed to account for the multiple sensor data and the correlated noise among them [1]. An example of the output pulse from the SuperCDMS DAQ is shown in Figure 3.1. The set of pulses from N channels is the input to the N×M analysis algorithm. Each pulse is fitted to a linear weighted sum of templates that are derived using the shape of simulated pulses. The output of the N×M event reconstruction is then a set of amplitudes for each template and each channel. These are the parameters for energy and position reconstruction.

3.1 N×M Reconstruction

The N×M event reconstruction algorithm is formulated to account for correlated noise and pulse shape variations for the signals from the detectors. It generalizes to fit for N channels and M linearly independent pulse templates in the frequency domain [1]. The general method for fitting a time series with an optimal filter is using a χ^2 optimization and is the basis of the N×M algorithm. Given an equally spaced time series d_j and a normalized template T_j of length S, we fit for the amplitude a that would minimize the χ^2, which is formulated as

$$\chi^2(a) = \sum_{j=1}^{S} \left(\frac{d_j - aT_j}{\sigma_j} \right)^2 . \tag{3.1}$$

The σ_j is the measurement uncertainty on the j^{th} point in the time series.

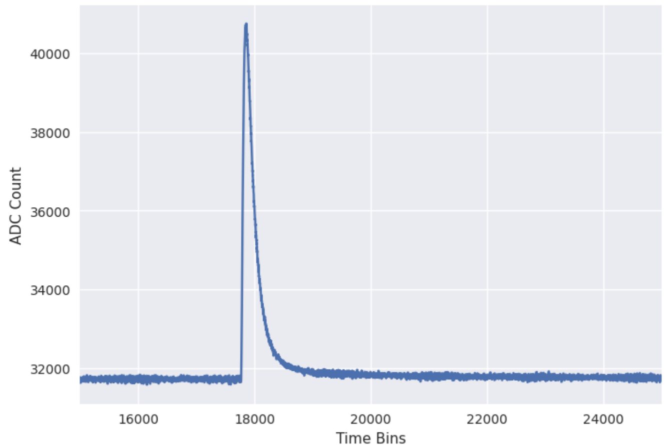


Figure 3.1: Example of a SuperCDMS detector pulse from a single channel, in units of ADC counts.

As written, Equation 3.1 assumes that the uncertainties on different points in the time series are statistically independent. This can be generalized to account for any correlation by replacing the uncertainty term with a covariance matrix V_{jk}. The generalized χ^2 is:

$$\chi^2(a) = \sum_{j=1}^{S} \sum_{k=1}^{S} (d_j - aT_j)(d_k - aT_k)(V^{-1})_{jk}. \tag{3.2}$$

This version of our χ^2 includes a S×S covariance matrix that needs to be inverted, this process is both numerically unstable and computationally expensive. A fit in the frequency domain solves this issue. Assuming stationary noise in our time series, the terms in the discrete Fourier transform of the time series will be statistically independent and thus the covariance matrix will be diagonal. The χ^2 in Fourier space is then a sum over discrete frequency f bins:

$$\chi^2(a) = \sum_{f \geq 0} \frac{|\tilde{D}_f - a\tilde{T}_f|^2}{\sigma_f^2}, \tag{3.3}$$

where $\tilde{D}_f$ and $\tilde{T}_f$ are the Fourier transformations of the time series and the template, and σ_f the uncertainty (i.e. the noise at frequency f).

N×M event reconstruction is then used to extend this to multiple channels and templates. In order to deal with multiple channels, the N×M formulation simply calculates the χ^2 for each of the channels and sums them all. Given a free parameter E that our amplitudes depend on, the χ^2 equation with the channel inclusion (c) is:

$$\chi^2(E) = \sum_{c=1}^{N} \chi_c^2 = \sum_{c=1}^{N} \sum_{f \geq 0} \frac{|\tilde{D}_{c,f} - a_c(E)\tilde{T}_{c,f}|^2}{\sigma_{c,f}^2}. \tag{3.4}$$

This equation is applicable to multiple statistically independent channels. However, the channels in the SuperCDMS detector have correlated noise between them so an N×N covariance matrix is used where $V_{c_1,c_2,f} = cov(\tilde{D}_{c_1,f}, \tilde{D}_{c_2,f})$ and thus the χ^2 form with correlated channels is

$$\chi^2(E) = \sum_f \sum_{c_1} \sum_{c_2} (\tilde{D}_{c_1,f} - a_{c1}\tilde{T}_{c_1,f})(\tilde{D}_{c_2,f} - a_{c2}\tilde{T}_{c_2,f}) V^{-1}_{c_1 c_2,f}. \tag{3.5}$$

The final stage in the N×M algorithm is accounting for multiple templates. At a given energy there are varying shapes for the pulses at different channels, influenced by factors like the position of energy deposition. Multiple templates allow for the extraction of pulse shape variation information from the different channels. The prediction for a pulse in the N×M algorithm is a linear sum of templates, each with its own amplitude:

$$\vec{p_c} = \vec{a}_c \cdot \mathbf{T}_c, \tag{3.6}$$

where $\vec{p_c}$ is the prediction, $\vec{a}_c$ the vector of amplitudes with the size 1×M and $\mathbf{T}_c$ an M×S matrix that accounts for M templates each with the size of the time series length S.

Until this point the pulse shape predictions were only described as functions of amplitude. However, there is also an alignment problem between the pulses and the templates, and so there needs to be a time shift for the templates to align with the pulse. In the Fourier space a time shift is achieved by a rotation, so with a given time delay δ the translated template is

$$\tilde{T}_f(\delta) = e^{-2\pi i f \delta} \tilde{T}_f(0). \tag{3.7}$$

In the final formulation, we require a time delay for each channel δ_c, and a rotation matrix $\mathbf{R}_c$ is used that shifts each of the templates. The complete form of the prediction used in N×M is:

$$\vec{p_c} = \vec{a}_c \cdot \mathbf{T}_c \cdot \mathbf{R}(\delta_c). \tag{3.8}$$

The sets of pulses and predictions can be written sequentially as a vector so that $\vec{D}_f$ and $\vec{P}_f$ are row vectors of size 2N (the 2 is due to separating the real and imaginary parts). A vector $\vec{a}$ of length NM can then account for all the amplitudes needed for all the templates and channels. The final form of the N×M algorithm χ^2 is:

$$\begin{aligned}\chi^2 &= \sum_f \chi_f^2 \\ &= \sum_f \left(\vec{D}_f - \vec{P}_f\right)^T \mathbf{V}_f^{-1} \left(\vec{D}_f - \vec{P}_f\right) \\ &= \sum_f \left(\vec{D}_f - \vec{a} \cdot \tilde{\mathbf{T}}_f \cdot \mathbf{R}_f(\vec{\delta})\right)^T \mathbf{V}_f^{-1} \left(\vec{D}_f - \vec{a} \cdot \tilde{\mathbf{T}}_f \cdot \mathbf{R}_f(\vec{\delta})\right).\end{aligned} \tag{3.9}$$

It is not possible to analytically solve for both amplitudes and time delay, so the time delays for each channel are found numerically before the N×M algorithm is used and they are taken to be constants in the final form. The χ^2 can then be minimized analytically with respect to the amplitudes such that:

$$\vec{a}^T = \left(\sum_f \left(\tilde{\mathbf{T}}_f \cdot \mathbf{R}_f\right) \mathbf{V}_f^{-1} \left(\tilde{\mathbf{T}}_f \cdot \mathbf{R}_f\right)^T\right)^{-1} \sum_f \left(\tilde{\mathbf{T}}_f \cdot \mathbf{R}_f\right) \mathbf{V}_f^{-1} \tilde{\vec{D}}_f^T. \tag{3.10}$$

3.2 Template Generation

N×M is designed to fit M linearly independent templates to each channel. The shapes of the pulses in the detector carry information about location and energy of events and thus any generated template needs to sufficiently capture the variations in pulse shape. The heights of the pulses are dependent on the energy deposited in that specific channel, and the rise and fall times provide information about the location of that energy deposition. Each event produces prompt phonons that propagate quickly from the interaction site and thermalized phonons that are created from the diffusion of that energy throughout the crystal. Channels closer to the energy deposition then experience faster rise time due to the larger fraction of prompt phonons while the falling tail, consists of thermalized phonons, has less position dependence. This necessitates a fit that allows for pulse shape variation. The most computationally efficient method is to assume linearity and produce a set of linearly independent templates that will approximately capture the most significant pulse shape variations.

The templates are initially created in time space and during the fitting algorithm are transformed into frequency space. The method used for template generation for SuperCDMS N×M is Principal Component Analysis [1]. PCA is a dimensionality reduction tool that transforms data into sets of orthogonal bases that maximize the variance of the data. The output is a set of orthogonal templates and their percent contribution to the variance of the data. The variance contribution is important in our analysis as it provides information that helps determine the number of templates needed. Theoretically a large number of templates could be used to perfectly capture the shape of the pulses, however computational limitations require a conservative approach in template selection, and understanding the contribution of each template from PCA allows for careful selection of the number of templates.

Running the PCA requires a large sample of events evenly selected from each channel of the detector. Whilst it is possible to use real data for template generation, the noise in the pulses starts dominating the higher-order components in the PCA, especially if the selected events are at lower energies and thus have a lower signal-to-noise ratio. SuperCDMS Detector Monte Carlo (DMC) simulation data allow for access to pure traces without noise and are thus the standard choice for creating N×M templates. Even though the shape of the pulse might not be identical to real data, the ability to have higher-order, low-noise templates allows for improved capture of pulse shape variations. Another major benefit in choosing simulation data is that this avoids the alignment problems with real data. To capture the exact bases that make up the data, we require all events in the sample to be perfectly lined up. While it is possible to achieve good alignment, the process will always be inferior to perfect alignment possible for simulated pulses. Figures 3.2 and 3.3 show PCA applied to simulated and real data respectively. For real data, we can see noise starting to appear on the third component, and component four loses many of its features as noise dominates. The deposition energy of the events is around 11 keV which is on the higher end of the range of our detectors and thus the best signal-to-noise ratio possible. This is in contrast to the simulated data principal components that provide clear templates up until component five. The general practice for the N×M template generation is using five templates. The PCA applied to real data shows that by component four noise is starting to interfere with the variance of the templates and this is with the highest signal-to-noise ratio possible. In practice, for most of the range of our detector, noise dominates the data by the third component and thus five templates should be sufficient in capturing pulse shape variations.

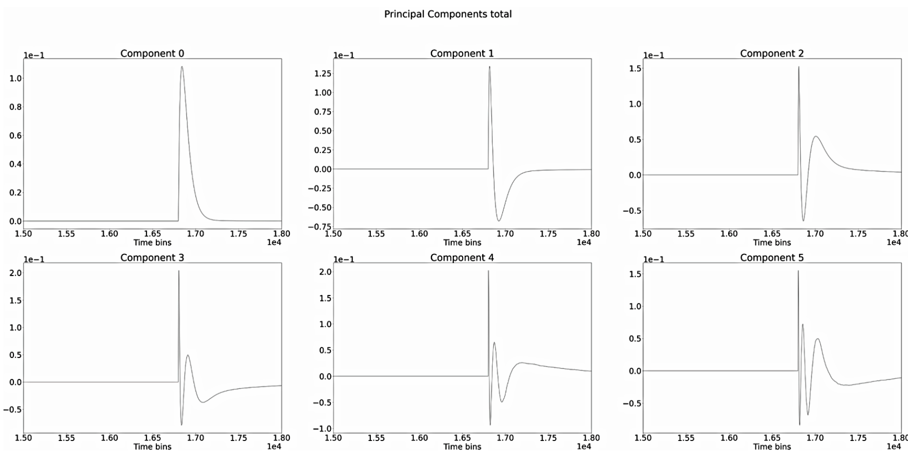


Figure 3.2: First six components from applying PCA to simulation data.

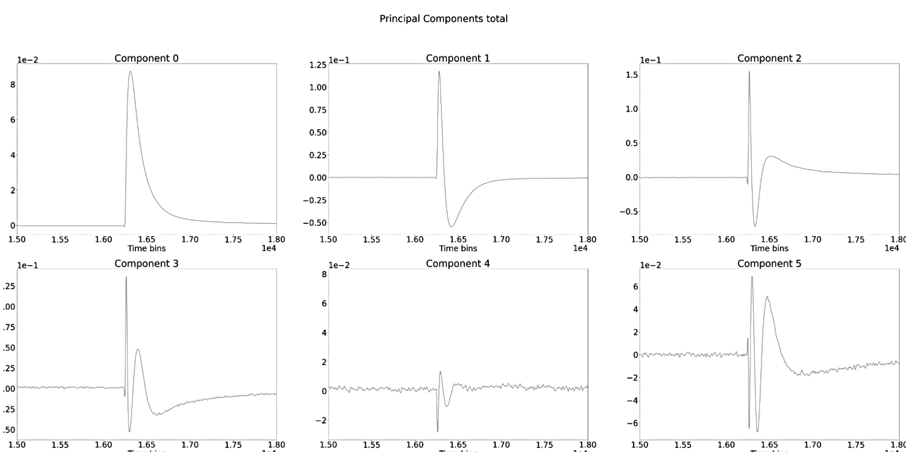


Figure 3.3: First six components from applying PCA to real data with energy of 11 keV.

3.3 Energy Reconstruction in SuperCDMS

The NM amplitudes from the N×M event reconstruction algorithm are the parameters used in the current SuperCDMS energy estimator. The energy is estimated using a combination of a linear estimator and a Gradient Boosted Decision Tree (GBDT).

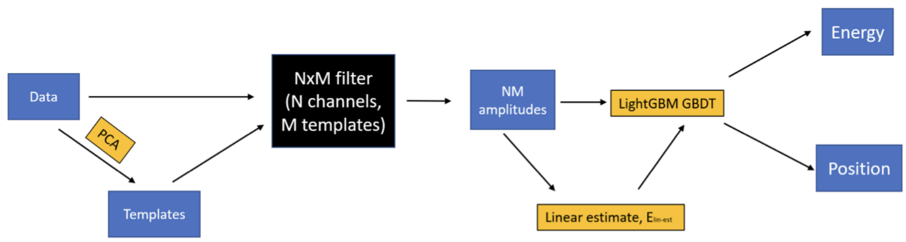


Figure 3.4: Flowchart of the current N×M data processing and energy-position reconstruction [1].

