

**Novel reconstruction techniques for detecting low mass dark matter in
the SuperCDMS experiment and characterization of SuperCDMS
SNOLAB detectors**

by

Aditi Pradeep

B. Tech, National Institute of Technology Calicut, 2019

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

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Physics)

The University of British Columbia
(Vancouver)

October 2024

© Aditi Pradeep, 2024

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

Novel reconstruction techniques for detecting low mass dark matter in the Super-CDMS experiment and characterization of SuperCDMS SNOLAB detectors

submitted by **Aditi Pradeep** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Physics**.

Examining Committee:

Scott Oser, Professor, Physics and Astronomy, UBC
Supervisor

Alison Lister, Professor, Physics and Astronomy, UBC
Supervisory Committee Member

Douglas Scott, Professor, Physics and Astronomy, UBC
Supervisory Committee Member

Christopher Hearty, Professor, Physics and Astronomy, UBC
University Examiner

Raymond Ng, Professor, Computer Science, UBC
University Examiner

Additional Supervisory Committee Members:

Ingrid Stairs, Professor, Physics and Astronomy, UBC
Supervisory Committee Member

Abstract

The existence of dark matter has been inferred through many pieces of astrophysical evidence. However, much about its nature is unknown to this day. The several decades-long search for dark matter has given rise to many experiments and even more dark matter candidates. SuperCDMS is a direct detection experiment that uses cryogenic detectors to probe interactions of dark matter particles with Standard Model particles. Work done towards advancing two frontiers of the SuperCDMS SNOLAB experiment will be the main points of discussion in this thesis – (a) the characterization of the new SuperCDMS SNOLAB detectors, and, (b) the development of novel event reconstruction techniques for the experiment.

After introducing the motivation for dark matter and the current experimental techniques used to detect it, the thesis will introduce SuperCDMS detection principles and pre-requisite knowledge to motivate and understand most of the work described within. First, preliminary results will be presented from testing and characterizing SuperCDMS detector towers at the Cryogenic Underground TEst facility, a low background test facility at SNOLAB in Sudbury, Canada. The remainder of this thesis will focus on novel reconstruction techniques critical to achieving the experiment’s projected sensitivity post commissioning. This discussion is divided into two. The first aspect will present a detailed discussion of advanced reconstruction algorithms to fit data sampled at non-uniform speeds to keep within the bandwidths of the readout electronics and maintain low trigger thresholds at SuperCDMS SNOLAB. The second major development discussed will be a novel reconstruction technique called the $N \times M$ filter, which fits N channels with M shapes/templates simultaneously and develops a pipeline that integrates machine learning to achieve excellent resolution improvement. The key outcomes of this thesis are (a) capability demonstration of the SuperCDMS SNOLAB detectors, (b) development of a less memory-intensive algorithm to process non-uniformly sampled data, and, (c) demonstration of a two-fold improvement and nearly a four-fold improvement in energy resolution in old and new data sets using the $N \times M$ filter, respectively.

Lay Summary

The existence of an unseen component in the Universe has been inferred through astrophysical and cosmological observations. This component is only known to exhibit gravitational effects and interacts very weakly with the ordinary matter that we see around us, reflecting no light, and hence is called “dark matter”. This thesis describes work done in a rare event search experiment called SuperCDMS, which looks for signals from the scattering of dark matter particles in semiconductor crystal detectors placed deep underground in a lab called SNOLAB in Sudbury, Canada. The thesis describes results from testing new detectors for the first time, and demonstrates improved energy resolution from novel ways to model the signal produced from the devices incorporating advanced statistics and machine-learning-based techniques.

Preface

This thesis predominantly shows work that is my own, but often builds on existing work or uses data and technology that were developed in collaboration with other SuperCDMS researchers. This is simply due to the scale of the experiment encompassing over 100 collaborators who contribute to everything, including but not limited to detector design, development, facility operations, detector operations, analysis, software, hardware, and engineering.

Chapter 1 is based on research by numerous scientists since the 1930s who played a significant role in formulating our understanding of dark matter to the extent we know today. Chapter 2 covers the legacy of the SuperCDMS experiment, which was conceived in the 1990s and has continued to set world-leading limits on dark matter sensitivity ever since. The experiment has risen to its current stature due to consistent contributions from many researchers across North America, Europe and Asia. The two initial chapters are included to provide context for my own research, which builds on past results and data.

The SLAC tower testing effort described in Sec. 3.2 was led by Francisco Ponce with me playing a prominent role in detector operations and data analysis. The results shown in the rest of Ch. 3 are primarily from the preliminary analysis of data taken during the five-month long test run at the Cryogenic Underground TEST facility at SNOLAB, Canada. The success of this test run is owed to many shifters, including myself, who volunteered their time for detector operation, on-site shifts, and data analysis. A special thanks is owed to Yan Liu who co-led the testing campaign with me, and to Wolfgang Rau and Ziqing Hong for their consistent support and timely feedback during this hectic time. Most of the results in this chapter are my own work, but I also present some key findings from a few other analyzers who have been awarded due credit for their work in the thesis body, namely Madeleine Zurowski's noise modeling, Joseph Mammo's noise PSD comparison plot, Shubham Pandey's Z partition parameter and Moaaz Elwan's single e^-/h^+ pair analysis. These points are included to provide a complete description of the detector and were performed under my supervision in my role as the CUTE-SuperCDMS analysis liaison. The noise

mitigation team mentioned in this chapter consisted of Yan Liu, Ziqing Hong and Tom Ren who went down the mine-shaft to figure out why our detectors appeared noisier than expected.

Chapter 4 briefly introduces work by William Page, who developed an algorithm to handle non-uniformly sampled data. The new algorithms that I present in this chapter are meant to primarily reduce computational time and memory usage, which were limitations of Page’s techniques.

The $N \times M$ filter studied in Ch. 5 and 6 was conceived many years ago by Matt Pyle and others, but has evolved significantly since its conception, primarily due to work by Yan Liu and Scott Oser. The work described in this chapter is currently being prepared for publication [1]. My contributions are the development of the machine learning parts, including PCA and gradient boosted decision trees, and algorithm validation. The idea to use gradient boosted decision trees rather than neural networks (which are severely data hungry), came from Wojtek Fedorko at TRIUMF, who was also kind enough to provide a setup on the TRIUMF computing cluster to perform my studies. The simulated data used in Ch. 6 was produced by Rik Bhattacharyya and Stefan Zatschler using the SuperCDMS simulation infrastructure, which is primarily supported by researchers from the University of Toronto and Texas A&M University. Much of the background modeling and cuts used for the sensitivity study in this chapter were originally the work of former SuperCDMS collaborators, William Page and others.

Implementation of the algorithms into CDMSBats (our data processing software) and data processing in general was supported by Bruno Serfass and Yan Liu, the CDMSBats package managers. Last but most importantly, all of my work in this thesis was supervised by Scott Oser, who is owed a great acknowledgment for his wholehearted support and feedback.

Table of Contents

Abstract	iii
Lay Summary	iv
Preface	v
Table of Contents	vii
List of Tables	xi
List of Figures	xiii
List of Abbreviations	xix
Glossary	xxiii
Acknowledgments	xxv
Dedication	xxvii
1 Dark matter and experimental searches	1
1.1 Evidence for dark matter	3
1.1.1 Astrophysical evidence	3
1.1.1.1 Rotational curves of galaxies	3
1.1.1.2 Anomalies in measurements from galaxy clusters	4
1.1.2 Cosmological evidence	6
1.1.2.1 Anisotropies in the Cosmic Microwave Background	7
1.1.2.2 Large scale structure formation in the universe	8

1.1.2.3	Measurements of matter density	10
1.2	Dark matter in particle physics	10
1.2.1	Dark matter candidates	12
1.2.2	Dark matter detection	14
2	Direct detection of dark matter with SuperCDMS	19
2.1	Detection principle	20
2.2	SuperCDMS detectors	25
2.3	Data acquisition	27
2.4	Event reconstruction	30
2.5	Detector calibration and dark matter sensitivity	35
3	Characterizing SuperCDMS SNOLAB detectors	39
3.1	The CUTE facility	39
3.2	Tower testing at SLAC	42
3.3	Tower testing at CUTE	48
3.3.1	Detector characterization and functionality	52
3.3.2	Noise environment	55
3.3.3	Detector calibration	61
3.3.3.1	Calibration preliminaries	61
3.3.3.2	Fiducialization	67
3.3.3.3	Relative calibration	72
3.3.3.4	Ge calibration	79
3.3.3.5	Si calibration	83
3.3.4	Neutralization	89
3.4	Summary and outlook	90
4	Event reconstruction at SuperCDMS SNOLAB	92
4.1	SuperCDMS SNOLAB data acquisition strategy	92
4.2	Hybrid readout – features and challenges	95
4.3	Event reconstruction with hybrid data	99
4.3.1	Hybrid Optimal Filter	99
4.3.2	2-speed Optimal Filter	102
4.3.2.1	Wrap-around effect	105
4.3.2.2	Combining amplitudes	109

4.3.2.3	Algorithm validation with toy models	112
4.3.2.4	Algorithm validation with real data	117
4.3.3	2-speed Minimizer	122
4.3.3.1	Pulse shape functions	123
4.3.3.2	Algorithm	124
4.3.3.3	Algorithm validation with toy models	125
4.4	Conclusion and future directions	128
5	$N \times M$ reconstruction and machine learning	130
5.1	Multi-channel reconstruction	130
5.2	The $N \times M$ filter	134
5.3	The $N \times M$ processing pipeline	138
5.3.1	Template generation with unsupervised learning	138
5.3.2	Reconstructing physics with supervised learning	142
5.4	$N \times M$ filter for hybrid data	149
5.4.1	2-Speed $N \times M$ filter	149
5.4.2	Algorithm validation	151
5.5	Summary	155
6	Application of $N \times M$ filter to SuperCDMS data	156
6.1	Validation with GEANT4 simulations	156
6.1.1	The data	156
6.1.2	Template generation	158
6.1.3	Energy reconstruction	160
6.1.4	Position reconstruction	165
6.2	Validation with CDMSLite data	168
6.2.1	CDMSLite experiment	168
6.2.2	Template generation	169
6.2.3	Energy reconstruction	171
6.2.4	Impact on dark matter sensitivity	180
6.3	Validation with data from the CUTE facility	185
6.3.1	Data preparation	186
6.3.2	Reconstructing the Ge spectrum	187
6.3.3	Effect of missing channels in the detector	190

6.3.4	Reconstructing the Si spectrum	193
6.4	Summary and future directions	194
7	Conclusions	196
	Bibliography	199
A	Deriving the $N \times M$ Filter amplitudes	211

List of Tables

Table 3.1	The measured mean resistance parameters of interest for SuperCDMS SNOLAB detectors.	46
Table 3.2	Total exposure for each detector type for the collected data in units of detector-days during CUTE tower testing.	50
Table 3.3	Tabulation of signal-to-noise ratio at various bias points for detector 2 in Tower 3.	54
Table 3.4	Electron capture probabilities for ^{71}Ge	62
Table 3.5	Expected location of Compton steps in SuperCDMS detectors.	66
Table 3.6	High energy electron capture γ peaks of ^{133}Ba	66
Table 3.7	The energy resolution of calibration peaks in T3Z1.	82
Table 3.8	The estimated baseline resolution of tested Ge detectors in Tower 3 from the CUTE facility.	82
Table 3.9	The estimated baseline resolution of tested Si detectors in Tower 3 from the CUTE facility.	89
Table 4.1	Estimated maximum data rates and throughput for SuperCDMS SNOLAB for fully sampled 52.43 ms waveforms.	94
Table 4.2	Table of pre-, on- and post- pulse lengths for a hybrid waveform for each channel type.	97
Table 4.3	Estimated maximum data rates and throughput for SuperCDMS SNOLAB for hybrid sampled 52.43 ms waveforms.	98
Table 4.4	A comparison of reconstruction resolution between various energy estimators for a mono-energetic toy spectrum.	116
Table 4.5	A comparison ^{55}Fe peak resolutions from various energy estimators.	121
Table 4.6	A comparison baseline resolutions from various energy estimators.	122

Table 4.7	A comparison of percentage error between various estimators for a mono-energetic toy spectrum.	126
Table 5.1	A comparison of widths of energy distributions of simulated toy pulses.	154
Table 6.1	A comparison of energy resolution between different energy estimators.	165
Table 6.2	A list of correction factors to correct the $N \times M$ energy estimator for CDMSLite Run 3 data.	175
Table 6.3	Comparison of energy resolution of the $N \times M$ filter with other energy estimators in SuperCDMS along with the 1σ statistical uncertainty from the Gaussian fit. .	176
Table 6.4	A comparison of baseline resolution for $N \times M$ and 1×1 OFs for each sub-run in CDMSLite.	178
Table 6.5	Estimates of the parameters in Eq. (6.4) for CDMSLite data.	179
Table 6.6	A table of resolution values chosen for demonstrating the effect of resolution on dark matter sensitivity. The two models differ only modestly in resolution at low energies, but show two-fold reduction at 10.37 keV. The difference in models roughly capture the improvement in resolution from performing a GBDT correction.	180
Table 6.7	A comparison of activation peak resolutions when various types of channels are “missing” in the detector.	192

List of Figures

Figure 1.1	Energy density $\rho(t)$ as a function of scale factor a for each component in a flat universe ($\Omega_k = 0$).	3
Figure 1.2	Universal rotation curve of spiral galaxies.	5
Figure 1.3	Image of the Bullet cluster obtained from the Chandra X-ray observatory and optical data.	6
Figure 1.4	The cosmic microwave background anisotropy power spectrum from Planck 2018 data.	8
Figure 1.5	Comparison of results from redshift surveys with simulations from the Millennium project.	9
Figure 1.6	Dark energy density, Ω_Λ vs matter density, Ω_m for the Λ CDM model from the Hubble Space Telescope cluster Supernova survey	11
Figure 1.7	Evolution of dark matter comoving number density as a function of dark matter mass to temperature.	13
Figure 1.8	Schematic representation of various dark matter detection channels.	15
Figure 1.9	Schematic representation of various direct detection strategies.	17
Figure 1.10	Comparison of the most recent exclusion limits from various direct detection experiments for spin-independent WIMP-nucleon scattering.	17
Figure 2.1	Schematic representation of photon absorption in direct vs indirect band gap materials.	22
Figure 2.2	Cartoon of a TES transition curve.	23
Figure 2.3	Schematic showing the channel map for SuperCDMS SNOLAB HV and iZIP detectors.	26
Figure 2.4	Fabricated SuperCDMS SNOLAB detectors.	27
Figure 2.5	Schematic of the experimental setup of SuperCDMS SNOLAB.	27

Figure 2.6	Schematic of the SuperCDMS trigger system.	28
Figure 2.7	Comparison of fully sampled waveforms with a hybrid sampled waveform. . .	30
Figure 2.8	Flow chart showing the workflow of the processing package, CDMSBats. . . .	31
Figure 2.9	Comparison of a toy pulse overlaid with a template. The simplest way of extracting meaningful information from pulses often involves fitting a mod- el/template to the signal.	32
Figure 2.10	Pulses for each channel from an example event decomposed into prompt (fast) and ballistic (slow) components.	34
Figure 2.11	Sensitivity projections for spin-independent nuclear recoil dark matter at Su- perCDMS SNOLAB.	37
Figure 2.12	Sensitivity projections for electron recoil dark matter at SuperCDMS SNOLAB.	38
Figure 3.1	Cartoon image showing the design of the CUTE facility.	41
Figure 3.2	Schematic of the CUTE facility’s shields.	42
Figure 3.3	Schematic of the Ba source deployment system at CUTE.	43
Figure 3.4	Images of the Bluefors fridge at SLAC with the detector tower being installed.	44
Figure 3.5	An example SQUID pattern for an iZIP detector obtained from tower testing at SLAC.	44
Figure 3.6	The TES sensor circuit which is current biased (left) and the Thevenin equiva- lent voltage circuit (right).	46
Figure 3.7	A comparison of the measured critical temperatures of SuperCDMS SNOLAB detectors.	47
Figure 3.8	Examples of a few types of events observed during tower testing at SLAC. . . .	49
Figure 3.9	Example charge channel pulses for a SuperCDMS SNOLAB iZIP detector. . .	50
Figure 3.10	Timeline for tower testing.	51
Figure 3.11	Installation of Tower 3 into CUTE.	53
Figure 3.12	A plot showing burst events observed on a single occasion during tower testing at CUTE on Z3.	56
Figure 3.13	Comparison of transition noise PSDs for Z1 and Z2 in Tower 3.	56
Figure 3.14	Example to demonstrate modeling of the various noise components for a single channel (PAS1) on Z1.	57
Figure 3.15	Comparison of noise PSDs between operating a single detector with two well- synced DCRCs and a single detector with an extra DCRC, synced by hand. . .	59

Figure 3.16	A comparison of the PSDs observed during CUTE tower testing when the SDU was on and off.	59
Figure 3.17	A cropped out circuit schematic of the VIB showing the arrangement of the HV circuit.	60
Figure 3.18	Comparison of noise PSD at various voltages.	61
Figure 3.19	Final energy spectrum for CDMSLite Run 3.	63
Figure 3.20	Schematic depiction of Compton scattering.	64
Figure 3.21	Compton spectra for a free electron and a bound electron.	65
Figure 3.22	The E-field contours near one end of a SuperCDMS SNOLAB iZIP detector. . .	68
Figure 3.23	Constructing $X - Y$ partitions for SuperCDMS SNOLAB iZIP detectors. . . .	70
Figure 3.24	Constructing $X - Y$ partitions for SuperCDMS SNOLAB HV detectors.	71
Figure 3.25	Constructing a Z-partition for SuperCDMS SNOLAB HV detectors.	73
Figure 3.26	The radial partition parameter.	74
Figure 3.27	Pulses from an example event from SuperCDMS Soudan for every channel in the detector post relative calibration.	75
Figure 3.28	Triangle alignment with scale factors for sector channels.	76
Figure 3.29	Triangle alignment with scale factors for the ring channels and one sector channel. . .	77
Figure 3.30	Triangle alignment with scale factors for the remaining ring channel.	77
Figure 3.31	Alignment of the two sides of the detector with scale factors.	78
Figure 3.32	The (preliminary) low energy spectrum for T3Z1 detector at 0 V.	79
Figure 3.33	The (preliminary) energy spectrum for T3Z1 detector at 50 V.	80
Figure 3.34	The high energy γ spectrum of T3Z1.	81
Figure 3.35	The K and L shell peak position as a function of applied voltage for T3Z1. . . .	84
Figure 3.36	The (preliminary) low energy calibration spectrum for T3Z2 detector at 0V. . .	85
Figure 3.37	An example radial cut to pick a fiducial volume for Z2.	86
Figure 3.38	The (preliminary) low energy calibration spectrum for T3Z2 detector at 90 V. .	87
Figure 3.39	An example trace with several minuscule pulses within.	88
Figure 3.40	The (preliminary) high energy spectrum for T3Z2 at 50 V.	88
Figure 3.41	A comparison of the K shell peak before and after neutralization.	90
Figure 4.1	Picture of a DCRC board.	93
Figure 4.2	A plot of an example hybrid pulse.	96
Figure 4.3	Plots showing the uniformly downsampled and on-pulse waveforms.	97

Figure 4.4	A comparison of noise PSD between the fully sampled, downsampled and on-pulse waveforms.	99
Figure 4.5	A cartoon comparing the true OF fit with a poor time delay fit.	103
Figure 4.6	A plot showing the effect of template shifting in optimal filtering.	104
Figure 4.7	A comparison of the fully sampled, downsampled and on-pulse templates used in the 2SOF algorithm.	105
Figure 4.8	An example of a hybrid pulse with a long fall time	106
Figure 4.9	A demonstration of the “wrap-around” effect in time shifted pulses.	107
Figure 4.10	A cartoon demonstration of generating sub-templates from an extended OP template.	108
Figure 4.11	A scatter plot between the estimated amplitudes from the DS pulse, DSamp, and the OP pulse, OPamp.	109
Figure 4.12	A flow chart outlining the steps in the 2-speed OF algorithm.	111
Figure 4.13	Example noise traces generated via Monte Carlo sampling.	113
Figure 4.14	A comparison of the averaged PSD between simulated noise samples and original noise samples.	113
Figure 4.15	A cartoon demonstration of simulating a signal from a noise trace and a pre-defined template.	114
Figure 4.16	A comparison of the energy spectra from various estimators for mono-energetic toy pulses.	115
Figure 4.17	The distribution of difference in time delays estimated from the OP pulse and the fully sampled trace.	117
Figure 4.18	An example pulse from the PD2 detector compared to the hybridized pulse. . .	118
Figure 4.19	A cartoon showing a typical SuperCDMS energy spectrum.	118
Figure 4.20	A plot showing the χ^2 vs amplitude plane in PD2.	120
Figure 4.21	A comparison of the two ^{55}Fe peaks obtained from various estimators.	120
Figure 4.22	The complete energy spectrum for PD2 detector.	121
Figure 4.23	A flowchart showing the workflow of the 2-speed Minimizer algorithm with a three parameter fit.	126
Figure 4.24	Distribution of fit parameters from applying the 2SM to toy pulses.	127
Figure 5.1	A schematic showing the construction of Principal Components in PCA.	142
Figure 5.2	An example scree plot for the first ten PCs.	142
Figure 5.3	A schematic of a binary decision stump.	144

Figure 5.4	A schematic of decision boundaries in a decision tree regressor.	145
Figure 5.5	The reconstruction pipeline for the $N \times M$ filter.	148
Figure 5.6	A plot of OPamps vs DSamps for one channel from a 2×1 fit for simulated toy data.	150
Figure 5.7	A flowchart showing the workflow of the 2-Speed $N \times M$ algorithm.	152
Figure 5.8	A plot of the correlations vs. frequency for 2 channels.	153
Figure 6.1	A schematic showing the workflow of the SuperCDMS detector simulations framework.	157
Figure 6.2	A selection of 100 simulated mono-energetic DMC events for a SNOLAB HV detector at 100 V for all 12 channels.	158
Figure 6.3	First ten Principal Components for simulated pulses.	160
Figure 6.4	Scree plot showing the amount of information/variance in the first ten PCs plotted against the component number.	161
Figure 6.5	The distribution of fitted template amplitudes as a function of the true event position in the detector.	162
Figure 6.6	A plot showing the result of tuning the GBDT hyperparameters.	164
Figure 6.7	A contour plot of the BDTGain in the horizontal plane of the detector.	165
Figure 6.8	A comparison of the reconstructed mono-energetic spectra of simulated events for a SNOLAB HV detector operated at 100 V.	166
Figure 6.9	A comparison of the GBDT predicted position and the true position.	167
Figure 6.10	A representation of the CDMSLite detector biasing scheme.	169
Figure 6.11	90% profile likelihood ratio confidence limit for nuclear-recoiling dark matter performed with CDMSLite Run 3 data.	170
Figure 6.12	The final selection of pulses used in template generation for the four channels that were read out in the CDMSLite detector.	171
Figure 6.13	The first four PCs derived from CDMSLite data normalized to a maximum height of 1.	172
Figure 6.14	The full sequence of corrections applied to the $N \times M$ energy estimator to account for environmental and event position effects.	175
Figure 6.15	A comparison of the resolution of calibration peaks from a CDMSLite Run3 data.	177
Figure 6.16	A comparison of resolution models from various energy estimators for CDMSLite Run3.	179

Figure 6.17	Example spectra for Run 3a and Run3b generated from our background and signal models for CDMSLite Run3, smeared by the resolution and scaled by cut efficiency and livetime.	182
Figure 6.18	An example limit derived from a randomly chosen Monte Carlo data set simulated from the summed background and signal model.	185
Figure 6.19	The χ^2 vs OF amplitude curve for detector T3Z1 from the SuperCDMS SNOLAB tower tested at CUTE.	187
Figure 6.20	The final selection of pulses used in template generation for T3Z1, a Super-CDMS SNOLAB HV detector tested at CUTE.	188
Figure 6.21	This figure shows (a) the first 10 PCs of the data set along with (b) the scree plot and (c) the final normalized and smoothened templates.	189
Figure 6.22	A comparison of the resolution of calibration peaks for T3Z1 detector from CUTE tower testing.	190

List of Abbreviations

2DFGRS 2-degree Field Galaxy RedShift Survey

2SM 2-Speed Minimizer

2SN $\times$ M 2-Speed N $\times$ M

2SOF 2-Speed OF

ADCS Analog-to-Digital Converters

ALPS Axion Like Particles

BAO Baryon Acoustic Oscillations

BBN Big Bang Nucleosynthesis

BSM Beyond the Standard Model

CFA2 Center for Astronomy Redshift Survey

CL Confidence Limit

CMB Cosmic Microwave Background

CRESST Cryogenic Rare Event Search with Superconducting Thermometers

CSD Cross Spectral Density

CUTE Cryogenic Underground TEst facility

DACS Digital-to-Analog Converters

DAQ Data AcQuisition

DCRCS Detector Control and Readout Cards

DELIGHT Direct search Experiment for LIGHT dark matter

DFT Discrete Fourier Transforms

DM dark matter

DS Downsampled

ER electron recoil

FERMI-LAT Fermi Large Area Telescope

GBDTS Gradient Boosted Decision Trees

HDPE High Density PolyEthelene

HEMTS High Electron Mobility Transistors

HOF Hybrid Optimal Filter

HV High Voltage - A type of SuperCDMS SNOLAB detector.

IZIP interleaved Z-dependent Ionization and Phonon

KPP Key Performance Parameter

Λ **CDM** Λ cold dark matter

LHC Large Hadron Collider

LNGS Laboratori Nazionali del Gran Sasso

MACHOS Massive Compact Halo Objects

MOND MOdified Newtonian Dynamics

MSE Mean Squared Error

NR nuclear recoil

NSOF Non-Stationary Optimal Filter

NTL Neganov-Luke-Trofimov

ODB Online DataBase

OF Optimal Filter

OI Optimum Interval

OP On-Pulse

PCA Principal Component Analysis

PCS Principal Components

PLR Profile Likelihood Ratio

POE Power-Over-Ethernet

PSD Power Spectral Density

PT Pulse Tube

QCD Quantum Chromodynamics

QETS Quasiparticle-trap-assisted Electrothermal-feedback Transition-edge-sensors

RQS Reduced Quantities

RRQS Reduced Reduced Quantities

SDSS Sloan Digital Sky Survey

SDU Signal Distribution Unit

SLAC Stanford Linear Accelerator Center

SM Standard Model

SNOLAB Sudbury Neutrino Observatory LAB

SUF Stanford Underground Facility

SUPERCDCMS Super Cryogenic Dark Matter Search

SUSY SuperSymmetry

TES Transition Edge Sensors

TEVES Tensor-Vector-Scalar gravity

VIB Vacuum Interface Board

WIMPS Weakly Interacting Massive Particles

WMAP Wilkinson Microwave Anisotropy Probe

Glossary

2SM 2-Speed Minimizer - A novel reconstruction algorithm developed for SuperCDMS SNOLAB to model non-uniformly sampled data with the pulse shape modelled using an analytic parametrization with free parameters.

2SN $\times$ M 2-Speed N $\times$ M - A novel reconstruction algorithm developed for SuperCDMS SNOLAB to model non-uniformly sampled data while also taking into account correlated noise and pulse shape variation in the detector.

2SOF 2-Speed OF - A novel reconstruction algorithm developed for SuperCDMS SNOLAB to model non-uniformly sampled data assuming a common pulse shape for every event.

CL Confidence Limit - An exclusion curve plotted to mark the regions of the parameter space of the model inconsistent with the data.

DCRCS Detector Control and Readout Cards - The circuit boards that interface with the readout cables and perform key data acquisition functions such as issuing triggers, digitizing signals and controlling certain hardware components.

HOF Hybrid Optimal Filter - A computationally intensive reconstruction technique to model non-uniformly sampled data for SuperCDMS SNOLAB.

HV High Voltage - A type of SuperCDMS SNOLAB detector.

IZIP interleaved Z-dependent Ionization and Phonon - A type of SuperCDMS SNOLAB detector.

NTL Neganov-Luke-Trofimov - An effect that results in the conversion of charge carrier energy to phonon energy under the application of an external electric field, resulting in an amplified phonon signal.

OF Optimal Filter - A matched filter used in SuperCDMS to model pulses.

OI Optimum Interval - A conservative limit-setting technique used when there is an unknown experimental background.

OP On-Pulse - The region of a non-uniformly sampled waveform that is fully sampled and contains the “peaky” part of the pulse.

PCA Principal Component Analysis - A technique used to mathematically extract information from a given data set by rotating the parameter space of the data in such a way that the most significant information is contained in a few significant directions in the rotated parameter space.

PCS Principal Components - Eigenvectors of the covariance matrix that represents variance in the input data.

PLR Profile Likelihood Ratio - A statistical quantity used to signify how well a particular model fits the given data.

PT Pulse Tube - A cooling mechanism used as a pre-cooler in dilution fridges.

QETS Quasiparticle-trap-assisted Electrothermal-feedback Transition-edge-sensors - SuperCDMS phonon detection technology that contains an aluminium trapping region to collect phonons and a tungsten TES for readout.

RQS Reduced Quantities - Physically interpretable information extracted from SuperCDMS raw data as the output of data processing.

RRQS Reduced Reduced Quantities - Quantities generated from Reduced Quantities that are often calibrated in a meaningful way.

SDU Signal Distribution Unit - An electronic unit used to synchronize the DCRCs and facilitate readout of auxillary devices such as accelerometers.

TES Transition Edge Sensors - SuperCDMS phonon readout technology, but having applications in many other fields.

VIB Vacuum Interface Board - The circuit board that connects the readout components on the vacuum side with the electronics mounted on the exterior of the vacuum chamber.

Acknowledgments

Understanding and learning how to operate a big experiment is no easy feat. I would not have accomplished what I have without the whole-hearted support of the SuperCDMS collaboration. I would like to extend my heartfelt gratitude to all collaborators for providing me with guidance, expert advice, tools and machinery when I needed it. Although I will try to extend thanks to individuals here, I may not be able to capture everyone who has helped me along the way – so I apologize in advance if I missed your name!

First and foremost, my supervisor, Scott Oser. Scott has constantly been my champion and an exceptional mentor. Five years ago, I would not have imagined myself accomplishing the scale of work that I did in research and leadership. I have grown a lot from the bumbling undergrad I was when I met Scott, to the (still kind of bumbling) particle physicist I am today. Scott made sure to understand exactly where I excelled and where I could potentially excel and directed his efforts to ensure that I built a very broad skillset and knowledge across all aspects of the experiment. He also let me pursue my interests independently within and outside the collaboration and was always around to bounce ideas off him. He believed in me when I did not and by being my guiding light, I have gained a lifelong mentor. I will truly miss working with you!

I would like to extend a personal note of gratitude, in no particular order, to my colleagues across the world with whom I have worked and exchanged ideas with: the current and former members of the UBC + TRIUMF/Queen’s group (Wolfgang Rau, Yan Liu, Emanuele Michielin, Musaab Al-Bakry, Eleanor Fascione, Richard Germond, Ryan Underwood, Ashley Li, Lucas Bezerra), the UofT group (Ziqing Hong, Pekka Sinervo, Miriam Diamond, Maddy Zurowski, Ariel Zuniga Reyes, Tyler Reynolds, Ata Sattari, Stefan Zatschler and many others) and my CUTE Analysis and data processing team (Maddy, Ariel, Joseph Mammo, Moaaz Elwan, Ruchi Soni, Ry Pressmon-Cyna, Shubham Pandey, Weigeng Peng and many others) and the DAQ team (Belina von Krosigk, Sukee Dharani and others). A special note of gratitude to Yan, Wolfgang and Ziqing who shared the burden of sleepless shifts every day during the tower testing run and for

being the best mentors and to the countless CUTE shifters who made the tower testing campaign a tremendous success! I also extend my thanks to Bruno Serfass and Tina Cartaro for consistently supporting the software infrastructure and processing of the data used throughout this thesis. The list goes on, but I will pause in the interest of time and pagecount.

Finally, coming to the most important people I'm grateful to have in my life. My partner, Connor Adair, who got me through the past five years and continues to support me in every possible way, every day. Thank you for picking up my slack when I sank deep into work! My family has continuously supported me in everything I have ever done, from school to university and then moving halfway across the world for the first time in my life, away from everything that both they and I have ever known. Thank you for being brave enough to send me out into the world and for instilling in me the strength and courage to persevere.

As a final word, a shout-out to the Universe – of all the places I could've been, I ended up in Vancouver! Let's believe the many-world interpretation of quantum mechanics is real for a minute. Then I'm sorry for all the other versions of me who did not end up at UBC!

This thesis is dedicated to my grandparents, Geetha and Chandrashekharan, who raised me to be the woman I am today...

Chapter 1

Dark matter and experimental searches

Nothing cannot exist forever. — Stephen Hawking

About 13.8 billion years ago, the Universe began with a “big bang”, expanding over the years into a network of filaments, voids and nodes. We know today that the evolution of the Universe into its present structure was facilitated by an array of complex processes. Inflation, which began perhaps about 10^{-36} s after the Big Bang, is believed to have pushed the Universe into a state of rapid expansion, establishing just the right initial conditions of density perturbations [2, 3] which acted as seeds for structure formation. Inflation left behind a mostly homogeneous universe, which was followed immediately by a period of adiabatic cooling that continues to this day. This slow expansion of the Universe is described by the Friedmann equation, which solves Einstein’s field equations under the assumptions of a nearly isotropic and homogeneous universe [4]:

$$H^2(t) = \frac{8\pi G}{3} \left[\rho(t) + \frac{\rho_{\text{cr}} - \rho_0}{a^2(t)} \right]. \quad (1.1)$$

Here, $H(t)$ is the Hubble rate given by $H(t) = \frac{\dot{a}}{a}$, where a is the scale factor that defines the expansion of the Universe, $\dot{a}$ is the time derivative of a and G is Newton’s constant. The quantity $\rho(t)$ is the time-dependent energy density of the Universe, which is made up of several components – radiation, matter, spatial curvature and an energy density with negative pressure (dark energy) whose nature is unknown, while ρ_{cr} denotes a critical density whose value happens to be just right so that our Universe remains spatially flat and ρ_0 is the energy density of the Universe today.