The linear estimator is simply the weighted sum of the N×M amplitudes calculated in Equation 3.10. The linear estimation of the energy is then:

$$E_{lin} = \sum_{i=1}^{NM} w_i a_i. \tag{3.11}$$

With a set of data with j events with known energies, a χ^2 minimization can be used to analytically derive these weights. The solution for each weight is:

$$w_i = \frac{\sum_j a_i^j E_{true}^j}{\sum_j (a_i^j)^2}. \tag{3.12}$$

The N×M amplitudes implicitly include position information; however, the position dependence of the energy has higher-order contributions and thus further analysis is required for better energy estimation. The current analysis framework uses GBDTs from the LightGBM package in order to account for these higher-order terms. The GBDT is trained on the fractional deviation of the linear energy estimator from the true energy so the correction factor G_{BDT} is defined as:

$$G_{BDT} = \frac{E_{true} - E_{lin}}{E_{lin}}. \tag{3.13}$$

The validation for this method was done using SuperCDMS Detector Monte Carlo (DMC) simulation data in order to train and test for the resolution of the energy reconstruction. Ten thousand mono-energetic events were simulated to be uniformly distributed inside the detector. The energy of the events was 500eV and the events were split 90/10 for training and testing.

Estimator	Peak resolution in %
1×1 OF	$2.56\% \pm 0.14\%$
N×M OF without position correction ($E_{\text{lin-est}}$)	$2.39\% \pm 0.13\%$
N×M OF with G_{BDT} position correction ($E_{\text{N}\times\text{M}}$)	$0.73\% \pm 0.04\%$

Table 3.1: The resolution of different event reconstruction and energy reconstruction methods [1].

Table 3.1 shows the resolution of the N×M amplitudes with GBDT method compared to previous methods. The 1×1 Optimal Filter (OF) uses a single template to fit a single time series summed across channels and thus lacks position and channel correlation correction and thus has the poorest resolution. GBDT method provides a significant improvement over just using N×M and it provides evidence for significant higher-order contributions to the energy [1].

3.4 N×M Limitations

There are two primary limitations that the N×M analysis and energy reconstruction framework fail to address. The first is the occurrence of pulse saturation in higher-energy detector data, which manifests in two forms: Analog-to-Digital Converter (ADC) saturation and TES saturation. ADC saturation arises from hardware constraints in the readout circuitry: past a specific threshold, further increases in current cannot be captured, resulting in a flat-topped trace. Because ADC saturation is a more complex and less frequent issue, it falls outside the scope of this work. Conversely, TES saturation occurs at lower thresholds than ADC saturation and stems from the non-linear behavior of the TES temperature-resistance curve at higher temperatures. The N×M algorithm is fundamentally designed under the assumption that raw pulses scale proportionally with their templates, reducing the fitting problem purely to an amplitude scaling factor. However, this linear relationship breaks down at energies above 10 keV, where the pulse amplitude rises at a slower rate than the deposited energy.

The second limitation involves the practical deployment of the N×M energy estimator. The weights required for optimal energy reconstruction vary across different energy scales. We will evaluate different methods for deriving these weights later in this work; however, this energy dependence introduces a circular problem: one must already possess some information regarding an event's energy to correctly apply the weights optimized for that specific range. Consequently, the energy resolutions shown in Table 3.1 represent the ideal, upper-bound performance, as every event is matched with its correct set of weights. Achieving this is impossible in practice; to select the appropriate weights for the estimator, an initial inference regarding the pulse's energy range must be made, making the final resolution highly dependent on that initial guess. The remainder of this paper addresses these two challenges: the treatment of TES saturation within the N×M framework, and the development of a generalized energy estimation methodology that functions for any event at any energy scale.

Chapter 4

Saturation in N×M Event Reconstruction

4.1 Saturated Template Generation

SuperCDMS Detector Monte Carlo (DMC) simulation data can be used to study the shape difference between non-saturated and saturated pulses from the SuperCDMS detectors. In case of near-linearly scaled pulses like the ones from the SuperCDMS detectors we can use PCA in order to understand the fundamental pulse shape from the detectors. A single pulse has the form:

$$\vec{x} = a\vec{S} + H.O.T + \vec{\epsilon}, \tag{4.1}$$

where $\vec{x}$ is a pulse written as a vector matching the dimensionality of our sample space (specifically 32768 samples long in the case of the SuperCDMS pulses), a is an arbitrary amplitude, $\vec{S}$ is the average shape of the pulses from the detector, $\vec{\epsilon}$ is the noise and the higher-order Terms (H.O.T) refer to variations of the pulse shape due to factors such as the position and energy of the event.

Principal Component Analysis, when applied to a sample of simulated pulses, can be used to study the higher-order terms. PCA removes the mean of the pulses and finds the direction of maximum variance. By averaging the pulses we are removing random contributions to the signal, like position dependence and noise. So we can model our average pulse as:

$$\langle\vec{x}\rangle = \overline{a}\vec{S}, \tag{4.2}$$

where $\overline{a}$ is the mean amplitude. Subtracting this mean from our pulses we get

$$residual = \vec{x} - \langle\vec{x}\rangle = (a - \overline{a})\vec{S} + H.O.T + \vec{\epsilon}. \tag{4.3}$$

The residual used for our PCA has the same form as the model for our original pulse. The mean pulse shape is ideally the direction of maximum variance for our pulses, and thus the first principal component. This first-order component allows for an initial energy estimate, while the higher-order terms are then used for position corrections and other secondary factors. The first principal component can also be used to understand the fundamental difference in the shape of saturated and non-saturated pulses, as it coincides with the average of the pulses. Figure 4.1 shows the plot of the first PCA component from a set of 1.3 keV pulses and a set of 10.3 keV pulses. The 10.3 keV pulses contain some saturation. For the saturated pulses, we see a lower pulse amplitude and a slower fall time, which are the shape signatures of TES saturation.

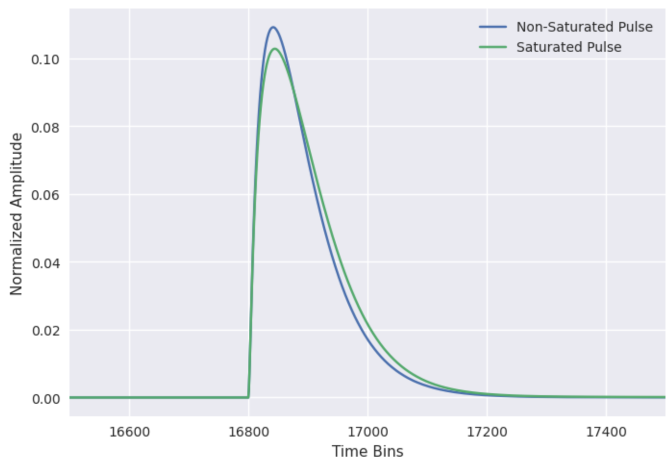


Figure 4.1: First PCA component of a set of non-saturated pulses at 1.3 keV (blue) and saturated pulses at 10.3 keV (green).

The goal is to capture these differences in shape between non-saturated and saturated pulses within the N×M framework without degrading the resolution at lower energies. The general approach is to create a hybrid set of templates: the first three derived from non-saturated pulses to maximize resolution, and the final two from saturated pulses to better model saturated shape variation. The process would then be to run PCA on a set of non-saturated events, sampled evenly from different channels, and take the first three components, which we label as $\vec{T}_1, \vec{T}_2, \vec{T}_3$. We can model a saturated pulse $\vec{X}_s$ as a linear combination of the non-saturated templates and two unknown saturated templates $\vec{T}_{s1}, \vec{T}_{s2}$. We can ignore higher-order terms in this model. Based on the construction of PCA, we know that these templates are orthonormal and we write our saturated pulse as:

$$\vec{X}_s = a_1\vec{T}_1 + a_2\vec{T}_2 + a_3\vec{T}_3 + a_{s1}\vec{T}_{s1} + a_{s2}\vec{T}_{s2}. \tag{4.4}$$

We can extract the amplitude a_1, a_2, a_3 by taking the dot product of the saturated pulse and our non-saturated components

$$\vec{T}_i \cdot \vec{X}_S = a_i\vec{T}_i \cdot \vec{T}_i = a_i. \tag{4.5}$$

We can now project out the contributions from the non-saturated templates by subtracting these from the data to yield Δ_s, the residual shape variations not captured by $\vec{T}_1, \vec{T}_2, \vec{T}_3$:

$$\Delta_s = \vec{X}_s - (a_1\vec{T}_1 + a_2\vec{T}_2 + a_3\vec{T}_3). \tag{4.6}$$

PCA is then applied to the variable Δ_s to extract $\vec{T}_{s1}$ and $\vec{T}_{s2}$. A dataset of SuperCDMS DMC pulses was utilized for this template generation. These pulses were categorized into non-saturated and saturated sets based on an energy threshold below and above 10 keV, respectively, with approximately 100 events per channel selected for PCA training.

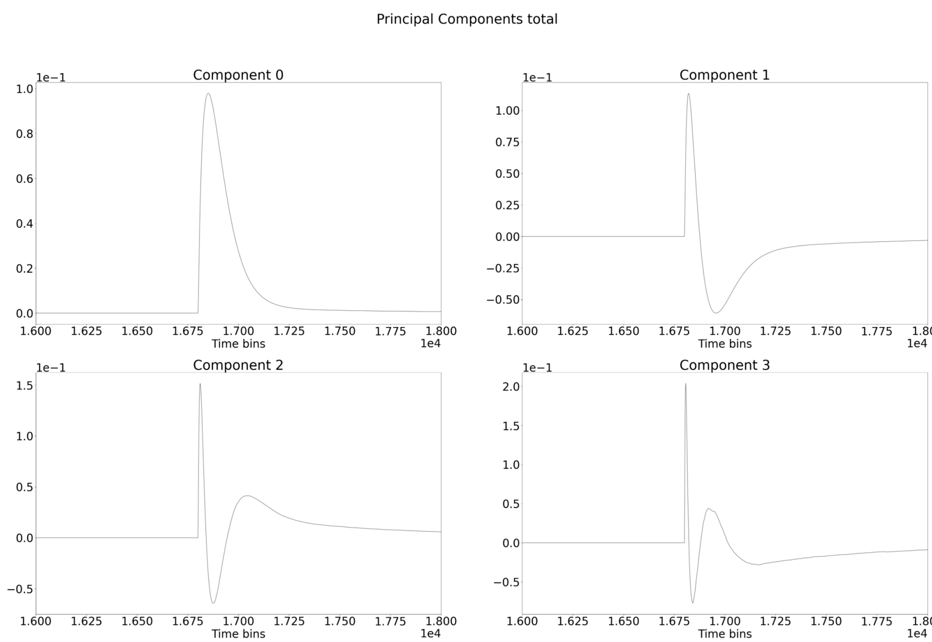


Figure 4.2: First four components of PCA applied to non-saturated simulated pulses.

Figure 4.2 illustrates the first four PCA components derived from the non-saturated pulses. To isolate the saturation effects, the first three components were projected onto the saturated pulses to determine their amplitudes, and the corresponding linear combination was subtracted from the raw pulses as in Equation 4.6. The PCA results for the saturated contribution Δ_s are shown in Figure 4.3. The two saturated components are combined with the three non-saturated templates to construct the complete template basis for the N×M analysis.

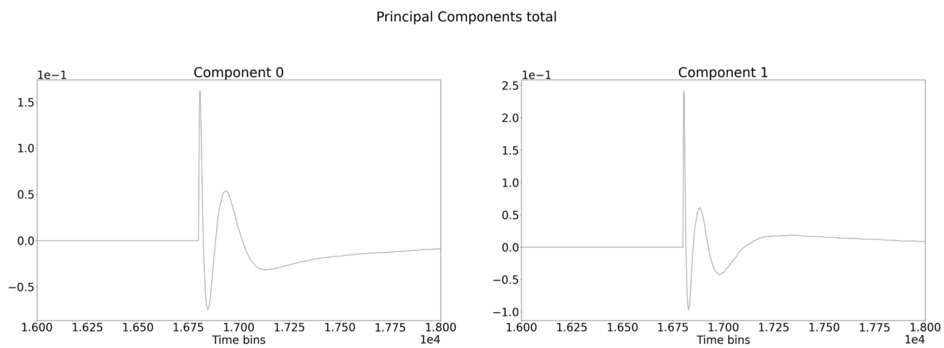


Figure 4.3: First two components of PCA applied to the saturated contribution of pulses.

A consideration when replacing higher-order non-saturated templates with saturated components is to minimize the resolution loss. Utilizing a template optimized for the pulse variations of saturated events risks degrading the fit quality for low-energy pulses. This degradation can be quantified by a scree plot illustrating the fractional variance contributed by each template for the non-saturated dataset, shown in Figure 4.4. As indicated by the scree plot, the fractional variance captured by the fourth non-saturated component is on the order of 10^{-4}. Maintaining a five-template basis remains crucial, as it provides more degrees of freedom for the N×M analysis. The scree plot, however, demonstrates that substituting the higher-order non-saturated templates with saturated ones should only incur a negligible loss in the resolution and the N×M fit. This exact resolution loss will be investigated further.