Eq. (1.1) can otherwise be written in terms of fractional densities of each component $\Omega_i = \frac{\rho_i}{\rho_{\text{cr}}}$, where ρ_i is the energy density of component i , as follows:

$$\frac{H^2(t)}{H_0^2} = \Omega_{r,0} \left(\frac{a_0}{a(t)} \right)^4 + \Omega_{m,0} \left(\frac{a_0}{a(t)} \right)^3 + \Omega_{k,0} \left(\frac{a_0}{a(t)} \right)^2 + \Omega_{\Lambda,0}. \quad (1.2)$$

Here, a subscript 0 is used to denote the present-day values of that parameter. The above equation shows the evolution of densities of four components with respect to the Hubble rate: radiation ($\Omega_{r,0}$); matter ($\Omega_{m,0}$); spatial curvature ($\Omega_{k,0}$) and the cosmological constant ($\Omega_{\Lambda,0}$)¹. Over the course of evolution of our Universe, various components have come to dominate its total energy density demarcating various eras in history. There is strong evidence to indicate that the Universe is very close to flat. In a flat Universe, the curvature parameter $k = 0$ and therefore $\Omega_k = 0$. The spatial curvature density parameter has been measured precisely by Planck [5], the Wilkinson Microwave Anisotropy Probe (WMAP) [6] and others to be consistent with a flat model. The very early Universe was a period too hot for particles to combine and matter to form. This period was, therefore, dominated by relativistic particles ($E \gg m$) which were effectively radiation. This was the radiation-dominated era (region with $a < a_{eq}$ in Fig. 1.1). Particle number density is inversely proportional to volume and hence scales as a^{-3} , while radiation energy is inversely proportional to wavelength and so scales as a^{-1} . This means that the radiation energy density scales as a^{-4} , as is seen in Eq. (1.2). With time, the radiation density fell and attained equality with matter. This brief period of matter-radiation equality is denoted by the equilibrium value of the scale factor, a_{eq} , in Fig. 1.1. Beyond this epoch, matter dominated the Universe until recently. Almost all matter is non-relativistic and therefore has an energy density that scales as a^{-3} . As the matter density decreased with expansion, the cosmological constant came to be the dominant contribution. The cosmological constant Λ , otherwise known as dark energy, is responsible for the accelerated expansion of the Universe. The nature of this dark energy remains a big mystery in modern physics.

Today of all the matter in the world, visible matter constitutes only about 4.6% [7]. The predominant contribution to the matter density comes from an unknown component whose effect is known to us only through astrophysical observations and gravitational effects. This strange “invisible” matter is called dark matter. The origin and nature of dark matter is one of the biggest questions in cosmology and particle physics. In this chapter, I will discuss some of the observational evidence and motivation for dark matter, a few popular dark matter candidates, and methods for detecting dark matter that are currently in use.

¹In Eq (1.2), dark energy takes the simplest possible form of a cosmological constant.

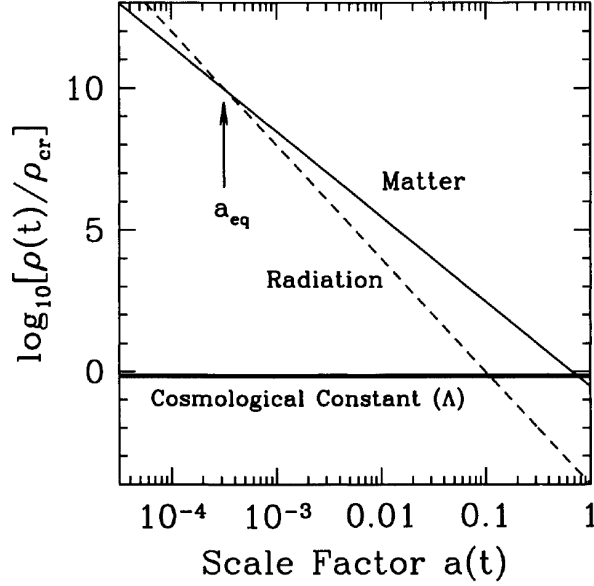


Figure 1.1: Energy density $\rho(t)$ as a function of scale factor a for each component in a flat universe ($\Omega_k = 0$). For component i , $\frac{\rho_i(t)}{\rho_{cr}} = \frac{\rho_i(t_0)}{\rho_{cr}} \frac{\rho_i(t)}{\rho_i(t_0)} = \Omega_{i,0} \frac{\rho_i(t)}{\rho_i(t_0)}$, where subscript 0 denotes present-day values. The quantity $\rho_i(t)$ scales as a function of scale factor as a^{-4} , a^{-3} and a^0 for radiation, matter and Λ , respectively, and a_{eq} refers to the scale factor at the time of matter-radiation equality. Figure taken from Ref. [4].

1.1 Evidence for dark matter

1.1.1 Astrophysical evidence

1.1.1.1 Rotational curves of galaxies

One of the earliest pieces of observational evidence for dark matter comes from galactic rotation curves. Spiral galaxies contain a spiral structure with most of their visible content (stars, gas, etc.) concentrated in their disks and central bulge. The visible mass decreases exponentially as a function of radius. Let us call the total mass within a distance R of the galactic center, $M(R)$. For a point mass at a distance R from the center of the galaxy, the gravitational acceleration can be equated to the centripetal acceleration to obtain a distribution of velocity, v :

$$v(R) = \sqrt{\frac{GM(R)}{R}}. \quad (1.3)$$

From Eq. (1.3), we expect the velocity distribution to decrease with radius, since $M(R)$ plateaus at large radii. Surprisingly, observational evidence suggests that the velocity remains constant well towards the edge of the galaxy. This effect is shown in Fig. 1.2. Rogstad [8], Rubin [9] and many others performed extensive studies of a host of galaxies, all reporting similar effects.

This observation can be explained if there is an otherwise invisible component of matter, dark matter, contributing to the mass distribution of the galaxy. From simple dimensional analysis, if $d(R)$ denotes density and V denotes volume, $M(R) \propto d(R)V(R) = d(R)R^3$. This means that the dark matter component would need to have a density distribution proportional to R^{-2} to maintain a roughly constant velocity at large radii, in contrast to the exponential decrease of visible mass. A more thorough derivation postulates a spherically distributed dark matter halo that surrounds every galaxy. In the next subsection, we will note how the presence of dark matter halos contributes to gravitational lensing effects and addresses X-ray anomalies in galaxy clusters, all of which constitute compelling evidence for the existence of dark matter.

1.1.1.2 Anomalies in measurements from galaxy clusters

Galaxy clusters often contain 100s to 1000s of galaxies, separated by a medium of hot gas. This gas in the intracluster medium acts as a rich source of X-rays from which the properties of the cluster can be probed. X-ray measurements can be used to obtain the mass-to-light ratio, Υ , of the cluster. The advantage of studying clusters is that their masses can be verified independently with different methods. The three common methods of estimating cluster mass are: (1) using the virial theorem; (2) using the principle of hydrostatic equilibrium and (3) via gravitational lensing [11]. In method 1, the virial mass is estimated from the velocity dispersion of the gas, which in turn is obtained from X-ray measurements. In method 2, the mass is derived by equating the hydrostatic pressure of the cluster to the opposing gravitational pull. Method 3 uses gravitational lensing, i.e., the distortion or bending of light by a massive object due to its gravity. The distortion spreads the lensed light in a circular region, often called the Einstein ring [12]. The measured angular radius of this Einstein circle is proportional to the mass of the object. All three methods consistently have revealed large mass-to-light ratios and shown that the amount of visible matter in galaxy clusters is, in fact, small relative to the total mass. So, where is the remaining mass coming from? The remaining mass is believed to come from dark matter in these clusters. Zwicky was one of the first to observe this effect in nebulae in the Coma cluster [13].

A strikingly apparent evidence for dark matter comes from the Bullet cluster. This structure consists of two individual clusters, passing through each other in opposite directions. The electro-

1.1. EVIDENCE FOR DARK MATTER

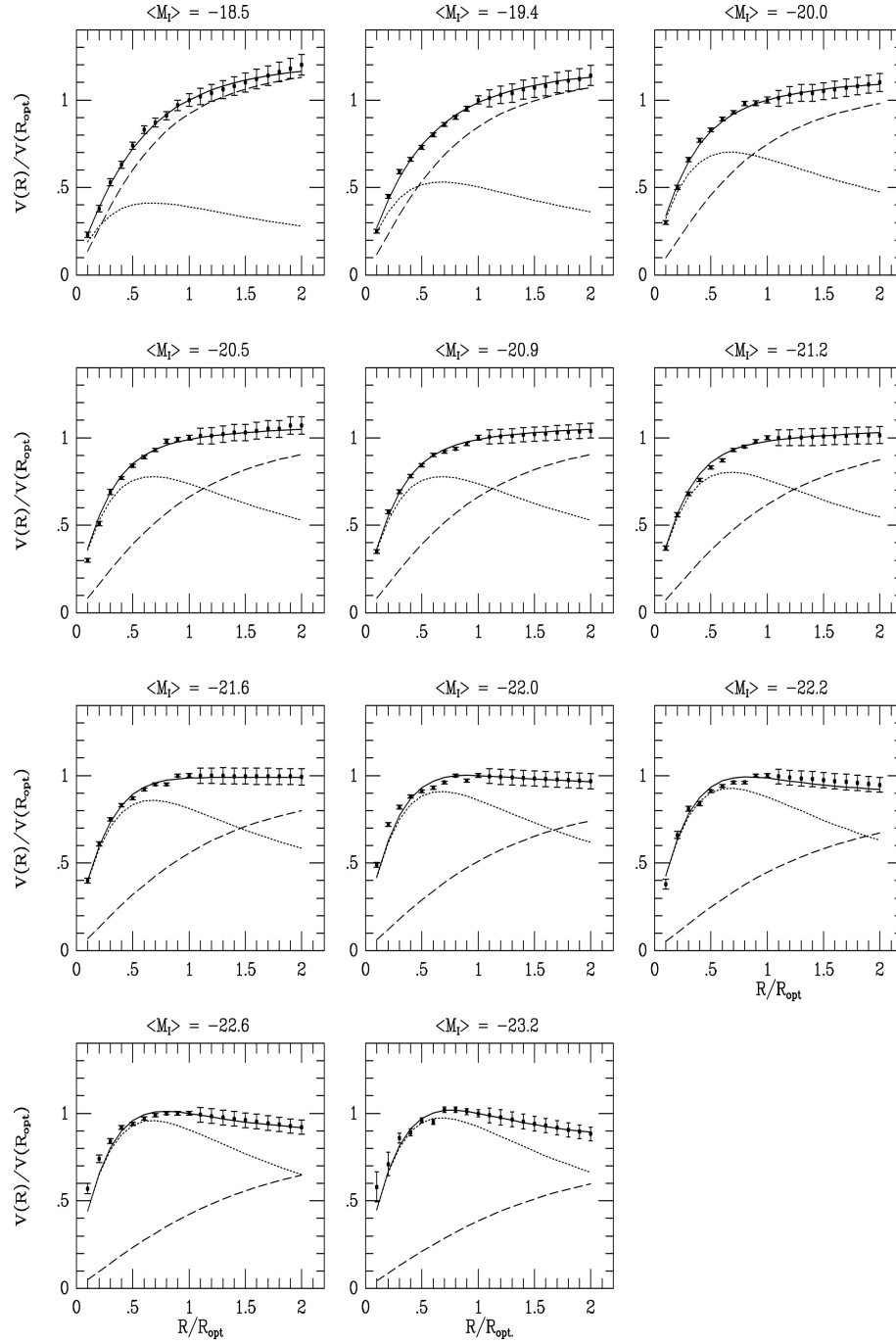


Figure 1.2: Universal rotation curve of spiral galaxies. Solid lines show the best fits to the data, dashed lines show the halo-induced velocity, dotted lines show the disk-induced velocity. This plot points to the existence of a dark matter halo surrounding every galaxy. Figure taken from Ref. [10]. Reprinted with permission.

magnetic interactions of matter in the cluster produce a drag force that gives rise to its characteristic bullet shape. A comparison of the gravitational lensing and luminosity distributions of this cluster indicates that the luminous components (gas and stellar objects) are located in distinct central regions of the cluster, while most of the mass is located far outside this luminous region. This is seen clearly in the combined data of this cluster shown in Fig. 1.3.

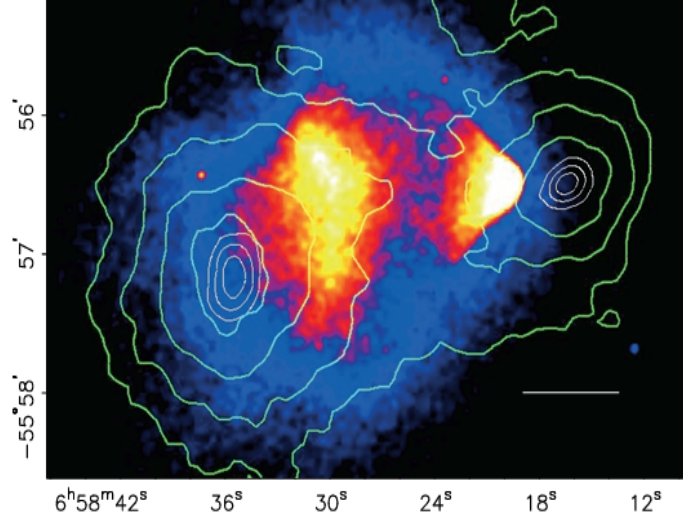


Figure 1.3: Image of the Bullet cluster obtained from the Chandra X-ray observatory and optical data. The green contour lines show the reconstructed matter distribution from lensing effects. The heat map shows the distribution of luminous matter in the cluster. Figure taken from Ref. [14]. Reprinted with permission.

Theories of modified gravity have been proposed to address some of the previously stated observations supporting dark matter, such as M^Odified Newtonian Dynamics (MOND) or Tensor-Vector-Scalar gravity (TeVeS). However, none of them offer a simultaneous explanation of all cosmological observations and often require unrealistic parameters or the addition of dark matter [15].

1.1.2 Cosmological evidence

In addition to the astrophysical evidence stated in the previous subsections, dark matter is well motivated by cosmology. Modern cosmology is described by the Λ Cold Dark Matter (Λ CDM)

model, which consistently explains the formation of large scale structures, accelerated expansion of the Universe and other phenomena. The model includes a dark matter component without which a sensible interpretation of data is not currently possible. In this subsection, I will discuss some of the smoking gun evidence from cosmology that supports dark matter.

1.1.2.1 Anisotropies in the Cosmic Microwave Background

The origin of the Cosmic Microwave Background (CMB) can be traced back to about 400,000 years after the Big Bang. Before this epoch, photons existed in equilibrium with electrons, protons, and other particles in a hot dense plasma, with everything scattering off of each other. As the Universe expanded, it cooled enough to allow protons, electrons and heavier nuclei to combine and form atoms. As particles recombined (termed the recombination era) and the universe further expanded, photons could no longer maintain equilibrium via scattering. They decouple and are red-shifted with the expanding Universe. Today this decoupled residual radiation from the past is what we see as the CMB. The CMB shows a blackbody spectrum with an average temperature of about 2.725 K and is isotropically distributed all around us. Being decoupled from the rest of the components, the CMB offers a window into the past to study how our Universe came to be the way we see it today.

Although the CMB is largely isotropic in temperature, there exist small fluctuations that are believed to have their origin in tiny perturbations in the primordial plasma during inflation. The power spectrum of the CMB anisotropies shows distinct peaks (called acoustic peaks, see Fig. 1.4) associated with evolution of the photon-matter interactions in the primordial plasma. Current measurements indicate that the CMB contains anisotropies at the 10^{-5} level. The relative amplitudes and locations of the acoustic peaks in the CMB anisotropy spectrum are determined jointly by the structure of the universe, the matter density (both baryonic and non-baryonic) and the radiation density at that epoch. The location of the first peak is dependent on the curvature of the Universe, while the heights of the odd peaks are enhanced by the baryon density relative to the even peaks. Thus, the ratio of the heights of the first two peaks can be used to estimate the baryon density of the Universe. The heights of the higher order peaks, and to some extent the first two peaks, are determined by the density of non-relativistic matter and radiation in the Universe, due to the effects of transition from a radiation to matter dominated era being felt more at smaller (earlier) wavelengths or higher acoustic peaks. Therefore, the height of the third peak relative to the second peak is sensitive to the presence of dark matter. The acoustic peaks in the CMB anisotropy spectra obtained from observational surveys, such as Planck [16], can be modeled precisely with the Λ CDM model

of the universe. The remarkable precision with which the Λ CDM model describes CMB structure strongly supports the existence of a non-relativistic dark matter component.

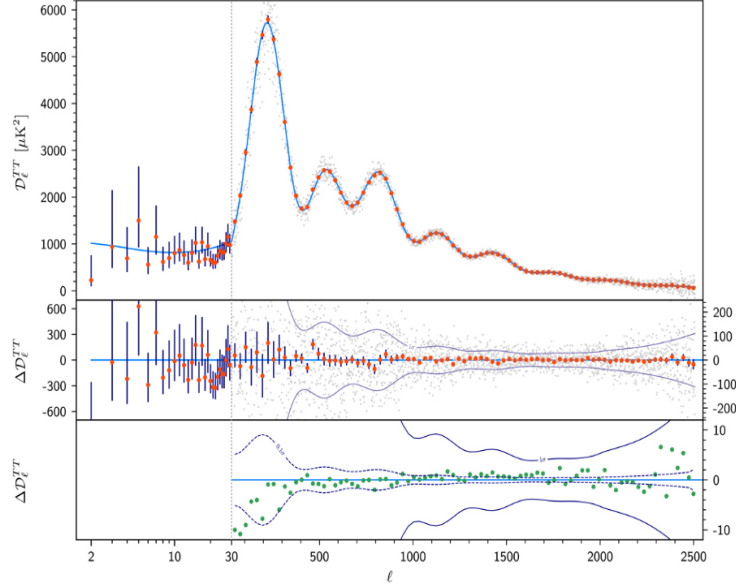


Figure 1.4: The cosmic microwave background anisotropy power spectrum from Planck 2018 data. In the top panel, the blue curve shows a fit to the data with the Λ CDM model. The middle panel shows residuals from the difference between the model and the fit, with the $\pm 1\sigma$ error curve in purple. The bottom panel shows the difference between the 2015 and 2018 Planck results. The relative heights of the first, second, and third peaks determine the relative amounts of dark matter vs. baryons. Figure taken from Ref. [16]. Reprinted with permission.

1.1.2.2 Large scale structure formation in the universe

The early Universe was very smooth and homogeneous. So it is puzzling that a homogeneous Universe would give rise to clumpy structures such as what we see today. In the expanding Universe, gravity pulls more mass into a region of larger than average density while the expansion tries to push away the mass. This results in the formation of structure in the Universe. Dark matter is only known to interact gravitationally with ordinary matter. This means that in the early Universe, dark matter would have been the first to collapse in on itself, creating over-densities that pulled more matter in. These over-densities evolve in time to produce strong gravitational wells and large structures billions of years later. If it were only baryons contributing to over-densities in the early

Universe, structure formation would have taken longer, since the baryon density at the time was not high enough to produce the level of over-densities required to give rise to the structures we see today [11].

To precisely model these effects, computationally intensive N-body simulations are performed, which have been successful in explaining observed data. Their key to success lies in the inclusion of dark matter in the simulations. The largest of such N-body simulations, the Millennium simulation, performed in the mid-2000s, has been able to successfully reproduce results from surveys such as the Center for Astronomy Redshift Survey (CfA2) [17], the 2-degree Field Galaxy RedShift Survey (2dFGRS) [18] and the Sloan Digital Sky Survey (SDSS) [19]. A comparison of the simulated structure from the Millennium simulations with the observed structure is shown in Fig. 1.5.

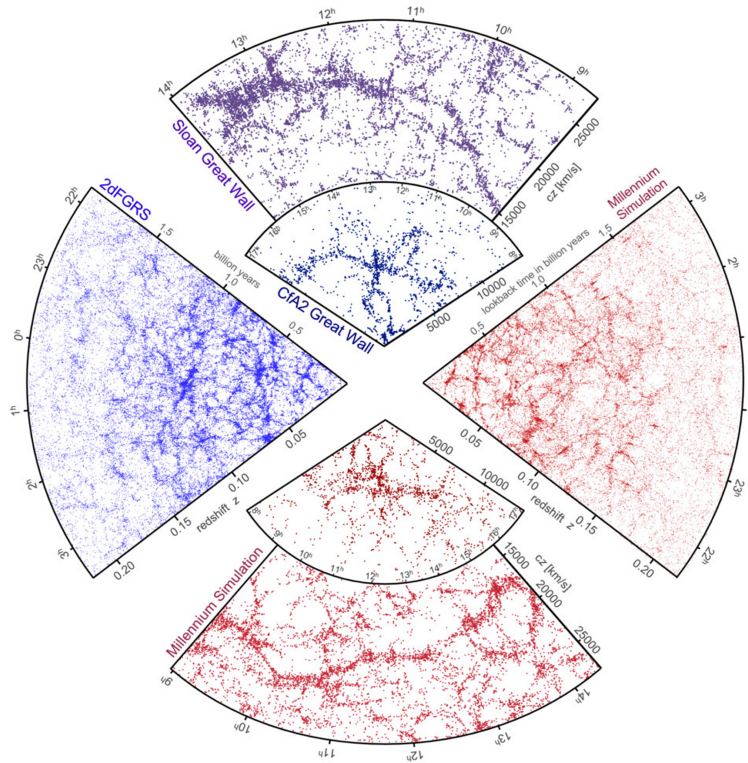


Figure 1.5: Comparison of results from redshift surveys with simulations from the Millennium project. The images in red in the bottom and right sectors show the simulated results. The images on the top and left in blue show data from redshift surveys. The left sector is from 2dFGRS, the top outer set is from SDSS and the inset on the top sector is from CfA2. Figure taken from Ref. [20]. Reprinted with permission.

1.1.2.3 Measurements of matter density

The baryon composition in the early Universe was dominated by protons and neutrons. An estimate of the baryon density at the time of recombination can be obtained from the second acoustic peak of the CMB spectrum or using theoretical estimates from Big Bang Nucleosynthesis (BBN). BBN refers to the array of processes that produced the first elements when the Universe was less than a few minutes old. It starts with a chain reaction that combines a proton and a neutron to produce ${}^2\text{H}$, which then produces ${}^3\text{He}$, ${}^3\text{H}$ and ${}^4\text{He}$. The absence of stable nuclei with $A = 5$ or 8 reduces the likelihood of producing even heavier elements. So the predominant elements in the early Universe were hydrogen and helium. Most BBN chain reactions terminate in ${}^4\text{He}$ and most of the He in the Universe today was produced during this early phase. In general, the abundances of these elements are dependent on the baryon-to-photon ratio in the early Universe, from which an estimate of the baryon density can be derived. At the same time, independent measurements of the total matter density can be obtained from studying the CMB power spectrum, supernovae, galaxies or galaxy clusters at large scales. Measurements show a value of $\Omega_{\text{m}}h^2 \sim 0.14$ and $\Omega_{\text{b}}h^2 \sim 0.022$ [5], where the unitless Hubble parameter, $h = H_0/(100 \text{ km/s/Mpc})$ and $H_0 \simeq 70 \text{ km/s/Mpc}$. These measurements and past ones not only show that baryons do not constitute the entirety of the matter density, but also that dark matter, whatever it may be, must be non-baryonic.

Figure 1.6 shows an example measurement of Ω_{m} and Ω_{Λ} from Type IA supernovae using data from the Hubble Space Telescope, overlaid on estimates from the CMB and Baryon Acoustic Oscillations (BAO) [21]. The intersection of all three estimates coincides nicely with the prediction of a flat Universe with $\Omega_{\text{m}} \sim 0.3$.

1.2 Dark matter in particle physics

Although the evidence for dark matter is entirely astrophysical and cosmological, it is also well motivated in the context of particle physics. It is generally accepted in the particle physics community that the Standard Model (SM) of particle physics is incomplete. The existence of neutrino mass, the inability to provide a consistent description of gravity, and the matter-antimatter asymmetry are some examples that call for the need for some “Beyond the Standard Model (BSM)” explanation. Dark matter fits naturally into the picture of BSM physics.

So far we have reviewed the empirical motivation behind the idea of dark matter. Based on this evidence, we can postulate certain properties for a potential dark matter candidate. The candidate must have a mass and number density sufficient to account for the dark matter density we measure today. Most importantly, dark matter must be weakly interacting to maintain this number density

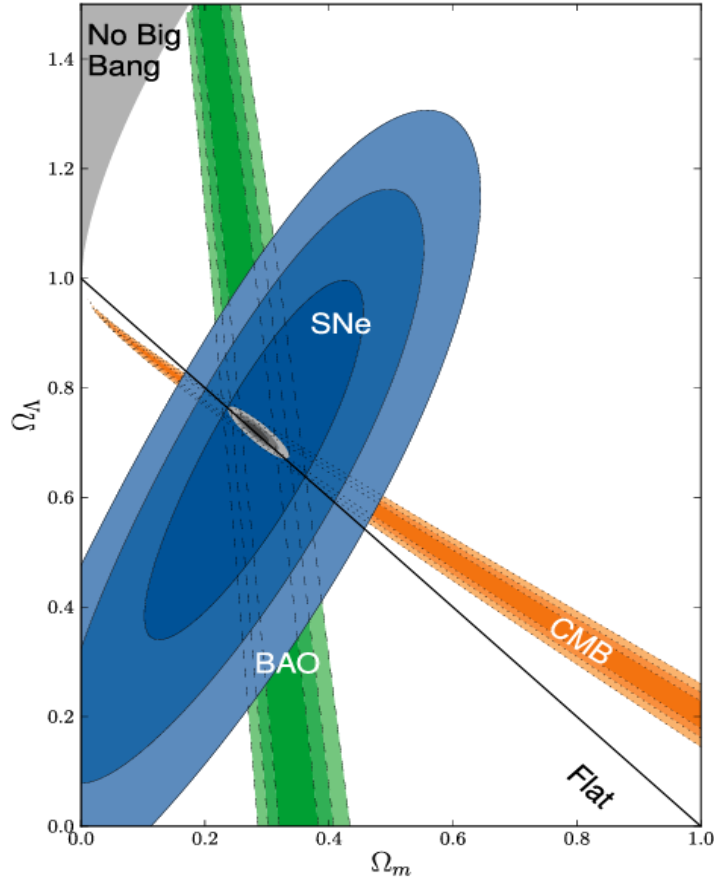


Figure 1.6: Dark energy density, Ω_Λ vs matter density, Ω_m for the Λ CDM model from the Hubble Space Telescope cluster Supernova survey. The black line shows the prediction for a flat universe. The blue, orange and green regions show the estimate from supernovae, CMB and BAO. The contours indicate 68.3%, 95.4% and 99.7% confidence regions. Figure taken from Ref. [21].

to present day. It must interact with SM particles through gravitational effects, predominantly, to account for all the gravitational evidence we have accumulated and potentially also couple to SM particles via some electroweak-scale mechanism [22]. Further, as discussed in the previous section, dark matter must be non-baryonic, since the total estimated baryon density is much lower than the estimated total matter density. “Cold” dark matter means that it consists of non-relativistic particles, while hot dark matter refers to ultra-relativistic particles. Hot dark matter would have had too high velocity dispersion to collapse into gravitational wells in the later Universe. Thus,

it is disfavoured by cosmology, since structure formation with cold dark matter has been able to successfully reproduce current large-scale structures [20, 23]. Therefore, dark matter must be “cold”. All these conditions can be used to define a broad category of dark matter candidates called Weakly Interacting Massive Particles (WIMPs).

Many particle physics experiments have been on the hunt for dark matter since the 20th century with primary focus on WIMPs. Over time, they have expanded their sensitivity to more exotic candidates. In the next two sections, I will provide a brief overview of some popular dark matter candidates and common methods of searching for them.

1.2.1 Dark matter candidates

Dark matter first attained popularity amongst particle physicists in the context of SuperSYmmetry (SUSY). SUSY addresses many shortcomings of the SM, like, e.g., the hierarchy problem and shows the path towards a Grand Unified Theory [22, 24]. SUSY provides an elegant mathematical framework that naturally predicts the existence of new particles at the electroweak scale. These new particles are SUSY partners to SM bosons and fermions, i.e., a SUSY transformation under the SUSY operator Q converts a SM boson (fermion) into a SUSY fermion (boson) and vice versa:

$$\begin{aligned} Q|\text{Boson}\rangle &= |\text{Fermion}\rangle, \\ Q^\dagger|\text{Fermion}\rangle &= |\text{Boson}\rangle. \end{aligned} \tag{1.4}$$

The existence of these particles could account for dark matter. Such particles would exist at the electroweak scale and fall under the umbrella of WIMPs. The motivation for WIMPs originally come from Big Bang cosmology. The dark matter density in the universe has been estimated to be about $\Omega_{\text{dm}}h^2 = 0.12$ [5]. In the SUSY model, in the early Universe, dark matter existed in equilibrium with the cosmic plasma, its number density unchanged and maintained by constant production and annihilation controlled by its cross section. As the Universe cooled, mechanisms that produced dark matter became energetically forbidden and so dark matter was no longer in equilibrium with the rest of the Universe. The remaining dark matter particles began annihilating with each other while the Universe’s expansion diluted their number density. Eventually the annihilation could no longer keep up with the expansion causing the dark matter density to “freeze out”. The dark matter annihilation cross section, σ , is inversely related to the dark matter abundance [25]:

$$\Omega_{\text{dm}}h^2 \propto \frac{1}{\langle\sigma v\rangle}, \tag{1.5}$$

where v is the dark matter velocity. This relationship is portrayed in Fig. 1.7 in terms of the comoving WIMP number density.

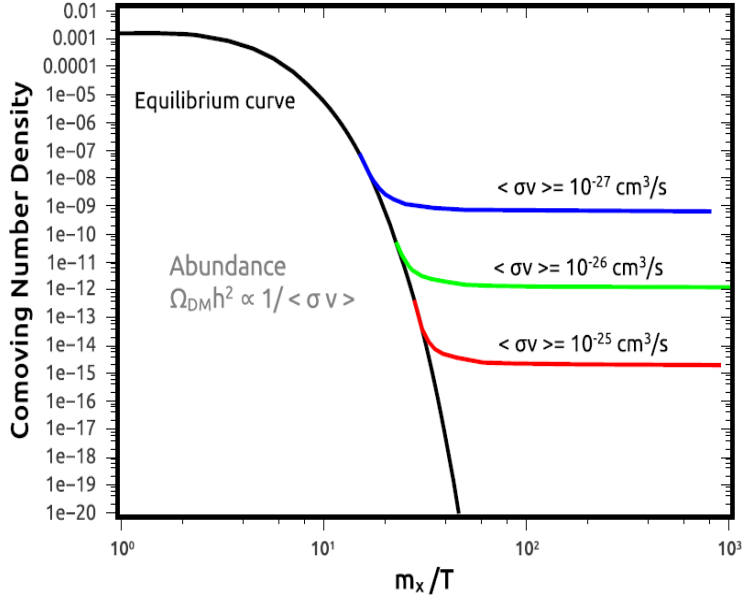


Figure 1.7: Evolution of dark matter comoving number density as a function of the ratio of dark matter mass to temperature. The value of $\langle \sigma v \rangle$, and by extension the expected WIMP mass necessary to produce the current abundance of dark matter, are decided by cosmology. The colored lines indicate the freeze-out abundance for different cross-sections. Figure taken from Ref. [25]. Reprinted with permission.

One can estimate this cross section and show that this corresponds to a particle of $\langle \sigma v \rangle \sim 10^{-26} \text{ cm}^3/\text{s}$ and mass $O(100 \text{ GeV}/c^2)$ – a perfect alignment with the electroweak scale where SUSY predicts new physics. This is called the “WIMP miracle” and it inspired the first dark matter experiments.

Decades of research has given rise to several candidates that could potentially be dark matter. This includes searches for SUSY partners such as neutralinos or gravitinos [26] or for sterile neutrinos [27] which are a product of some minimal extensions to the SM. The predicted WIMP mass of $O(100 \text{ GeV}/c^2)$ with a cross-section at the electroweak scale has already been ruled out by early experiments [28]. Nevertheless, search for heavier WIMPs continues at GeV and TeV scales, with colliders and astrophysical sources. The term WIMPs is typically used for dark matter with a thermal production mechanism, such as freeze-out, as described earlier. However, non-thermal production mechanisms can result in dark matter with masses of the order of $10^{12} - 10^{16} \text{ GeV}/c^2$

[29]. These are typically termed “WIMPZillas” due to their high mass and present as an exotic WIMP-like dark matter candidate.

As conventional WIMPs fell increasingly out of favour, many sub-GeV candidates have gained general interest. Some attribute the absence of dark matter signatures in experiments to be due to indirect coupling of dark matter particles to SM particles via mediators, which together constitute a “dark sector” [30, 31]. Dark photons are a popular mediator which are massive, gaining their mass from a dark Higgs. Dark photon coupling to dark matter is a common search theme amongst dark matter experiments [32] and is something that can even be probed gravitationally [33].

Another popular dark matter candidate is the axion. Axions were originally proposed as a solution to the strong CP problem [34], but could at the same time be dark matter. Axion searches typically try to detect axion-to-photon conversions in the presence of an electromagnetic field. These photons can be detected via “light shining through a wall” experiments, which look for photons produced from axions on the other side of a barrier [35], or cavity searches, which look for resonant peaks in the photon spectra [36], or even solid-state detectors such as Ge or Si [37]. Quantum Chromodynamics (QCD) axions are expected to have a mass of $\sim 10^{-5} \text{ eV}/c^2$. A whole class of Axion Like Particles (ALPs) has been proposed as extensions to the axion model, spanning a range of masses (typically $\lesssim 1 \text{ keV}/c^2$) [38].

Axions and ALPs are examples of ultra-light dark matter, while dark photons and asymmetric dark matter [39] are examples of light dark matter. There are also many interesting candidates in the higher mass range such as primordial black holes [40] and Massive Compact Halo Objects (MACHOs). The former are non-baryonic, where as the latter are potentially baryonic. MACHOs have been found to constitute only a negligible portion of the dark matter halo [41] meaning that they have effectively been ruled out as a dark matter candidate.

Thus, particle physics converts the idea of an invisible thermal relic into a search for a detectable, weakly interacting particle. In the next sub section, I will discuss dark matter detection methods in this context.

1.2.2 Dark matter detection

Particle physics experiments for dark matter searches look for some form of coupling between Dark Matter (DM) and SM particles. This is demonstrated in Fig. 1.8. There are three broad categories of dark matter searches:

- **Collider search:** Looks for signatures of DM production from relativistic SM particles in colliders.

- **Indirect detection:** Looks for signatures of SM particle production from DM annihilation.
- **Direct detection:** Looks for signatures of direct DM interactions with SM particles.

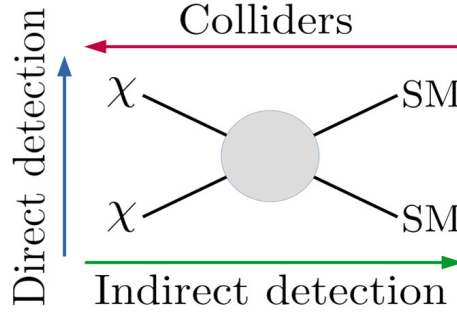


Figure 1.8: Schematic representation of various dark matter detection channels. Figure taken from Ref. [12]. Reprinted with permission.

Colliders, such as the Large Hadron Collider (LHC), look for signatures of new physics in particle collisions. They operate at high energies (e.g. the LHC operates at an energy scale of $\sqrt{s} \sim 14$ TeV) which makes them an ideal test bed for SUSY predictions and BSM physics. Searches for Higgs couplings [42] have been of particular interest recently with the discovery of the Higgs Boson [43, 44]. The “Higgs portal” refers to a plethora of mechanisms by which DM could interact with Higgs sector particles [45] producing potentially detectable missing energy or momentum. So far no such hints have been observed, but searches are still underway.