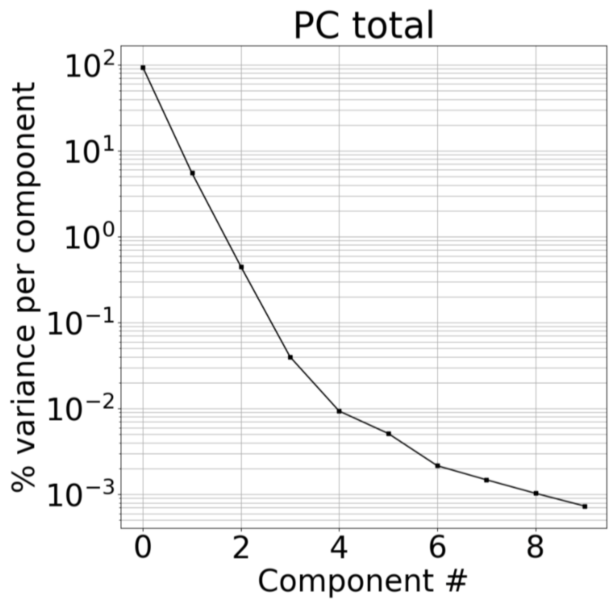


Figure 4.4: Scree plot of the variance contribution of non-saturated Principal Components (PC).

4.2 Saturated Template Evaluation

The aim of using saturated templates is to ensure better energy resolution at higher energies. The first way to quantify this improvement is by comparing the N×M χ^2 values of two different sets of templates. The first set consists of conventional templates derived from pulses at lower energies, and the second set includes templates with saturation accounted for. Figure 4.5 shows a scatter plot of the N×M χ^2 values for both template sets applied to 1,000 simulated pulses with energies ranging from 500 eV to 12 keV. At lower energies, both sets have roughly the same χ^2 values, which implies that the inclusion of saturated templates does not deteriorate the shape variation that we can capture. At higher energies, we see a large improvement in the pulse variation that we are able to capture. The more important factor, however, is evaluating how these templates affect the resolution of our energy estimator.

For the preliminary quantification of the energy estimation using the saturated templates, we use the N×M linear estimator combined with a Boosted Decision Tree (BDT). We select arbitrary regions of low and high energies (1-1.5 keV and 11-11.5 keV in this case) to train and test our en-

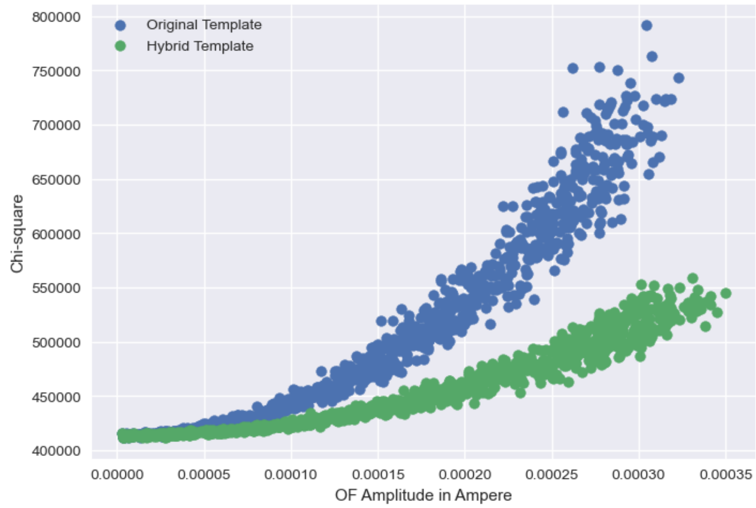


Figure 4.5: Comparison of the χ^2 fit of conventional templates (blue) and templates with saturated effects included (green).

ergy estimator. This process involves binning our data and filtering for the appropriate energy range (which is possible because we are using simulation data), and then using a subset of the training data to calculate the weights for the linear estimator using Equation 3.12. The remainder of the training subset is then used to train a BDT. This entire process is repeated for both the hybrid and original template sets across each energy range. Figures 4.6 and 4.7 show the distributions of the energy residuals for 1 keV and 11 keV, respectively.

The preliminary study shows an improvement at both energy ranges. Fitting the distributions with a Gaussian allows us to derive a standard deviation to quantify the energy resolution. For the 1 keV range, the standard deviations are 25 $\pm$ 2 and 24 $\pm$ 2 eV for the original and hybrid templates, respectively. The resolutions are within one standard deviation of one another and thus show that the inclusion of higher energy pulses in our template generation does not degrade the resolution at lower energies. At the 11 keV range, the resolutions are 77 $\pm$ 1 and 75 $\pm$ 1 eV for the original and hybrid templates, respectively, showing a slight improvement on the resolutions. This motivates the use of hybrid templates as the basis for N$\times$M moving forward in this paper.

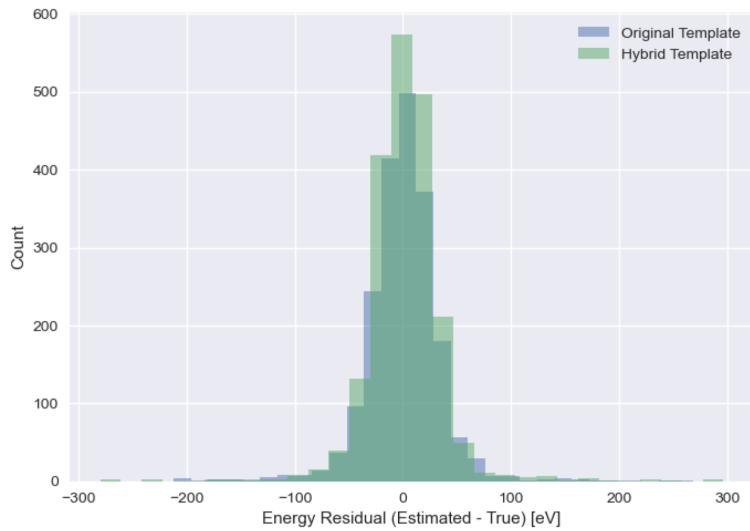


Figure 4.6: Energy residual for N×M energy estimator with hybrid and original template amplitudes at 1-1.5 keV.

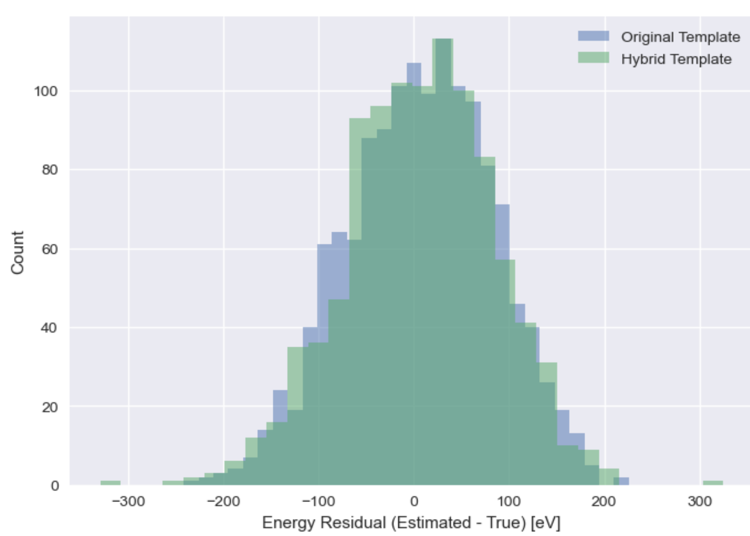


Figure 4.7: Energy Residual for N×M energy estimator with hybrid and original template amplitudes at 11-11.5 keV.

Chapter 5

General N×M Energy Estimator

5.1 Different Energy Reconstruction Algorithms with N×M Amplitudes

The previous chapter showed that incorporating saturation effects into our N×M templates yields improvements in the resolution of the N×M energy reconstruction. Moving beyond template generation, we can study alternative analytical techniques to determine whether reconstruction resolution can be improved beyond the baselines established by linear estimators and Boosted Decision Trees (BDT). To achieve this, we examine the relationship between true energy and total signal amplitudes for each template, characterizing the amplitude response across the full energy domain. Figure 5.1 shows five relevant plots. The first three amplitudes appear mostly linear, but the amplitudes associated with templates four and five exhibit some non-linearity. We can propose an analytical method that uses a quadratic model to parametrize the energy-amplitude relationship to capture some of this non-linearity .

First, to quantify the possible improvement from a quadratic form, we look at a set of simulations for the energy spectra from decays of Germanium-71, specifically the K and L shell energies at 10.3 and 1.3 keV, respectively. We can compare three different methods of reconstruction: linear, BDT and quadratic. For the linear method, a training set of N×M amplitudes and corresponding energies are used to derive the linear weights shown in Equation 3.12. The BDT method takes those linear estimates, calculates the correction factors and trains a decision tree. The quadratic method works similarly to a linear estimator, however now we can write our energy as

$$E_{quad} = \sum_{i=1}^{NM} w_i a_i + \sum_{i=1}^{NM} \sum_{j=1}^{NM} v_{ij} a_i a_j. \tag{5.1}$$

where v_{ij} and w_i are the weights for the quadratic estimator. The benefit

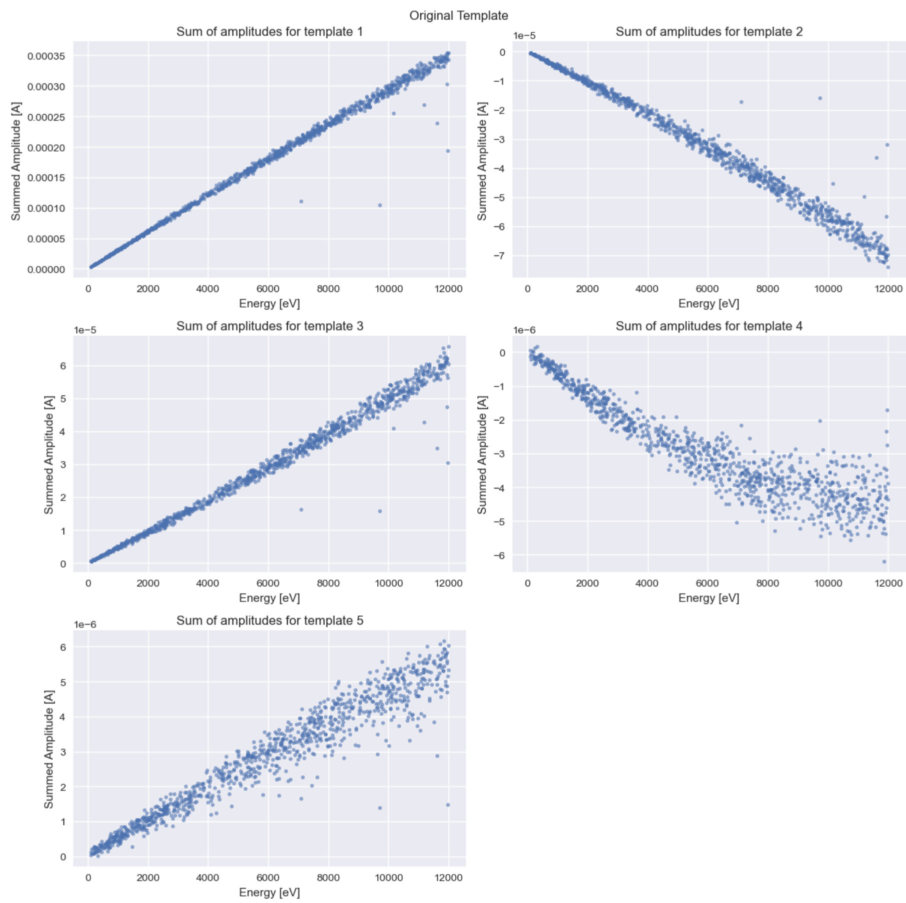


Figure 5.1: Scatter plots of amplitudes for each template summed over all channels for 1000 simulated events.

of this quadratic estimator is that the weights are a solution to a linear equation and thus computationally efficient to calculate. If we have a vector of true energies with the length of our training data sample space n, we can write the energies as

$$\vec{E} = \mathbf{A} \cdot \vec{\omega}. \tag{5.2}$$

Here, each row in the design matrix $\mathbf{A}$ consists of the N×M amplitudes, their cross-terms, and their squares, while $\vec{\omega}$ represents the vector containing all parameters w_i and v_{ij}. Thus, the optimal weights can be determined by finding the values that minimize the cost χ^2 function:

$$\chi^2(\vec{\omega}) = (\mathbf{A} \cdot \vec{\omega} - \vec{E})^T (\mathbf{A} \cdot \vec{\omega} - \vec{E}) \tag{5.3}$$

Normalizing our amplitudes becomes very important in our quadratic model. The N×M amplitudes can be on the orders of 10^{-5} to 10^{-7} and these amplitudes are then squared in the quadratic model and squared once again computationally in linear solvers. therefore, we may encounter floating point errors when solving for our quadratic weights. The solution to this problem is scaling the quadratic amplitudes. With 12 channels and 5 templates for each channel, we have 60 N×M amplitudes; if we calculate their cross-terms and squares, we will end up with 1890 unique amplitudes for each event. We have n (number of samples) of each quadratic amplitude, and we can get the average of each over n and scale them by dividing by the average. This removes any problem with floating point errors and the smaller orders of magnitude help with the stability of the quadratic estimator.

Another important consideration is the treatment of events near the edge of our detector. If an electron-hole pair is created near the edge of the detector, there is a probability of absorption by the sides of the detector and thus the electron-hole pair does not experience the complete NTL gain. The result is an apparent lower energy pulse in our detector. Figure 5.2 shows the distribution of the sum of the amplitudes from all the templates at 10.3 keV.