Dark matter annihilation in galactic centers or galaxy clusters could produce γ -rays, neutrinos and other byproducts [15]. Indirect detection experiments look for excess signals beyond what is expected from visible matter in nearby extra-terrestrial sources. Neutrino experiments such as IceCube [46] and Super Kamiokande [47] are well suited for identifying neutrino products from dark matter annihilation. No such hints have been reported by these experiments so far. The Fermi Large Area Telescope (FLAT), a γ -ray telescope, had previously reported hints of DM annihilations from the Galactic center [48, 49] but they have been explained away with alternate models such as unresolved γ -ray point sources [50].

Direct detection experiments search for direct interactions of DM with a SM particle. This is done by searching for nuclear recoil (NR) or electron recoil (ER) from elastic scattering with an incoming dark matter particle as dark matter winds from the galaxy sweep through the Earth. WIMPs are typically expected to produce nuclear recoils, whereas lighter dark matter candidates, such as axions and dark photons, kinematically favor electron recoils. A detailed derivation of the

differential scattering rate R can be found in a review report by Lewin and Smith [51] – the bible of direct detection. The differential scattering rate is given by,

$$\frac{dR}{dE_{\text{recoil}}} = \frac{\rho_\chi}{m_\chi m_T} \int_{v_{\min}}^{\infty} v \cdot f(\vec{v}) \cdot \frac{d\sigma_\chi}{dE_{\text{recoil}}} d\vec{v}. \quad (1.6)$$

Here, E_{recoil} is the recoil energy of the nucleus, ρ_χ is the local dark matter density, m_T is the mass of the target nucleus or electron, v is the dark matter velocity, $f(\vec{v})$ is the dark matter velocity distribution, and σ_χ is the dark matter-nucleon interaction cross-section and m_χ is the mass of the dark matter particle. The differential cross section $\frac{d\sigma_\chi}{dE_{\text{recoil}}}$ can be split into a spin-dependent and spin-independent term. The spin-independent cross section scales with the square of the atomic mass, A , of the target and so heavier nuclei in targets give a better chance of spin-independent interactions [15]. Spin-dependent dark matter interaction detection depends on the spin of the target and presents different challenges in terms of the need for isotopically pure targets. Thus, spin-dependent dark matter detection requires specially chosen target materials for optimal sensitivity.

In general, many target materials exist for dark matter detection, including liquid noble detectors such as XENON [52] and DarkSide [53], cryogenic bolometers with semiconductor substrates such as Super Cryogenic Dark Matter Search (SuperCDMS) [32], and Cryogenic Rare Event Search with Superconducting Thermometers (CRESST) [54] and superheated liquid bubble chambers like PICO [55] and Direct search Experiment for LIGHT dark matter (DELIGHT) [56]. The various direct detection technologies currently in use today are shown in Fig. 1.9. Liquid noble detectors are easier to purify and cost effective to scale to larger sizes whereas semiconductor detectors, owing to their small bandgaps, are much easier to produce excitations in and are hence more sensitive detectors at the lowest dark matter masses [57]. The former have most sensitivity to electroweak-scale dark matter while the latter have most sensitivity to few GeV to sub-GeV dark matter. Experiments with spin-sensitive target nuclei (e.g. ^{19}F in C_3F_8 used in the PICO bubble chamber) are suitable for spin-dependent dark matter interactions.

In the absence of a dark matter signal, the general strategy is to construct an exclusion curve that constrains the region of parameter space defined by the dark matter mass and cross section. Such a limit curve represents the smallest cross section at a given dark matter mass that can be excluded given the data. Experiments often construct limit curves with the same underlying assumptions about the dark matter distribution, interaction rates and properties to compare their sensitivities against other experiments. The standard assumptions are that WIMPs have a Maxwellian velocity distribution in the galaxy's rest frame, and that they scatter off of nucleons, resulting in an expo-

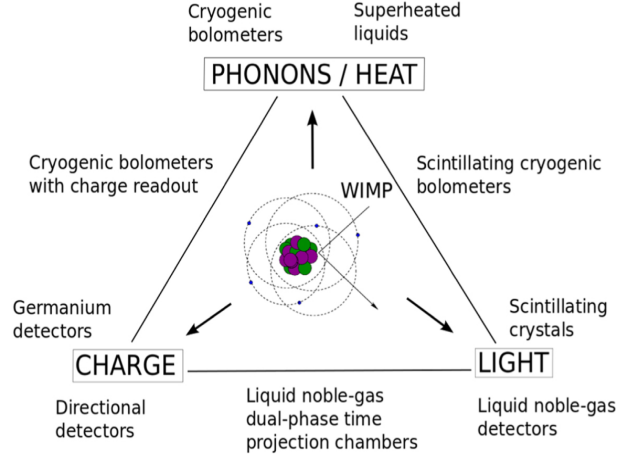


Figure 1.9: Schematic representation of various direct detection strategies. Figure taken from Ref. [15]. Reprinted with permission.

nentially decreasing event rate as a function of recoil energy. The most recent result comparing limits from all the major direct detection experiments for spin-independent dark matter scatters is shown in Fig. 1.10.

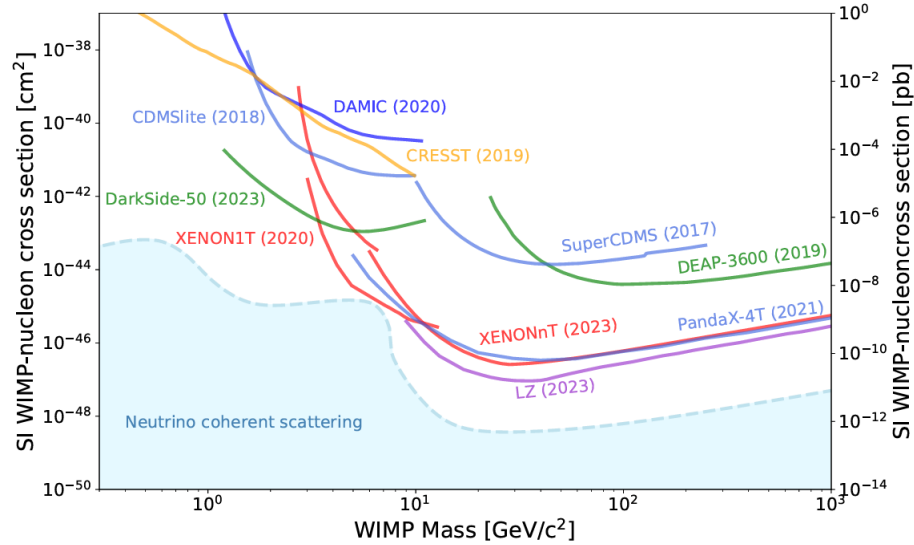


Figure 1.10: Comparison of the most recent exclusion limits from various direct detection experiments for spin-independent WIMP-nucleon scattering. Figure taken from Ref. [58]. Reprinted with permission.

Unlike collider searches and indirect detection experiments, direct detection experiments conduct searches at a significantly lower energy scale. Therefore, performing such experiments on the surface would make them extremely background limited due to exposure to higher amount of surrounding radiation. These experiments are, therefore, often performed in underground labs such as the Sudbury Neutrino Observatory LAB (SNOLAB) or Laboratori Nazionali del Gran Sasso (LNGS) to shield against cosmic radiation.

SuperCDMS SNOLAB is a planned 4th generation direct detection experiment that is expected to begin commissioning sometime this year (2024) at SNOLAB (one of the deepest underground facilities in the world). In the next chapter, I will discuss the motivation, detection principles and experimental techniques used in SuperCDMS and introduce the motivation for my research.

Chapter 2

Direct detection of dark matter with SuperCDMS

CDMS detectors are cryogenic devices with semiconductor crystals as target materials. An energy deposition in the crystal produces ionization (charges) and quantized vibrations (phonons). The charges are drifted to the end of the detectors and collected there, while the phonons are collected via a bolometric mechanism. We will visit the details of how this works as we progress through this chapter.

Direct dark matter detection with CDMS commenced in the 1990s with the first version of the experiment, CDMS I, at the Stanford Underground Facility (SUF) [59, 60]. CDMS I utilized a simultaneous charge and phonon readout mechanism. Most background radiation produces electron recoils (ERs), which gives them a bigger ionization signal than phonon signal while nuclear recoils (NRs) produce the opposite effect. This allowed the CDMS detectors to achieve excellent background rejection using NR-ER discrimination. Nuclear recoils are the primary mode of interaction for heavy WIMPs and gave the experiment sensitivity to candidates with a mass of 50 to a few 100 GeV/c^2 [61]. In order to reduce cosmogenic background which limited experimental sensitivity, CDMS II moved to the Soudan mine in 2002, a deeper underground facility, to achieve a factor of 10^4 reduction in muon flux [62]. CDMS II set the most stringent limits on spin-independent WIMP masses of 40 GeV/c^2 and higher at the time [63] and for the first time extended its search to axions using electron recoils [37]. CDMS II also saw the experiment's first attempt to search for lighter dark matter candidates of masses $< 10 \text{ GeV}/c^2$ with a lower trigger threshold [64]. CDMS II was succeeded by SuperCDMS Soudan which used interleaved ionization electrodes to improve NR-ER discrimination at the detector surface, thus improving the experiment's

fiducialization capability [65, 66]. These detectors were called interleaved Z-dependent Ionization and Phonon (iZIP) detectors. SuperCDMS Soudan, with its advanced background rejection capabilities, performed searches for WIMPs [67] and other exotic particles, such as dark photons [32]. SuperCDMS Soudan was the first in CDMS history to leverage the Neganov-Luke-Trofimov (NTL) effect [68, 69] to produce amplified phonon signals and significantly improve the sensitivity of the experiment to low mass dark matter candidates [67]. This was done during three runs of a subsidiary experiment called CDMSLite, which will be a topic of focus in the final chapters of this thesis. The demonstrated success of CDMSLite [70, 71] inspired the development of High Voltage (HV) detectors for the next generation of the SuperCDMS experiment, SuperCDMS SNOLAB, which is under construction at SNOLAB near Sudbury, Canada. In this chapter, I will explain the detection principle behind SuperCDMS detectors with a focus on SuperCDMS SNOLAB detectors and provide an overview of the planned experiment in terms of hardware, reconstruction techniques and sensitivity expectations.

2.1 Detection principle

SuperCDMS uses cryogenic detection technology to detect particle interactions. The detectors used are ultra-low radioactivity Si and Ge crystals equipped with Transition Edge Sensors (TES)¹. Semiconductor crystals, such as those used by SuperCDMS, present as an attractive candidate to build cryogenic detectors for dark matter searches, owing to their small band gaps, excellent background discrimination power and relative ease of obtaining radio-pure substrates [72, 73]. The description of semiconductor physics in this section follows similar discussions as in Ref. [74–76].

Semiconductors contain a valence band and a conduction band, with electrons typically occupying a given state in either band at room temperature, with some probability, f , described by the Fermi-Dirac distribution:

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_f}{kT}\right)}. \quad (2.1)$$

Here, E is the energy of the electron, T is the temperature of the system, k is the Boltzmann constant and E_f is the energy of the Fermi state or the hypothetical state that has a 50% chance of occupancy at thermal equilibrium. At low temperatures, such as the operational point of cryogenic detectors, electrons are confined to the valence band. The difference in energy between the highest state in the valence band and the lowest state in the conduction band, called the band gap, E_{gap} ,

¹TES technology will be explained later

2.1. DETECTION PRINCIPLE

gives a sense of the minimum required energy to excite an electron from the valence band to the conduction band. For Si and Ge, the band gaps at 0 K are 0.74 eV and 1.17 eV, respectively [77]. The excitation of an electron from the valence band to the conduction band produces a hole in the valence band. The hole constitutes a quasi-particle of equal and opposite charge to the electron. Every excited electron in the conduction band thus has a complementary hole in the valence band, together forming an e^-h^+ pair. The minimum energy required to produce a single e^-h^+ pair is at least equal to the band gap energy. An average particle interaction can produce one or more e^-h^+ pairs, depending on the energy of the interaction. Every interaction could produce secondary ionizations, rendering the process of pair-creation non-Poissonian [78]. Si has an average pair-creation energy, ϵ , of 3.8 eV, while it is about 3 eV for Ge [79].

Particle interactions deposit some energy in the crystal, thus producing e^-h^+ pairs. In the presence of an electric field, the free electrons travel towards the positive potential and the holes propagate in the opposite direction in an “apparent motion”. The charges that arrive at the ends of the detector can be collected directly to produce a corresponding ionization signal for the interaction that produced the excitation.

Quantized vibrations of the crystal lattice are called phonons. Particle interactions can produce phonons, either via electron recoils or nuclear recoils, and those have energies about three orders of magnitude lower than e^-h^+ pairs [80]. The amount of energy that goes towards generation of phonons depends on the nature of the interaction and the energy of the incoming particle. In typical electron interactions, an ionization signal is more common. In nuclear interactions, such as those produced by incoming WIMPs or neutrons, about 60–90% of the energy goes towards phonon production [80, 81]. Si and Ge are examples of indirect band-gap materials, which means that the highest state in the valence band does not vertically align in momentum (see Fig 2.1) with the lowest state of the conduction band. In indirect band-gap materials, a particle with energy less than the band gap can still excite an electron into the conduction band. The incident particle transfers energy to produce a change in energy, while a phonon absorption causes a change in momentum sufficient to place the electron in the conduction band. This is shown in Fig. 2.1. Whether the phonon is absorbed or emitted depends on the energy of the incident particle and the momentum transfer required to attain stable equilibrium.

Phonons are of two types: acoustic and optical. The former are produced from in-phase vibrations, where as the latter are products of out-of-phase vibrations of the atoms in the crystal lattice. In addition, phonons can be grouped based on thermal properties into thermal and athermal phonons. Thermal phonons arise from thermal fluctuations in the equilibrium energy state of the lattice at non-zero temperatures. Athermal phonons, on the other hand, are produced from the

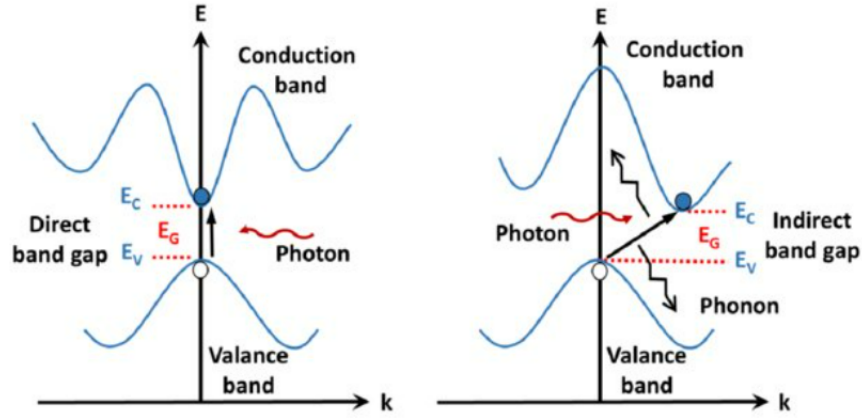


Figure 2.1: Schematic representation of photon absorption in direct (left) versus indirect (right) band-gap materials. E_G is the band gap, E_C the bottom of the conduction band and E_V the top of the valence band. The k axis is the momentum and the y -axis shows energy. Figure taken from Ref. [82].

thermal non-equilibrium in the lattice, such as by a particle scattering. For example, an incoming particle could excite an electron into the conduction band to some energy, which will then relax into a stable valley in the conduction band via athermal phonon emission. Athermal phonons can be acoustic or optical.

The preliminary athermal phonons produced by a particle interaction constitute “prompt” phonons. The application of an electric field to drift charges across the crystal produces secondary phonons called NTL phonons or Luke phonons [68, 69]. This is caused by the charges quickly getting accelerated to their terminal velocity by the applied electric field and additional energy gained after this point being transferred to the lattice. Furthermore, charges collected at the ends of the crystal can recombine producing “recombination” phonons. These three kinds of athermal phonons produce a phonon signal that is measured by SuperCDMS detectors.

Athermal phonon energy can be measured via a bolometric mechanism, since its production yields a measurable change in temperature. SuperCDMS measures phonon energy using tungsten Quasiparticle-trap-assisted Electrothermal-feedback Transition-edge-sensors (QETs). QETs contain aluminum fins which improve phonon collection efficiency and operate based on an electrothermal feedback mechanism [83]. The QETs are deposited on the surface of the crystal lattice, distributed in such a way that they partition the crystal volume into sections. Each section collec-

2.1. DETECTION PRINCIPLE

tively constitutes a channel with a signal collected and read out in every channel for each event. TESs are superconductors operated at a transition point between their superconducting and normal states. A thermal excursion in the TES produces a change in resistance which is converted to a measurable current. Tungsten is a popular choice as a superconducting material for TESs due to the tunable nature of its critical temperature, T_c [84]. In SuperCDMS, the baseline energy resolution, σ , and by extension the minimum achievable experimental threshold is closely tied to T_c , since $\sigma \propto T_c^3$ [83].

Superconductors are materials that exhibit zero resistance below their critical temperature. Figure 2.2 shows a cartoon of the resistance profile of a TES as a function of its temperature. The “normal” resistance of the TES is its resistance at room temperature. As the TES is cooled it “transitions” into a superconducting phase. It is important to note that the transition does not happen in a direct step from normal to zero resistance, but rather over a width of a few tens of mK, marked by the gray box. This grayed area represents the operational region of the TES in cryogenic experiments. The TES is typically held half-way between its superconducting and normal state by the application of a bias current (QET bias). A particle interaction generates phonons that are collected by the Al fins, where the phonon energy breaks up Cooper pairs whose energy heats up the TES. Note that the phonons themselves could directly heat up the TES, but this is not the predominant effect. The heating pushes the TES upwards in transition, resulting in a drop in current (for a constant voltage applied across the TES), which we measure as a signal or pulse. Further, electrothermal feedback cools the TES, returning it to its equilibrium state.

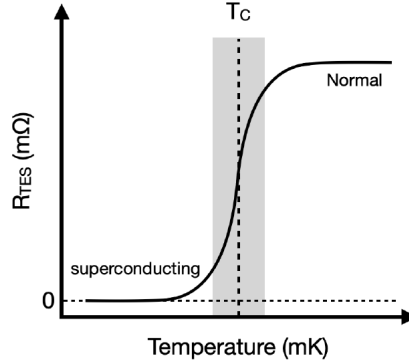


Figure 2.2: Cartoon of a TES transition curve. This transition curve shows the resistance profile of the TES (R_{TES}) as a function of temperature. The normal, transition and superconducting phases are clearly marked. The gray shaded box shows the transition width or window of the TES. Figure taken from Ref. [75]. Reprinted with permission.

2.1. DETECTION PRINCIPLE

The total energy collected by the TES (E_t) is a combination of all three types of phonons. From [75],

$$E_t = E_{\text{prompt}} + E_{\text{recomb}} + E_{\text{NTL}}, \quad (2.2)$$

which can alternatively expressed as

$$E_t = E_r + neV_b, \quad (2.3)$$

where E_r is the energy deposited by the recoiling particle, n is the average number of e^-h^+ pairs produced and V_b is the applied bias voltage. The first term is the total deposited energy and the second term is the amplification obtained from the NTL effect. The average number of e^-h^+ pairs produced depends on the pair-creation energy and is given by $n = \frac{Y(E_r)E_r}{\epsilon}$, where $Y(E_r)$ is the ionization yield or the fraction of recoil energy that goes towards charge production. This substitution simplifies Eq. (2.3) to give

$$E_t = E_r \left(1 + \frac{Y(E_r)eV_b}{\epsilon} \right). \quad (2.4)$$

One can derive the energy of the recoiling entity with simple kinematic considerations. If an incoming particle of mass m_χ and velocity v scatters off a target of mass m_T , then the energy of the target recoiling at an angle θ_R is

$$E_r = \left(\frac{m_\chi m_T}{m_\chi + m_T} \right)^2 \frac{v^2}{m_T} (1 - \cos \theta_R). \quad (2.5)$$

The mass of a Si or Ge nucleus is about $O(50 \text{ GeV}/c^2)$. For lighter incoming particles with masses in the keV to MeV range recoiling off nuclei in these materials, in Eq. (2.5), $m_T \gg m_\chi$. Then E_r , which is proportional to m_χ/m_T , is suppressed for nuclear recoils from lighter incoming particles in these targets. On the other hand, the mass of an electron is only $0.51 \text{ MeV}/c^2$, which makes it a more favorable target for energy transfer for lighter incoming particles. Similarly, particles with larger masses kinematically favor nuclear recoils. Electron recoils produce more ionization than nuclear recoils. The ionization yield is defined with respect to the ionization produced by an electron of the same energy, thus giving a yield of 1 for electron recoils. Substituting this value for the yield in Eq. (2.4), gives the energy in electron-equivalent units,

$$E_{ee} = E_r \left(1 + \frac{eV_b}{\epsilon} \right). \quad (2.6)$$

We will use Eqs. (2.4) and (2.6) in later chapters to denote energies with corresponding definitions. Nuclear recoils have an energy-dependent yield that is described by the Lindhard model [85, 86]. This is often approximated by an empirical formula that depends on the atomic number Z and atomic mass A of the target material:

$$Y(E_r) = \frac{k \cdot g(\epsilon)}{1 + k \cdot g(\epsilon)}, \quad (2.7)$$

where

$$\begin{aligned} \epsilon &= 11.5 E_r \cdot Z^{-\frac{7}{3}}, \\ g(\epsilon) &= 3\epsilon^{0.15} + 0.7\epsilon^{0.6} + \epsilon, \\ k &= 0.133 \cdot Z^{\frac{2}{3}} \cdot A^{-\frac{1}{2}}. \end{aligned} \quad (2.8)$$

Equation (2.7) for the yield agrees with measurements only for high recoil energies, while the measurements deviate from the Lindhard prediction at lower energies. The measurement of the yield at low energies is an active area of research [87–89].

2.2 SuperCDMS detectors

SuperCDMS employs two detector types. The iZIP detector operates with both charge and phonon readout. This allows this detector to achieve excellent discrimination between electron and nuclear recoils due to the significantly different amount of ionization produced by these two interaction types. These detectors are designed to tolerate only low voltages, which means that the contribution from the second term in Eq. (2.3) is small. This limits the lowest energy one can measure with these detectors, resulting in higher thresholds. Therefore, these detectors are more sensitive to dark matter masses in the range of 5 to 10 GeV/ c^2 . HV detectors, on the other hand, extensively use the NTL effect and are designed to tolerate up to 100 V, thus being capable of producing up to a factor of ~ 34 (27) amplification in Ge (Si). This allows HV detectors to achieve low thresholds and be sensitive to 0.5 to 5 GeV/ c^2 in dark matter mass, and even probe sub-GeV scale dark matter particles. Note that events recorded in multiple detectors from the same particle scatter can be used to isolate backgrounds since dark matter interactions, being rare events, are unlikely to produce multiple scatters.

The detectors will be arranged in stacks called towers, each containing six detectors. Although SuperCDMS SNOLAB facility has the capability to host up to seven towers, the initial payload will contain four towers, two of which contain iZIP detectors only and the other two with HV detectors only. The towers will be housed in a SNOBOX which contains seven nested copper cans, with

2.2. SUPERCDMS DETECTORS

each can thermally linked to a stage of the dilution fridge. The SNOBOX is enclosed by various layers of shielding to reduce the ambient neutron and gamma background in the cavern and sits on a seismic platform which reduces the impact of vibrations coupling into the detectors. The dilution fridge is placed outside on one side of the shield (in the C-tank), with a stem that connects the various cooling stages of the fridge to the SNOBOX cans. This is meant to vibrationally isolate the detectors from the noise sources in the dilution fridge. Similarly, the electronics are separately housed in the E-tank. The initial payload features a total of eighteen Ge and six Si detectors. All towers contain four Ge and two Si detectors with the exception of tower one hosting six Ge iZIP detectors. Both iZIP and HV detector types are segmented into twelve phonon readout channels, as compared to 8 channels in previous detector designs (see Figs. 2.3 and 2.4). The iZIP additionally comes equipped with four charge readout channels. The new detectors will have a size of 10 cm in diameter, 3.33 cm in thickness and weigh 1.39 kg and 0.61 kg for Ge and Si, respectively. The increase in size and weight compared to their predecessors allows these state-of-the-art devices to gain in net exposure and achieve a higher overall sensitivity. A schematic of the SuperCDMS SNOLAB facility is shown in Fig. 2.5.

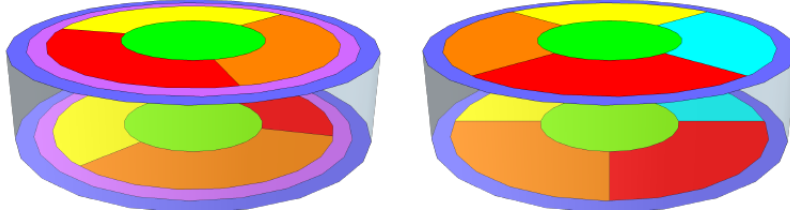
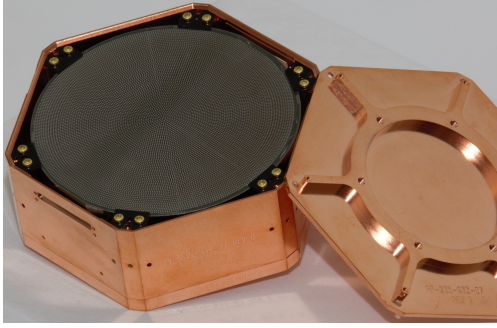
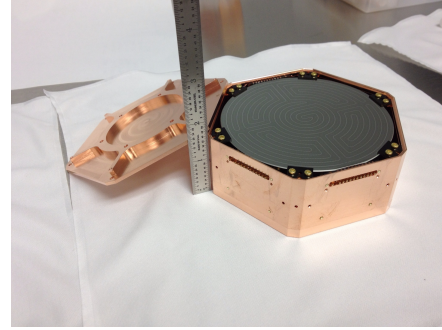


Figure 2.3: Schematic showing the channel map for SuperCDMS SNOLAB HV (left) and iZIP (right) detectors. Both detector types have 12 phonon channels arranged to maximize position sensitivity. Figure taken from the SuperCDMS Technical Design Report [90].

These new detectors are designed for a specific performance level and extensive testing is required to ascertain this. “Tower testing” is an effort that was conducted at the Stanford Linear Accelerator Center (SLAC) and then again at the Cryogenic Underground TEST facility (CUTE) in SNOLAB. The CUTE setup will be discussed in more detail in the next chapter. This will be a topic of focus in my thesis and major results will be shown in the next chapter.



(a) HV detector



(b) iZIP detector

Figure 2.4: Fabricated SuperCDMS SNOLAB detectors.

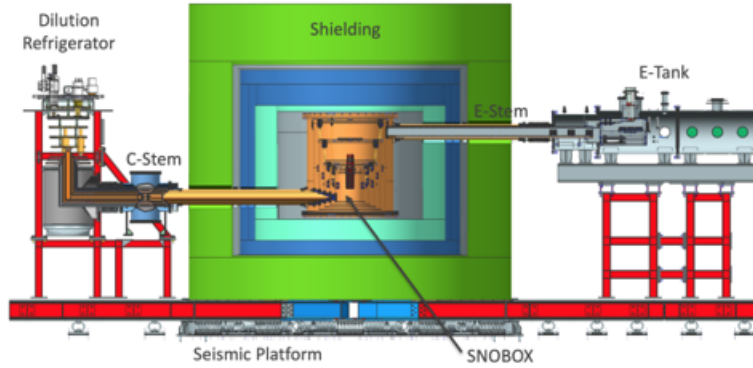


Figure 2.5: Schematic of the experimental setup of SuperCDMS SNOLAB. Figure taken from the SuperCDMS Technical Design Report [90].

2.3 Data acquisition

The next point in the data chain is the SuperCDMS Data Acquisition (DAQ) system. SuperCDMS uses a multi-level trigger system. My collaborators provide a detailed discussion of the SuperCDMS SNOLAB Level 1 trigger system in Ref. [91]. I will attempt to provide an executive summary here. The Detector Control and Readout Cards (DCRCs) are at the core of our data acquisition program and are responsible for issuing triggers. Every DCRC is capable of twelve phonon channel and four charge channel readout. A single DCRC is used to read out iZIP detectors, i.e., all phonon and charge channels are used. An HV detector only contains phonon channels, six on each side. Each side is read out by one DCRC, half of whose phonon channels are used. This

2.3. DATA ACQUISITION

allows the reference ground for each DCRC to be independently floated, permitting a larger voltage difference across the detector. Fig. 2.6 shows a schematic of the trigger system.

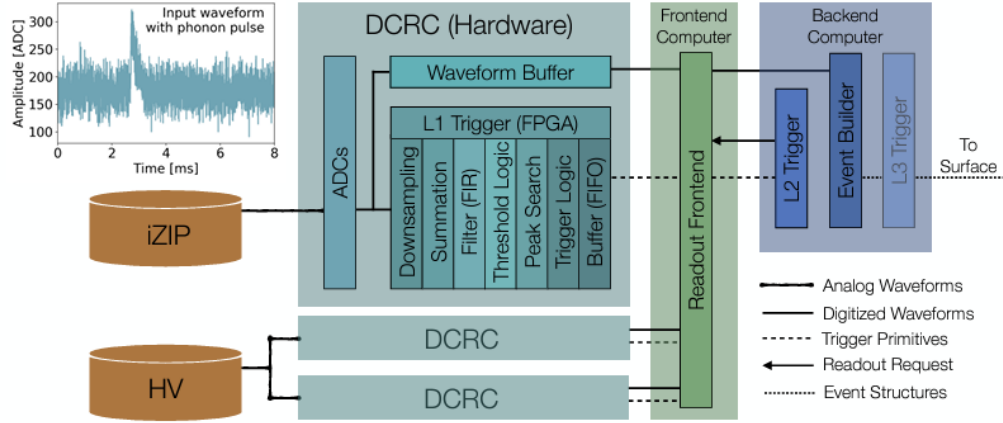


Figure 2.6: Schematic of the SuperCDMS trigger system. Figure taken from Ref. [91].

The waveforms are stored in a circular buffer (~ 6.7 s long) till they are passed to the DCRC's FPGA. DCRCs have a digitization rate of 625 kHz for phonons and 2.5 MHz for charge waveforms. The FPGA operates an L1 trigger containing a filter which downsamples the waveform by a factor of 16. A linear combination module performs a weighted sum of the downsampled waveforms and a finite impulse response filter filters out signals of interest from the noisy pulse. The Level 1 trigger is issued by the FPGA of the DCRC when the weighted sum of the downsampled, filtered, input waveforms exceeds a pre-determined threshold. A peak search is performed to produce trigger primitives that describe the location of the peak and its properties. A trigger logic module then uses these trigger primitives to make a trigger decision. The Level 1 decisions are stored in a FIFO buffer and passed to the software front-ends at a rate of 1 Hz. Note that upgrades to the trigger system that yield more efficient triggering have also been a topic of interest (e.g., see Ref. [92] for a study on a neural-network-based trigger).

Level 2 decisions are made by MIDASDAQ – the SuperCDMS DAQ software. MIDASDAQ is coded in C++ and uses MIDAS, which is a DAQ software developed by TRIUMF. It is capable of reading and writing to the DCRCs via TCP/IP sockets. MIDAS allows one to set up a web interface where all experimental settings stored in the Online DataBase (ODB) can be viewed and changed. Level 1 trigger decisions intercepted by the MIDAS frontends are used to issue a Level 2 trigger

2.3. DATA ACQUISITION

and construct an “event”. On top of Level 1 information, the software trigger algorithm determines if there are pile-up triggers (multiple events in a single pulse window of one detector) or coincident triggers (the same event triggered on multiple detectors) and merges or discards these to give good events that can be analyzed meaningfully. A Level 2 trigger causes the recorded waveforms to be read out from the DCRCs. The data for all triggered events are written to disk and transferred to the surface computers and further on to SLAC for data processing and permanent storage.

In addition to MIDASDAQ, other helper packages have been developed to enable automation of various detector testing tasks. For example, the DCRC UI is one such software package which allows the automation of SQUID tuning, resistance measurements and T_c measurements, automatically starting pre-defined runs, saving ODB settings to a database, etc. Similarly, MIDASTools provides a GUI where one can view pulses (and their power spectral densities) in real time, quickly change certain frequently used settings and save pulses.

MIDASDAQ is special because it allows interfacing with the MPOD as well. The MPOD is a high voltage power supply which is used for the operation of HV detectors. It contains four modules – half with positive polarity and the other half with negative polarity. One side of the HV detector is hooked up to a positive output of the MPOD while the other is hooked up to the negative output. MIDASDAQ can communicate with this HV supply using its custom-designed High Voltage Driver. Thus, MIDASDAQ enables a single point of operation for all detector settings using its ODB.

SuperCDMS SNOLAB will have an initial payload of four detector towers with six detectors each. The iZIP towers will use a total of 12 DCRCs and the HV towers will use a total of 24 DCRCs bringing the final count to 36 DCRCs. Each DCRC has a maximum throughput of $\sim 5 \text{ MBs}^{-1}$. SuperCDMS SNOLAB will read out 52 ms waveforms for a standard WIMP search analysis. A typical dark matter data set can go up to a trigger rate of approximately 2 Hz. Calibration data sets often have much higher rates close to 5 Hz per detector. A typical SuperCDMS event has an average size of 1.8 MB when using the standard tracelength. This means we require a throughput roughly in the range of 3.6 to 9 MBs^{-1} per DCRC for calibration data, which exceeds the network capacity on these cards. In order to overcome this limitation, SuperCDMS SNOLAB will downsample the pre- and post-pulse regions of every trace (which is typically just the baseline current, i.e., noise) by a factor of 16 (see Fig. 2.7). This “hybrid” mode allows us to keep within our hardware and software abilities, while taking high quality data sets. This complicated new data format affects our data processing, limiting our reconstruction capabilities. Part of the work detailed in this thesis will describe new reconstruction techniques which were specifically designed to process hybrid data and demonstrate how these methods could help the experiment maintain a lower threshold.

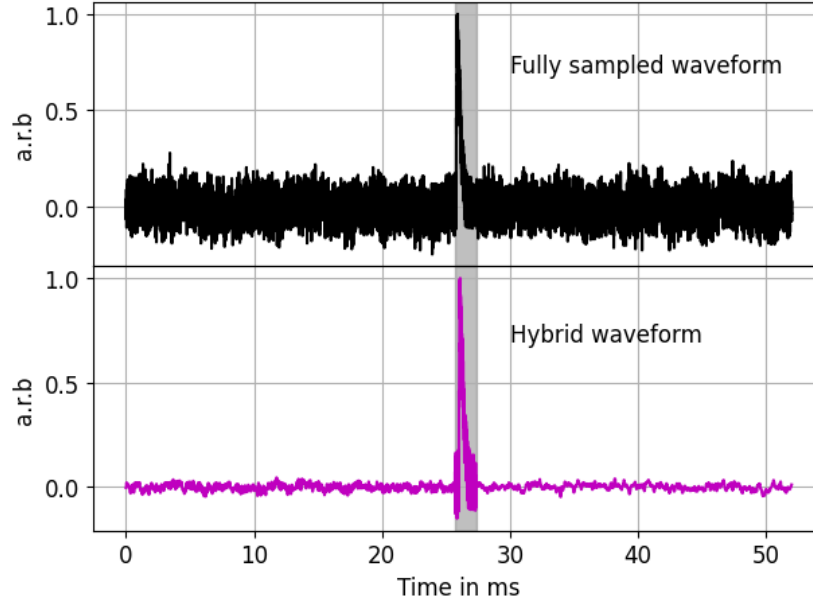


Figure 2.7: Comparison of fully sampled waveforms with a hybrid sampled waveform. The top panel shows a typical fully sampled waveform. The bottom panel shows the same waveform with its tails downsampled by a factor of 16 by averaging every 16 time bins. The gray shaded region shows the fully sampled region and where it appears in the fully sampled trace. “a.r.b” stands for arbitrary units, since the y-axis is scaled by an arbitrary normalization factor.

2.4 Event reconstruction

The data arriving at SLAC are run through a data processing pipeline, which passes it to our core processing software. CDMSBats is the main processing package which contains five sub-packages. Two of these sub-packages, BatViewer (for viewing pulses) and BatFaker (for generating events with artificial pulses), are utility packages and are not worth a mention in this thesis. The remaining three sequentially process data to convert raw pulses into Reduced Quantities (RQs) which characterize physically interpretable information such as the baseline of the pulse, the height or amplitude of the pulse and energy collected in the pulse, and then the RQs are converted into Reduced Reduced Quantities (RRQs), which are calibrated quantities. The full CDMSBats pipeline is shown in Fig. 2.8.

SuperCDMS raw data consists of digitized waveforms. The primary goal of reconstruction is to extract the energy of the recoiling nucleus or electron. The recoil energy is kinematically related

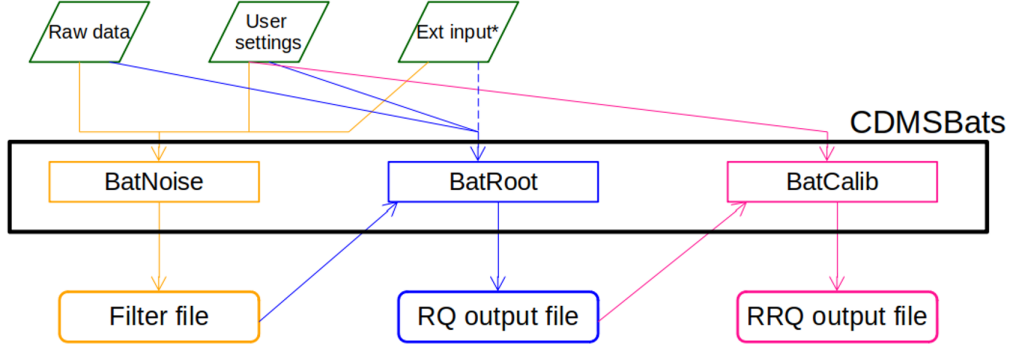


Figure 2.8: Flow chart showing the workflow of the processing package, CDMSBats. CDMSBats has three main sub-packages that handle data. BatNoise takes the raw data, user settings and other external quantities, such as template files, as input and outputs a filter file containing quantities required for BatRoot processing. BatRoot takes the filter file, raw data and other optional external quantities as input and produces an RQ file in ROOT format. BatCalib applies calibration factors to the BatRoot outputs.

to the the energy of the incoming particle and by extension its mass and cross-section. The height or amplitude of the pulse (also called trace or waveform in SuperCDMS) is directly proportional to the collected energy. The easiest way to extract this information is to fit a model to each pulse via a χ^2 fit (see Fig. 2.9). If the noise is not white noise, the time-domain covariance matrix of the noise data is non-diagonal. On the other hand, the frequency-domain covariance matrix, V , of the data is diagonal. This makes it advantageous to perform the χ^2 fit in frequency space by working with the Discrete Fourier Transform (DFT) of the pulses. One can minimize the following χ^2 summed over frequency, j , with respect to the pulse amplitude, $\hat{A}$:

$$\chi^2 = \frac{\sum_j |s_j - e^{i\omega_j \hat{t}_0} \cdot \hat{A} \cdot v_j|^2}{J_j}, \quad (2.9)$$

where at frequency index j , s_j is the signal or trace DFT, v_j is the DFT of the template or model which captures the shape of the signal (normalized to 1), J_j is the squared noise Power Spectral Density (PSD), i.e., the diagonal elements of V , and $\hat{t}_0$ is the time delay between the template and the triggered signal. For a fixed value of t_0 , the above χ^2 can be minimized analytically to obtain the amplitude, $\hat{A}$:

$$\hat{A} = \frac{\sum_j e^{i\omega_j \hat{t}_0} \frac{v_j^*}{J_j} s_j}{\sum_j \frac{|v_j|^2}{J_j}}. \quad (2.10)$$

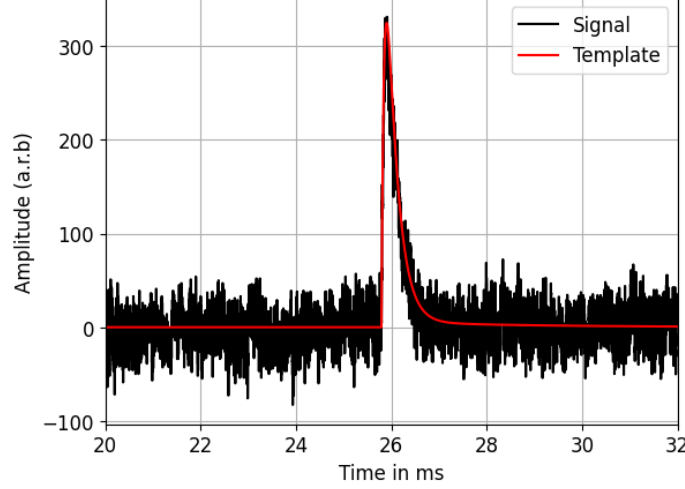


Figure 2.9: Comparison of a toy pulse overlaid with a template. The simplest way of extracting meaningful information from pulses often involves fitting a model/template to the signal. This can be done by minimizing the simple χ^2 given in Eq. (2.9).

Here, s_j is the Fourier transformed signal, the quantity $\frac{v_j^*}{J_j}$ is a filter that optimally filters the signal from the noise and the quantity $\sum_j \frac{|v_j|^2}{J_j}$ is the signal-to-noise ratio. This algorithm is often called the Optimal Filter (OF). To distinguish this from the new algorithms that will be introduced in this thesis, I will refer to this as the 1×1 OF from now on, since it fits one template to one waveform. Note that the value of the time offset t_0 is not determined by Eq. (2.10). Analytically minimizing the χ^2 in Eq. (2.9) with respect to $\hat{t}_0$ is not trivial, since it is not a simple linear relationship. However, it turns out that the optimal time delay corresponds to the value of t_0 that maximizes $\hat{A}$, which in turn is the value that minimizes the χ^2 . Therefore, the simplest algorithm simply scans over a range of $\hat{t}_0$ values to find the value which gives the maximum amplitude.

This simple fit, although it works well, does not provide the best possible energy estimator for the data. This is because the model being assumed fails to consider effects of correlated noise and position dependence. Correlated noise refers to any kind of noise which affects multiple detector channels. This could originate from unwanted vibrational couplings to the detector's surroundings

or could be an artifact of common channel readout lines in the electronics. Position dependence arises from the difference in pulse shapes and energy collected due to the event location in the crystal. Not accounting for these two effects worsens the resolution of our energy estimators. Sophisticated algorithms need to be developed to effectively correct for correlated noise and position dependence. A part of this thesis will focus on some novel approaches to accomplish this. More details will follow in later chapters. Coming back to OFs, Eq. (2.9) assumes a simplistic model, v_j , which no longer fits the data when the shape of the signal varies. Thus, it ends up broadening the distribution of $\hat{A}$ in Eq. (2.10). Further, the denominator is only valid if our noise is truly uncorrelated. However, correlations between channels can turn the covariance matrix non-diagonal which means that the χ^2 defined in Eq. (2.9) needs to be generalized to account for this effect.

Over the years, SuperCDMS has developed a variety of reconstruction algorithms to deal with these two situations. One such algorithm is called the two-template OF or the 1×2 OF. The initial prompt phonons, discussed in Sec. 2.1, travel with a mean free path much shorter than the size of the crystal, scattering and downconverting to lower energy phonons via a few frequency-dependent processes [93]. Prompt phonons arrive first and therefore contain strong position information about where the recoil happened in the detector. The final low energy phonons produced are called ballistic and have a mean free path longer than the size of the crystal. They scatter less, effectively travelling in straight lines until they are absorbed by the QETs. The ballistic phonons contribute a slow component to the pulse shape, which mostly contains energy information. The prompt phonons, on the other hand, contribute a fast component to the pulse shape, which begins and ends faster. These “slow” and “fast” components can be extracted by studying a collection of pulses. The average of a collection of pulses can be a rough estimator of the slow component. The residuals obtained from subtracting the slow component from the pulses could be averaged to obtain the fast component. With these two components, one can perform a modified χ^2 minimization with two templates:

$$\chi^2 = \sum_j \frac{|s_j - e^{i\omega_j t_0} \cdot A_1 \cdot v_j^{(1)} - e^{i\omega_j t_0} \cdot A_2 \cdot v_j^{(2)}|^2}{J_j}, \quad (2.11)$$

where we have simply added a second amplitude and template component to Eq. (2.9). Quantities $v_j^{(1)}$ and $v_j^{(2)}$ are templates that correspond to the mean fast and slow components of the pulse shape. The corresponding amplitudes, A_1 and A_2 , contain predominantly position and energy information, respectively. The fast amplitudes, in the past, have been used to construct parameters to fiducialize the detector [94].

In Fig. 2.10, I show an example event from a detector with four channels. The slow and fast components are shown in cyan and pink. The pulse in the channel closest to the location of the event has the fastest rise time, because it has the highest fraction of prompt phonons. The ballistic phonons bring most of the energy information to the sensors and therefore shows a roughly similar shape across all four channels. This particular event happened closest to channels C and A, as indicated by the sharp rise time on these channels. The sign of the fast component is indicative of the closeness of the event to a particular channel. Negative fast amplitudes usually indicate events that are farther from the channel, as is evident for channels B and D in Fig. 2.10.

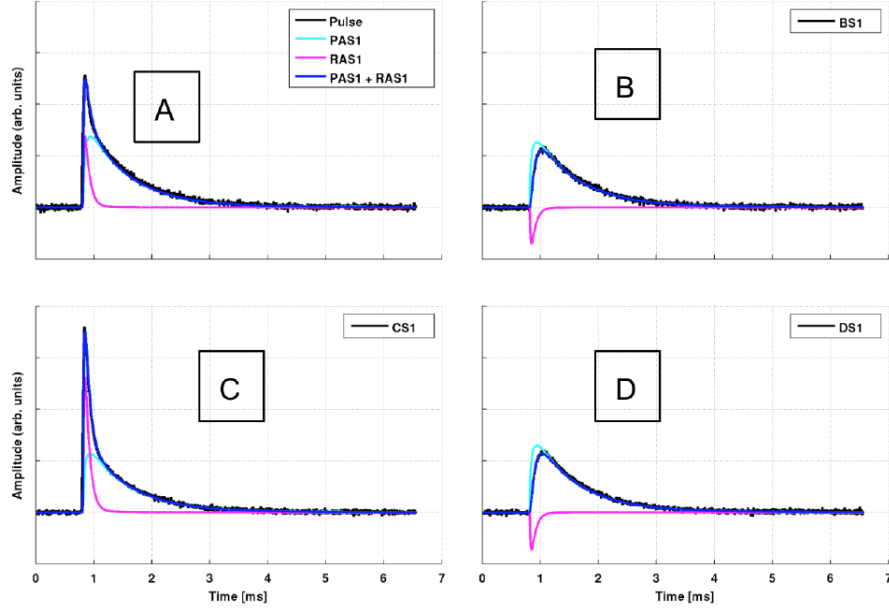


Figure 2.10: Pulses for each channel from an example event decomposed into prompt (fast) and ballistic (slow) components. The pulse is shown in black, the fast and slow components obtained from the 1×2 fit are shown in pink and cyan respectively. The sum of the fast and slow component is shown in blue, which looks like a good fit to the pulse in all four channels. Figure taken from Ref. [76].

A third energy estimator of note is the Non-Stationary Optimal Filter (NSOF). The stationary noise assumption is what allowed us to write the denominator in Eq. (2.9) with simply the noise PSD or the diagonal elements of the covariance matrix. But what if we considered the position dependence business to be a form of noise due to unmodelled variations in the template's pulse shape, which can be absorbed into the covariance matrix? This is the idea behind the NSOF,

where the position dependence is accounted for as non-stationary noise, which makes the frequency domain covariance matrix, V , non-diagonal. The χ^2 to be minimized for this redefined problem becomes

$$\chi^2 = \sum_{j,k} [s_j - e^{i\omega_j t_0} \cdot A \cdot v_j]^\dagger V_{j,k}^{-1} [s_k - e^{i\omega_k t_0} \cdot A \cdot v_k], \quad (2.12)$$

where we have replaced the denominator with a full covariance matrix whose elements now describe correlations introduced from position dependence and $\dagger$ denotes complex conjugate.

It is important to note that while both the 1×2 OF and NSOF correct for position dependence, neither of them correct for correlated noise between channels. As you will see in Ch. 3, correlated noise produces a significant smearing of the estimated baseline resolution and correcting for this is essential for the experiment to achieve its required performance level. A big part of this thesis will be dedicated to novel reconstruction techniques, which not only address the issues of correlated noise and position dependence, but also address reconstruction of hybrid data discussed at the end of Sec. 2.3.

2.5 Detector calibration and dark matter sensitivity

Detector calibration is the process of establishing an energy scale for your device. The quantity measured by the TES is a current. Calibration is the process of converting that current into a corresponding energy. Some of the energy scales used are the total phonon energy (E_t), the electron equivalent energy (E_{ee}) and the nuclear recoil energy (E_{NR}), all of which were discussed in Sec. 2.1. Calibration is done via the use of one or more peaks whose energies are known a priori.

SuperCDMS SNOLAB will use two main calibration strategies: neutron and gamma calibration. A ^{252}Cf neutron source will be the primary means of calibrating Ge detectors in the energy range of interest. Ge undergoes neutron capture to produce an isotope, ^{71}Ge , which decays via electron capture to ^{71}Ga , releasing Auger electrons and X-rays with energies 10.4 keV, 1.3 keV and 0.16 keV [95].

HV detectors are especially sensitive to lower energies and could potentially show features in the energy spectra of electron recoils from γ -ray scatters. These correspond to the binding energies of atomic electrons. Traditional Compton scattering refers to the scattering of high-energy photons off free electrons, but for low-energy scatters off atomic electrons the binding energies of the electrons may not be negligible. These features are called Compton steps. A γ source could be used to knock out atomic electrons producing steps that could be measured in our detectors.

These Compton steps, and by extension their Compton edges, could be used for calibration in the < 10 keV energy range. The Ge Compton steps are expected at 11.1 keV, 1.4 keV and 1.3 keV. A few experiments have observed the steps in Ge such as CDMSLite [96]. Compared to Ge which can be calibrated with its internal neutron capture lines, Si does not have any such activation lines. Si calibration is expected to be performed, almost exclusively, with Compton steps. Similar to Ge, Si Compton steps can be produced with the γ source to give steps at 1.8 keV, 0.15 keV and 0.10 keV. Compton steps require a large amount of accumulated data to be observed. Experiments such as DAMIC [97] have spotted the steps in Si devices, which provides a positive outlook for SuperCDMS SNOLAB.

A ^{133}Ba gamma source provides multiple lines in the hundreds of keV range. ^{133}Ba decays via electron capture to give a prominent peak at 356 keV and other smaller peaks in the broad range of 80–390 keV. The 356 keV peak is often used as a calibration point at higher energies, however HV detectors saturate immensely in the 100s of keV range and therefore require additional considerations (like a saturation model). Nevertheless, one could potentially keep an eye out for a broadened high energy γ peak in HV detectors.

One could also consider calibrating Si detectors with shared gamma-ray lines. These are produced by multiple scatters in adjacent detectors, and the calibration is performed by summing the two detectors to get the total spectrum. Another possibility for calibrating Si detectors, identified from simulations recently at CUTE, is using K_α, K_β lines from the Cu detector housing (8 keV, 8.9 keV). These peaks would be dominated by events at the edge or face of the detector, which are often cut out by fiducialization², but can nevertheless provide an energy scale.

Another calibration source which has been used by SuperCDMS in the past is ^{55}Fe . This source has been successfully used in previous CUTE runs to calibrate small, single channel Si R&D³ detectors. The ^{55}Fe source provides two peaks corresponding to K_α (5.9 keV) and K_β (6.5 keV) decay. This source is, however, not very penetrating in large detectors and therefore difficult to simulate and study in practice. The Al in the QET could also produce a fluorescence peak at 1.5 keV, which is an additional low energy calibration point.

Calibration of SuperCDMS HV detectors was a high priority item at the CUTE tower testing project. Neutron and γ calibration along with potential shared line calibration options are being explored. Results from this endeavor will follow in the coming chapters.

SuperCDMS SNOLAB detectors have been designed with an average critical temperature of $T_c \sim 50$ (40) mK in Ge (Si). The reduction in critical temperature along with other factors reduces

²The process of defining a search volume within the detector

³Research & Development

the effective baseline resolution of these devices. Baseline resolution is critical in determining the dark matter sensitivity of the experiment. CDMSLite Run 3 demonstrated a baseline resolution of about 260 eV_t (or $\sim 10 \text{ eV}_{ee}$) at a voltage bias of 75 V [71]. The new devices are projected to achieve a baseline resolution of about 50 eV_t ($\sim 1 \text{ eV}_{ee}$) at 100 V . Searching for spin-independent nuclear recoil dark matter is one of the primary science goals of SuperCDMS. Figure 2.11 shows the latest sensitivity projections for SuperCDMS SNOLAB. Predictions using both Optimum Interval (OI) [98] and Profile Likelihood Ratio (PLR) methods are shown.

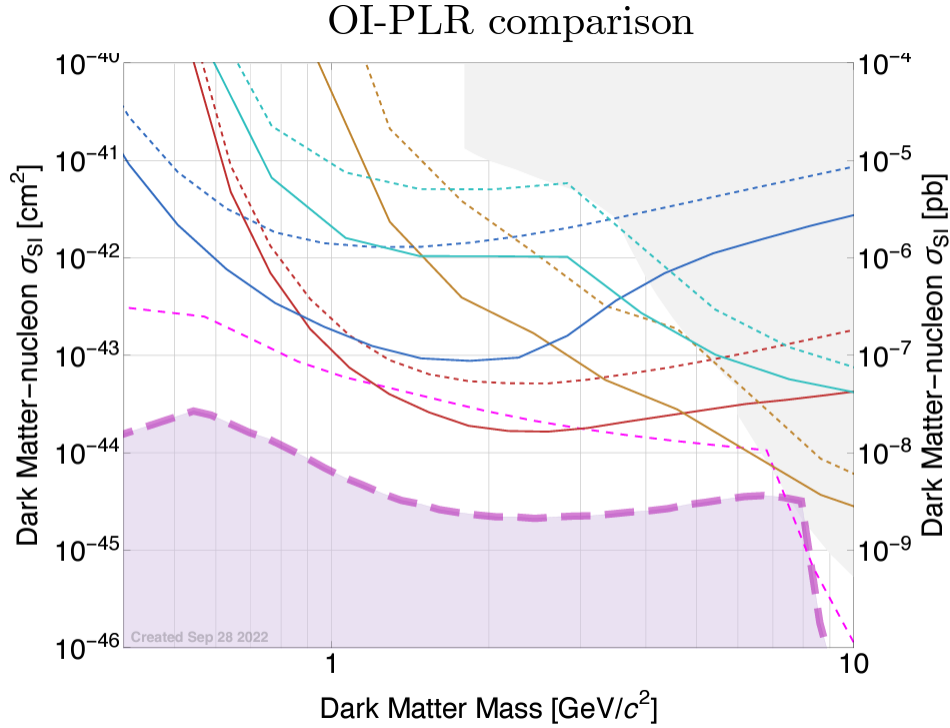


Figure 2.11: Sensitivity projections for spin-independent, nuclear recoil dark matter at SuperCDMS SNOLAB. The solid line show results from PLR and the dashed line shows OI results. The legend is as follows: **Ge HV**, **Si HV**, **Si iZIP**, **Ge iZIP**, **Neutrino fog** (shaded region) and **single neutrino sensitivity** (short dashed line). The gray shaded area marks the boundary of current limits. Figure taken from Ref. [99].

From Fig. 2.11, the SuperCDMS Si detectors clearly display an improved sensitivity at the lowest masses ($< 0.5 \text{ GeV}/c^2$), while Ge performs better in the 1 to $10 \text{ GeV}/c^2$ mass range. The advantage of achieving lower thresholds from increased NTL amplification is made explicit by the world-leading sensitivity that the HV detectors are projected to achieve. The pink region indicates

the boundary beyond which neutrinos become a persistent background. The “neutrino fog” is a theoretical lower bound below which coherent scattering of neutrinos from nuclei becomes a dominant effect that mimics WIMP signals. SuperCDMS SNOLAB also has secondary science goals of detecting solar axions, Axion-Like Particles and Lightly Ionizing Particles. In addition to nuclear recoil dark matter, Ref. [99] also contains extensive projections for these electron recoil dark matter candidates. An example sensitivity projection for electron recoiling dark matter interactions is shown in Fig. 2.12, which once again shows the ability of Si detectors to push to lower masses. The paper also details potential future upgrade scenarios under various background assumptions to demonstrate the ability of the experiment to probe the most interesting regions of the parameter space.

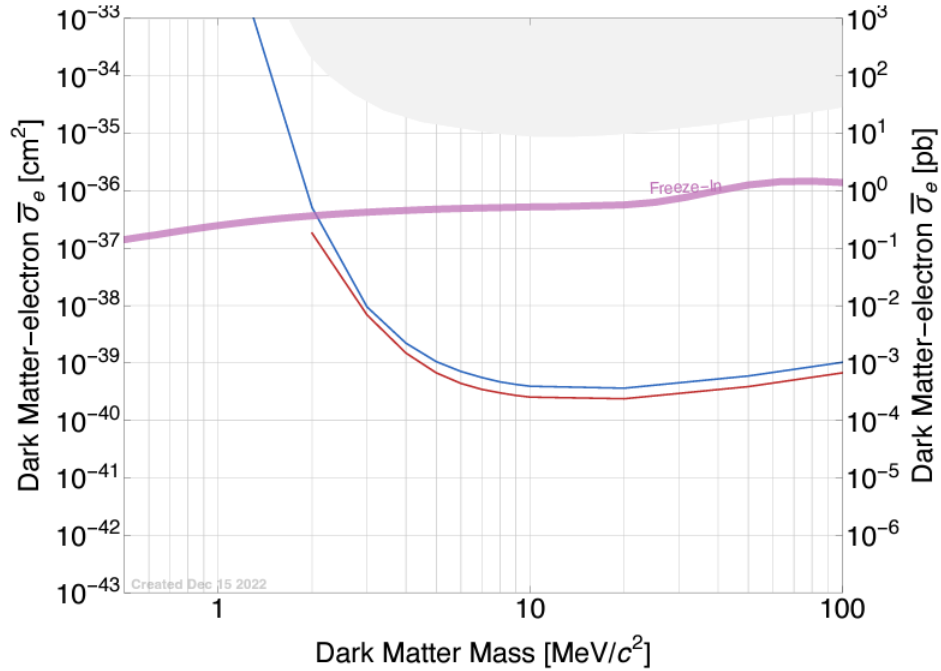


Figure 2.12: Sensitivity projections for electron recoil dark matter at SuperCDMS SNOLAB using PLR. The legend is as follows: Ge, Si. The gray shaded area marks the boundary of current limits. Figure taken from Ref. [99].

In the next chapter, I will describe the first results from operating some of these sensitive devices in a low background environment at CUTE.

Chapter 3

Characterizing SuperCDMS SNOLAB detectors

SuperCDMS SNOLAB is designed to house up to seven detector towers, each capable of hosting six SuperCDMS-style detectors. The initial payload will consist of four towers – two iZIP towers and two HV towers. This totals to twenty-four detectors in the initial payload, six of which are Si detectors and the remaining eighteen are Ge detectors. All four towers were briefly tested for basic functionality and operability at the Stanford Linear Accelerator Center (SLAC) in the summer of 2022. The testing objectives were defined with limited scope for this effort due to the high risk of extended cosmic exposure on the surface, which diminished the ability to perform meaningful science analyses with these data sets. A second testing effort was carried out for one of the detector towers, Tower 3, at the Cryogenic Underground TEst facility at SNOLAB in Sudbury, Canada from October 2023 to March 2024. This marked the first operation of the new SuperCDMS detectors in a low background environment with a broadened scope for many science analyses.

This chapter will provide an overview of the CUTE facility, major observations from testing the towers at SLAC and a select list of interesting results from testing the towers at CUTE. The author would like to emphasize that in-depth analysis of the data is still underway. The goal behind this chapter is to provide a preliminary assessment of the quality of the detectors and their performance.

3.1 The CUTE facility

The Cryogenic Underground TEst facility is a low background testing site for sensitive low energy detectors at SNOLAB. The facility is equipped with a cryogen-free dilution refrigerator designed

3.1. THE CUTE FACILITY

with specialized technology that minimizes vibrations. A dilution refrigerator cools devices to ultra-low temperatures of the order of tens of milli-Kelvin via a staged cooling mechanism that circulates a mixture of ^3He / ^4He [100]. The mixture of $^3\text{He}/^4\text{He}$ exists in two phases separated by a physical phase boundary: (a) a dilute phase that is nearly 100% ^3He , and, (b) a concentrated phase that is mostly ^4He with some ^3He . The cooling power is provided by the enthalpy of mixing of ^3He and ^4He from the concentrated to the dilute phase at the phase boundary. As the concentration of ^3He in the dilute phase increases, in order to maintain the mixing of the two phases, ^3He is regularly pumped out of the dilute phase with external pumps. This technique is able to achieve temperatures as low as 10 mK with the typical base temperature at CUTE being somewhere between 10 and 15 mK. The ^3He that is circulated is pre-cooled in stages with the help of a pulse tube cooler [101]. The use of a pulse-tube cooler introduces mechanical vibrations into the system which couple in as noise into phonon-mediated detectors. In addition to the vibration-suppression technology that comes built-in in the CUTE fridge, a special vibration isolation system is deployed at CUTE to vibrationally isolate the detectors. Details of this vibration isolation system are available in Refs. [102–104].

The CUTE facility is capable of housing a payload of up to 20 kg [103]. In the context of SuperCDMS, CUTE is capable of hosting a single SuperCDMS SNOLAB tower at a time containing six detectors. While the SNOLAB rock overburden shields against most cosmic radiation, the cavern housing the CUTE facility contains many naturally occurring radioactive sources, predominantly potassium (^{40}K), uranium and thorium. These produce γ -rays and neutrons which are the dominant backgrounds. The CUTE facility comes with many layers of shielding which control the net background that the payload is exposed to to just a few events $\cdot \text{keV}^{-1} \cdot \text{kg}^{-1} \cdot \text{day}^{-1}$ in the 10 – 100 keV energy range. The fridge is placed inside a stainless steel drywell surrounded by a layer of ultra-pure water which acts as a neutron and γ shield. In addition, two layers of low-activity and very low-activity lead separated by a layer of μ -metal are present around the cryostat providing γ and magnetic shielding, respectively. The drywell is flushed with low-radon air to prevent exposure to the high radon environment in the lab. An internal lead block is present within the cryostat to shield from vertical influx of background radiation. The drywell cavity is sealed by a High Density PolyEthelene (HDPE) shield on the top to block vertical neutron influx. The full schematic of the setup of the CUTE facility is shown in Fig. 3.1. The full schematic of the shield is shown in Fig. 3.2.

The CUTE facility also allows the use of three possible calibration sources – a ^{133}Ba gamma source, an ^{55}Fe low energy γ source (which is typically inserted internally below the vertical lead shield), and, a ^{252}Cf neutron source borrowed from SuperCDMS. While the ^{55}Fe source, once

3.1. THE CUTE FACILITY

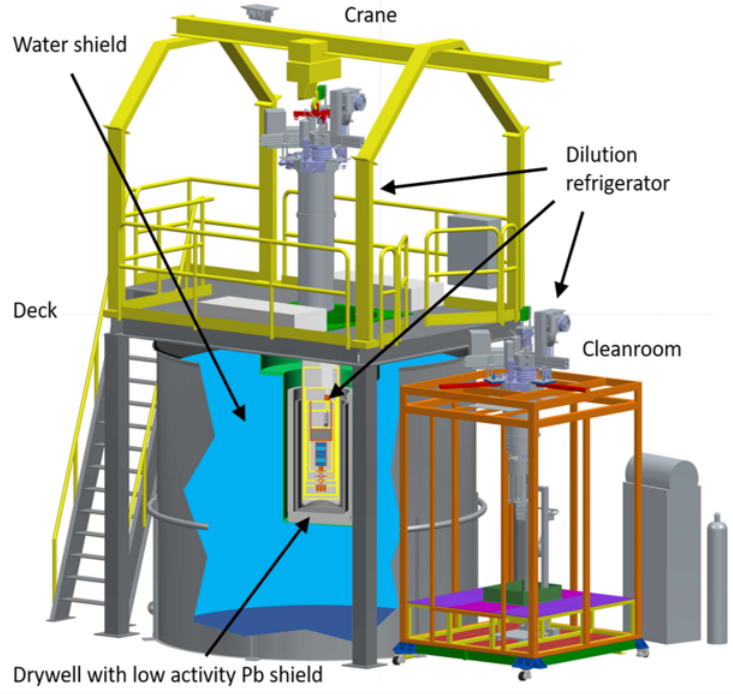


Figure 3.1: Cartoon image showing the design of the CUTE facility. The facility is equipped with a cleanroom where the devices are handled and installed in the cryostat. The cryostat and the fridge are then lifted via a crane into the drywell. The overall structure of the facility is shown. Figure taken from Ref. [103].

installed, can only be removed after the end of the run, the Ba source can be deployed and controlled remotely with the help of a stepper motor. The neutron source is deployed by hand with the help of an on-site shift operator. The Ba calibration source is guided into the space between the internal lead shield and the cryostat can with copper tubes as shown in Fig. 3.3. The Ba source produces γ s of varying energy, the most prominent of which is seen as a peak at 356 keV in the energy spectrum of our detectors. The ^{55}Fe source produces two calibration peaks at 5.9 keV and 6.5 keV, the presence of both of which has been confirmed in the data obtained from many detectors tested at CUTE. Additionally, the ejection of 1s electrons from the Al fins in the QETs causing the drop of a 2p electron to fill the gap can release photons of 1.5 keV in our detectors. Signs of these peaks have been seen during the testing of small R&D devices at CUTE. Neutron calibration is primarily intended for Ge detectors. More details will follow in subsequent sections of this chapter.

Several SuperCDMS devices have been tested at CUTE over the past six years. This includes

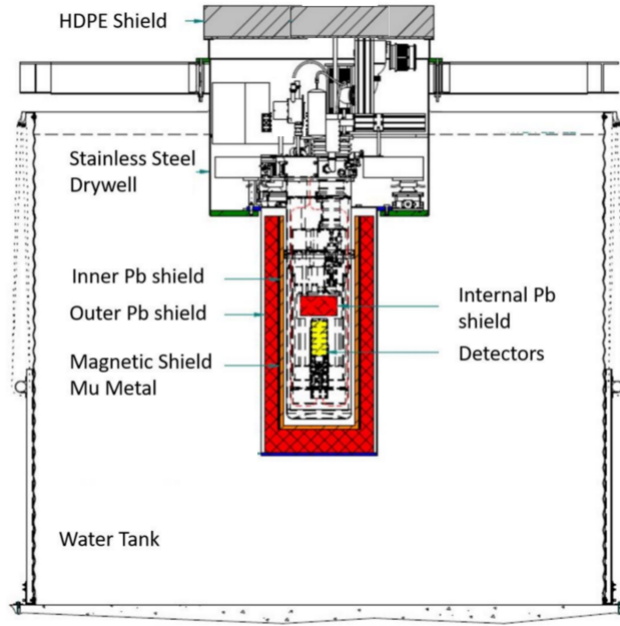


Figure 3.2: Schematic of the CUTE facility’s shields. The various layers of shielding are shown with respect to the cryostat in the middle. The payload is installed in the region colored in yellow. The outermost structure, the water tank, is roughly 3.6 m in diameter.

testing of SuperCDMS Soudan detectors (iZIP), prototype SuperCDMS SNOLAB detectors (HV), smaller R&D devices and, more recently, a full stack of SuperCDMS SNOLAB HV detectors.

3.2 Tower testing at SLAC

The Stanford Linear Accelerator Center is home to a test site equipped with a Bluefors dilution refrigerator also capable of housing a single SuperCDMS-style tower containing six detectors. The facility comes with the capability to install the same readout electronics that will be used in the main experiment at SNOLAB. Testing of all four SuperCDMS SNOLAB towers was performed at this site over the span of four runs, each spanning a timescale of a week in the summer of 2022. The fridge and the detectors installed are shown in Fig. 3.4.

The detectors were tested for basic functionality and operability. The phonon channels are read out with TESs (see Sec. 2.1) while the ionization signal is read out by collecting charges drifted to the ends of the detector with High Electron Mobility Transistors (HEMTs). An array of SQUIDs is used to read out the response of each phonon channel corresponding to a thermal incursion on

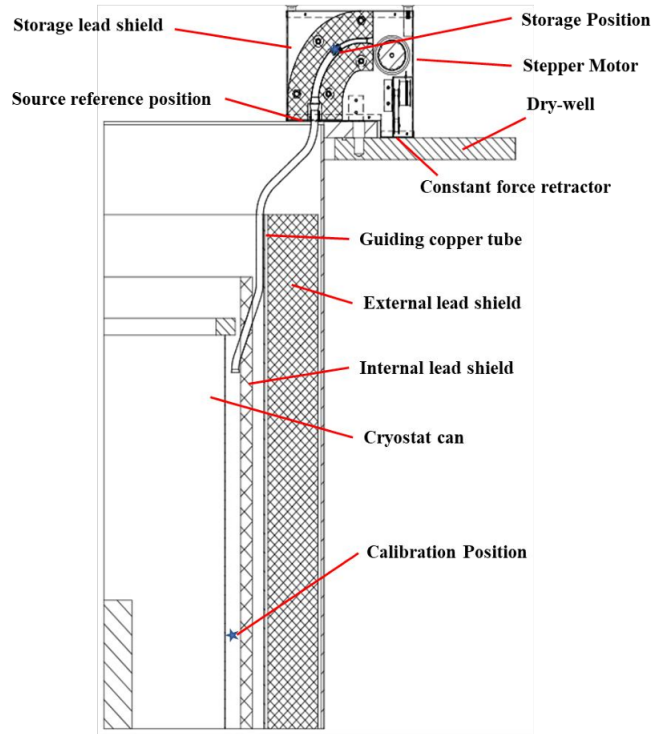
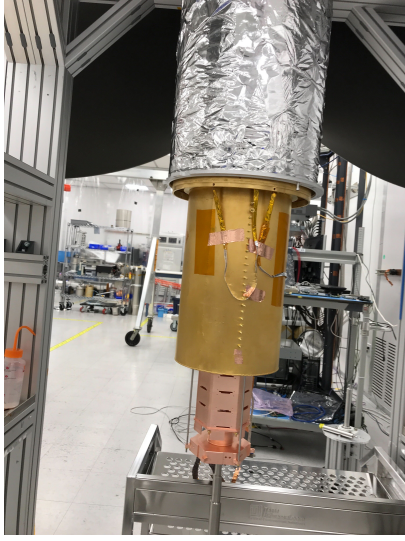


Figure 3.3: Schematic of the Ba source deployment system at CUTE. The source is guided into the space between the cryostat cans and the internal lead with copper tubes. A stepper motor is used to remotely control the location of the source.

the TES. The SQUIDs, made of niobium, attain superconductivity near 1 K. The application of a triangle wave to the input circuit shows up as a distinctive oscillating curve in the SQUID response. SQUIDs detect a change in magnetic flux as a change in voltage. A typical “good” voltage response curve to the passage of a triangle wave is shown in Fig. 3.5. The bias point of the SQUID is set at that value that gives maximum change in voltage for small changes in the flux through the SQUID. The process of selecting this optimum bias point for the SQUID is called “SQUID tuning”.