Since all of these events had the same energy, we would expect comparable amplitudes, but we can see a long tail. About 7% of the data fall in this tail and we can plot the position of these events. Figure 5.3 shows the position of energy deposition of the tail events and Figure 5.4 shows the X-Y slice of the distribution. As expected, most of these events lie in the edges of the detector. For now, we will ignore the events in the tail by doing a radial cut based on the deposition position of the simulated pulses, removing events outside the central 40 mm radius of the detector. This will superficially improve the resolution of our models but it will give a baseline

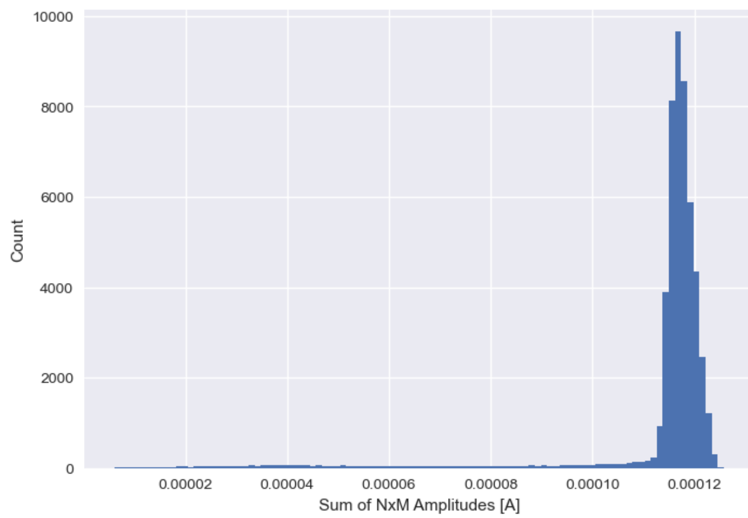


Figure 5.2: Histogram of the sum of the amplitudes from all 5 templates for 5000 events. All events are at simulated at 10.3 keV.

to compare all the methods. The treatment of these events will be discussed later.

To compare all of the reconstruction methods, testing and training sets were kept identical and each method used the scaled amplitudes. Figure 5.5 shows the result of the 3 different methods on a set of simulated Germanium-71 K-shell data at 10.3 keV. The BDT has an expected improvement over the linear estimator but the quadratic model provides a substantial improvement over both methods. We can fit each distribution to a Gaussian to calculate the standard deviation and percent resolution of our models. At 10.3 keV the linear model has a standard deviation of 27 eV and resolution of roughly 0.3%. The BDT performs almost twice as well with a standard deviation of 15 eV and resolution of 0.15%. The quadratic model finally had a standard deviation of 5 eV and resolution of 0.05%. These values are substantially better than previously calculated ones shown in Table 3.1 but most of that is likely due to the absence of edge events and the better signal-to-noise ratio at 10.3 keV compared to 500 eV.

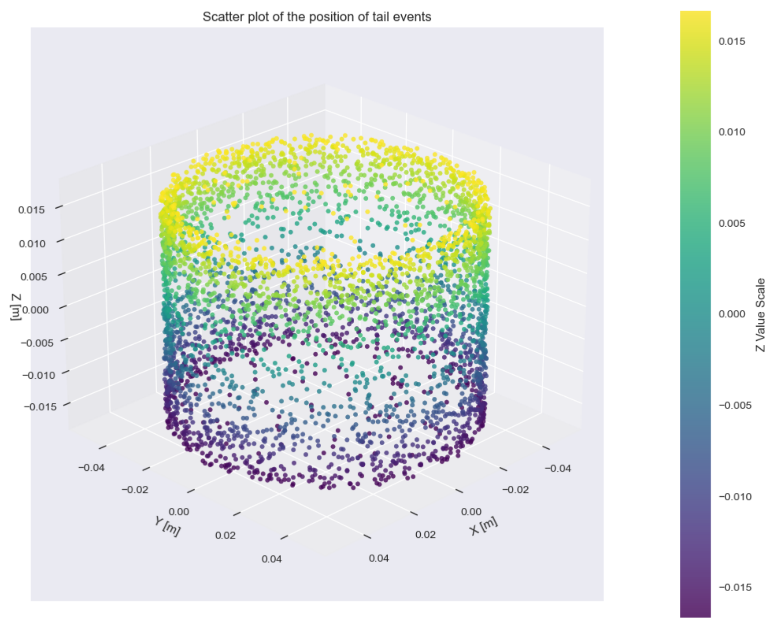


Figure 5.3: 3D scatter plot of the spatial distribution of events in the tail of the amplitude distribution.

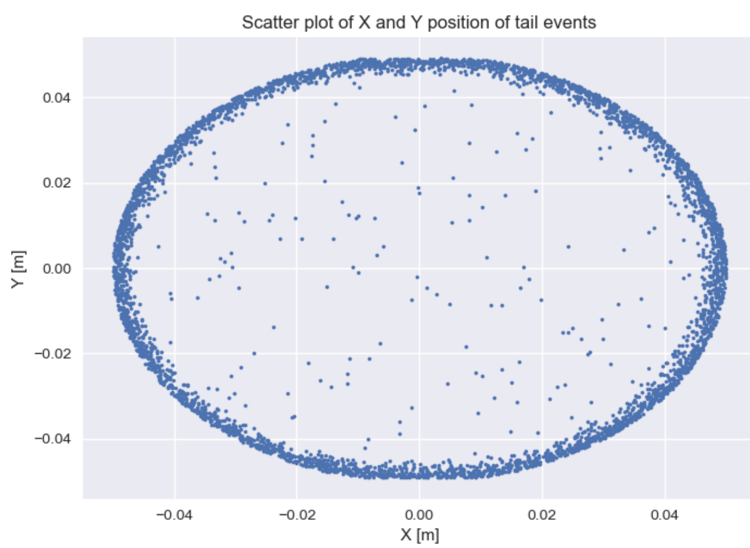


Figure 5.4: X and Y scatter plot of the spatial distribution of events in the tail of amplitude distribution.

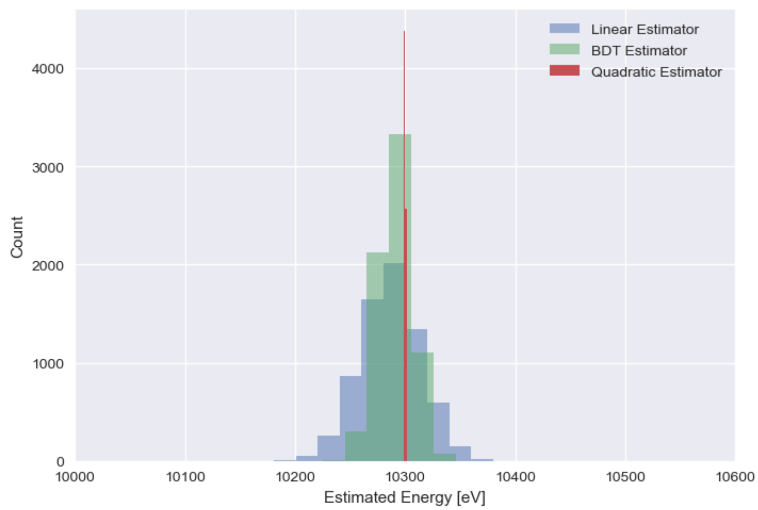


Figure 5.5: Distribution of 5000 10.3 keV events reconstructed using linear, BDT and quadratic methods. The events outside of the central 40 mm radius of the detector were removed.

Beyond ignoring the edge events, these results are impossible to achieve in practice due to the energy banding required to calculate them. We optimized our weights perfectly for events at 10.3 keV and applied them to the same energy events. In reality, we will not know the exact energy of an event prior to applying the weights, meaning weights must be created for a range of energies rather than one specific value.

5.2 Methods of Quadratic Weight Calculation

We have shown the efficacy of applying quadratic weights to a specific energy range of events for energy reconstruction. The challenge is now to find a way to reconstruct energies without the knowledge of the range of energies that the event falls in. The two approaches are to either look at a general set of weights that apply to all the energies, or create multiple sets of weights for different energy ranges.

We can start with general weights that either apply to or scale with all energies. Looking at Equation 5.1, the linear weights w_i are energy independent as our amplitudes contain the energy information. The quadratic weights v_{ij}, however, must contain some energy dependence, on dimensional grounds. The most basic approach for a general energy estimator is to include a factor that scales with energy. With basic dimensional analysis of energy and amplitude, we can conclude that the quadratic weights vary as $v_{ij} \propto 1/E_0$ where E_0 is the energy it is trained on. To obtain a new weight v'_{ij} for our new energy E', we then need to include a factor of E_0/E' for our original weight. This introduces a problem as we don't have knowledge of E'. However, we can introduce an approximate scaling by using the linear estimator to replace our E' term.

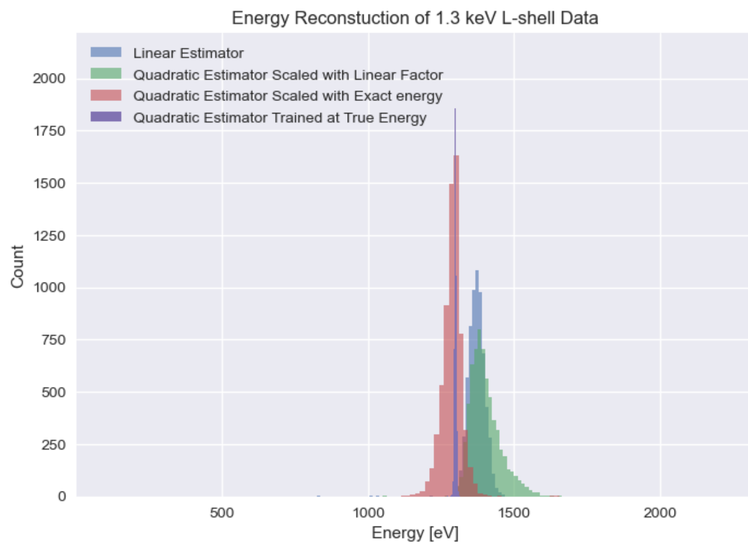


Figure 5.6: Comparison of a linear estimator, quadratic estimator with weights scaled by exact energy, quadratic estimator with weights scaled by a linear estimator and quadratic estimator with weights trained on the same energy applied on simulated data. Only central events inside the 40 mm radius of the detector are selected.

To compare the efficacy of each method we apply the estimator for 10.3 keV to L-shell events at 1.3 keV. Figure 5.6 shows the resolution of different methods of weight calculation. The linear estimator is trained using events at 10.3 keV and those weights are then applied to our 1.3 keV N×M amplitudes. The standard deviation of the resulting distribution is 26 eV. The quadratic estimator with a linear factor takes our quadratic weights calculated at 10.3 keV and multiplies the second-order weights with E_0/E_{lin}, where E_{lin} is the energy calculated from the linear estimator. The second quadratic estimator applies the same E_0/E' factor but here E' is 1.3 keV, the true energy of the events. This approach is not possible with real data but provides a point of comparison for our approximate model with the linear energy factor. Finally we have a third quadratic estimator where the weights are calculated with events at 1.3 keV and applied to a different set of 1.3 keV data. The linear estimator and the quadratic with an exact factor have standard deviations of 26 and 28 eV but average energies of 1370 and 1289 eV. The deviation is interesting as the weights are calculated at energies that might include some TES saturation and thus they would have a bias toward higher energies. The average from the linear estimator shows that the model overestimates the energy of 1.3 keV events by nearly 70 eV. The quadratic estimator seems to then correct for this bias but it overshoots the target energy. The quadratic with a linear estimate in the energy term performs the worst with a resolution of 38 eV and an average energy of 1400 eV; thus the error in the linear estimator seems to get further propagated when then used as an energy factor to scale the quadratic weights. The quadratic estimator with weights calculated at 1.3 keV performs the best with a standard deviation of 3 eV and an average energy of 1300 eV. This first test motivates the use of binned weights for energy estimation, where we split our data into smaller energy intervals and train a distinct set of weights at each energy bin. We can then use a general estimator, like a linear model to select the bins that each event goes through. However, we require further investigation of how both the parametrized and binned weight estimators can reconstruct events over a range of energies.

5.3 Initial Energy Determination with Linear Estimator

Both parametrized or binned quadratic energy estimators require an initial guess of energy, either to use as the dependent term or to select bins in each respective method. The most basic approach for determining an initial guess for each event is the use of a linear estimator calculated on the whole range of data. The energy independence of each linear weight further encourages this approach. Figure 5.6 shows that the linear estimator trained at 10.3 keV is able to achieve a narrow distribution of energies when applied to events of 1.3 keV, but it has an offset from the true energy. Therefore, for a general linear estimator, the best approach is to train the weights on the entire range of energies. Going forward, we will use a set of SuperCDMS simulated events with energies uniformly distributed from 0.5 to 12 keV. For an initial evaluation of the linear estimator, we can run a test on the subset of simulated data and calculate the residual of the linear energy estimate and the true energy. For the linear estimator the unscaled values of the amplitudes are used, as they are still large enough to avoid floating-point errors. Figure 5.7 shows the average of the residuals for events in each 500 eV bin. The residual varies between 20 and 350 eV and scales with energy. This motivates the use of varying bin sizes when training the binned quadratic estimator.

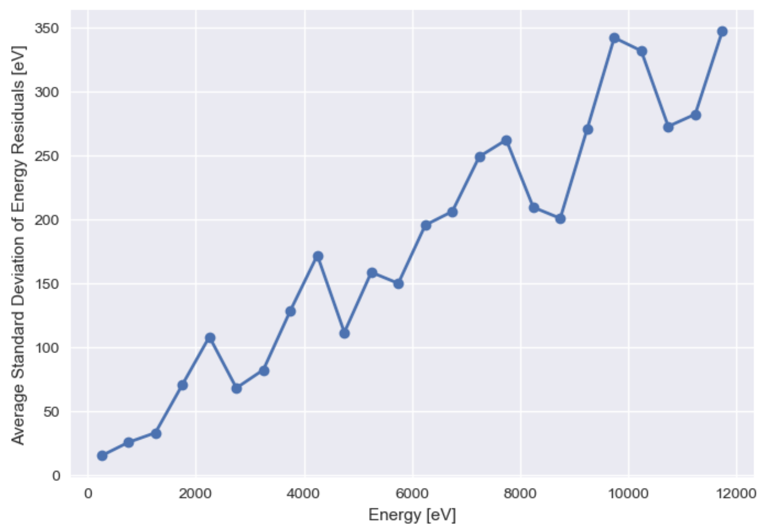


Figure 5.7: Average standard deviation of energy residuals in each 500 eV interval from a general linear estimator applied to events of energy 0.5 to 12 keV.

5.4 Comparing Quadratic Energy Estimators for Uniformly Distributed Events

The parametric energy estimator modifies our quadratic equation to:

$$E_{quad} = \sum_{i=0}^{NM} w_i a_i + \sum_{i=0}^{NM} \sum_{j=0}^{NM} \frac{v'_{ij} a_i a_j}{E_{lin}} + c. \tag{5.4}$$

We substitute the real energy with the linear estimate and we are able to add a constant c as we are now training our data on a range of energies, not just a single value. The binned quadratic estimator uses Equation 5.1, calculated in 23 bins between energies of 0.5 and 12 keV. As discussed before we expect the weights in the quadratic terms to go as $1/E$, thus meaning at low energies the adjacent energy bands should have substantially different weights. This could cause a lower resolution at low energies for both quadratic estimators. The parametric estimator has the benefit of being continuous over the energy range, likely improving the resolution of energies near the edges of the 500 eV bandwidth.

Creating a binned quadratic estimator starts with our training set of N×M amplitudes and the corresponding energies. The training set is separated into two groups: one set of roughly 10,000 events for the linear estimator and a set of 300,000 for the quadratic estimator. The optimized weights for the linear estimator are calculated using Equation 3.12 on our first set of training data. The second set of training data is then expanded into a quadratic form by calculating the cross-terms and squares of each. The quadratic amplitudes are then scaled by the event averages. Each event is separated into different bins based on its initial linear estimate. The energy bins start from 0.5 keV and go to 12 keV starting in intervals of 0.2 keV until 2 keV, then with intervals of 0.3 keV until 5 keV, and the rest of the range is with 0.5 keV intervals. For each bin, the optimum least square quadratic weights are calculated based on Equation 5.1 along with an additional constant term. The parametric quadratic estimator starts in the same manner with the initial linear estimate. The amplitudes are then expanded to include their cross-terms and squares and the second-order terms are divided by the linear energy estimate. The sets of quadratic terms have the form of Equation 5.4, where the weights are optimized using a least squares calculator. These weights are then applied to our test set of 20,000 events to compare both methods.

There are two ways of comparing our methods that give insight into different strengths of our analysis. When looking at the residual of the estimated and real energy, we can look at the absolute standard deviation of the distribution but we can also fit the distribution with a Gaussian for the fitted standard deviation. The absolute standard deviation is an indicator of the accuracy of our model on the entire set of data, and it accounts for the large outliers in the set. The fitted standard deviation is an indicator of the accuracy of the central bulk of data, while the information in the tail is suppressed. Both of these methods will be used for a comprehensive comparison of different methods and will be compared to the linear estimator used as the basis.

Figures 5.8 and 5.9 highlight different strengths of the different energy estimators. The common feature in both is the performance of the linear estimator at lower energies. As expected based on energy dependence, the weights are more sensitive at lower energies, and this can be seen in both Figures 5.8 and 5.9. This motivates the use of a pure linear estimator at lower energies and using a quadratic for higher-energy corrections. Looking at the absolute standard deviations in Figure 5.8 for energies above 2 keV, the quadratic estimators perform better than the linear. Between 2 keV and 7 keV the performance of both quadratic estimators are similar with the binned estimator having the lowest absolute standard deviation after 7 keV. With the fitted resolution in Figure 5.9 the linear model outperforms

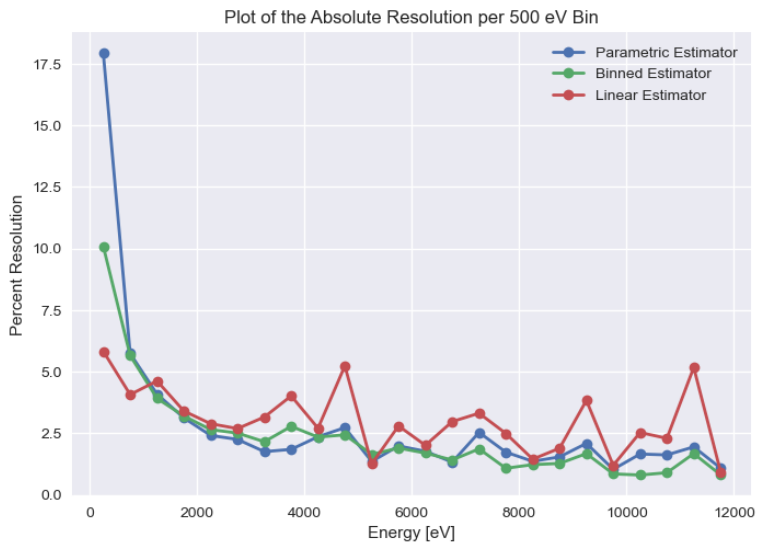


Figure 5.8: Comparison of the absolute standard deviation of the residual of the estimated and true energies for linear (red), parametric (blue) and binned quadratic estimators (green).

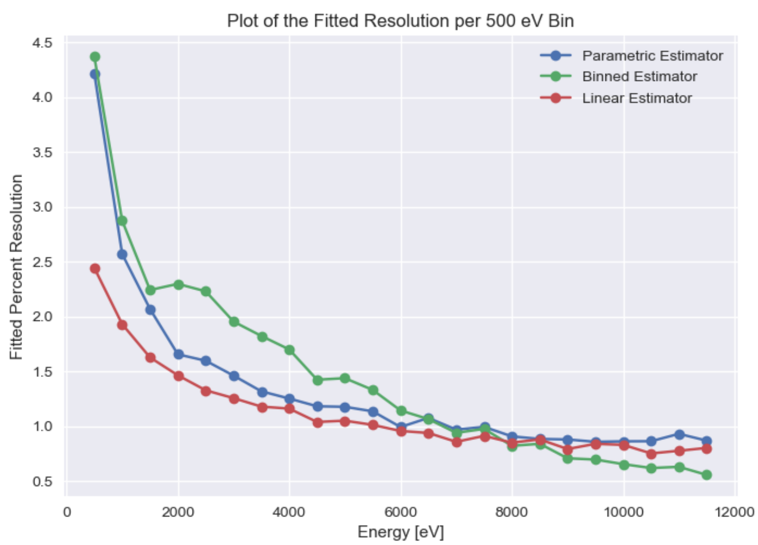
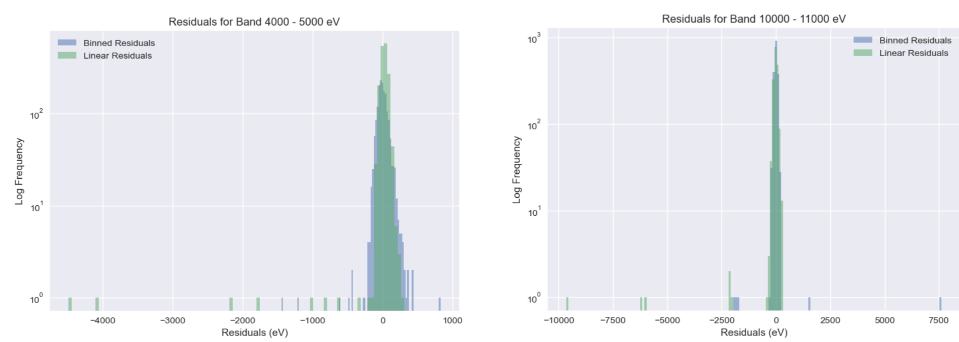


Figure 5.9: Comparison of the fitted resolution estimated from a Gaussian fit to the residual of the estimated and true energies for linear (red), parametric (blue) and binned quadratic estimators (green).

the binned estimator up to 6 keV, and, more notably, it consistently outperforms the parametric estimator. The parametric estimator is the weakest method: whilst being comparable to the other estimators for absolute and fitted resolution, it consistently performs more poorly than either of the other model and thus we can shift our focus to the linear and binned quadratic estimators.

To demonstrate the interpretation of the absolute and fitted resolutions, we can look at two examples of the distribution of the residuals from the linear and binned quadratic estimators at two ends of the energy spectrum.



(a) Histogram of the distribution of the residuals for events between 4 and 5 keV for linear and binned quadratic estimators on a log scale.

(b) Histogram of the distribution of the residuals for events between 10 and 11 keV for linear and binned quadratic estimators on a log scale.

Figure 5.10: Comparison of the distribution of residuals for linear and binned quadratic estimators.

Figure 5.10 showcases the takeaway from the absolute and fitted resolution plots. At 4-5 keV, the linear residual has the bulk of the data centralized; however, it also contains significant outliers with multiple events being mapped to energies 4 keV away. The binned quadratic estimator, however, has a broader distribution but with a more limited tail. However, at 10 keV, the binned quadratic estimator possesses a narrower distribution along with better suppression of outliers as was indicated in Figures 5.8 and 5.9.

This result showcases the limitations of extending our quadratic estimator to a broad range of energies. As previously shown in Figure 5.5, the quadratic estimator was a substantial improvement over the linear and the BDT energy estimates. However, when extending our energies, with either a parametric or a binned method, our resolution will always be limited by the error in our linear estimator. This can also be seen without the BDT estimator. We can create the same binning algorithm for the BDT by splitting our events into 0.5 keV intervals, and for each interval training a BDT. The amplitude scaling for the BDT includes dividing the 60 N×M amplitudes by the largest value. Figure 5.11 shows the comparison of the BDT method to the linear and binned quadratic estimator. Even though we have previously shown that for mono-energetic events a BDT can improve the resolution, on a broadband spectrum the limiting factor continues to be the initial energy guess. Thus, the comparison of the possible broad energy estimators positions the binned quadratic models as the optimum reconstruction solution, at least for specific ranges of energy. There are, however, many parameters inside the binned quadratic estimator that can be optimized to improve the performance.

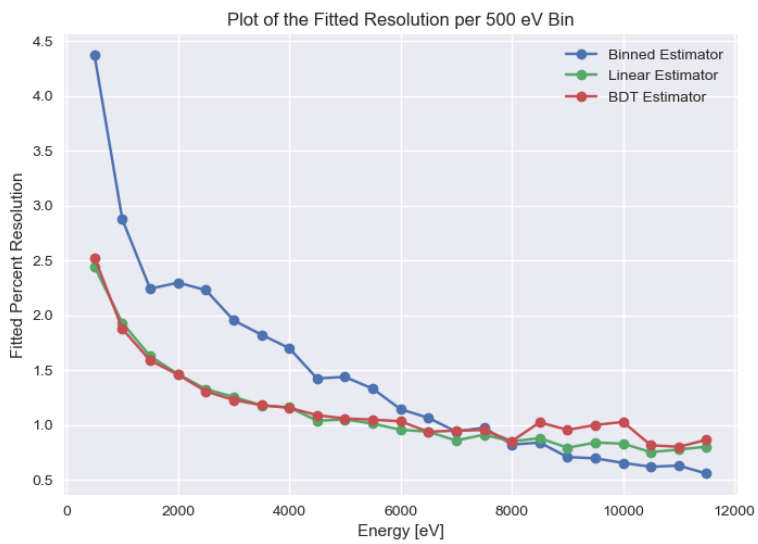


Figure 5.11: Fitted resolution plot for the residuals from linear (green), BDT (red) and binned quadratic estimator (blue).

5.5 Optimization of the Binned Quadratic Estimator

The main issue with the binned quadratic model is the limitation of its performance based on the selection quality of our linear estimator. While we are limited to using the linear estimate in our final estimator, there are two ways of selecting the energy bins in our quadratic model training. The two options are to bin our training data by either the true energy or the linear estimated energy. We can compare the resolutions of both of these approaches to decide on the optimum binning method.

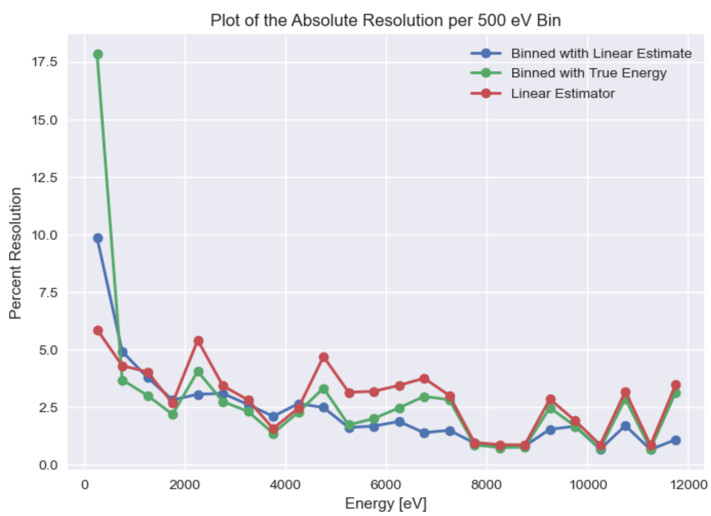


Figure 5.12: Comparison of the absolute resolution of the residual of the estimated and true energies for linear (red), binned quadratic estimator with linear estimated bins (blue) and binned quadratic estimator with true energy bins (green).