The second step after SQUID tuning is to assess the performance of the TES itself and study properties of the detector. Properties such as the critical temperature and operational point of the TES have significant effects on the energy resolution and sensitivity of the detectors. The main properties measured are the resistances: R_p , the parasitic resistance of the circuit measured in superconducting state; R_n , the TES resistance in normal state; and the TES critical temperature, T_c , which marks the transition point into the superconducting state (see Fig. 3.6). The signal

3.2. TOWER TESTING AT SLAC



(a) The Bluefors fridge at SLAC being prepared to install the detectors.



(b) The detector tower installed into the Bluefors fridge.

Figure 3.4: Images of the Bluefors fridge at SLAC with the detector tower being installed. Both images were taken by on-site shifters at SLAC.

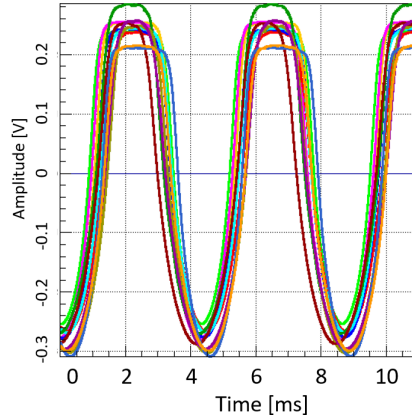


Figure 3.5: An example SQUID pattern for an iZIP detector obtained from tower testing at SLAC. Each curve is for a single channel.

energy collected is physically the heat lost by Joule heating in the TES. Thus, the TES resistances factor into signal calibration. The T_c is a quality factor fixed by design but crucial to understanding our expected energy resolution. This is because energy resolution scales as T_c^3 . The goal for

SuperCDMS SNOLAB was to restrict $T_c < 60$ mK by design. The TES resistances are calculated via a measurement we call “ $I_b I_s$ ”. The process involves measuring the current through the TES at various bias voltages. By sending either a slow triangular test signal to change the bias voltage or manually changing the operational bias in steps using the DCRC, one can estimate the current passing through the TES and relate it to the TES resistance using simple circuit equations [83] for the circuit shown in Fig. 3.6:

$$I_s = \frac{R_s}{R_s + R_{\text{TES}}} I_b. \quad (3.1)$$

R_s is a known shunt resistance in the circuit, I_s is the TES current and I_b is the applied external bias current. For a little context, zero bias voltage implies the TES is in its superconducting state and at ~ 200 (~ 60) μA , the Ge (Si) TES is in its normal state. The TES resistance has two components as previously mentioned: $R_{\text{TES}} = R_p + R_{\text{transition}}$. In the superconducting state, $R_{\text{transition}} = 0$ and so the R_{TES} measured from the procedure mentioned above is equal to R_p . By placing the TES in its normal state by increasing its bias current after R_p is measured, one can measure the TES normal resistance, $R_n = R_{\text{transition}}$, using the same equation. Once the TES resistances are known the IV (current versus voltage) curves can be used to map the entire resistance profile of the TES at different points in transition. This allows us to determine the exact 50% R_n operational bias value for the TES. The $I_b I_s$ measurement is typically performed via automation using the data acquisition system. This code has been expanded to include calculation of power loss at the TES and current offsets, with detailed tabulation of each variable at selected bias values. Simpler versions of resistance measurements are also utilized at discretion, where one applies a time varying voltage signal to the TES circuit, measures the amplitude of the current and divides the input voltage by measured current to get the resistance. These methods, however, are exclusively resistance measurements and cannot be extrapolated to obtain other useful parameters, like with $I_b I_s$.

The following quantities were determined as part of the testing procedure:

- the parasitic resistance (R_p) and the normal resistance (R_n) of each TES,
- the bias profile of the TES from $I_b I_s$ measurements, and,
- the critical temperature (T_c) of the TES.

The expected values for the first and second in the list were set by design and act as a key performance indicator for the device. The parasitic and normal resistances were designed to be under 7.5 m Ω and 100–150 m Ω , respectively. The average values of R_p and R_n with the standard deviation for each crystal type are shown in Table 3.1. All measured parasitic resistances exceeded

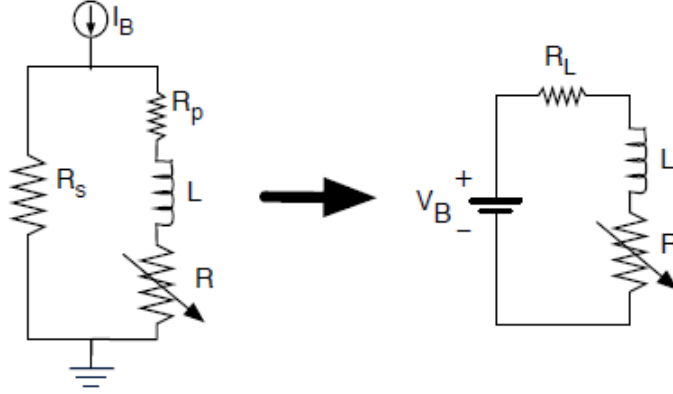


Figure 3.6: The TES sensor circuit which is current biased (left) and the Thevenin equivalent voltage circuit (right). Figure taken from Ref. [83].

expectations on a channel-by-channel basis, showing values near or below 1 m Ω . Similarly the normal resistances also exceeded requirements showing mean values less than the required 100 m Ω .

Detector type	R_p (m Ω)	R_n (m Ω)
Si	0.79 ± 0.21	86.29 ± 13.92
Ge	0.69 ± 0.15	93.90 ± 14.42

Table 3.1: The measured mean resistance parameters of interest for SuperCDMS SNOLAB detectors. The values were averaged across all detectors in all four towers that were tested at SLAC.

The bias profile determines the ideal operation point of the TES, although sometimes this is used in conjunction with an analysis of the signal to noise ratio. The typical TES operation point is the bias current that holds the TES in transition at 50% R_n . This number is tied to the critical temperature of the TES, which is estimated by measuring the resistance of the TES as a function of its temperature. The temperature of the system is raised in steps using an external heater, measuring R_{TES} at each step. The critical temperature is measured as the point where the TES resistance slides from transition to normal. Due to limitations in increasing temperature in tiny steps, the measured T_c from this technique is not exact although it comes close to the true T_c of the device. The mean T_c for each SuperCDMS SNOLAB detector is shown in Fig. 3.7. One detector in Tower 4 was not measured at the time due to schedule limitations and therefore measurements for 23 out of 24

3.2. TOWER TESTING AT SLAC

detectors are shown in the plot. One of the Ge detectors in Tower 4 measured a lower T_c on par with the Si detectors.

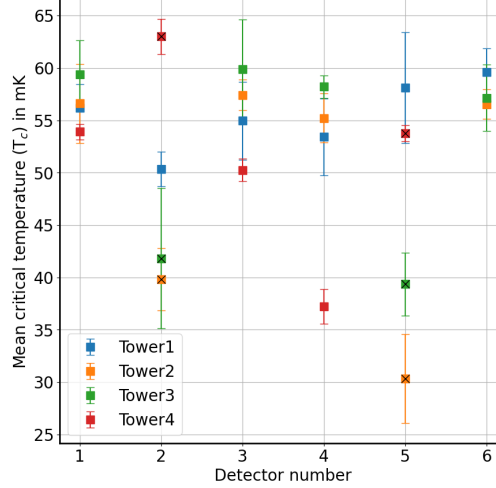


Figure 3.7: A comparison of the measured critical temperatures of SuperCDMS SNOLAB detectors. Most Si detectors marked with an ‘x’ show low T_c ’s consistent with design, with the exception being Tower 4 Si detectors. Ge detectors were seen to show higher T_c on average, with the exception of detector 4 in Tower 4.

The low critical temperature of the Si TESs means that lower resistive heating is required to hold the TES in transition which reduces the operation point of these detectors. This results in bias powers of the order of just a few picowatts in Si, contrary to bias powers of the order of a few tens of picowatts for Ge detectors. The delicacy of the TES design ought to be highlighted, since any higher T_c would have pushed the operation points closer to that of previous generation detectors, reducing the net gain in dark matter sensitivity achievable from the new design. On the other hand, any lower T_c would have made it tedious to cool the detectors down to superconducting temperatures using the available cryogenic systems at SLAC and CUTE.

Operation of these detectors at the surface facility was kept to a minimum since it poses the risk of exposure to cosmic radiation which produce radioactive isotopes in the detector materials. At present, the largest source of background for the experiment is expected to be tritium (^3H) and, particularly for Si detectors, the naturally occurring isotope of Si, ^{32}Si [90]. Exposure to high energy cosmic-ray secondaries, such as neutrons, protons or muons, produce tritium as a spallation product in the material [105]. ^{32}Si , on the other hand, is endemic to Si and produced as a spallation product from cosmic secondaries produced from atmospheric argon [106]. Further, the installation

and extraction procedures for the towers risk dust and other accumulative contamination (^{238}U , ^{232}Th and ^{40}K) every time.

The large background also meant that we saw significantly large number of events, marked by a high trigger rate. This produced many unwanted effects, both technical and physical, which limited the diversity of tasks that could be accomplished at this site. The readout typically saw rates of the order of 10 Hz per detector. The DAQ readout bandwidth is capable of handling much less, which resulted in what we call “stale triggers”. This simply refers to triggers being lost due to the large event rate combined with limited bandwidth. This often resulted in large effective “dead times” (times of no DAQ triggers) and an inability to acquire meaningful triggered data.

Furthermore the lack of any shielding except for a simple layer of copper permits a large rate of highly energetic muon events. These often partially or fully saturate the TES and produce poor quality traces such as pile-up events (events with multiple pulses in a single trace), pulses with very long fall times (times that are not entirely captured in a single trace) and pulses that sit on the falling tail of another pulse. Examples of such pulses are shown in Fig. 3.8. Note that tracelength had to be kept short during these tower testing runs, half the full length to be exact, due to the large event rates in order to reduce the high pile-up event rate.

The iZIP detectors were only tested at SLAC and therefore, any inspection of the charge channels is only possible from this data. Example charge pulses are shown in Fig. 3.9. While phonons propagate at the speed of sound, electrons and holes propagate much faster. This causes the latter to have sharper pulses compared to the former. Design flaws in the readout electronics produced oscillations in the charge channel baselines and coupling of noise from the charge channels into the phonon channels, degrading performance when operating a full detector. The tower testing effort provided an opportunity to identify such issues and promptly address them prior to the main experiment. Despite exposure, background and hardware limitations, the project helped demonstrate the operability and quality of these devices in terms of key performance parameters.

Testing a tower at the CUTE facility was a means to understand the energy scale and sensitivity of the detectors. The major highlights from this effort are described in the next section.

3.3 Tower testing at CUTE

A single SuperCDMS detector tower was tested at CUTE – the first operation of this tower in a low background environment. The proximity of the CUTE facility to the SuperCDMS site at SNOLAB implies that the cavern backgrounds in both cases are bound to be similar, although the difference in shielding between CUTE and SuperCDMS does mean that the former will see more backgrounds

3.3. TOWER TESTING AT CUTE

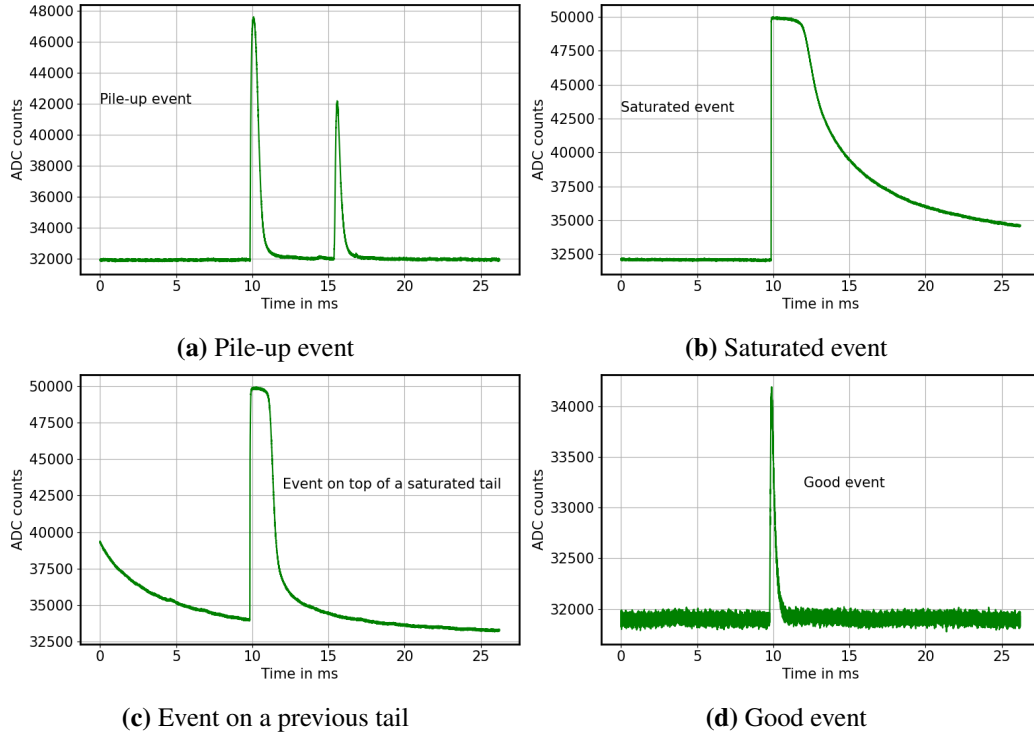


Figure 3.8: Examples of a few types of events observed during tower testing at SLAC. Example of (a) a pile-up event, (b) a saturated event with tails that go beyond the recorded trace, (c) an event that sits on the tail of a previous energy deposition, and (d) a typical “good” event are shown. Although events of type (d) are seen in the data set, the lack of sufficient shielding on the surface resulted in an elevated rate of events of the types (a), (b) and (c). This limited the data quality and the amount of useful data acquired within the short run times at SLAC.

in the detector than the latter. Tower 3, an HV tower, was chosen due to the greater opportunity for competitive science offered by the lower threshold of these detectors. The detectors were tested over a period of a little over five months, commencing in October 2023 and ending in early March 2024. The full timetable of activities that took place during this international effort is shown in Fig. 3.10.

In Fig. 3.10, various sub-activities are shown, such as preliminary tuning of the detectors, application of high voltage and acquisition of calibration data with periods of neutron activation to enable these. The color gradients in the legend show the range of values or proportion of effort on that topic over that period of time. Of the total run time, we estimate about 117 days at a base

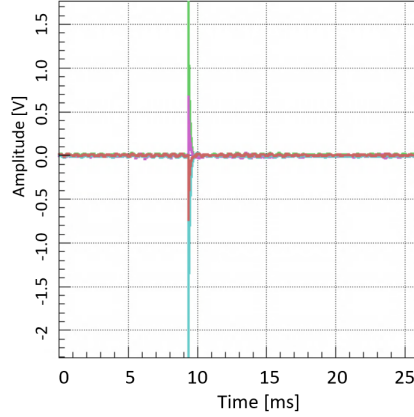


Figure 3.9: Example charge channel pulses for a SuperCDMS SNOLAB iZIP detector. The sides are biased at a small voltage on either side, positive on one side and negative on the other side. Depending on the polarity, the pulses can be either positive or negative.

temperature of ~ 13 mK, which is the amount of time that was available for useful data taking. A total of over 400 detector-days of data was acquired over the course of the testing effort, which includes about 208 detector-days of calibration data. A total of 109.2 detector-days of triggered data without any calibration sources was also acquired to assess background rates and potentially perform dark matter searches and/or sensitivity studies. The detailed amount of data collected for each detector type is shown in Table 3.2.

Detector type	Low background (detector-days)	Calibration (detector-days)
Si	44.8	58.5
Ge	64.4	40.5

Table 3.2: Total exposure for each detector type for the collected data in units of detector-days during CUTE tower testing.

Installation of the tower into the CUTE fridge involved several steps. The tower was removed from the transport containers and prepared in the SuperCDMS low-radon cleanroom as a background control measure. The prepared tower was transported on a cart from the SuperCDMS cleanroom to the CUTE cleanroom where it was installed into the CUTE fridge. The readout cables and other services were connected within the CUTE clean room and the fridge was closed up. The fridge was then lifted with the overhead crane and into the CUTE drywell following which necessary connections to the DAQ and slow control systems were made. After all connections were

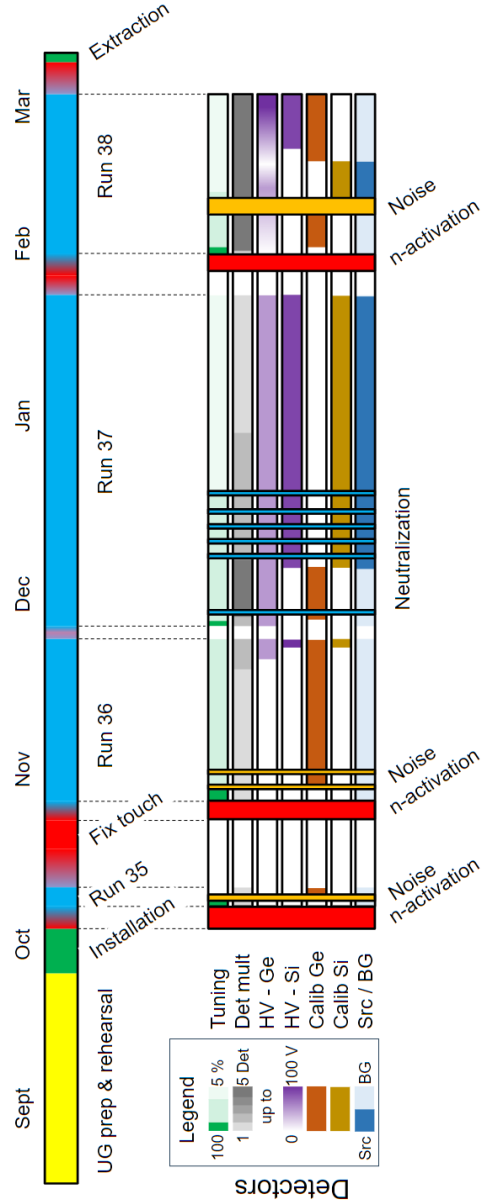


Figure 3.10: Timeline for tower testing. On the left bar, periods of installation and removal are marked in green, blue indicates periods at base temperature and red indicates warming periods that typically end a run. The right side shows details of activities during each cold period, as described in the legend. “Tuning” refers to optimizing the detector’s operation points, “det-mult” refers to data taken with multiple detectors, “HV-Ge” and “HV-Si” are operations of the Ge and Si detectors at high voltage, “Calib” refers to calibration data taking and “Src/ BG” refers to data taken with a source. Image courtesy: Wolfgang Rau.

made, the HDPE lid on top of the drywell was closed and locked in place. All shifters were off deck when taking science data and WiFi and Bluetooth are avoided when working near the DCRCs on the deck. The presence of people or disturbances on the deck can couple into the readout as vibrational noise at low frequencies where our signal-to-noise ratio is high. Presence of WiFi and Bluetooth near the DCRCs has been picked up in the past in the readout as well.

Figure 3.11 shows the installation procedure in a few images capturing the key points for interested readers. The diligent installation of the tower was a week-long process that was preceded by a preparation phase to uncover any gaps in the procedure that potentially risk either damage to the tower or unwanted contaminant exposure. The first cooldown of the fridge was achieved in October 2023, with all six detectors entering the superconducting state almost immediately. This first run was termed Run 35, which was cut short by the need to fix thermal links between the 4 K and 60 K stages of the fridge that caused an unstable operation condition for the detectors. The run was resumed as Run 36, and progressed steadily until late December, when a warm-up was prompted by a power outage at SNOLAB resulting in a loss of compressed air which affected fridge operations. This was resolved quickly, requiring only a partial warm-up and marked the beginning of Run 37 which continued until February 2024. Another quick warm-up was performed to clean impurities from the circulation lines in the fridge that caused pressure instability, marking the end of Run 37. The final run, Run 38, continued till early March with the tower testing concluding in the week of March 11th 2024. The major results from this effort are summarized below. Most of the work in this section are analyses in progress and do not represent final results from the SuperCDMS collaboration. In the following sections, I will use the terms T3Z# and Z# interchangeably to denote detector number.

3.3.1 Detector characterization and functionality

The first step taken was to ensure reproducibility of the results obtained for this tower at SLAC. This included re-measuring the TES and crystal parameters. SQUID tuning and TES characterization were performed to find the optimal bias currents for the SQUIDs and the TESs. The bias powers were found to be similar to those observed at SLAC. In fact, the detectors arrived at a lower base temperature at CUTE, and monitoring the power required to bias the TES revealed lower and lower bias power as the run progressed. Initially all detectors were operated at the 50% transition point. But the operation points of detectors two and three were later switched to 30% transition based on the signal-to-noise ratio. LEDs were used to illuminate the detector with mono-energetic photons, producing a single peak in the energy spectrum. The location of this peak divided by the baseline

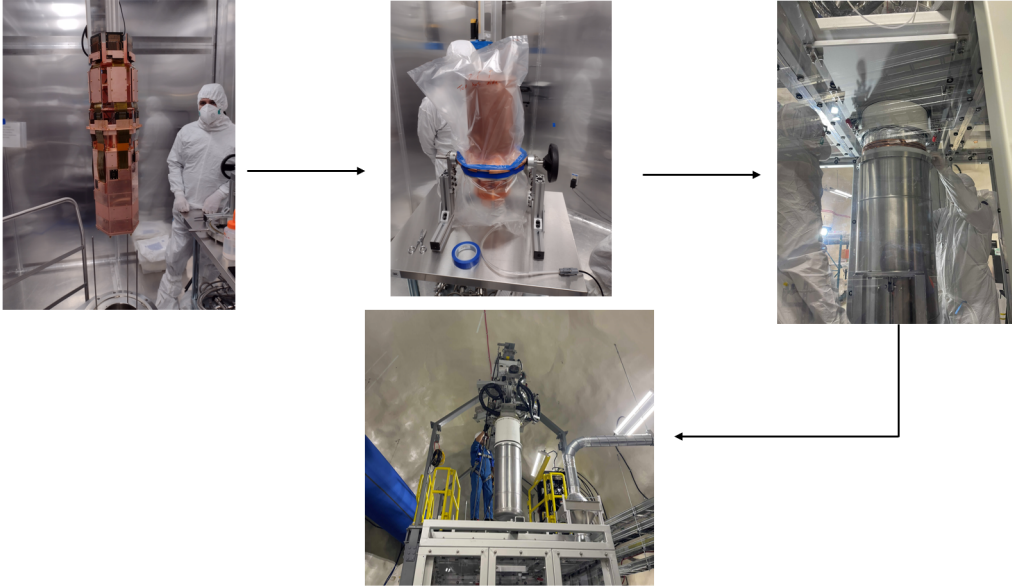


Figure 3.11: Installation of Tower 3 into CUTE. From left to right, the images show the preparation of the tower in the SuperCDMS cleanroom, its loading on a cart for transportation to the adjacent CUTE cleanroom, the installation of the tower in CUTE and final transport of the tower into the CUTE drywell using the overhead crane. In the second image, the payload is covered with a purge bag flushed with low-radon air to prevent exposure to the radon-rich air in the lab during transport. All images in this figure were taken by CUTE on-site shifters.

resolution was used as a measure of the signal-to-noise. The measurement of signal-to-noise ratio for detector 2 is shown in Table 3.3 as an example.

These optimal operation points will serve as reference values to guide detector tuning at SuperCDMS SNOLAB. The basic operability of every channel in every detector was ascertained by visual inspection. Here, “basic operability” refers to the ability to (a) observe a tunable SQUID pattern in the PulseViewer, (b) observe pulses on that channel, and (c) tune the QET bias to hold the TES in transition at a current that is similar to other channels on that detector. With this criteria in mind, three of the six detectors (detectors Z1, Z2, Z6) were found to be fully functional – all twelve channels satisfy criteria (a), (b) and (c). In both detectors Z3 and Z5, one of the outermost channels, Channel A, was found to be problematic. On Z3, this channel required a large bias current of the order of 2 mA to hold it in transition, while on Z5 this channel seemed to have an open connection. Both these observations were consistent with what was observed at SLAC. As you will see in Ch.

3.3. TOWER TESTING AT CUTE

Bias point (% of R_n)	Signal-to-noise ratio
50%	506
40%	524
35%	547
30%	596
25%	384

Table 3.3: Tabulation of signal-to-noise ratio at various bias points for detector 2 in Tower 3 which was used to determine the operational bias point of this detector. The signal-to-noise ratio was calculated as the mean amplitude of mono-energetic LED photon signals in the detector divided by the baseline resolution. The table reveals that the optimum bias point that maximizes the signal-to-noise ratio is at 30% R_n and therefore, this detector was operated at this optimal bias point. This measurement was performed by Yan Liu.

6, the loss of an outer channel has minimal impact on the reconstructed energy resolution, and therefore, we do not expect this to have a significant impact on the detectors' performance. These detectors are, therefore, still science grade detectors and perfectly usable. Z4 presents a different scenario, with four channels that were either consistently or intermittently unusable. While two of these problematic channels had connection issues, the other two displayed trouble in producing a SQUID pattern consistently. This detector was therefore not tested beyond this stage, since the lack of four channels would potentially affect reconstruction for the worse in a noticeable manner. It is to be noted that only two channels on this detector showed issues when operated at SLAC and therefore it is still possible that at least two of these channels are recoverable. We also believe that one of the connection issues may be fixable, even though this was not done at CUTE. Only for the purpose of CUTE tower testing, this detector was considered not to be science grade and left aside for the whole effort.

Another performance factor for the detector is the ability to hold high voltages. The maximum safe operation voltage by design for these detectors is 100 V with a ± 50 V limit on each side. Application of high voltages to the crystal over long periods of time can induce current flow across the crystal such that the crystal starts conducting. This is termed "breakdown" and poses a barrier to long-term operation of the detectors at high voltages. Breakdown also has the effect of producing small leakage currents that stochastically create phonons that degrade the energy resolution. The ability to operate at the highest possible voltage is critical to the science goals of the experiments, especially with regards to probing the lowest dark matter masses. All detectors, except Z4 which was not tested for HV, demonstrated the ability to operate at 100 V, with the breakdown voltage at

some point larger than 100 V. The longest tested operation time was for Z2, a Si detector, which was placed at 100 V for ~ 13.9 hours without breakdown. Breakdown is typically observed with a draw of a large current (10s of μA) from the HV power supply which is used as a way to monitor detector health. The detectors displayed good health overall when operating at high voltages, but drew gradually increasing currents of the order of a few μA as time progressed.

An isolated example of a time where a burst of events were observed accompanied with a raised current draw from the supply is shown in Fig. 3.12. The detector had been operated only for ~ 1.6 hours at 94 V before the burst occurred as an elevated number of events at that time, lasting well into the end of the run at a decreasing rate. Although this was initially suspected to be from breakdown, the presence of a stable stream of LED events between 20–30 μA after the initial burst shows that the crystal itself had not broken down. The current hypothesis to explain these events is the rare occurrence of a highly energetic muon that deposited large enough energies that a burst of events was induced briefly replicating the effect of a breakdown, but temporarily. The rarity of this event, with this being the only incident, is also supportive of this hypothesis.

The key highlight from testing HV operations is that no temperature or baseline changes were observed while ramping up the voltage. This is a sign that the application of HV does not heat up the crystal, which would have been perilous to operations.

3.3.2 Noise environment

A preliminary assessment of the noise environment was made by studying the PSDs of noise traces. Z1 showed the best transition noise performance in terms of PSD, with the lowest noise level in the low to mid frequency regions. Detectors Z2, Z3, and Z5 showed satisfactory noise levels in transition while Z6 displayed a comparatively worse noise PSD. A comparison of Z1 and Z2 noise PSDs is shown in Fig. 3.13 to provide some context on the range of values expected for typical PSDs. Acceptable transition noise levels fall in the range of 10^{-10} – 10^{-9} Amperes/ $\sqrt{\text{Hz}}$. Most of the high frequency components in the PSD are contributed by the electronics. The PSD in Fig. 3.13 below is averaged over many traces obtained by summing up the contribution of every channel.

Similar measurements of noise are performed in the superconducting and normal state of the detector as well. The PSDs are then fit with noise models to understand the contribution of sources of noise that can be modeled. This includes Johnson noise¹ from the shunt resistors and other passive elements in the circuit (passive Johnson noise), Johnson noise from the TES itself, SQUID noise and thermal fluctuation noise in the link between the TES and the cooling bath [83]. Analysis

¹Johnson noise is the thermal noise in electrical components from thermal agitation of electrons.

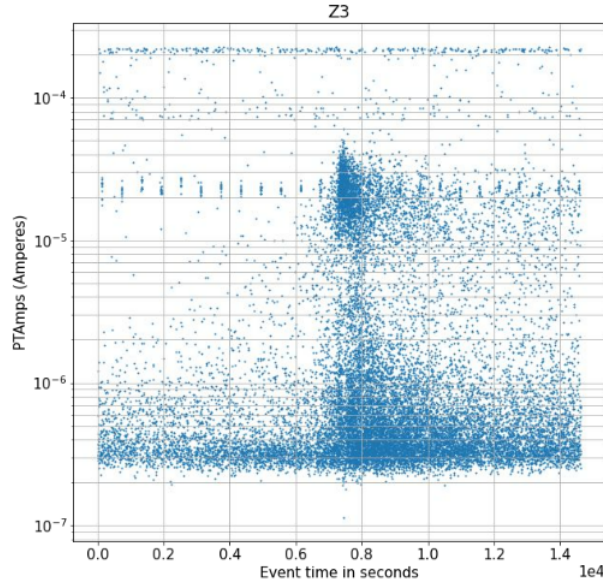


Figure 3.12: A plot showing burst events observed on a single occasion during tower testing at CUTE on Z3. The plot shows the event energy as a function of time. PTOFamps refers to the OF amplitudes (see Sec. 2.4) of the summed pulse from all channels, with PT standing for Phonon Total. The burst events were observed about 1.67 hours into the start of the run. The physical understanding of these events is a current topic of study.

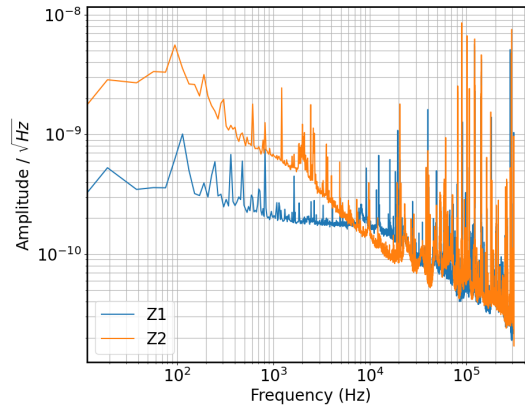


Figure 3.13: Comparison of transition noise PSDs for Z1 and Z2 in Tower 3.

is in progress to model noise on superconducting, normal and transition noise PSDs, as shown in Fig. 3.14.

3.3. TOWER TESTING AT CUTE

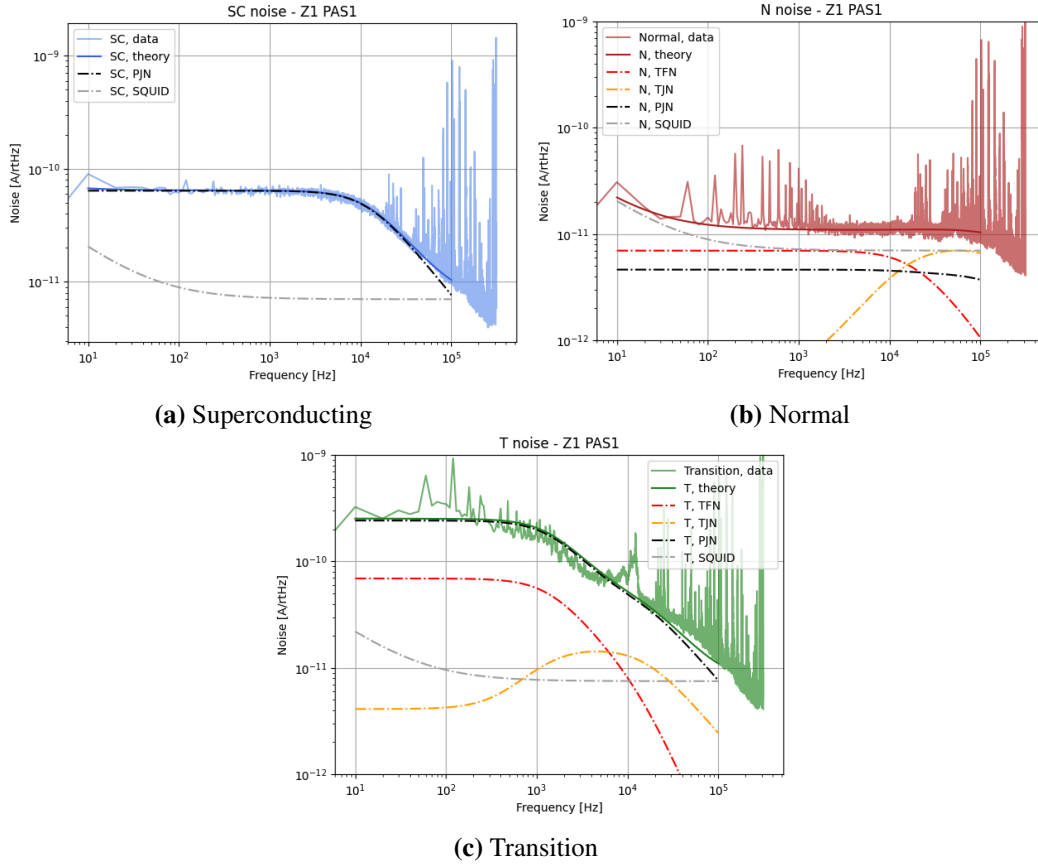


Figure 3.14: Example to demonstrate modeling of the various noise components for a single channel (PAS1) on Z1. Similar models are constructed for every channel in every detector. The theoretical models are overlaid on the measured PSD. PJN, TFN, TJN and SQUID stand for passive Johnson noise, thermal fluctuation noise, TES Johnson noise and SQUID noise, respectively. Images were produced by Madeleine Zurowski as part of her noise-modeling analysis and are used with permission.

From Fig. 3.14, it is clear that passive Johnson noise dominates in the superconducting state. The SQUID noise dominates at the lowest frequency and is a significant contributor in the normal regime. In the transition and normal regime of the TES, TES Johnson noise makes significant contributions at the highest frequencies. While the noise in the particular channel shown in Fig. 3.14 was well-described by the four sources that can be modeled sans the peaks, many other channels showed elevated noise levels that could not be explained with these four sources.

A dedicated noise-mitigation team was rapidly put together to investigate the unexpectedly

elevated noise seen at CUTE. Three broad categories of noise sources were identified by this team and are listed below.

1. Clock noise from DCRCs coupling into TES readout lines: Each side of an HV detector is read out by a single DCRC. So each HV detector requires two DCRCs and thus a total of twelve DCRCs were in use for tower testing. Each DCRC has a clock that it uses to assess timing information for every event. This means that the clocks on all the DCRCs must be synchronized with one another to ensure triggers are issued and recorded appropriately. Every DCRC comes with a “clock-in” and “clock-out” port through which they are “daisy-chained” to one another by connecting a cable from the clock-out port of one DCRC to the clock-in port of the next DCRC. Six DCRCs are chained together at a time in this fashion. The Signal Distribution Unit (SDU) is a device that acts as the leader to both the daisy chains, sending out a sync signal that aligns the clocks of all the DCRCs. Initially, the CUTE team attempted to sync the DCRCs without any SDU, i.e., by making one DCRC the leader and the rest the followers. This required detector operations shifters to sync the DCRCs from a terminal interface with the speed often depending on the efficiency of the script being used or on the shifter’s own typing and tab-shifting skills! It was found that this procedure was often not fast enough, resulting in the DCRC clocks not being synced to sufficient precision. This produced a beating effect between the clock frequencies, causing a cluster of peaks to appear in the mid-frequency region of the noise PSD, as shown in Fig. 3.15.
2. High frequency noise from the SDU: It was found that turning on the SDU radiates mid to high frequency noise producing several peaks in the 100 Hz to kHz range. In addition, the proximity of the ethernet cables from the SDU to the DCRC produces a strong peak at 1.8 kHz. A comparison of the noise PSD with the SDU on and off is shown in Fig. 3.16.
3. Noise from the HV supply: Plugging in the molex connector for the HV supply produces elevated noise at the lowest frequencies. Very often, peaks at 60 Hz and its harmonics are seen in the PSD through coupling from the standard power supply. This noise source is sub-dominant compared to the other two peaks.

These issues ultimately limit our ability to acquire high quality data. Operating multiple DCRCs simultaneously proved to be difficult owing to the above described issues. The collaboration hopes to address all three listed sources of noise in time for the main experiment. The design of a filter board is underway at SLAC, to serve as a barrier between the VIB² and the read-

²Vacuum Interface Board, where DCRC meets the readout wiring from the detectors.

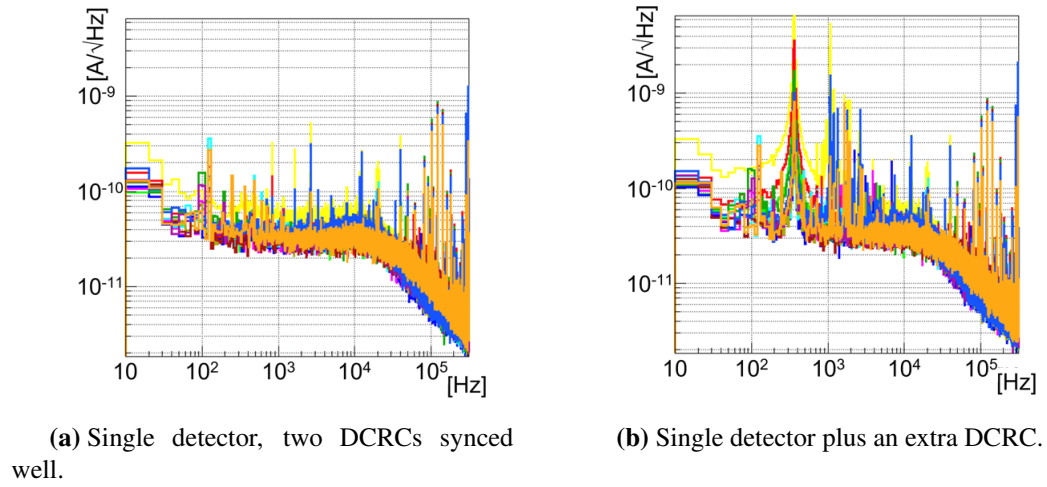


Figure 3.15: Comparison of noise PSDs between operating a single detector with two well-synced DCRCs and a single detector with an extra DCRC, synced by hand.

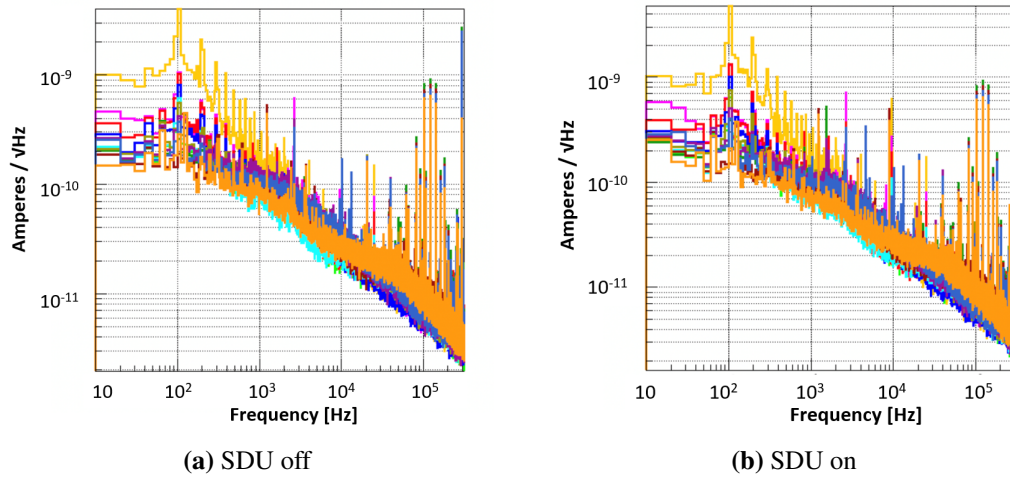


Figure 3.16: A comparison of the PSDs observed during CUTE tower testing when the SDU was on and off. Turning on the SDU causes the DCRCs to pick up noise in the Hz to kHz range.

out wiring, to avoid coupling of the clock noise into the readout. Options to physically isolate the SDU from the e-tank (the structure where the readout electronics for SuperCDMS will be placed) are being explored in addition to using shielded ethernet cables in the daisy chain.

The HV, on the other hand, is applied as a positive bias on one side and a negative bias on the other side with respect to the global ground (the chassis of the VIB). A 10 k Ω resistor exists between the ground and the 0 V point in the circuit. Replacing this 10 k Ω resistor with a 0 Ω one was found effective, during tower testing, in reducing the HV induced noise that was seen. Figure 3.17 provides a visual aid for the circuit.

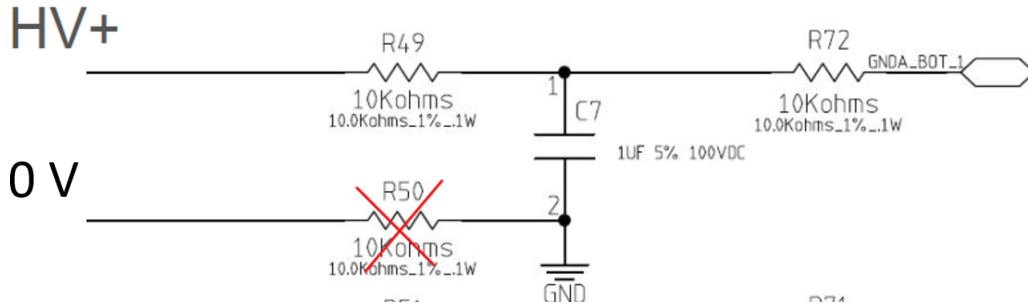


Figure 3.17: A cropped out circuit schematic of the VIB showing the arrangement of the HV circuit. The replacement of the resistor marked in the image with a 0 Ω equivalent proved to be effective in addressing HV-supply-induced noise.

Another prevalent (but easily removed) contamination in a PSD is a randomly coincident pulse. The noise data are randomly recorded traces, read out at a user-defined rate that is typically a couple of Hz or less. So it is naturally expected that a phonon pulse initiated from a particle interaction could coincidentally occur in the moment of acquisition. The pulses produce a broadband increase in the PSD level, of the form shown in Fig. 3.18. These pulses are not a source of noise per se, but bias our interpretation of the actual noise in the system. The particular example shown in Fig. 3.18 was a study performed to understand our leakage rate when applying high voltage. As the voltage is increased gradually from zero, charge carriers are injected in the crystal (termed leakage) producing trains of minuscule pulses that contaminate the PSD. Higher voltages can also liberate loosely bound electrons in the crystal and drift them. The band of PSDs at voltages between 0 and ± 50 V are at the nominal level for this detector. Increasing the voltage to or beyond ± 50 V causes leakage to become dominant for this detector, primarily contributed from loosely bound charges getting ejected and drifted. The tiny pulses produced from them create an elevated PSD. Pre-biasing is a technique that helps overcome leakage currents such as this. The voltage is ramped

up to a higher value, ± 60 V in the figure, and held here for 10 mins or so in order to bring the detector to equilibrium. Decreasing of the voltage after pre-biasing then places the detector in a stable, leakage-free state that is free of these pulses. Thus, at voltages of ± 55 V and ± 50 V, where leakage was previously seen, we no longer see an elevated PSD after pre-biasing. Pre-biasing has also been found to enhance the breakdown voltage for this type of device [90].

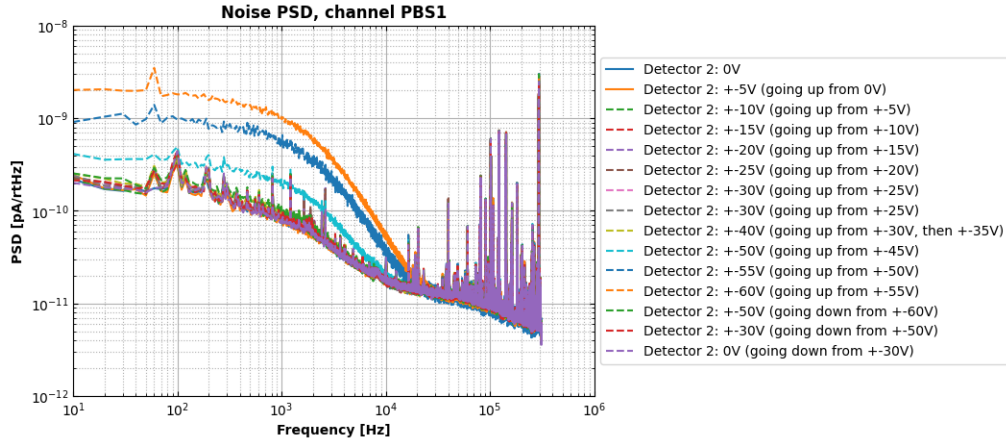


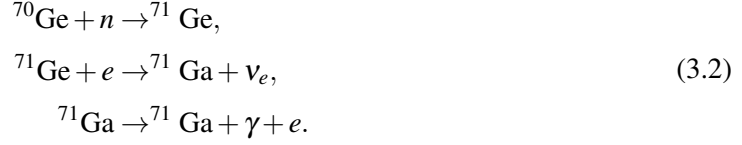
Figure 3.18: Comparison of noise PSD at various voltages for a single channel (PBS1) in detector Z2. Image courtesy: Joseph Mammo.

3.3.3 Detector calibration

A large part of the tower testing runs was dedicated to calibrating the detectors. About 99 detector-days of calibration data were acquired over the four tower testing runs at CUTE. The two detector types require different types of calibration, as mentioned earlier.

3.3.3.1 Calibration preliminaries

The primary mode of calibration for Ge detectors is via internal activation peaks. Controlled exposure to a neutron source, such as ^{252}Cf , which produces neutrons in the MeV scale, creates ^{71}Ge from ^{70}Ge , which decays to ^{71}Ga via electron capture, whose subsequent de-excitation produces gammas and electrons. This decay sequence is explicitly:



${}^{71}\text{Ge}$ decays to ${}^{71}\text{Ga}$ with a half life of 11.43 days [107], which necessitated reactivating the detectors after two weeks of data taking to maintain the electron-capture rate in the crystal.

The electron capture produces peaks in the energy spectrum corresponding to the atomic shell from which the electron is captured by the nucleus. The most probable captures are from the K, L and M shells of the atom, with every subsequent outer shell requiring more data to obtain sufficient statistics to discern the peaks. The electron capture probabilities for the major peaks in the spectrum are shown in Table 3.4. To offer some perspective on the requisite amount of data, at a rate of 0.012 Hz, 11.5 hours of calibration data will provide approximately 500 K-shell events, 60 L-shell events and 10 M-shell events. The observation of the M shell is dependent on the low energy background rate and the energy scale of the device which can be tuned using NTL gain.

Peak / Shell	Peak position	Probability
K	10.37 keV _{ee}	87.6%
L	1.3 keV _{ee}	10.5%
M	0.16 keV _{ee}	1.78%
N	0.0015 keV _{ee}	0.14%

Table 3.4: Electron capture probabilities for ${}^{71}\text{Ge}$. This table is based on values from Ref. [108].

In CDMSLite (see Ch. 6), operated at a high voltage of 7 V, the position of the M shell was quite close to the analysis threshold of 70 eV_{ee}, as seen in the spectrum shown in Fig. 3.19. The SuperCDMS SNOLAB detectors, due to their ability to tolerate up to 100 V, are capable of producing an amplification of 34.33, placing the K, L and M shell peaks at a maximum energy of 356.0 keV, 44.6 keV and 5.5 keV, respectively. At these voltages, the K shell events are expected to be fully saturated and the L shell events at least partially saturated which makes reconstruction of these pulses challenging. This is not a significant worry, but would force us to rely on the M shell alone for low energy calibration at 100 V. The N shell peak, due to its low probability of occurrence, will likely not be visible even at 100 V even if the threshold allowed it. Note that at a constant decay rate of 0.012 Hz, it will take roughly 15 hours to produce a single N shell event, which will be below the trigger threshold of our DAQ.

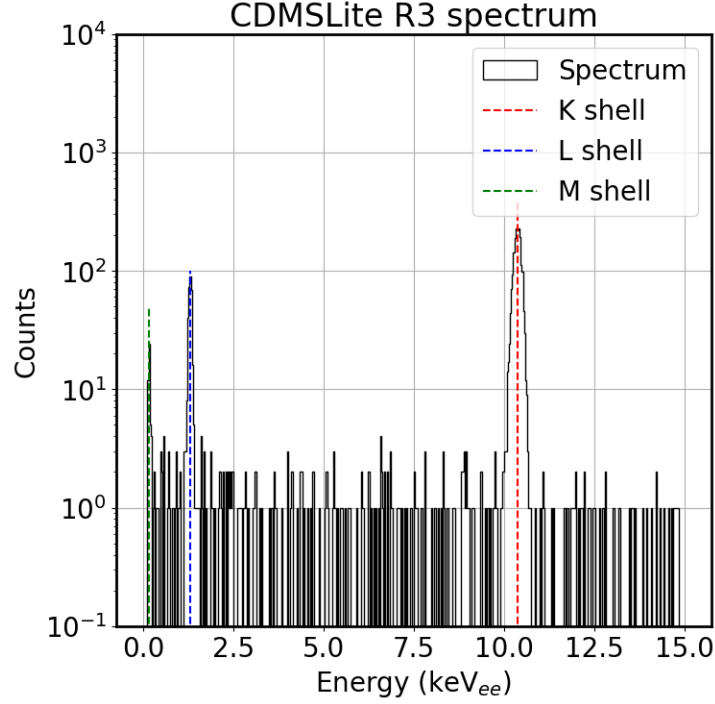


Figure 3.19: Final energy spectrum for CDMSLite Run 3. All the electron capture peaks are visible in the plot. The number of events observed was found to be consistent with the expected rate shown in Table 3.4.

Contrary to the abundant availability of low energy calibration points in Ge, Si does not possess any internal activation lines. At the lowest energies of interest, electron recoils from Compton scattering remain a strong background and a potential calibration feature. Compton scattering is the process by which an incoming photon transfers momentum to a stationary electron. The electron scatters at an angle θ relative to the incident direction of the photon, while the photon wavelength is red-shifted depending on the scattering angle as shown in Fig. 3.20. For a free electron, the energy of the scattered photon is

$$E_\gamma = \frac{E_\gamma^0}{1 + \frac{E_\gamma^0}{m_e c^2} (1 - \cos \theta)}, \quad (3.3)$$

where E_γ^0 is the energy of the incident photon and m_e is the electron mass.

The above equation defines a continuum spectrum for Compton scatters corresponding to all possible values of θ with $\theta = \pi$ corresponding to complete backscatter of the photon and maximum

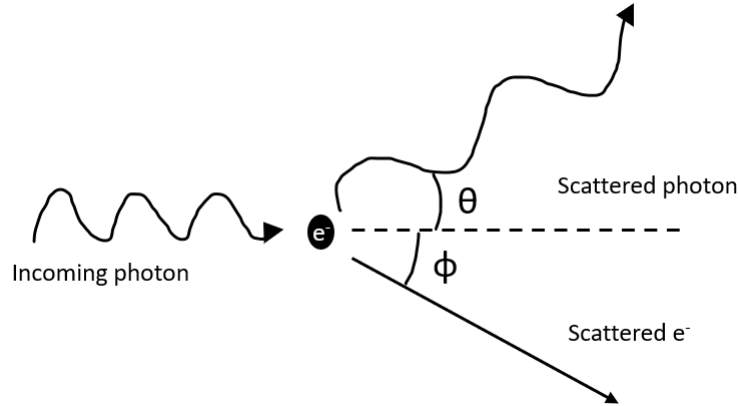


Figure 3.20: Schematic depiction of Compton scattering. An incoming photon scatters off a target electron at rest, transferring momentum to the electron. The electron scatters at an angle θ and the photon changes direction by an angle ϕ with its wavelength shifted.

energy transfer to the electron. The differential cross section for the scatter is given by the Klein-Nishina formula which gives the familiar shape of the Compton spectrum shown in Fig. 3.21a. For bound electrons, the observations deviate from the Klein-Nishina formula. The description is altered to represent the bound electrons as free electrons with a constrained momentum derived from the bound state wavefunction [109] – called the relativistic impulse approximation. With this modification, it turns out that the minimum energy required for a photon to interact is equal to the binding energy of the electron in the shell. This creates step-like features at the lower end of the Compton continuum as shown in Fig. 3.21b.

In an experimental setup, Compton steps are generated by exposing the detector to a γ source such as ^{133}Ba . Observation of steps typically require large exposures of the order of several weeks to a couple of months to generate sufficient statistics. Compton steps are present in both the Ge and Si spectra, but are hidden under the more prominent electron capture lines in Ge. Evidence of Compton steps in Ge was found in calibration data obtained from CDMSLite [96]. Compton steps will allow low energy calibration in Si and are expected to be the primary mode of calibration for Si detectors at SuperCDMS SNOLAB. Compton steps in Si have been observed by both SuperCDMS in prototype R&D devices [111] and by DAMIC (a dark matter experiment that uses skipper CCDs) [97] with several months of data. The expected locations of the steps in the spectrum are shown in Table 3.5. The height of the step or how distinguishable it will be is proportional to the number of electrons in the shell. For, example with a total of 32 electrons and 2 in the K shell, Ge would have

3.3. TOWER TESTING AT CUTE

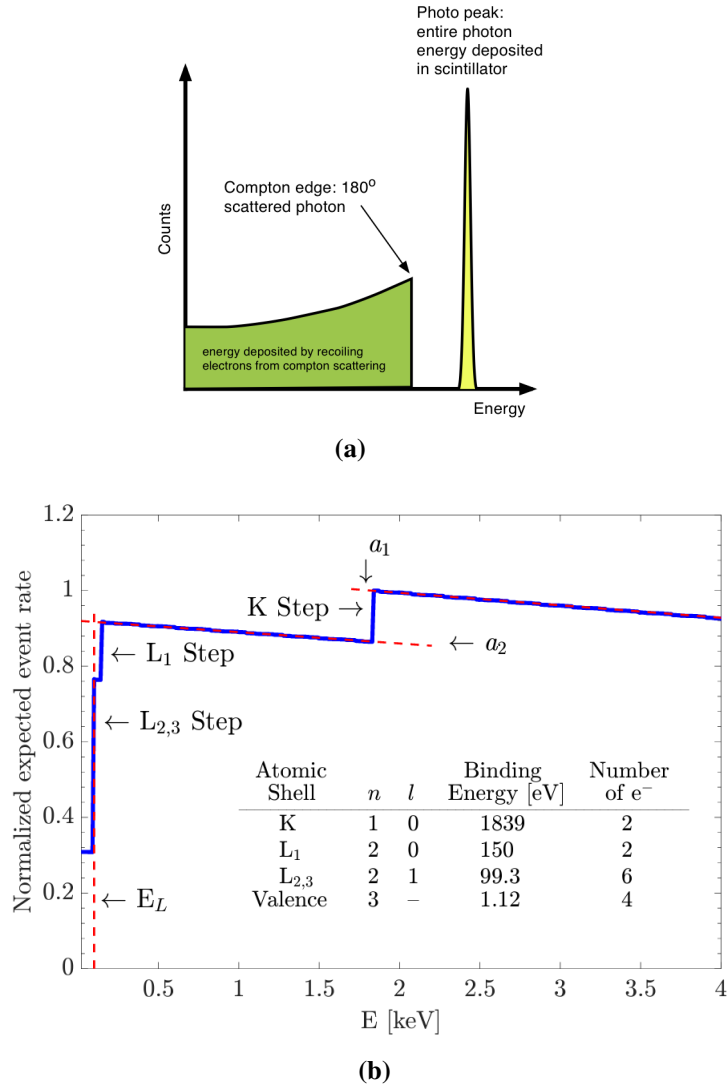


Figure 3.21: Compton spectra for a free electron and a bound electron. Atomic binding of electrons results in steps at the lower end of the Compton spectrum corresponding to the binding energies of the atomic shells. (a) Schematic of Compton spectrum for a free electron. Figure taken from Ref. [110]. (b) Compton spectrum of electrons bound in atomic shells. Figure taken from Ref. [109]. The example in this figure is for a Si spectrum which is of interest to SuperCDMS.

a K step with a height of $2/32 \times 100\%$ or 6.25%.

The potential to see single e^-/h^+ pair creation events (see Ch. 2) in large detectors such as

3.3. TOWER TESTING AT CUTE

Material	K step	L step	M step
Ge	11.100 keV _{ee}	1.260 keV _{ee}	0.130 keV _{ee}
Material	K step	L1 step	L2 step
Si	1.839 keV _{ee}	0.150 keV _{ee}	0.099 keV _{ee}

Table 3.5: Expected location of Compton steps in SuperCDMS detectors. The values for Ge were taken from Ref. [112], while the values for Si were taken from Ref. [109].

these speaks volumes to the sensitivity of the devices. The average pair creation energy in Si is 3.8 eV, which translates to a total phonon energy of 103.8 eV at an applied voltage of 100 V. This is much larger than the threshold of these Si devices (note that the L2 step is expected at an energy of roughly 100 eV) and we expect to observe these events in the low background scenario. If present, these events will serve as an additional calibration point at the lowest energies and a testament to the sensitivity of our detectors.

High energy calibration is limited by our ability to reconstruct saturated events. In addition to the amplified K shell peak, high energy calibration is made possible with the help of a ^{133}Ba γ source. ^{133}Ba decays via electron capture with a half life of 10.5 years, producing γ -rays with several discrete energies. A list of the most prominent high energy peaks are shown in Table 3.6. These peaks, although not in our region of interest in the spectrum, provide points of known energy in the spectrum that can be used to develop reconstruction algorithms. The presence of these peaks is common to both Ge and Si and can be used for shared calibration of a Si detector using a Ge detector or vice versa.

Energy in keV	Probability
276.4	7.5%
302.9	19.1%
356.0	63.6%
383.9	9.1%

Table 3.6: High energy electron capture γ peaks of ^{133}Ba . The values in this table were taken from Ref. [113].

Finally, shared line calibration is a tactic that could be explored in the scenario where two detectors share energies from the same source. Shared events are those events in which the gamma deposits energy in one detector then scatters and deposits the rest of its energy in a neighboring detector. Energetic events, such as those from the Ba high energy peaks, could produce a large

deposit in one detector and a comparatively small deposit in an adjacent detector. The summed spectrum from the two detectors will then produce the total spectrum of Ba events which can be calibrated. Such calibration techniques can only be applied in the scenario where (a) two adjacent detectors share events, and (b) the shared events have a majority energy deposition in the detector adjacent to our detector of interest. This technique has been demonstrated on CDMS data in the past but could be explored with tower testing data at CUTE.

3.3.3.2 Fiducialization

Fiducialization refers to the process of establishing a detector volume, called the fiducial volume, from within which science-grade events are selected. In SuperCDMS, the fiducial volume generally excludes surface and edge events that either do not produce sufficient ionization or do not receive the full NTL gain due to edge effects in the electric field. The former is applied within the context of iZIP detectors while the latter is applied within the context of HV detectors.

In iZIP detectors, the interleaved electrodes in the crystal produce an vertical electric field of ~ 1 V/cm. The two surfaces are held at a constant, small voltage. Events that occur in the bulk are drifted to the ends by this modest electric field and receive full NTL amplification. Close to the surface, however, the field is non-uniform causing events in this region to not undergo full amplification, and thereby causing a mis-estimation of the amount of ionization produced in the recoil.. This is a major source of background in iZIP detectors which mimics nuclear recoiling dark matter that do not produce significant ionization. At SuperCDMS SNOLAB, the design incorporates a larger transverse field at the surface (see Fig. 3.22) to drift surface events in that direction, resulting in an asymmetric collection of the ionization signal on the side closer to the event, thus allowing easy rejection of surface events [90].

In HV detectors at CUTE and SNOLAB, the non-uniform electric field near the surface creates amplification-reduced events from only partial Luke amplification. This has the effect of smearing the resolution of activation peaks and diminishing the heights of steps in the measured energy spectra. In iZIP detectors, the availability and use of charge channel data provides a straightforward technique of fiducializing these detectors. This is not the case for HV detectors and highlights the need for alternate techniques.

In the past, construction of “partition” parameters that project the events onto an X-Y partition plane has been used to identify spatially clustered sources of noise [70]. Similarly, radial (R) and azimuthal (Z) partitions have been constructed to fiducialize the detector volume, see Ref. [70, 71] for examples. These partitions are constructed as some combination of the OF amplitudes and/or

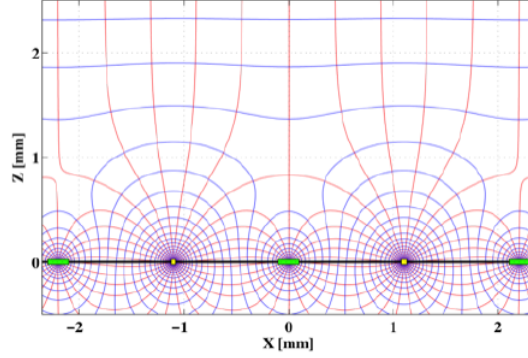


Figure 3.22: The E-field contours near one end of a SuperCDMS SNOLAB iZIP detector. The phonon sensors are shown in green and the ionization electrodes in yellow. The electric field lines are shown in red and equipotential surfaces are marked in blue. The strong transverse electric field at the surface drifts the charges from near-surface events in this direction, causing asymmetric collection of charges on one side of the detector. This allows easy rejection of surface backgrounds and better fiducialization capability. Figure taken from Ref. [90].

OF time delay. Note that these parameters do not represent the true position of the events in the detector but rather some empirical estimate that non-linearly correlates with the true event position. For example, in CDMSLite, the X and Y partition parameters were constructed as follows:

$$\begin{aligned} X_{\text{part}} &= \frac{\cos(30^\circ) \cdot D + \cos(150^\circ) \cdot B + \cos(270^\circ) \cdot C}{B + C + D}, \\ Y_{\text{part}} &= \frac{\sin(30^\circ) \cdot D + \sin(150^\circ) \cdot B + \sin(270^\circ) \cdot C}{B + C + D}. \end{aligned} \quad (3.4)$$

In Eq. 3.4, B , C , D refer to the channel amplitudes obtained from an OF fit. The partition is constructed taking into account the channel geometry of the detector where B , C and D are sectors spanning 120° in angle (see Fig. 6.10a). Z partitioning for this detector was relatively straightforward due to the availability of charge readout. A radial partition parameter was constructed as the difference in OF amplitudes of the event's closest inner channel and the outer ring A . The partitions are constructed with OF amplitudes, following the idea that events closer to a channel deposit higher energy in that channel. Similarly, radial parameters based on time delays have also been tested which follow from the idea that events closer to a channel produce sharper rising pulses on that channel (or in other words smaller time delays) since they trigger closest to the preset trigger point.

SuperCDMS SNOLAB detectors have very different channel geometries. The iZIP detectors have four inner channels which will need to be projected onto a quadrilateral shape in the $X - Y$ partition plane. The iZIP X and Y partitions will be defined as

$$\begin{aligned} X_{\text{part}} &= \frac{\cos(45^\circ) \cdot E + \cos(135^\circ) \cdot B + \cos(225^\circ) \cdot C + \cos(315^\circ) \cdot D}{B + C + D + E}, \\ Y_{\text{part}} &= \frac{\sin(45^\circ) \cdot E + \sin(135^\circ) \cdot B + \sin(225^\circ) \cdot C + \sin(315^\circ) \cdot D}{B + C + D + E}. \end{aligned} \quad (3.5)$$

In Eq. (3.5), B, C, D and E are the OF amplitudes (1×1 or 1×2 residual amplitudes) for that channel. The partition plane using Eq. (3.5) for a prototype SuperCDMS SNOLAB iZIP detector is shown in Fig. 3.23. The four inner sectors are projected to the corners of a square using the residual/fast amplitudes from the 1×2 OF.

The HV detectors have three sector-like channels and a central ring. The sector channels can be used to produce the $X - Y$ partition while the combination of the central and outer ring will create the radial partition. The HV partition parameters for side 1 and side 2 are defined as follows:

$$\begin{aligned} X1_{\text{part}} &= \frac{\cos(0^\circ) \cdot E1 + \cos(120^\circ) \cdot C1 + \cos(240^\circ) \cdot D1}{E1 + C1 + D1}, \\ Y1_{\text{part}} &= \frac{\sin(0^\circ) \cdot E1 + \sin(120^\circ) \cdot C1 + \sin(240^\circ) \cdot D1}{E1 + C1 + D1}, \\ X2_{\text{part}} &= \frac{\cos(0^\circ) \cdot C2 + \cos(120^\circ) \cdot E2 + \cos(240^\circ) \cdot D2}{E2 + C2 + D2}, \\ Y2_{\text{part}} &= \frac{\sin(0^\circ) \cdot C2 + \sin(120^\circ) \cdot E2 + \sin(240^\circ) \cdot D2}{E2 + C2 + D2}, \end{aligned} \quad (3.6)$$

where each channel name denotes the OF amplitudes for that channel. In Eq. (3.6), I have defined two sets of partition parameters – one pair per side. This is done so that the final triangles from each side, when overlaid, accurately portray the orientation of the channels on the detector. Once again, the orientation of the channels is inspired by the channel geometries. The partition plot for an HV detector tested at CUTE is shown in Fig. 3.24. For HV detectors, in addition to creating a spatial representation of the events, the partition plots were also used to determine the relative calibration factors for HV detectors at CUTE. More details on this are given in the next sub-section.

For SuperCDMS SNOLAB HV detectors, there is no charge readout. Therefore, a Z -partition parameter is warranted in this case. Z dependency is difficult to characterize in HV detectors due to large secondary phonon production in the Z direction which washes out strong dependencies. The hope, however, is that the non-uniform E-field close to the edges which produces reduced NTL gain in edge events preserves some Z dependency which may be used to tag them. Simple Z

3.3. TOWER TESTING AT CUTE

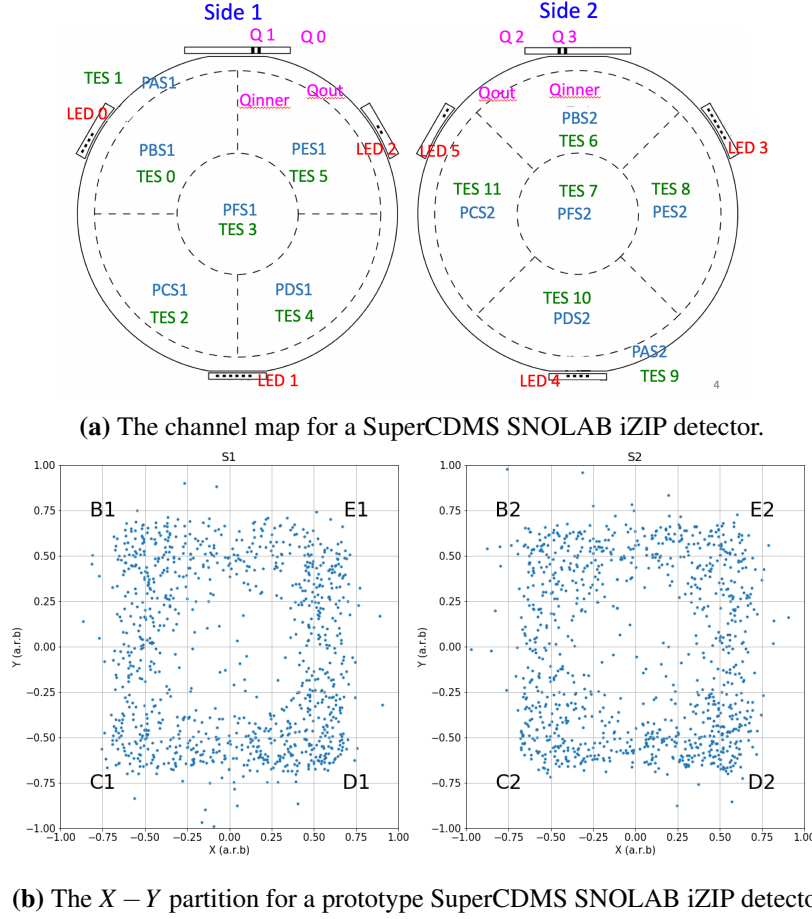
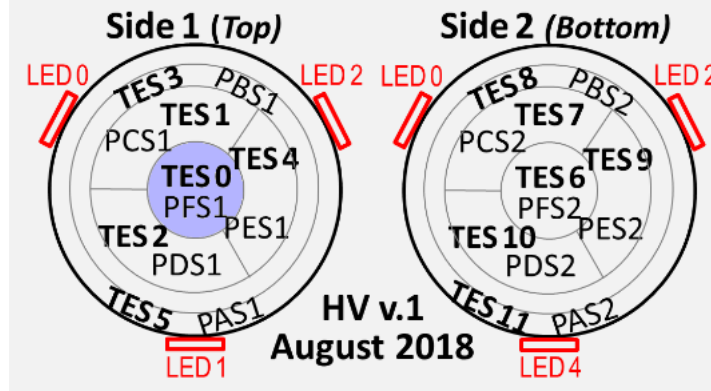


Figure 3.23: Constructing $X - Y$ partitions for SuperCDMS SNOLAB iZIP detectors. The partitions defined in Eq. (3.5) were used to construct the X and Y partitions from the 2-template residual amplitudes. The data was obtained from the testing of this device at SLAC. Note that one could rotate the S2 partitions clockwise by 45° and flip about the vertical axis by 180° to replicate the physical orientation of the channels on the detector. The physical orientation of the channels on the detector corresponds to folding the channel map about the center like a book such that PES2 is oriented in between PBS1 and PCS1.

parameters can be constructed for iZIPs such as the difference between the energies collected on each side or the difference in time delays between two vertically aligned channels on either side. For HV detectors, since we hope to see reduced NTL gains at the edge, using the difference in energy or time delay between the outer ring channels, A and B, on either side proved to be a good



(a) The channel map for a SuperCDMS SNOLAB HV detector.