Figures 5.12 and 5.13 demonstrate the strengths of either approach. Using our linear estimate for binning allows us to provide better correction to the outliers produced by the linear estimator. During training, those outliers will be mapped to the correct energies and this translates to our testing trials. This can be seen in Figure 5.12 where the binned quadratic estimator with linear binning had the best absolute resolution above 5 keV. This factor however has the inverse influence on the distribution of the bulk of events. By training on the outliers from the linear estimator the quadratic weight has to account for a larger correction which ends up broadening the peak as

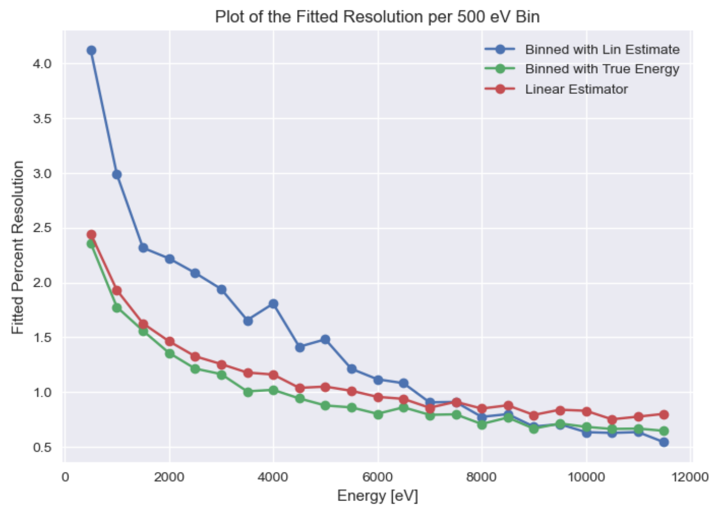


Figure 5.13: Comparison of the fitted resolution of the residual of the estimated and true energies for linear (red), binned quadratic estimator with linear estimated bins (blue) and binned quadratic estimator with true energy bins (green).

we can see in Figure 5.13. Figure 5.13 however shows the expected results from our previous trial on 10.3 keV mono-energetic events where, by using a quadratic fit over a linear one, we can achieve a better fitted resolution. This motivates the use of the true energies for training as it benefits from an improvement in handling large outliers from our linear estimator and it has the best performance in estimating the bulk of the events. The second technique applied in Figures 5.12 and 5.13 to improve our resolution is by extending our bins. It is inevitable that events will sometimes be misclassified into an adjacent bin; to counteract the error caused by this misalignment we have extended the width of each bin. For example for events with energies between 0.7-0.9 keV, the weights are trained on events with true energies of 0.6 - 1 keV where half a bin width is systematically added to each training set. This not only provides a more general quadratic weight to reduce errors, but it also allows for a larger training set. Figure 5.14 shows the fitted resolution of the binned quadratic estimator without the use of extensions in the bins and whilst it does perform better than the linear estimator it consistently has a resolution roughly 0.1% poorer than the quadratic with extensions bins.

The second issue to be addressed is the non-continuity of the energy

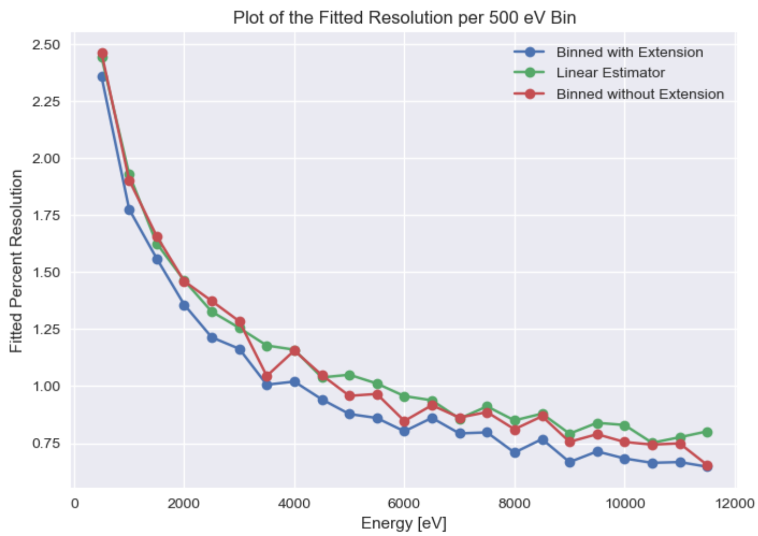


Figure 5.14: Comparison of the fitted resolution of the residual of the estimated and true energies for linear (green), binned quadratic estimator with extended bins (blue) and binned quadratic estimator with standard bins (red).

estimator. Regardless of where an event falls within a bin, the same set of weights gets applied, and while our bin extensions improve the performance for events near the edge of a bin, it helps to add a simple continuity measure in the inference step of the binned quadratic estimator. The easiest solution is a weighted average of our original energy estimate and an energy estimate from an adjacent band. We can select the adjacent band using the distance of the linear estimate from the center of the quadratic band. For example, an event with energy 0.75 keV energy will fall in the 0.7-0.9 keV band. As it falls on the lower end of this band, we will calculate E_{center}, an estimate with weights from the 0.7-0.9 keV band and E_{adj}, and estimate with weights from the 0.5-0.7 keV band. We can define our fractional parameter f as:

$$f = \frac{1}{2}\frac{|B - E_{lin}|}{\Delta_{Bin}/2}, \tag{5.5}$$

where B is the mid-point energy of the bin, E_{lin} is the linear estimate and Δ_{Bin} is the size of the bin. This is simply the measure of the fractional distance of our linear estimate and the bin center from 0 to 0.5. Using f we can define our final energy as:

$$E_{final} = (1 - f)E_{center} + fE_{adj}. \tag{5.6}$$

Thus our final energies will vary from a purely central estimate to a 50-50 estimate with the adjacent bin depending on the linear estimate. Figure 5.15 demonstrates the improvement made by using a weighted average to make our final energy predictions. The weighted estimator consistently improves on the central energy estimates as expected and thus is an important inclusion to the optimization of the binned quadratic estimator.

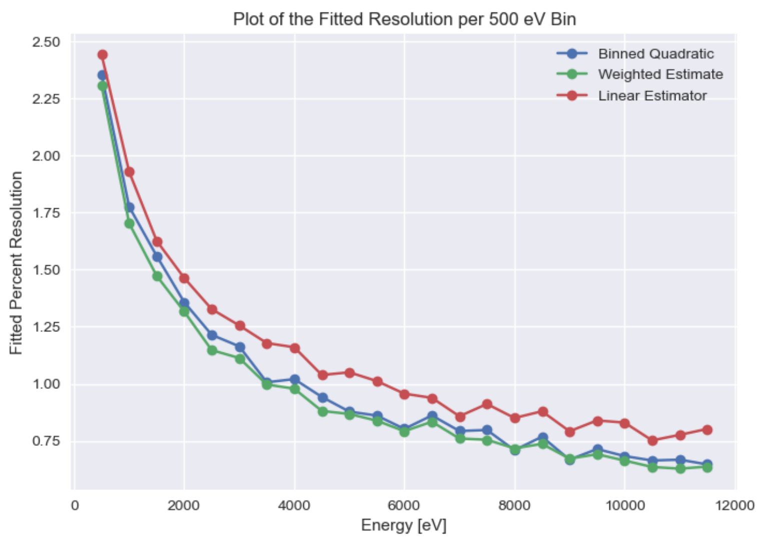


Figure 5.15: Comparison of the fitted resolution of the residual between the estimated and true energies. The models are linear (red), binned quadratic estimator with weighted output (green) and binned quadratic estimator standard output (blue).

As a final comparison to the BDT algorithm, we can include all the optimizations from our binned quadratic estimator in this section to the BDT method. This involves training on the true energies and extending the bins. Figure 5.16 shows that these optimizations do improve the BDT estimator at lower energies, but it still provides less of a resolution improvement compared to the binned quadratic estimator.

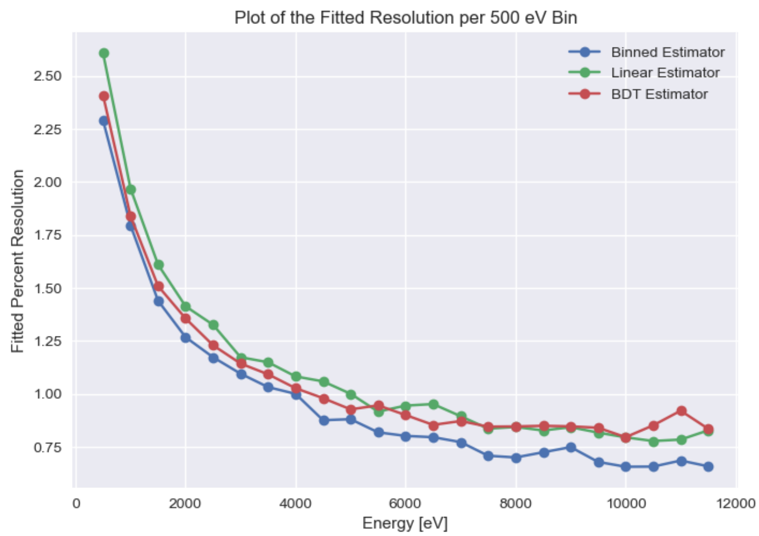


Figure 5.16: Comparison of the fitted resolution of the residual between the estimated and true energies. The models are linear (green), binned quadratic estimator (blue) and optimized BDT estimator (red).

5.6 Edge Event Analysis

The preceding analysis was concerned with the events in the central bulk of the detector. We can quantify if it is possible to extend this to the entire detector without a significant resolution loss from the edge events. We can calculate new quadratic weights, this time with the entire set of events including the edge events. We then apply those weights to the entire detector and to only central events and compare them to a set of central events with centrally calculated weights. Figure 5.17 demonstrates these three different applications. The best performance as expected is the centrally calculated weights applied to central events. This serves as the baseline resolution we were able to previously achieve. We can now look at weights calculated on the entire set of events and we can see how when applied to the entire detector we experience a substantial resolution loss, specifically at lower energies. We have shown that the events have asymmetric tails towards the lower energies, causing a compounding resolution loss towards 0.5 keV. Calculating weights on the entire range of the radius also reduces the resolution we can achieve for the central events. The quadratic model was aimed mainly at the bulk, and as we have seen in Figure 5.2, the edge events are distributed differently compared to the Gaussian central events and thus it is worthwhile focusing separately on the edge events.

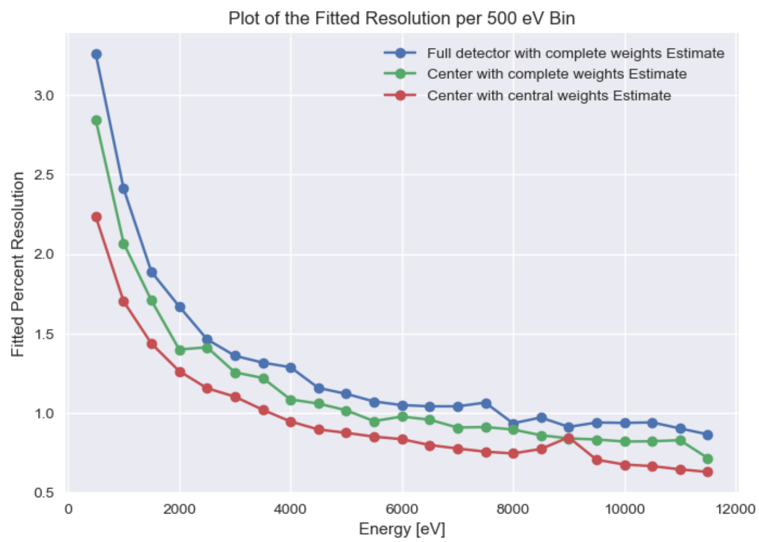


Figure 5.17: Comparison of the fitted resolution for the binned quadratic estimator. Weights calculated on the whole set are applied to all events (blue) and central events (green). This is contrasted with weights calculated on the central events applied on the center events (red).

The channel design of the SuperCDMS detectors provides a possible remedy to this issue. Shown in Figure 5.18, channels PAS1, PAS2, PBS1 and PBS2 are situated on the edge of the HV detectors. This design is foundational to position reconstruction for the detectors as the ratio of the signals from each channel contains spatial information for the event. This can also serve as a method of identification for events near the edge of the detectors.

The 40 mm radius initially selected was a conservative estimate of where the edge effects started. In reality, we can go to a radius of 46 mm without any impact on resolution. The goal is then to separate the events outside of the 46 mm radius and analyze them separately. A simple decision tree can be used to partition central and edge events. Due to the position of PA and PB channels, edge events will dominate those channels, making a decision tree trained on the ratio of the individual amplitudes to the amplitude of the channel with the largest amplitude an optimal partitioning tool. For an edge event the PA and PB amplitudes will be the largest and thus have

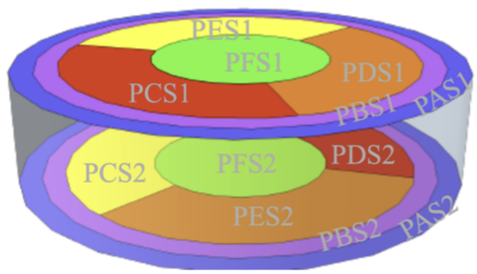


Figure 5.18: Channel partitioning of the SuperCDMS SNOLAB HV detectors [1].

a ratio close to one as opposed to the central channels that have a larger spread. Figure 5.19 demonstrates this effect for the first template of the PA channels. The decision tree can then make further cuts to separate any overlap of central and edge data. For the purposes of a radial cut, a decision tree with a depth of six is enough to achieve a 98% separation accuracy.

The second step is to design an analysis method for the energy estimation of the edge events. We can continue to use our binned quadratic estimator; however, for the edge cases, the distribution of our domain is largely different compared to the central case. To see the distribution difference, we can take the events between 6 and 6.2 keV and plot the distribution of the sum of the first 12 template amplitudes. Figure 5.20 shows the difference of event distribution for events inside and outside of the 46 mm radius. The central events mostly follow a Gaussian distribution over the mean but the edge events appear much more uniformly distributed over the entire range of energies. This causes a major overlap of events from different energies on the tail and thus more complex machine learning techniques can be more appropriate in edge event energy reconstruction. A basic but effective method that we will test is a MultiLayer Perceptron (MLP) neural network. MLP is a simple feedforward neural network, but the main advantage it possesses is the ability to use a non-linear activation function that the weighted outputs at each node can be fed through. Thus this universal network can model more complex functions than a quadratic [25]. This will raise the computational cost of our estimator; however, the edge cases are a limited sample of our entire events and thus it is a proportionally lower computational bottleneck for a full energy estimator.