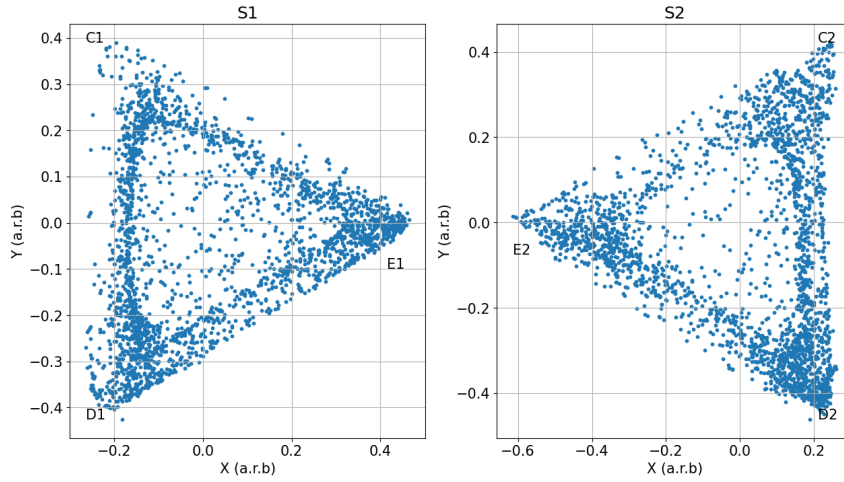

 (b) The $X - Y$ partition for a Tower 3 Ge HV detector tested at CUTE.

Figure 3.24: Constructing $X - Y$ partitions for SuperCDMS SNOLAB HV detectors. The partitions are defined in Eq. (3.6) using only the sector-like channels in the middle and constructed using the 1×1 OF amplitudes for each channel. The data was obtained from the testing of this device at CUTE.

Z parameter. While this is still a work in progress led by Shubham Pandey at U. Minnesota, I show preliminary plot of the Z -asymmetry parameter in Fig. 3.25. The parameter is defined as:

$$Z_{part} = \frac{(A2 + B2) - (A1 + B1)}{A1 + B1 + A2 + B2}, \quad (3.7)$$

where the channel names represent 1×1 OF amplitudes. In this definition, more positive Z implies

proximity to side 2 of the detector. The Z parameter in Eq. (3.7) is plotted as a function of OF amplitudes in Fig. 3.25. Subfigure (b) shows the distribution of events and their dependence on the Z -partition parameter. The cluster of events between $3\text{--}4\mu\text{A}$ are the L shell events, around $30\mu\text{A}$ are the K shell events, and the high energy clusters are mis-reconstructed saturated events. The central band represents events with symmetric distribution of energy on both sides. The deviation of the band from zero at the highest energies is due to saturation. The key points of note are the events in the “wings” of the central band. Consider the K shell peak. The winged extensions from the main cluster are clearly events from the K shell that recorded reduced amplitudes from reduced NTL gain. This is evidence for them being events near the surface which can be removed by a cut on this Z parameter.

Finally, the radial partition parameter is more straightforward to construct in the present geometry, simply by taking the difference in amplitudes or time delay between the central ring (PF) and the outermost ring (PA). The time delay was found to be more effective than the amplitudes since this quantity is less susceptible to saturation effects at medium to high energies. At the lowest energies, the timing parameter is noise limited and a combination of amplitudes and timing information must be employed. The definition of the time delay radial parameter at mid to high energies is:

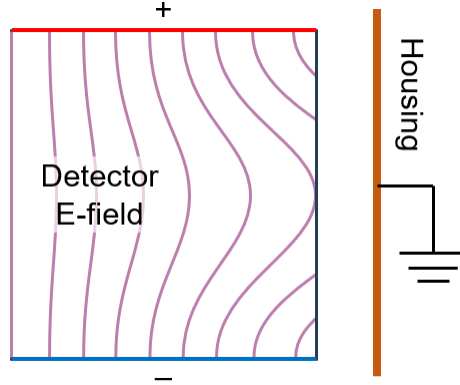
$$R_{part} = PF_{\text{delay}} - PA_{\text{delay}}. \quad (3.8)$$

In the above definition of the R partition, more negative R correspond to bulk events with full NTL gain. One can label the primary channel for each event by looking at the time delay for that event across all channels. The channel with the minimum time delay will be the primary channel. Labeling the distribution of radii by the primary channel provides an intuitive picture of the radial dependence of the radial partition parameter we defined in Eq. (3.8). The distribution of the radial parameter colored by primary channel shows the transition from inner to outer radii in Fig. 3.26. In Fig. 3.26b, in addition to the band of high radii events which constitute our surface background, the K shell peak is anti-correlated with the radial parameter. This relationship could be used to compute a position correction for energy to enhance the resolution of the observed peak.

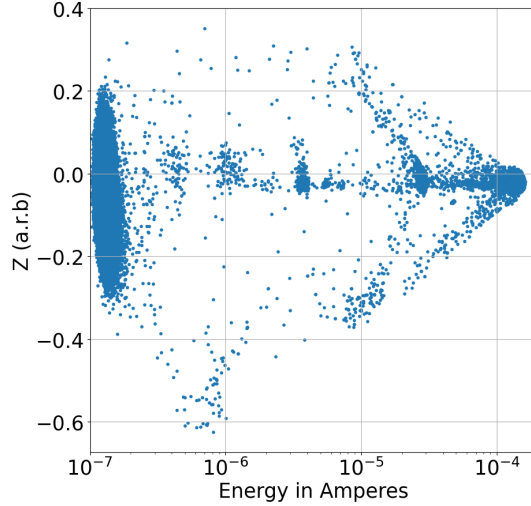
A parallel effort is the reconstruction of event position with the help of machine learning tools. Some results in this direction will be discussed in Ch. 6 using the $N \times M$ filter.

3.3.3.3 Relative calibration

The contribution of each channel in a detector is dependent on the QET distribution in the respective channel. The detectors are designed to have roughly equal distribution of QETs for this purpose.

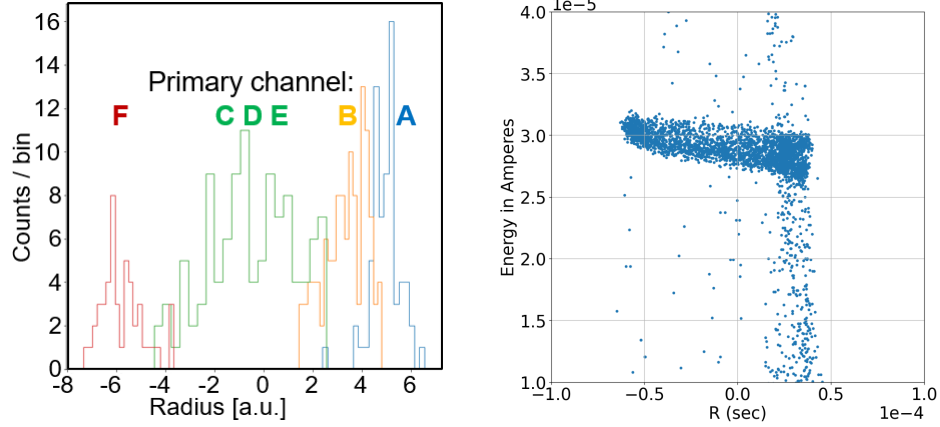


(a) Cartoon schematic of the E-field contour lines at the edge of an HV detector biased vertically at a voltage. The non-uniform E-field contours near the edges produce reduced NTL gain. Image courtesy: Wolfgang Rau



(b) Z partition parameter vs measured energy for T3Z1 Ge HV detector tested at CUTE.

Figure 3.25: Constructing a Z-partition for SuperCDMS SNOLAB HV detectors. In the definition in Eq. (3.7), more positive Z implies proximity to side 2 of the detector. Sub-figure (a) is a cartoon depiction of the electric field near the edge and (b) shows the distribution of events as a function of the Z-partition parameter. The cluster of events around $0.3\mu\text{A}$ are the M-shell events, around $1\mu\text{A}$ are LED events, between $3\text{--}4\mu\text{A}$ are the L shell events, around $30\mu\text{A}$ are the K shell events and the high energy clusters are mis-reconstructed saturated events. The “wings” extending from the central band are events that recorded reduced amplitudes from reduced NTL gain at the edges.



(a) The distribution of radii for events in the K shell for a Ge detector colored by the primary channel of the event.

(b) The energy as a function of the R-partition for the K shell peak in Ge.

Figure 3.26: The radial partition parameter. (a) shows the distribution of radial parameters colored by channels from innermost to outermost. The events tagged under PFS1 as the primary channel correspond to the lowest radii as expected while those with PAS1 as the primary channel has the largest radii. In (b) the distribution of the K shell peak as a function of the radial partition is shown. The band of outer events at high radii is clearly seen.

The initial phonons produced in an interaction bounce around a few times before they begin to thermalize. This typically happens after about 1 ms. Thermalized phonons have an even distribution of energy across all detector channels. At this stage, the energy collected in each channel is proportional to the collection area which is equal by design in SuperCDMS detectors. Nevertheless, factors such as the channel noise, differences in bias points etc can cause some channels to collect more or less energy than other channels. It is important to distinguish this from the difference in energy collected in a channel due to position dependence and the intrinsic energy collection differences between channels. The total energy deposition by an interaction in the detector is the sum of the energies from all the detection channels provided they are calibrated to produce equivalent response.

Relative calibrations are weights assigned to each channel relative to an arbitrarily chosen channel that I will refer to as the main channel here for convenience. The weights or relative calibration factors are normalized such that they sum to twelve (the total number of phonon channels in the detector). The main channel has a reference value of one before normalization and the relative

calibrations for other channels are a reflection of how different their responses are from the main channel. A larger factor indicates lower efficiency with respect to the main channel and vice versa.

SuperCDMS Soudan detectors had a lower QET density and therefore longer collection times which gave rise to long thermal tails in the pulses from thermalized phonons. The long tails after say 1 ms could be used to calculate the relative calibration factor. The factors are estimated as normalizations to be applied to each channel that equalize the integrals of the thermal tails. This makes the thermal tails align well in the pulses as shown in Fig. 3.27.

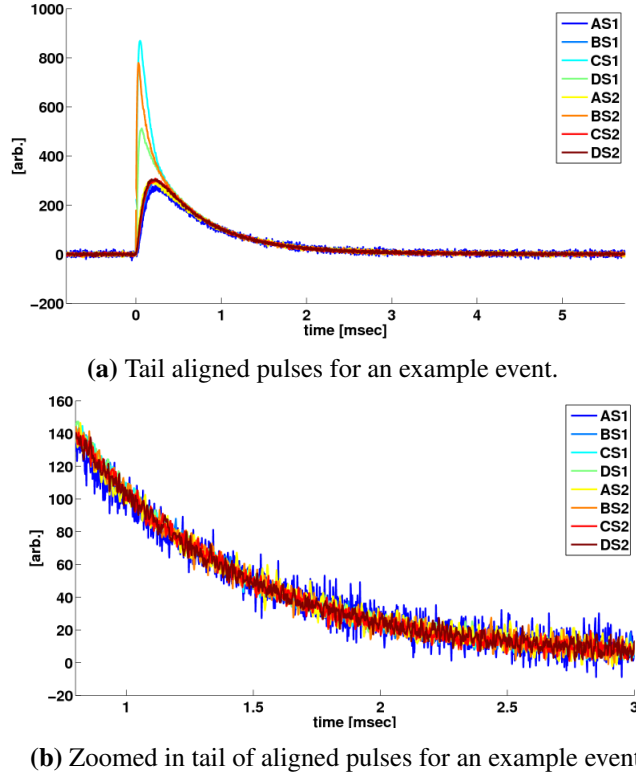


Figure 3.27: Pulses from an example event for every channel in the detector post relative calibration. Figures taken from Ref. [114].

This technique was practical at SuperCDMS Soudan due to a lower QET coverage, but this may not be the case for SuperCDMS SNOLAB HV detectors. The SuperCDMS SNOLAB detectors will require development of alternate relative calibration methodologies. Although the employability of older techniques remains to be verified, I will describe below one relative calibration technique (and in fact the only one at hand at the time this thesis was written), which was developed as part

of the prompt analysis performed during tower testing.

The technique is iterative and uses the triangle plots in Fig. 3.24 in conjunction with a couple of different versions of them to estimate scale factors for each channel. The basic principle in this analysis is that the equal partition of energy in each channel will centralize the triangle in the $X - Y$ partition plane. The procedure is as follows:

1. The sector channels (E, C and D) are used to construct the triangle plot defined in Eq. (3.6). The OF amplitudes of each channel, which are proportional to the energy collected in that channel, are scaled by weights such that the corners of the triangle pass through 60 degree lines drawn from the (0,0) as shown in Fig. 3.28.

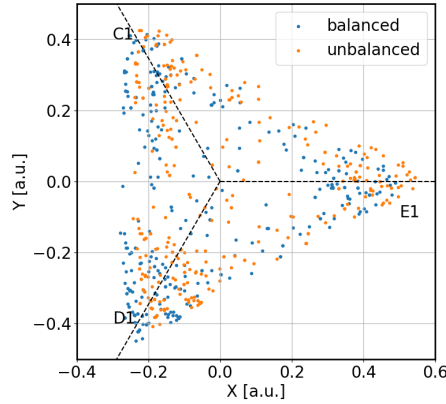


Figure 3.28: Triangle alignment with scale factors for sector channels. Unbalanced and balanced show the triangles before and after the relative calibrations are applied. The black lines are 60 degree markers for reference. X and Y are unitless.

2. Next we move to the ring channels. A second triangle is constructed but this time replacing two of the three scaled channels from step one. Here, we produce a “triangle” with channels F, B and D. Note that the plot in Fig. 3.29 is only referred to as a triangle since it contains three defined corners and resembles a deformed triangle of sorts – it is not a real triangle! The scale factor for D is fixed from step one. F and B are scaled so that the reference lines pass through the corners approximately. The technique is based on the idea that the corner events are quasi-localized to that channel and therefore, represent most of the energy being collected in that channel. The distribution of such events should possess a symmetry across all channels. Since one of the corners is fixed, there is not a huge wiggle-room to fine-tune the weights. This is a limitation of estimating relative calibration in this scheme.

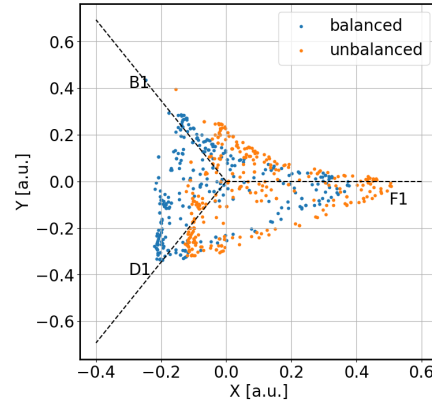


Figure 3.29: Triangle alignment with scale factors for the ring channels and one sector channel. X and Y are unitless.

3. A third triangle is constructed but this time replacing one of the three scaled channels from step two with the remaining unscaled channel on that side. We produce the “triangle” with channels F, A and D as shown in Fig. 3.30.

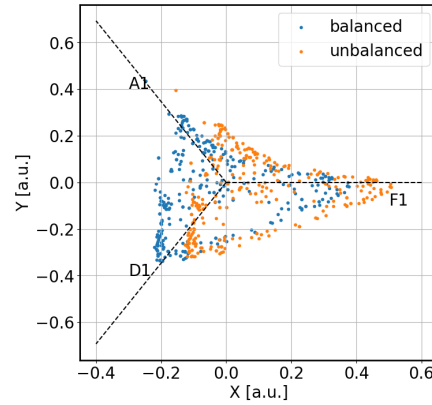


Figure 3.30: Triangle alignment with scale factors for the remaining ring channel. X and Y are unitless.

4. Steps 1, 2 and 3 are repeated for the other side of the detector.
5. Finally, a correction to equalize the two sides of the detector is applied. A ‘ Z ’ parameter is

constructed as the difference in scaled energy collected in each side:

$$Z = \frac{w_{s1} \cdot \sum_{c1} w_{c1} A_{c1} - \sum_{c2} w_{c2} A_{c2}}{w_{s1} \cdot \sum_{c1} w_{c1} A_{c1} + \sum_{c2} w_{c2} A_{c2}}, \quad (3.9)$$

where A refers to the channel amplitudes and c1 and c2 indices denote channels on side 1 and side 2 of the detector, respectively. An overall scale factor, w_{s1} , is applied to side 1 to reduce the asymmetry between side 1 and side 2 by placing the peak at zero as shown in Fig. 3.31.

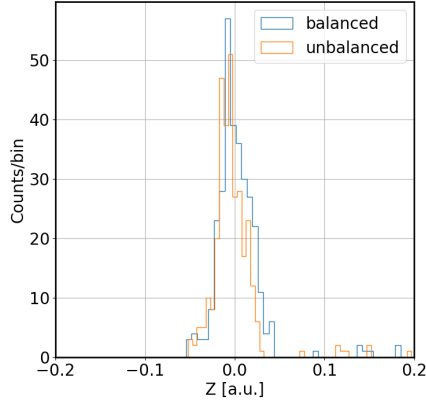


Figure 3.31: Alignment of the two sides of the detector with scale factors. Z is unitless.

6. As one last finishing step, the scale factors calculated in the previous steps are normalized by the total number of usable channels in the detector to get the relative calibration. The average relative calibration factor is then unity.

Note that the choice of channels in step one can, in principle, be any three channels. It is preferred to begin with the three sector channels since they have the most symmetry in terms of shape. This methodology clearly has limitations and is heavily reliant on OF reconstruction which in turn is noise dependent. The method also only offers limited tuning of part of the channels. Alternate techniques need to be studied with simulated data for a more standard methodology and development will continue as the analyses of these data progress. For the purpose of the results shown in the remainder of this chapter, we will use relative calibration factors computed using this method.

3.3.3.4 Ge calibration

With all the important tools in hand we proceed to the actual calibration of the detectors. Ge calibration data was acquired at a number of different voltages to (a) study the amplification gain from the NTL effect as a function of voltage, and, (b) understand the effects of saturation and build a saturation model.

The calibration involved finding the location of the three calibration peaks described in Sec. 3.3.3.1 in the spectrum. At 0V, the L and K shell peaks were visible in detectors Z1 and Z3, while only the K shell peak was visible in Z6. Z6 had a higher threshold from elevated noise seen in this detector which likely hid the L shell peak within the noise-dominated region of the energy spectrum. The 0V spectrum for Z1 is shown in Fig. 3.32.

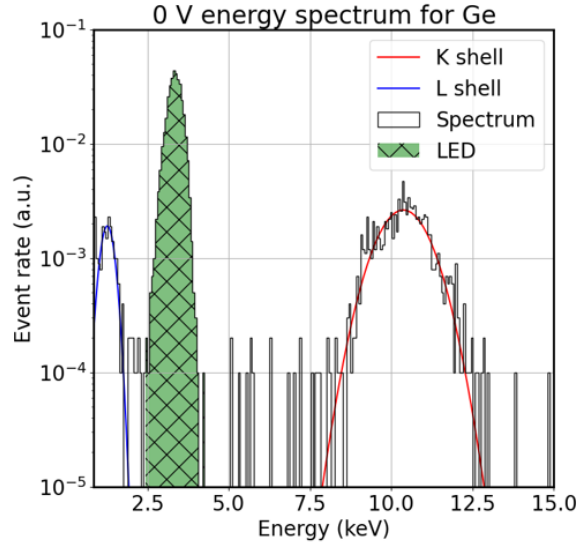


Figure 3.32: The (preliminary) energy spectrum for T3Z1 detector at 0 V. The K and L shell peaks are visible, marked with the corresponding Gaussian fits to label the peaks. The detector was periodically flashed with an LED during data taking which produced a mono-energetic peak in the spectrum from the LED photons. LEDs are used to monitor the stability of the detector over time.

The M shell peak begins to show at a voltage of ± 25 V at which voltage the NTL amplification scaled the total phonon energy to 2.8 keV. At 0V, none of the calibration peaks are saturated, however at 50 V the K shell peak begins to saturate. Reconstruction of saturated events is difficult with the OF, and therefore the OF amplitudes from this peak can no longer be used to calibrate the energy spectrum at this voltage. The L shell peak is used for calibration at this voltage and then the

M shell once the L shell is pushed into saturation with higher voltages.

The activation peaks are first fit with Gaussians to extract the mean location and the widths. An absolute calibration factor is calculated as (true energy)/(Gaussian mean) which is then used to scale the full spectrum and convert it from amplitudes to energy units. For the 0V spectrum in Fig. 3.32, the K shell was used as the calibration point while the 50 V spectrum in Fig. 3.33 used the L shell peak as the calibration point. The K shell peak displays lower energy due to effects of TES saturation which limits the full energy from being collected. The construction of a saturation model is usually the technique used to correct the energy scale for saturation effects. The development of an appropriate saturation model for the SuperCDMS SNOLAB HV detectors is actively being pursued by the collaboration which I leave for future publications to discuss in detail.

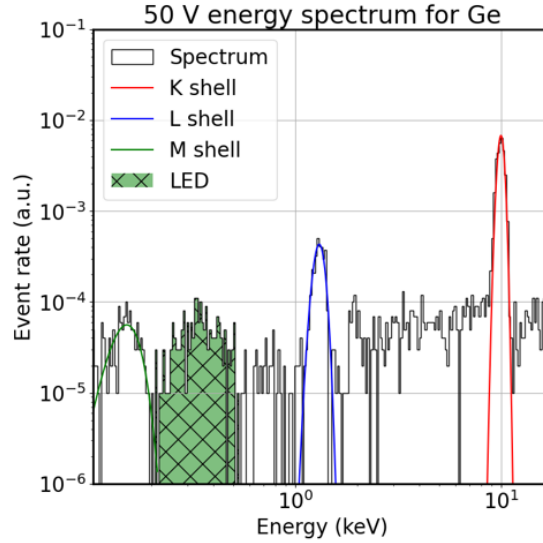


Figure 3.33: The (preliminary) energy spectrum for T3Z1 detector at 50 V. The K, L and M shell peaks are visible, marked with the corresponding Gaussian fits to label the peaks. The detector was periodically flashed with an LED during data taking which produced a mono-energetic peak in the spectrum from the LED photons.

The spectra shown in Fig. 3.32 and 3.33 were generated with simple cuts on the data that select only triggered events on T3Z1, a χ^2 cut to remove poorly fitted events from the OF and a cut on the Z_{part} (see Eq. 3.7) to remove reduced NTL events that smear the resolution of the peaks.

In addition to this, γ calibration data was also acquired during the tower testing period. Signatures from the Ba γ peaks mentioned in Table 3.6 were spotted and are awaiting further analysis. This analysis will involve development of reconstruction techniques capable of fitting saturated

pulses with flat tops and longer fall times and is being done by Ariel Zuniga Reyes. Apart from the Optimal Filter fits, integrating the total current in the pulse is also a measure of the amount of energy collected. Fig. 3.34 shows the high energy γ spectrum observed in T3Z1 obtained from the integral quantity of the pulse.

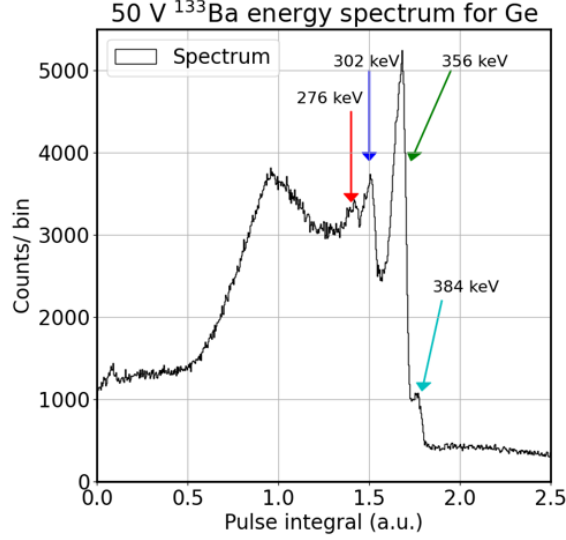


Figure 3.34: The high energy gamma spectrum of T3Z1. The spectrum is reconstructed from the integral of the pulse.

The resolution of the reconstructed spectrum in SuperCDMS varies as a function of energy (see Ref. [70, 71] or Ch. 6). In order to construct a full resolution model, the resolutions at different calibration points are calculated with a Gaussian fit. I do not digress into resolution modeling of these detectors in this thesis since they are analyses in progress, but rather will show the current best estimates for the resolution of each peak as of the time of writing this thesis. So, don't be too surprised if the numbers in a publication from early 2025 look better! I compare the resolution estimates from two algorithms: the simple 1×1 OF, and the $N \times M$ OF which will be a major point of focus of Ch. 5 and 6. To give the reader a head start, the $N \times M$ filter is an advanced reconstruction algorithm which will be the major driving force of SuperCDMS SNOLAB event reconstruction. The algorithm fits pulses from multiple channels simultaneously with multiple templates/models accounting for correlated noise between channels. If the templates are constructed such that they empirically capture position dependence, then a combination of standard statistical and machine learning techniques can yield significantly enhanced energy resolution. I do not attempt to perform machine learning techniques on this data set due to limited data size. The results from the $N \times M$

3.3. TOWER TESTING AT CUTE

filter compared to the 1×1 OF are shown in Table 3.7. By accounting for correlated noise and linear effects of position dependence, nearly a four-fold improvement is seen at the K shell and about a 31% improvement at the M shell. The numbers in the table are conservative estimates based on bootstrapping the data 500 times and extracting a resolution from each sample (see Ch. 6 for further explanation). The resolutions are calculated as the mean resolutions from all bootstrapped samples.

Estimator	K shell (eV _{ee})	L shell (eV _{ee})	M shell (eV _{ee})
1×1 OF	380.81 ± 4.11	67.62 ± 0.41	41.17 ± 0.92
$N \times M$ OF	99.29 ± 0.48	38.02 ± 0.19	31.50 ± 0.50

Table 3.7: The energy resolution of calibration peaks in T3Z1. Two estimates, one accounting for correlated noise and linear position dependence ($N \times M$) and one not accounting for these (1×1), are shown in the table.

The baseline resolution is calculated from noise traces using the 1×1 OF amplitudes. Baseline calculation from the $N \times M$ filter will be discussed in Ch. 6. The noise environment during the testing campaign, unfortunately, turned out to be quite turbulent due to various sources of noise described earlier in this chapter. In addition, one of the cooling stages of the fridge was at a significantly higher temperature than expected. With these limitations, our baseline resolutions varied over a range of values over the course of the run. Our best estimates for the resolution of most of the Ge detectors in tower 3, based on the range of values observed over the full testing period, are shown in Table 3.8.

Detector	Baseline resolution (eV _{pt})
Z1	60 – 90
Z3	50 – 70
Z6	130 – 200

Table 3.8: The estimated baseline resolution of tested Ge detectors in Tower 3 from the CUTE facility. The key driver for these resolution numbers was the variable noise in the detector and readout environment, a significant portion of which is expected to be addressed in the main experiment.

Finally, a verification of the amplification factor was also performed with data acquired at several voltages. The variation in location of the peak as a function of voltage can be used to estimate the gain factor. The amplification is expected to be linear with voltage from the NTL Eq.

(2.4) which I repeat below for your convenience:

$$E_t = E_r \left(1 + \frac{Y(E_r)eV_b}{\epsilon} \right).$$

Fig. 3.35 shows the position of the K and L shell peaks for T3Z1. The positions of the peaks agree well with the NTL gain at all voltages for the L shell peak. The L peak is not saturated even at 60 V and therefore does not see any non-linear energy variation. The K shell peak on the other hand enters the saturation regime of the TES after about 30 V. This reduces the amount of energy reconstructed with the OF since it fails to fit saturated pulses. A verification performed with the integral of the pulse showed that in the saturation regime the energy of the K shell peak does not deviate from the expected linear behavior. Thus, we see good agreement with the expected gain from NTL amplification in our Ge detectors.

3.3.3.5 Si calibration

Multiple avenues for Si calibration are being explored currently. I'll begin with Compton step calibration. At 0 V, the highest energy step (K) was visible after just a couple of days of gamma calibration data acquisition. The step was accompanied by an X-Ray fluorescence peak appearing from the Cu in the detector housing whose true energy is at 8.1 keV_{ee}. The location of the peak is lower than what one would expect relative to the location of the K step, the likely cause of which is TES saturation. Si detectors appear to saturate earlier than Ge detectors as indicated by the rising spectrum at these comparatively lower energies. The 0V spectrum for Z2 is shown in Fig. 3.36. The red line is drawn from a sum of an error function, straight line and a Gaussian to guide the viewer across the spectral features— the function was not fit and only drawn by manual tuning. An error function will be fit to the step in an unbinned likelihood fit to extract the location and resolution parameters of the step, while the straight line accounts for the saturation rise of the spectrum and the Gaussian fits the Cu peak:

$$\begin{aligned} f(E|\mu_{err}, \sigma_{err}, h_{err}; a, b; \mu_{gaus}, \sigma_{gaus}, h_{gaus}) &= f_{error} + f_{line} + f_{gaus} \\ &= \left[h_{err} \cdot \text{erf} \left(\frac{E - \mu_{err}}{\sigma_{err}} \right) \right] + [a \cdot E + b] \\ &\quad + \left[h_{gaus} \cdot \exp \left(-\frac{(E - \mu_{gaus})^2}{2\sigma_{gaus}^2} \right) \right]. \end{aligned} \quad (3.10)$$

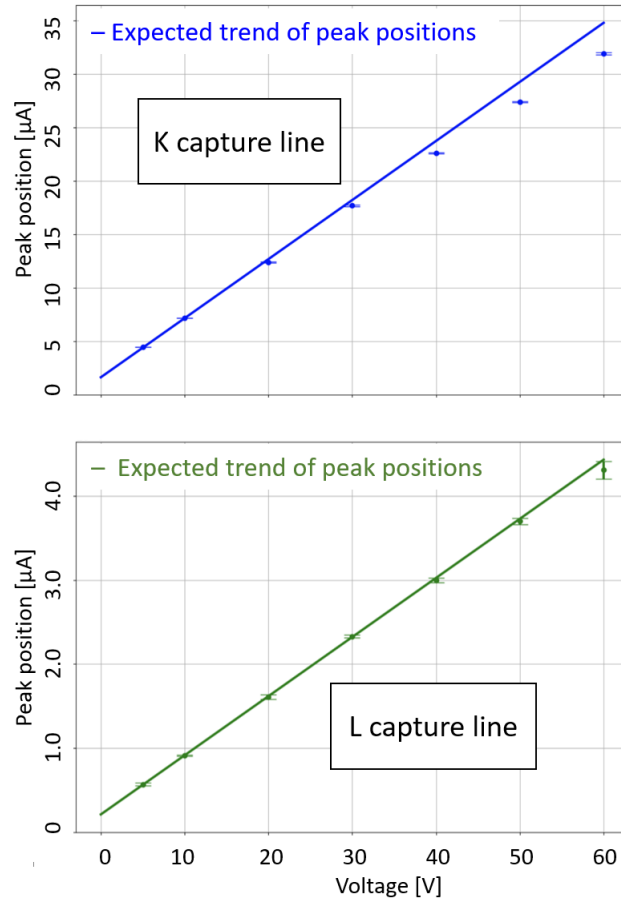


Figure 3.35: The K and L shell peak position as a function of applied voltage for T3Z1. The position of the L shell peak scales linearly with voltage, however the K shell peak deviates from linearity after about 30 V. This is, however, not an indication of under-amplification but rather an artifact of saturation effects at high energies. Image courtesy: Moaaz Elwan.

The HV calibration of the Si detectors was more challenging. The unavailability of peaks of known energy prevents us from procuring a clear understanding of the relationship between the energy and the fiducial parameters defined in Sec. 3.3.3.2. Although a plot such as Fig. 3.26b can be produced for the Si detector, the amount of information encoded in the plot is difficult to gauge in the absence of clear peaks. At a HV of 90 V (± 45 V) the Cu peak is pushed deep into the saturated region that limits our ability to reconstruct the peak. Further, even if we were able to perfectly reconstruct the Cu peaks at high voltages, these events are surface/outer events in the

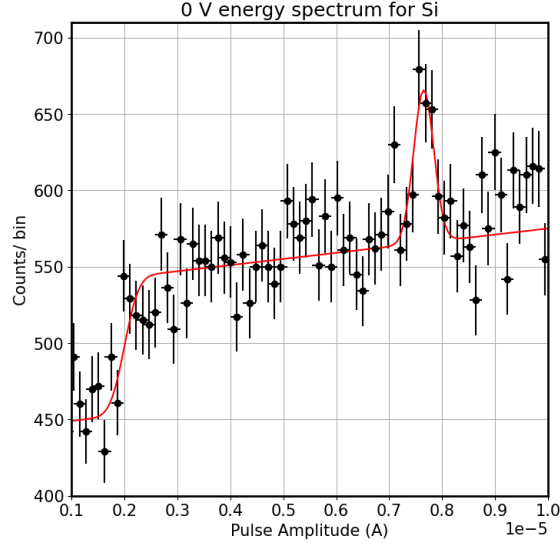


Figure 3.36: The (preliminary) energy spectrum for T3Z2 detector at 0V. The K step is clearly visible and marked in red. The Cu peak is also outlined with a Gaussian. The red line is meant only to guide the eye and is not a fit.

detector due to the low penetration power of the external X-rays. This means that even if these peaks were seen in the HV data, they could not be useful to us in understanding the behavior at the bulk of the crystal.