For the binned quadratic estimator, some of the parameters had to be

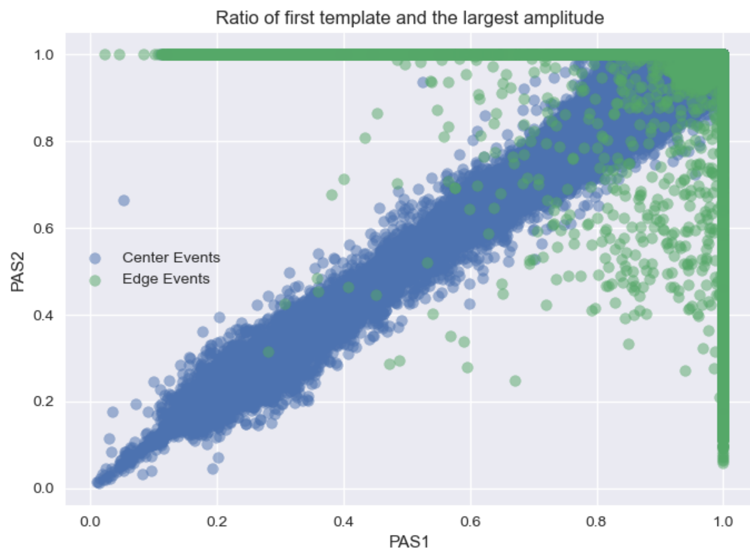


Figure 5.19: Comparison of the amplitude ratios (ratio of amplitude to the maximum N×M amplitude) from PAS1 and PAS2 for central and edge events.

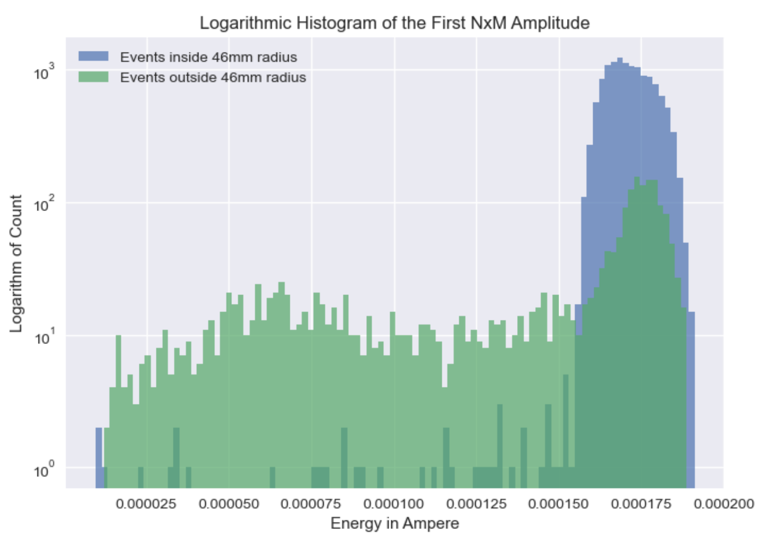


Figure 5.20: Logarithmic distribution of the total amplitude of events inside and outside the 46 mm detector radius.

adjusted, namely the binning size. At 200-500 eV binning range, we were limited in the number of events for training and thus every bin had to be doubled in size for a sufficient training sample size. This also proportionally increased the bin extension used during training. For the MLP we used a three-layer network and a non-linear activation function.

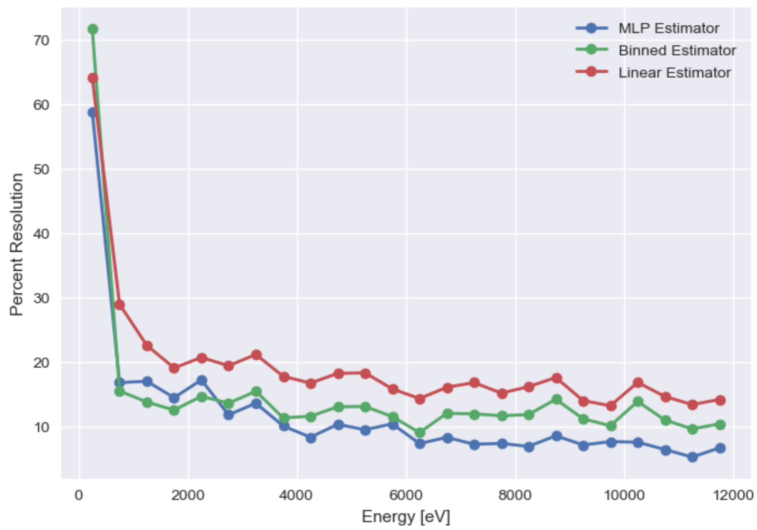


Figure 5.21: Comparison of the absolute resolution of the residual between the estimated and true energies for edge events. The models are linear, binned quadratic estimator and MLP neural network.

Figures 5.21 and 5.22 show the comparison of the absolute and fitted resolutions of the different energy estimators. In both cases, the MLP neural network has the strongest performance, as we would expect. The most interesting feature is in Figure 5.21, at the lowest energy. Due to the range of energies where the tail falls into the lowest bin, we see a large number of outliers at the lowest bin. The MLP neural network is then the best estimator for the edge events and the optimal selection for our complete estimator.

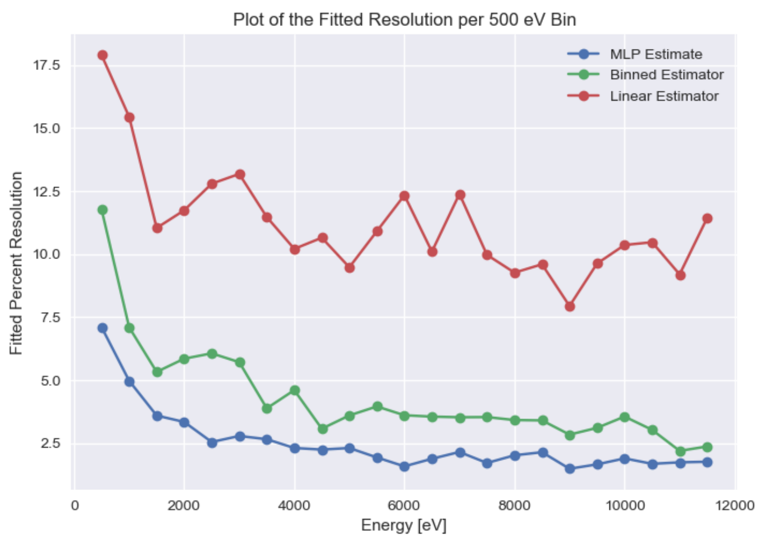


Figure 5.22: Comparison of the fitted resolution of the residual between the estimated and true energies for edge events. The models are linear, binned quadratic estimator and MLP neural network.

Chapter 6

Conclusion

This section contains the summary of the most significant conclusions from this thesis.

6.1 Template Generation

The standard template generation procedure for the N×M algorithm consisted of using Principal Component Analysis (PCA) on a set of non-saturated events in order to derive our templates. The proposed modification of the template generation procedure includes the integration of saturated templates into the higher-order templates. Using a five-template setup the first three templates are generated using the standard procedure for non-saturated templates. The last two templates however are trained using saturated events. For each saturated event the three non-saturated templates are projected out and subtracted from the signal and the residual of the signal and the three templates is run through PCA to attain the last two templates. As shown in Figures 4.6 and 4.7 we don't lose any resolution in the lower spectrum of energies but gain a modest improvement at saturated energy ranges.

6.2 Energy Reconstruction Algorithm

The N×M amplitudes obtained based on our templates are then the basis of the energy reconstruction algorithm. Figure 6.1 outlines the suggested reconstruction process in this paper. A simple decision tree radially separates the N×M amplitudes into central events with a radius less than 46 mm and edge events with a radius above that. A linear energy estimator, trained on the whole energy range, is used to give an initial energy estimate for each center event. Based on the linear estimate, each event goes through the appropriate binned quadratic estimator. The energy intervals range from 200 eV to 500 eV throughout the 12 keV energy range. The increasing interval allows us to accommodate the resolution of the linear estimator. For each event, two energy estimates are made: one from the selected bin and another from the closest adjacent bin to the linear estimate. The final output is the weighted average of those two values. The edge events go through a three-layer MLP neural network prediction. The result is a combined estimator capable of handling both bulk and edge events.

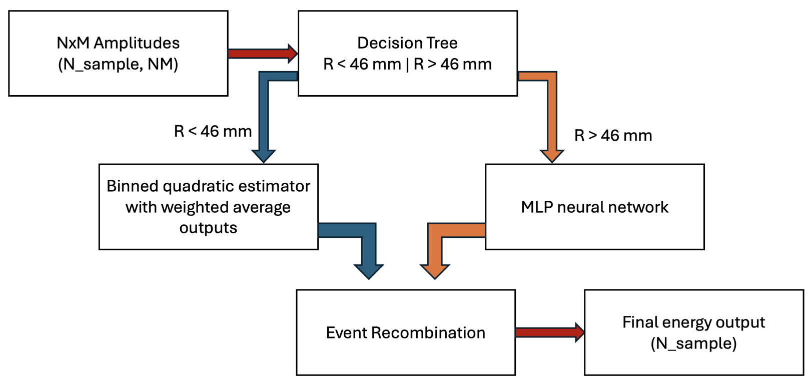


Figure 6.1: Data pipeline for energy reconstruction by $N \times M$ amplitudes. The data are radially separated by a decision tree. The central data then go through a binned quadratic estimator and the edge events are reconstructed using a MLP neural network.

We can compare this combined method with the binned quadratic estimator with weights calculated on the entire detector radius shown in Figure 5.17. Figure 6.2 shows the resolution of our final reconstruction algorithm. We were able to improve the model compared to a pure quadratic one and align it more with the resolution of the central quadratic estimate. Whilst some resolution loss was inevitable when looking at the events near the edge of the detector, the complex and universal nature of the MLP neural network and the radial separation of our data allowed us to minimize this resolution loss. The final resolution of the algorithm ranges from 2.3% at the lowest energy bin to a resolution of 0.7% at the highest end of the energy spectrum.

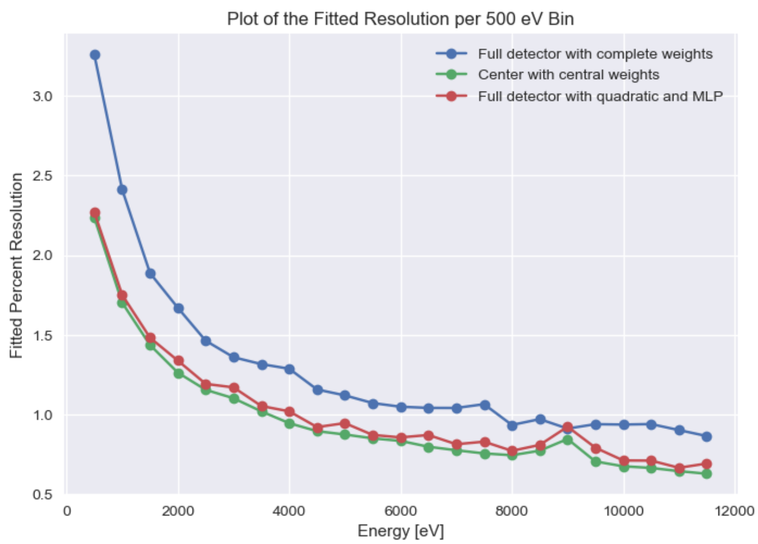


Figure 6.2: Comparison of the fitted resolution for full detector events. Weights calculated on the whole set are applied to evenly distributed events (blue). This is contrasted with estimates produced from the combination of the binned quadratic and MLP neural network algorithm (red). The green curve shows the resolution for only center events.

6.3 Future Work

The first major step in pushing forward this work is its application to real SuperCDMS SNOLAB data. Germanium-71 activation data is the best start as it has peaks at known energies and thus can be the first step in quantifying the energy resolution of our algorithm on a real set of experimental data. The SuperCDMS experiment is completing its commissioning in summer 2026, and science-quality data for this work will soon become available.

We can also investigate the integration of a continuous energy estimator within each bin. The parametric energy estimator was omitted due to the poor resolution compared to the binned model. The parametric estimator, however, had a major benefit in its formulation, which was its continuity. An expansion of the binned quadratic estimator could include the use of Equation 5.4 within each bin. The binning would allow for a more precise set of weights at each energy but that energy scaling would distinguish between energies at the extreme ends of each energy bin. The current solution to this discontinuity is the use of a weighted average, but using a scaled quadratic model could create a more robust and elegant solution to the continuity problem.

Bibliography

[1] M. F. Albakry, I. Alkhatib, D. Alonso-Gonzalez, J. Anczarski, T. Aralis, T. Aramaki, I. Ataee Langroudy, C. Bathurst, R. Bhattacharyya, A. J. Biff, P. L. Brink, M. Buchanan, R. Bunker, B. Cabrera, R. Calkins, R. A. Cameron, C. Cartaro, D. G. Cerdeno, Y. Y. Chang, M. Chaudhuri, J. H. Chen, R. Chen, N. Chott, J. Cooley, H. Coombes, P. Cushman, R. Cyna, S. Das, S. Dharani, M. L. di Vacri, M. D. Diamond, M. Elwan, S. Fallows, E. Fascione, E. Figueroa-Feliciano, S. L. Franzen, A. Gevorgian, M. Ghaith, G. Godden, J. Golatkara, S. R. Golwala, R. Gualtieri, J. Hall, S. A. S. Harms, C. Hays, B. A. Hines, Z. Hong, L. Hsu, M. E. Huber, V. Iyer, V. K. S. Kashyap, S. T. D. Keller, M. H. Kelsey, K. T. Kennard, Z. Kromer, A. Kubik, N. A. Kurinsky, M. Lee, J. Leyva, B. Lichtenberga, J. Liu, Y. Liu, E. Lopez Asamard, P. Lukens, R. Lopez Noe, D. B. MacFarlane, R. Mahapatra, J. S. Mammo, A. J. Mayer, P. C. McNamara, E. Michaud, E. Michielin, K. Mickelson, N. Mirabolfathi, M. Mirzakhani, B. Mohanty, D. Mondal, D. Monteiro, J. Nelson, H. Neog, J. L. Orrell, M. D. Osborne, S. M. Oser, L. Pandey, S. Pandey, R. Partridge, P. K. Patel, D. S. Pedrerosa, W. Peng, W. L. Perry, R. Podviianiuk, M. Potts, S. S. Poudel, A. Pradeep, M. Pyle, W. Rau, T. Reynold, M. Rios, A. Roberts, A. E. Robinson, L. Rosado, J. L. Ryan, T. Saab, D. Sadek, B. Sadoulet, S. P. Sahoo, I. Saikia, S. Salehi, J. Sander, B. Sandoval, A. Sattari, R. W. Schnee, B. Serfass, A. E. Sharbaugh, R. S. Shenoy, A. Simchony, P. Sinervo, Z. J. Smith, R. Soni, K. Stifter, J. Street, M. Stukel, H. Sun, E. Tanner, N. Tenpas, D. Toback, A. N. Villano, J. Viola, B. von Krosigk, O. Wen, Z. William, M. J. Wilson, J. Winchell, S. Yellin, B. A. Young, B. Zatschler, S. Zatschler, A. Zaytsev, E. Zhang, L. Zheng, A. Zuniga, and M. J. Zurowski. Multi-channel, multi-template event reconstruction for supercdms data using machine learning. pages 2–17, 2025. URL `https://arxiv.org/abs/2508.20090`.

[2] Aditi Pradeep. *Novel reconstruction techniques for detecting low mass dark matter in the SuperCDMS experiment and characterization of*

SuperCDMS SNOLAB detectors. PhD thesis, University of British Columbia, 2024. URL `https://open.library.ubc.ca/collections/ubctheses/24/items/1.0445508`.

[3] NASA Universe Web Team and Chelsea Gohd. Dark matter. *NASA Science*, March 2025. URL `https://science.nasa.gov/dark-matter/`.

[4] Marlene Gotz. Proceedings astronomy from 4 perspectives 1. cosmology dark matter. *Dark matter*, pages 1–3. URL `https://www.haus-der-astronomie.de/3440769/darkmatter.pdf`.

[5] K. Freese. Review of observational evidence for dark matter in the universe and in upcoming searches for dark stars. *EAS Publications Series*, 36:113–126, 2009. ISSN 1638-1963. doi: 10.1051/eas/0936016. URL `http://dx.doi.org/10.1051/eas/0936016`.

[6] Goddard Space Flight Center. Nasa webb 'pierces' bullet cluster, refines its mass, Jan 2026. URL `https://science.nasa.gov/missions/webb/nasa-webb-pierces-bullet-cluster-refines-its-mass/`.

[7] Anthony Challinor, Hiranya Peiris, Mario Novello, and Santiago Perez. Lecture notes on the physics of cosmic microwave background anisotropies. In *AIP Conference Proceedings*, page 1–5. AIP, 2009. doi: 10.1063/1.3151849. URL `http://dx.doi.org/10.1063/1.3151849`.

[8] David N. Schramm and Michael S. Turner. Big-bang nucleosynthesis enters the precision era. *Reviews of Modern Physics*, 70(1):308–312, January 1998. ISSN 1539-0756. doi: 10.1103/revmodphys.70.303. URL `http://dx.doi.org/10.1103/RevModPhys.70.303`.

[9] John Ellis. Limits of the standard model. pages 1–3, 2002. URL `https://arxiv.org/abs/hep-ph/0211168`.

[10] Gianfranco Bertone and Dan Hooper. History of dark matter. *Rev. Mod. Phys.*, 90:045002, Oct 2018. doi: 10.1103/RevModPhys.90.045002. URL `https://link.aps.org/doi/10.1103/RevModPhys.90.045002`.

[11] Nassim Bozorgnia, Joseph Bramante, James M. Cline, David Curtin, David McKeen, David E. Morrissey, Adam Ritz, Simon Viel, Aaron C. Vincent, and Yue Zhang. Dark matter candidates and searches. *Canadian Journal of Physics*, 103(8):671–673, August 2025. ISSN 1208-6045. doi: 10.1139/cjp-2024-0128. URL `http://dx.doi.org/10.1139/cjp-2024-0128`.

[12] Clara Murgui. Introducing the axion, Jan 2026. URL https://cerncourier.com/introducing-the-axion/.

[13] Teresa Marrodán Undagoitia and Ludwig Rauch. Dark matter direct-detection experiments. *Journal of Physics G: Nuclear and Particle Physics*, 43(1):5–10,28–29, 35–36, December 2015. ISSN 1361-6471. doi: 10.1088/0954-3899/43/1/013001. URL http://dx.doi.org/10.1088/0954-3899/43/1/013001.

[14] Taylor Rose Gray. Producing dark matter at particle accelerators. *Chalmers University of Technology*, pages 18–23, 2024. URL https://research.chalmers.se/publication/542372/file/542372_Fulltext.pdf.

[15] Marco Cirelli. Dark matter indirect detection. *Mémoire d'Habilitation à Diriger les Recherches*, pages 27–30, Oct 2014. URL https://www.marcocirelli.net/otherworks/HDR.pdf.

[16] A. Morselli, B. Canadas, and V. Vitale. The indirect search for dark matter from the centre of the galaxy with the fermi lat. *Il Nuovo Cimento C*, 34(3):312–315, June 2011. ISSN 03905551, 03905551. doi: 10.1393/ncc/i2011-10896-7. URL https://doi.org/10.1393/ncc/i2011-10896-7.

[17] R. Agnese, A. J. Anderson, T. Aramaki, I. Arnquist, W. Baker, D. Barker, R. Basu Thakur, D. A. Bauer, A. Borgland, M. A. Bowles, P. L. Brink, R. Bunker, B. Cabrera, D. O. Caldwell, R. Calkins, C. Cartaro, D. G. Cerdeño, H. Chagani, Y. Chen, J. Cooley, B. Cornell, P. Cushman, M. Daal, P. C. F. Di Stefano, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, M. Fritts, G. Gerbier, M. Ghaith, G. L. Godfrey, S. R. Golwala, J. Hall, H. R. Harris, T. Hofer, D. Holmgren, Z. Hong, E. Hoppe, L. Hsu, M. E. Huber, V. Iyer, D. Jardin, A. Jastram, M. H. Kelsey, A. Kennedy, A. Kubik, N. A. Kurinsky, A. Leder, B. Loer, E. Lopez Asamar, P. Lukens, R. Mahapatra, V. Mandic, N. Mast, N. Mirabolfathi, R. A. Moffatt, J. D. Morales Mendoza, J. L. Orrell, S. M. Oser, K. Page, W. A. Page, R. Partridge, M. Pepin, A. Phipps, S. Poudel, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, A. Roberts, A. E. Robinson, H. E. Rogers, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R. W. Schnee, B. Serfass, D. Speller, M. Stein, J. Street, H. A. Tanaka, D. Toback, R. Underwood, A. N. Villano, B. von Krosigk, B. Welliver, J. S. Wilson, D. H. Wright, S. Yellin, J. J. Yen, B. A. Young, X. Zhang, and

X. Zhao. Projected sensitivity of the supercdms snolab experiment. *Phys. Rev. D*, 95:1–4, Apr 2017. doi: 10.1103/PhysRevD.95.082002. URL https://link.aps.org/doi/10.1103/PhysRevD.95.082002.

[18] SuperCDMS Collaboration, M. F. Albakry, I. Alkhatib, D. W. P. Amaral, T. Aralis, T. Aramaki, I. J. Arnquist, I. Ataee Langroudy, E. Azadbakht, S. Banik, C. Bathurst, D. A. Bauer, R. Bhattacharyya, P. L. Brink, R. Bunker, B. Cabrera, R. Calkins, R. A. Cameron, C. Cartaro, D. G. Cerdeno, Y. Y. Chang, M. Chaudhuri, R. Chen, N. Chott, J. Cooley, H. Coombes, J. Corbett, P. Cushman, F. De Brienne, S. Dharani, M. L. di Vacri, M. D. Diamond, E. Fascione, E. Figueroa-Feliciano, C. W. Fink, K. Fouts, M. Fritts, G. Gerbier, R. Germond, M. Ghaith, S. R. Golwala, J. Hall, N. Hassan, B. A. Hines, M. I. Hollister, Z. Hong, E. W. Hoppe, L. Hsu, M. E. Huber, V. Iyer, A. Jastram, V. K. S. Kashyap, M. H. Kelsey, A. Kubik, N. A. Kurinsky, R. E. Lawrence, M. Lee, A. Li, J. Liu, Y. Liu, B. Loer, P. Lukens, D. B. MacFarlane, R. Mahapatra, V. Mandic, N. Mast, A. J. Mayer, H. Meyer zu Theenhausen, E. Michaud, E. Michielin, N. Mirabolfathi, B. Mohanty, S. Nagorny, J. Nelson, H. Neog, V. Novati, J. L. Orrell, M. D. Osborne, S. M. Oser, W. A. Page, R. Partridge, D. S. Pedreros, R. Podviianiuk, F. Ponce, S. Poudel, A. Pradeep, M. Pyle, W. Rau, E. Reid, R. Ren, T. Reynolds, A. Roberts, A. E. Robinson, T. Saab, B. Sadoulet, I. Saikia, J. Sander, A. Sattari, B. Schmidt, R. W. Schnee, S. Scorza, B. Serfass, S. S. Poudel, D. J. Sincavage, C. Stanford, J. Street, H. Sun, F. K. Thasrawala, D. Toback, R. Underwood, S. Verma, A. N. Villano, B. von Krosigk, S. L. Watkins, O. Wen, Z. Williams, M. J. Wilson, J. Winchell, K. Wyko, S. Yellin, B. A. Young, T. C. Yu, B. Zatschler, S. Zatschler, A. Zaytsev, E. Zhang, L. Zheng, and S. Zuber. A strategy for low-mass dark matter searches with cryogenic detectors in the supercdms snolab facility. pages 1–5, 2023. URL https://arxiv.org/abs/2203.08463.

[19] M I Hollister, D A Bauer, R C Dhuley, P Lukens, L D Martin, M K Ruschman, R L Schmitt, and G L Tatkowski. The cryogenics design of the supercdms snolab experiment. *IOP Conference Series: Materials Science and Engineering*, 278(1):1–2, dec 2017. doi: 10.1088/1757-899X/278/1/012118. URL https://doi.org/10.1088/1757-899X/278/1/012118.

[20] R. Agnese, A.J. Anderson, T. Aramaki, I. Arnquist, W. Baker, D. Barker, R. Basu Thakur, D.A. Bauer, A. Borgland, M.A. Bowles,

P.L. Brink, R. Bunker, B. Cabrera, D.O. Caldwell, R. Calkins, C. Cartaro, D.G. Cerdeño, H. Chagani, Y. Chen, J. Cooley, B. Cornell, P. Cushman, M. Daal, P.C.F. Di Stefano, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, M. Fritts, G. Gerbier, M. Ghaith, G.L. Godfrey, S.R. Golwala, J. Hall, H.R. Harris, T. Hofer, D. Holmgren, Z. Hong, E. Hoppe, L. Hsu, M.E. Huber, V. Iyer, D. Jardin, A. Jastram, M.H. Kelsey, A. Kennedy, A. Kubik, N.A. Kurinsky, A. Leder, B. Loer, E. Lopez Asamar, P. Lukens, R. Mahapatra, V. Mandic, N. Mast, N. Mirabolfathi, R.A. Moffatt, J.D. Morales Mendoza, J.L. Orrell, S.M. Oser, K. Page, W.A. Page, R. Partridge, M. Pepin, A. Phipps, S. Poudel, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, A. Roberts, A.E. Robinson, H.E. Rogers, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R.W. Schnee, B. Serfass, D. Speller, M. Stein, J. Street, H.A. Tanaka, D. Toback, R. Underwood, A.N. Villano, B. von Krosigk, B. Welliver, J.S. Wilson, D.H. Wright, S. Yellin, J.J. Yen, B.A. Young, X. Zhang, and X. and Zhao. Projected sensitivity of the supercdms snolab experiment. *Physical Review D*, 95(8):2–4,8–10, April 2017. ISSN 2470-0029. doi: 10.1103/physrevd.95.082002. URL `http://dx.doi.org/10.1103/PhysRevD.95.082002`.

[21] Robert Agnese. Simulating the SuperCDMS Dark Matter Detector Response and Readout. pages 32–34, 2017. doi: 10.2172/1457150. URL `https://inspirehep.net/literature/1679322`.

[22] R. Partridge. Ionization Readout Electronics for SuperCDMS SNOLAB Employing a HEMT Front-End. *Journal of Low Temperature Physics*, 176(5-6):932, September 2014. doi: 10.1007/s10909-014-1193-6. URL `https://ui.adsabs.harvard.edu/abs/2014JLTP..176..911P`.

[23] J.S. Wilson, H. Meyer zu Theenhausen, B. von Krosigk, E. Azadbakht, R. Bunker, J. Hall, S. Hansen, B. Hines, B. Loer, J.T. Olsen, S.M. Oser, R. Partridge, M. Pyle, J. Sander, B. Serfass, D. Toback, S.L. Watkins, and X. Zhao. The level-1 trigger for the supercdms experiment at snolab. *Journal of Instrumentation*, 17(07):2, 2022. ISSN 1748-0221. doi: 10.1088/1748-0221/17/07/p07010. URL `http://dx.doi.org/10.1088/1748-0221/17/07/P07010`.

[24] The SuperCDMS Collaboration. Supercdms snolab technical design report. pages 250–252, 2017.

[25] M.W Gardner and S.R Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the at-

mospheric sciences. *Atmospheric Environment*, 32(14):2627–2631, 1998. ISSN 1352-2310. doi: https://doi.org/10.1016/S1352-2310(97)00447-0. URL https://www.sciencedirect.com/science/article/pii/S1352231097004470.