A very harsh fiducial cut had to be applied to remove most surface backgrounds. From Fig. 3.37, the harsh cut rejects nearly 50% of events as surface backgrounds. This has the consequence of requiring twice as much data than previously determined from detector simulations to accumulate sufficient statistics to see the L steps in the spectrum. As a result, several weeks of data were acquired to search for the L steps at a voltage of 90 V. This also likely reflects the amount of data acquisition time that will be required for an L step calibration at high voltage during SuperCDMS SNOLAB.

The 90 V spectrum for Z2 is shown in Fig. 3.38. The spectrum was constructed after the fiducial cut shown in Fig. 3.37, a χ^2 cut to remove poorly fitted events and selection of events triggering only on T3Z2. In addition, relative calibration factors were also applied. Two features that are consistent with the L steps are observed and marked in red. Once again, the red curve is drawn to guide the eye and is not fitted to the data. An interesting point is the potential mismatch in the observed and expected location of the steps. The location of the K step in the 0 V data can be scaled by the NTL gain to estimate the expected location of the L step at 90 V, shown in green in

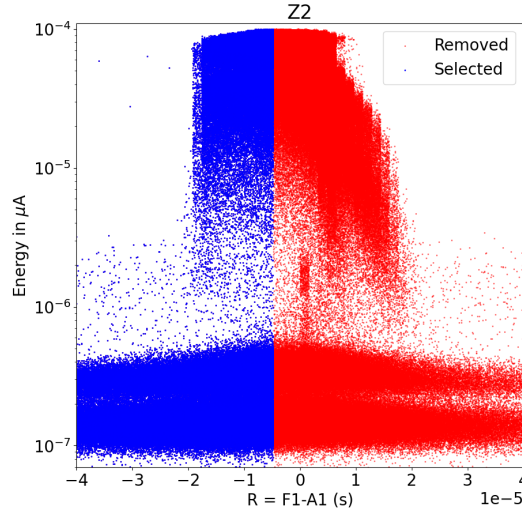


Figure 3.37: An example radial cut to pick a fiducial volume for Z2. The radial parameter is constructed from one side with time delay. As is evident from the image the selection of a cut value is deemed quite difficult from the lack of clear observable features. Other than a cluster around zero radius from a time dependent noise source, there are no discernible features in the scatter plot unlike in Ge.

the figure. The blue line marks where the feature is observed in the data. The potential mismatch in calibration factors at 0 V and 90 V seen in Fig. 3.38 has been confirmed in prototype single charge-sensitive devices tested by SuperCDMS. Although the preliminary results show signs of over-amplification at high voltages in the Si devices, a thorough modeling of the detector backgrounds is essential to properly ascertain and statistically quantify the effect. The effect, if confirmed, will have significant implications for SuperCDMS's future dark matter searches since a higher amplification than expected will further lower the effective threshold by about 10 – 15 %.

Small pulses of constant height were observed in the Si spectrum, most clearly at a voltage of 100 V ($\times 27$ amplification of energy). An example pulse is shown in Fig. 3.39. These tiny pulses are hypothesized to be originating from single e^-/h^+ events from infrared radiation leaking into the detector environment. A large number of noise traces (untriggered traces) were collected at voltages close to 100 V and each trace contained several such minuscule pulses. These events are reconstructed by calculating the height of each optimal filtered pulse in a trace that exceed a certain threshold. This technique of reconstructing these pulses has the disadvantage of picking small random excursions that exceed a threshold even though they may not be from a pulse. This gives the reconstructed spectrum of these events a broad resolution. Other better techniques to

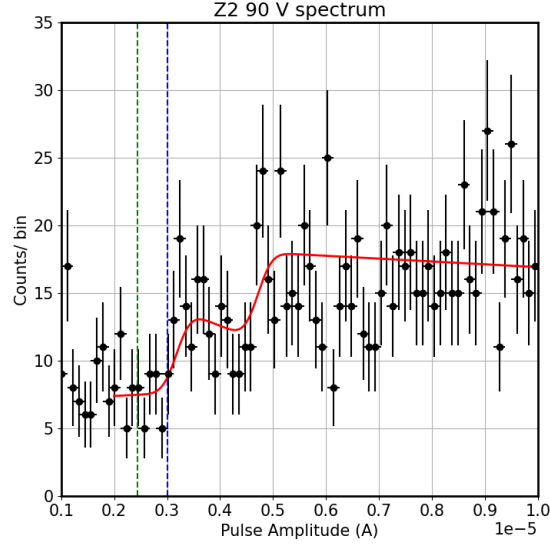


Figure 3.38: The (preliminary) low energy calibration spectrum for T3Z2 detector at 90 V. The features believed to be the L1 and L2 steps are marked with the red curve (not a fit, but tuned by eye to guide the reader). An interesting mystery is the apparent difference in location of the observed feature at 90 V (blue) and what is expected from the location of the K step in the 0 V data (green). We observe roughly 10-15% difference between the location of the step-features at 0 V and 90 V.

reconstruct these tiny events without bias are being explored by the collaboration. Preliminary reconstruction of the energy spectrum for these events indicate the presence of a secondary peak, distinct from noise at the lowest energies, close to the single e^-/h^+ pair production energy in Si. A comparison of the spectrum near threshold at various voltages will be made for further analysis of event rates. The mounting evidence points to these being single e^-/h^+ events, but further analysis is needed before confirmation. The origin of these events have been identified to be infrared leakage from the surrounding environment. At SuperCDMS SNOLAB, the occurrence rate of these events will be lower than what was observed at CUTE due to better infrared shielding around the detectors.

At higher energies, only the 356 keV peak from the gamma energy spectrum is observed in the Si detector as shown in Fig. 3.40. The spectrum is constructed from the integral quantity similar to the Ge high energy spectrum due to its capability to overcome saturation effects to a good extent. However, the lowest energies are of most interest in Si detectors since most of SuperCDMS's sensitivity reach to the lowest dark matter masses is expected to come from Si detectors.

A fitting of the Compton steps will require proper modeling of the low energy backgrounds

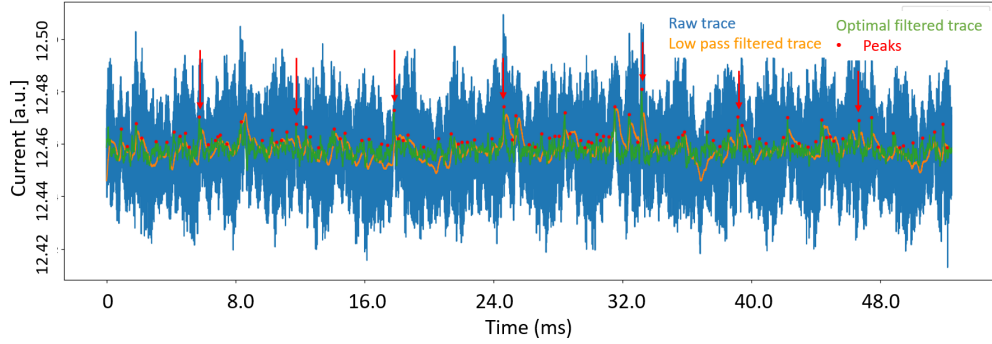


Figure 3.39: An example trace with several minuscule pulses within. A couple of the pulses in the trace are highlighted with arrows. Image courtesy: Moaaz Elwan.

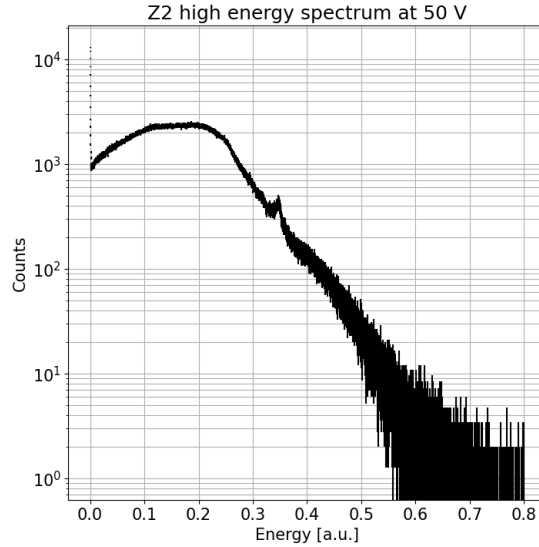


Figure 3.40: The (preliminary) high energy spectrum for T3Z2 at 50 V. The 356 keV peak from Ba is observed in the spectrum near 0.34.

and rates. Nevertheless, an approximate location of the steps, deduced from observation, was used to obtain a first estimate of the baseline energy resolutions of the Si detectors. Note that here I quote the 0 V resolutions for these devices, since one of the Si detectors, Z5, was only calibrated at 0 V with the K step. HV data for this detector was not collected due to schedule limitations. Table 3.9 shows the range of resolutions observed for the Si detector over the entire course of the run. A range is quoted rather than a value for the same reasons as Ge – the variation in the noise

environment caused the resolution to drift around at various points during detector testing.

Detector	Baseline resolution (eV_{pt})
Z2	20 – 33
Z5	20 – 35

Table 3.9: The estimated baseline resolution of tested Si detectors in Tower 3 from the CUTE facility. The key driver for these resolution numbers was the variable noise in the detector and readout environment, a significant portion of which is expected to be addressed in the main experiment.

3.3.4 Neutralization

All crystals have impurity sites and SuperCDMS detectors are no exception. Semiconductor crystals at room temperature have sufficient thermal energy to ionize these impurity sites creating free donors (D^+) and acceptors (A^-). The crystal also has an ensemble of thermalized e^-/h^+ pairs. As the detector cools, the e^-/h^+ pairs recombine with each other rather than with the free donors and acceptors. Therefore, at low temperatures, such as the operational temperatures for SuperCDMS detectors, these free donors and acceptors act as traps for drifting charges. This severely limits the iZIP charge collection efficiency and hinders proper amplification from the NTL effect in HV detectors. Therefore, it is necessary to neutralize cooled detectors before data taking.

The goal is to convert (D^+/A^-) impurity states into (D^0/A^0) impurity states. The simplest way to do this is to flood the crystal with excess free charge carriers before data taking, in the cooled down state. The free charge carriers would bind to the impurity sites, thus neutralizing them. Previously at SuperCDMS Soudan, it was found sufficient to ground the detectors and expose them to a radioactive source or flash an appropriate wavelength LED. This procedure seemed to generate enough charges to hold the detectors neutralized for a few hours or so underground.

The neutralization procedure for SuperCDMS SNOLAB towers was tested at CUTE. This is a difficult processes to test on the surface where cosmic activity can produce impurity (D^+/A^-) sites at a faster rate than underground, requiring a neutralization every few minutes to every few tens of minutes.

At CUTE the neutralization status of the Ge detector was assessed using the calibration lines. The K shell peak was completely washed out by the impurity states in the detector prior to neutralization. Flashing the detectors with LED photons for 5 – 10 hours, a process termed “baking”, was found effective in neutralizing the impurity states and restoring the high voltage resolution of

the calibration peaks. A comparison of the spectra before and after neutralization is shown in Fig. 3.41.

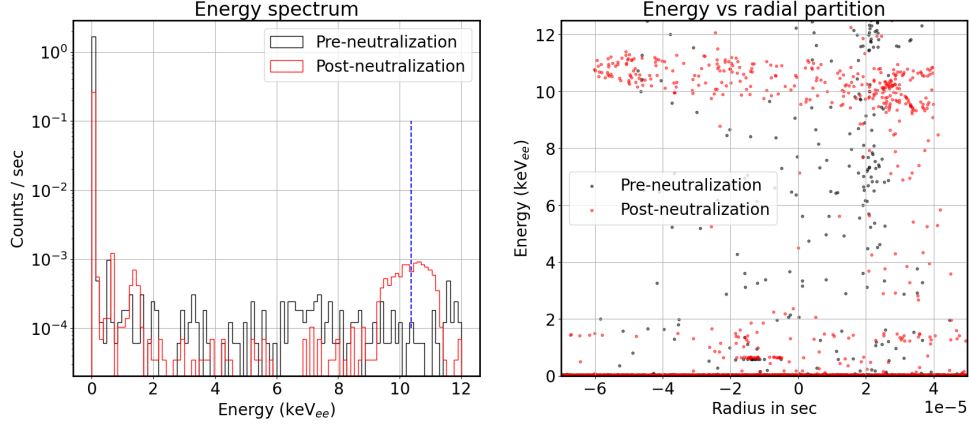


Figure 3.41: A comparison of the K shell peak before and after neutralization. The K shell peak is poorly resolved prior to neutralization, but well resolved after. The blue line marks the location of the 10 keV peak. The left plot compares the energy spectra directly while the right plot compares the event distribution in the energy - radius partition plane.

Assessing the neutralization status of the Si detectors was non-trivial. However, similar bake times as Ge were used on both Si detectors to achieve neutralization. The observation of the L steps was taken as a sign of proper neutralization of the Si detectors.

It was also found necessary to flash the detectors with LED after long operation times at high voltages. These “mini-flashes” were carried out at the same LED currents but only for 15 – 20 mins. This was found sufficient to neutralize the detector and restore it to its steady state.

3.4 Summary and outlook

In this chapter, the results from testing the new SuperCDMS SNOLAB detectors at two test facilities – SLAC and CUTE – were presented. While the testing was limited at SLAC, extensive study of the detectors and their properties was carried out for an HV tower at CUTE. These results are actively being analyzed and preliminary findings are reported in this chapter. This tower testing campaign marked the first operation of SuperCDMS SNOLAB detectors in a low background environment. The testing campaign proved fruitful in revealing hardware limitations and better understanding of our detectors in preparation for the commissioning phase. As more data is ana-

3.4. SUMMARY AND OUTLOOK

lyzed, some of the numbers in this chapter may change slightly for the better. The demonstration of our capability to calibrate SuperCDMS SNOLAB detectors at the lowest energy projects excellent outcomes for the experiment. Several analyses, hardware development projects, software improvements and analysis plans were also kick-started in preparation for SuperCDMS commissioning and science operations, supported by findings from this tower testing campaign.

Chapter 4

Event reconstruction at SuperCDMS SNOLAB

SuperCDMS raw data consists of time-varying current pulses. Parameters of interest such as the recoil energies and the event location in the crystal are encoded in the amplitude and shape of the pulses. Event reconstruction is a broad term which encompasses methods and strategies that extract these quantities of interest from the raw data.

SuperCDMS employs a variety of reconstruction algorithms to build a set of analyzable quantities called RQs. The most commonly utilized are the Optimal Filters (OFs) which fit signal-like events whose pulse shapes match some pre-defined template. The details of how this works were discussed briefly in Sec. 2.4. We reviewed some basic concepts necessary to understand the Optimal filtering formalism and introduced a few commonly used kinds of OFs like the Non-Stationary Optimal Filter (NSOF) and the 1×2 OF.

This chapter will focus primarily on novel reconstruction techniques for SuperCDMS SNOLAB. I will begin by motivating the need for new techniques and provide a detailed explanation of two new algorithms. I will also cover the validation of these new methods with simulated and real world data sets and provide an outlook for the future.

4.1 SuperCDMS SNOLAB data acquisition strategy

The Detector Control and Readout Cards (DCRCs) are at the heart of SuperCDMS SNOLAB's data acquisition scheme. A picture of a DCRC is shown in Fig. 4.1. The 100-pin connector on the left is used to mount the DCRC on a Vacuum Interface Board (VIB). The VIB is maintained

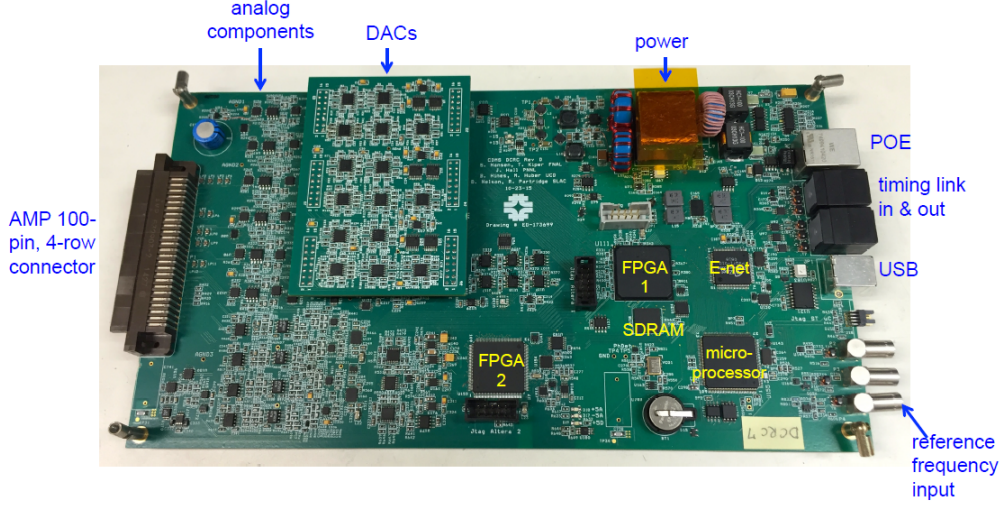


Figure 4.1: Picture of a DCRC board. Various components of the DCRC are labelled in the image. Figure taken from [90].

at room temperature and signals are passed to the VIB from the detector via flexible cables that are appropriately heat-sunk at various stages of the dilution fridge. The DCRCs contain Digital to Analog converters (DACs), which convert settings demanded by users into output voltages to control the detectors. The DCRCs also contain Analog-to-Digital Converters (ADCs) which digitize phonon/charge waveforms from the detector and store them in a circular buffer ~ 6.7 s long. The L1 trigger system detailed in Sec. 2.3 is implemented within the DCRC's FPGA. In addition, the DCRCs also contain a micro-processor which configures the DCRC's network and communication. The DCRCs are powered through a Power Over Ethernet (POE) supply which, as the name suggests, provides both a route for communication as well as power. The POE connects via an ethernet port, adjacent to which two RJ-45 connectors are available for timing and clock synchronization of the DCRC. Further, USB ports and LEMO connectors on the DCRC provide avenues for direct communication and interfacing for bench-top testing, respectively.

The DCRC consists of 5 ADC chips of which three collect phonon signals and two collect charge signals. The DCRC's FPGA sends periodic signals at a rate of 10 MHz (5 MHz) to the phonon (charge) ADCs to digitize waveforms. These frequencies are greater than the required sampling frequency of the ADC. Over-sampling in this manner at frequencies greater than or equal to the Nyquist frequency allows the DCRC to achieve a higher signal-to-noise ratio while being able to fully recover the signal [115] at lower cost. The digitized waveforms are downsampled

4.1. SUPERCDMS SNOLAB DATA ACQUISITION STRATEGY

by averaging every 16 points for the phonon channels and every 2 points for the charge channels, producing effective sampling rates of 625 kHz and 2.5 MHz for phonon and charge channels, respectively. These downsampled waveforms are stored in a buffer while waiting for the L1 trigger system to issue a trigger before they can be moved to permanent disk storage.

The maximum throughput of a DCRC was measured during a hardware testing effort conducted at TRIUMF and was found to be ~ 4.9 MB/s. Table 4.1 shows the estimated event rate and size for different kinds of data sets. Calibration data sets often produce large number of triggers owing to the detector’s direct exposure to a radioactive source. Dark matter data sets are termed “low background”, i.e., the number of triggers expected from potential dark matter signals is low. In addition, the DAQ also allows us to read out waveforms selected at random times at a specified rate which constitute noise triggers. The trigger rate is tunable in the DAQ, but certain prescriptions for optimal rates exist to allow the experiment to achieve its projected sensitivity. The maximum expected trigger rates are shown below. From the table, it is evident that our DAQ is limited by the data bandwidth of the DCRCs.

Data type	Expected max event rate	Max event size per detector	Max throughput
Calibration	5 Hz	1.8 MB	9 MB/s
Dark matter	2 Hz	1.8 MB	3.6 MB/s
Noise	0.5 - 2 Hz	1.8 MB	0.9 - 3.6 MB/s

Table 4.1: Estimated maximum data rates and throughput for SuperCDMS SNOLAB for fully sampled 52.43 ms long waveforms.

In addition to the readout limitation, dark matter data taken at a throughput of 3.6 MB/s translates to about 90 TB/year of data volume assuming an 80% livetime fraction. With a potential compression of data as a secondary step, this could go down to 45 TB/year. Although this number is significantly lower than SNOLAB’s network bandwidth to the outside world of 1 to 10 Gb/s (or 3942 to 39420 TB/year), disk space for data storage at scientific computing facilities is not infinite. Calibration data sets could span the order of months and noise data sets could total up to a month’s worth in the time scale of a year. In addition to raw data files, CDMSBats processing produces processed outputs which also need to be taken under consideration. Thus, data storage for an experiment hoping to run for a decade could pose an issue in the future.

Our experimental threshold is closely tied with our ability to trigger efficiently without any time gaps (“deadtime”) at the lowest energies. If the hardware cannot keep up with the trigger rate, then the experiment becomes threshold-limited. In order to overcome these limitations, SuperCDMS

SNOLAB will take data in hybrid format. In hybrid format, a single waveform is sampled at two different sample rates by keeping the signal portion intact and downsampling the non-signal-like tails. The effective data rate is significantly reduced in this scheme, as you will see in the following sections.

4.2 Hybrid readout – features and challenges

SuperCDMS waveforms are typically 52.4288 ms long. For phonon channels, this translates to 32768 samples per trace, while for charge channels the number is 131072 samples per trace. Typical pulses seen in SuperCDMS have a pulse width not larger than ~ 2 ms. Of the 52 ms waveform, the remaining 25 ms tails on each side predominantly contain the ambient noise in the system. A typical SuperCDMS phonon pulse was presented in the top panel of Fig 2.7 in Ch. 2. Charge pulses have a much shorter rise and fall time compared to phonon pulses and event information is predominantly contained in the pulse-like region of the trace. In this scenario, there seems to be an obvious solution to our predicament. Why don't we read out shorter traces if most of the event information is in the central 2 ms?

There are three reasons why this would not work. Our trigger system filters the input waveforms and assesses whether the resulting waveform exceeds a given threshold. In the case where a threshold is breached, a trigger is issued and a certain number of time bins before and after the maximum of the peak which breached the threshold are recorded. Depending on the event location and energy, the start and the end of the pulse varies from channel to channel for each event in each detector. Therefore, a shorter tracelength doesn't necessarily guarantee that the start of the pulse would be accurately recorded. The start bin of the pulse is a necessity for determining time delay in the optimal filtering mechanism, without which reconstruction is severely impacted. The second reason is closely related. Like the left end of the pulse, the falling tail need not be contained fully within a 2 ms waveform. Depending on the fall time of the pulse, part of the energy brought by the slow component could extend beyond the 2 ms estimate. The third reason is the ambient noise. Every waveform is recorded in units of ADC counts with the pulse sitting on top of the ambient noise. Effective optimal filtering requires a pulse with the baseline centered at zero. However, the baseline is not guaranteed to be a constant and varies from pulse to pulse. The usual reconstruction strategy is to collect a "pre-pulse" region in the trace to estimate the mean baseline which is then subtracted to get a zero-centered baseline. Previous studies done of charge channel reconstruction show in particular that using longer traces gives better resolution by giving access to lower frequency components. Thus, by keeping longer tracelengths we gain the advantage of

enhanced resolution of low frequency noise components. Given these reasons, our requirements become complicated. We need to maintain the ability to collect longer waveforms with high trigger efficiency while keeping up with hardware limitations.

The hybrid readout scheme is meant to address this. In this scheme, the “pre-pulse” and the “post-pulse” tails are downsampled by a factor of 16 by averaging every 16 time bins. The central region which contains most or all of the pulse, which we will refer to from now on as the “on-pulse” region, is fully sampled. Thus, the downsampled tails are maintained to provide the necessary baseline information. In addition, the tails provide the extended length necessary to guarantee the inclusion of the rising and falling edges of the pulse while staying below the maximum throughput of the DCRC. An example of a hybrid waveform was introduced in the bottom panel of Fig. 2.7 in Ch. 2. Figure 4.2 below shows the various regions in a hybrid pulse. The terms in quotes introduced in this paragraph will be used throughout this chapter, especially when dealing with reconstruction techniques for hybrid waveforms.

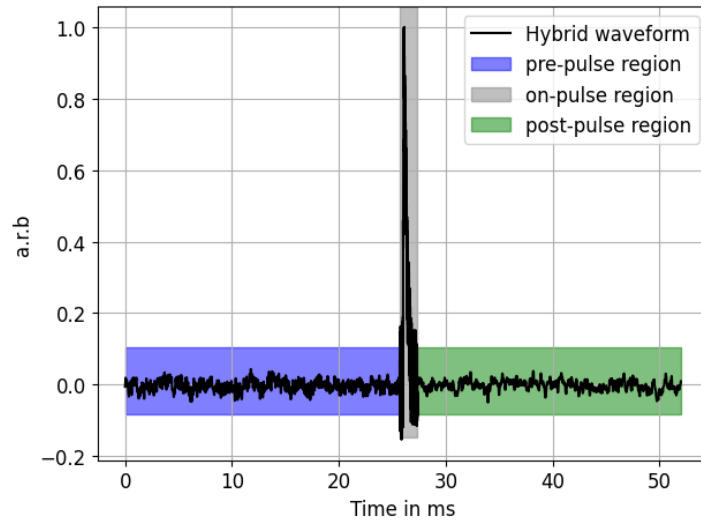


Figure 4.2: A plot of an example hybrid pulse. The pre-pulse, on-pulse and post-pulse regions are clearly marked in the plot.

The on-pulse region is sampled at the usual DCRC sampling frequency of 625 kHz (2.5 MHz) for phonon (charge) channels. The downsampling by a factor of 16 produces an effective sampling frequency of 39.0625 KHz (156.25 kHz) for phonon (charge) channels in the pre- and post- pulse regions. The proposed hybrid sampling scheme for phonon and charge channels is shown in Table 4.2.

An additional term to familiarize yourself with when dealing with hybrid waveforms is the “uniformly downsampled waveform”. This is the waveform that you obtain if you downsample the whole waveform by a factor of 16 including the on-pulse region. This waveform is 52.4288 ms long but contains only 2048 data points. The Downsampled (DS) and On-Pulse (OP) waveforms are shown in Fig. 4.3.

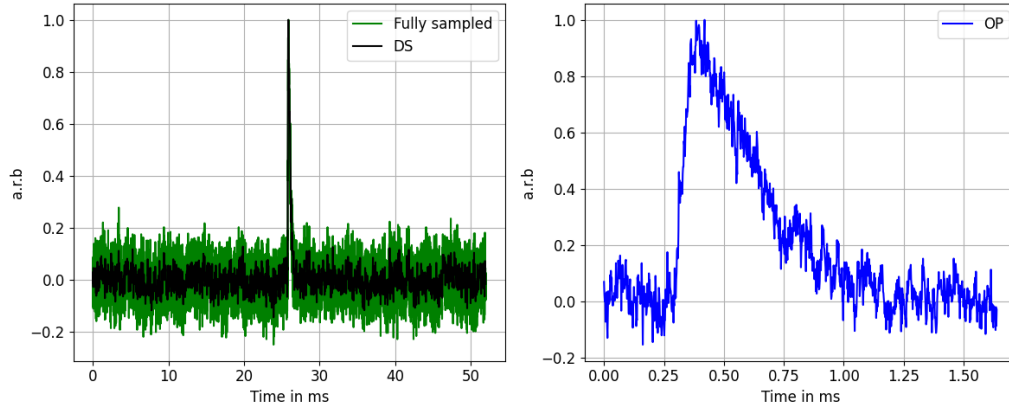


Figure 4.3: Plots showing the uniformly downsampled (DS) and on-pulse (OP) waveforms. The DS pulse contains 2048 data points obtained by replacing every 16 points in the 32768 bins long, fully sampled waveform with their average. The OP pulse is obtained by extracting the central 1024 time bins from the fully sampled waveform.

We will focus primarily on a discussion of phonon channels in this thesis. The reader should note that the same discussion holds for charge channels in terms of reconstruction methods and generic features, even though they are not explicitly discussed in this thesis.

Channel	Pre-pulse length (ms)	On-pulse length (ms)	Post-pulse length (ms)
Phonon	26.2145	1.6384	24.5761
Charge	26.2144	0.4096	25.8048

Table 4.2: Table of pre-, on- and post- pulse lengths for a hybrid waveform for each channel type.

With this hybrid readout scheme, the data throughput per DCRC that we expect for data sets with pulses is reduced by a factor of roughly 12 (see Table 4.3). Under this scheme, noise data will still be fully sampled to allow us to reconstruct PSDs without aliasing effects which will be discussed in the next section.

The 625 kHz sampling frequency in the on-pulse region of the phonon trace offers a superior

4.2. HYBRID READOUT – FEATURES AND CHALLENGES

Data type	Expected max event rate	Max event size per detector	Max throughput
Calibration	5 Hz	0.14 MB	0.7 MB/s
Dark matter	2 Hz	0.14 MB	0.28 MB/s
Noise	0.5 - 2 Hz	1.8 MB	0.9 - 3.6 MB/s

Table 4.3: Estimated maximum data rates and throughput for SuperCDMS SNOLAB for hybrid sampled 52.43 ms waveforms.

time resolution of $1.6 \mu\text{s}$. This allows us to pin down the time delay parameter in an optimal filter with great precision. The downsampled regions, on the other hand, have their time resolution at $25.6 \mu\text{s}$, worsened by a factor of 16. This means that time delay estimates must be performed exclusively on the on-pulse region. Contrary to the time resolution, the DS waveform has a superior frequency resolution of roughly $\sim 19.0735 \text{ Hz}$ compared to the OP waveform, which has a frequency resolution of only $\sim 610.35 \text{ Hz}$. Due to the differences in their sampling rates, the uniformly downsampled and on-pulse waveforms span the lower and higher frequency regions of the PSD with a small overlap. This is shown in Fig. 4.4 and will be explained in more detail in the next section. From the figure, it is evident that the downsampling procedure acts effectively like a low pass filter. Our pulses peak in the region $< 10 \text{ kHz}$, which means that the DS pulse contains predominantly energy information.

While the hybrid readout scheme allows our experiment to meet its targets, it introduces new challenges in the area of event reconstruction. Optimal filters (OFs) are our primary reconstruction techniques due to their relatively better performance at the lowest energies. In the discussion in Sec. 2.4, it was noted that optimal filtering is performed in the frequency domain to avoid correlations between different time bins. Due to the difference in sampling frequency in the full waveform, it is non-trivial to perform a Fourier transform of a hybrid-sampled waveform. Methods for non-uniform discrete Fourier transforms exist [116, 117] which could potentially be applied but are computationally expensive. William Page derived a slightly more robust technique to deal with this which will be discussed in the next section [118]. Even with Page’s improved algorithm, this computation is still expensive for the large amounts of data we plan to take at SuperCDMS SNOLAB. Other semi-empirical ways of fitting hybrid waveforms were studied as part of this thesis and were found to perform well while reducing computational complexity. These new techniques will be described in detail in the rest of this chapter.

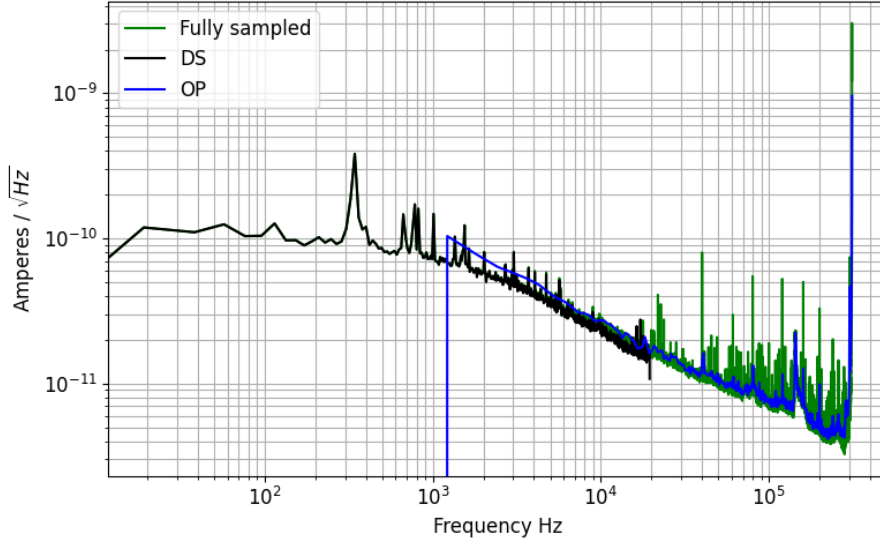


Figure 4.4: A comparison of noise PSD between the fully sampled, downsampled (DS) and on-pulse (OP) waveforms.

4.3 Event reconstruction with hybrid data

In this section, I will present reconstruction algorithms developed to process hybrid pulses.

4.3.1 Hybrid Optimal Filter

According to the Nyquist theorem, any waveform sampled at twice the Nyquist frequency can be fully reconstructed [115]. SuperCDMS waveforms satisfy this criterion with a sampling rate of 625 kHz. All frequencies from 0 to 312.5 kHz are perfectly reconstructed in our fully sampled noise PSD. For a fully sampled waveform with stationary noise¹, the inter-frequency correlations are zero which allows us to conveniently perform an optimal filter in the frequency domain via Fourier transformed signals and templates. In hybrid data, a single waveform contains three different regions sampled at two different rates (see Fig. 4.2). The on-pulse region sampled at 625 kHz spans the same range (0 to 312.5 kHz) as the fully sampled 52 ms waveform but has a frequency resolution of 610.35 Hz, about 32 times worse than the resolution of 19.0735 Hz obtained from the full waveform. The downsampled regions, on the other hand, equal the fully sampled waveform in terms of frequency resolution, but due to the low sample rate of 39.0625 kHz can reconstruct only up to 19.53125 kHz. This is also clearly shown in the comparison of the OP and DS PSDs in

¹noise whose statistical properties are time invariant

Fig. 4.4. From the figure and from noting the frequency span of the on-pulse and downsampled regions, there is an overlap from about 610.35 Hz to 19.53125 kHz. Further, any frequencies larger than 19.53125 kHz in the downsampled region alias power into lower frequencies. Thus, frequency correlations between the on-pulse and downsampled regions arise and they limit our reconstruction capacity.

In the presence of any kind of correlated data, the χ^2 in Eq. (2.9) must take into account a covariance matrix that captures the correlations, in place of the noise PSD. Then, Eq. (2.9) becomes:

$$\chi^2 = \sum_{i,j} [s_i - A \cdot v_i]^\dagger V_{i,j}^{-1} [s_j - A \cdot v_j], \quad (4.1)$$

where $V_{i,j}^{-1}$ represents the inverse of the covariance matrix which captures the frequency correlations between frequency indices i and j . The superscript $\dagger$ represents complex conjugation. Note that we have left out the time delay term in Eq. (4.1) for now. A formal investigation of methods to perform this χ^2 minimization was performed by William Page [118].

In order to motivate the need for new methods it is necessary to discuss some of the features and limitations of Page's Hybrid Optimal Filter (HOF). The χ^2 in Eq. (4.1) in matrix notation gives:

$$\chi^2 = [\vec{s} - A\vec{v}]^\dagger \cdot \mathbf{V}^{-1} \cdot [\vec{s} - A\vec{v}], \quad (4.2)$$

which can be minimized analytically to obtain the amplitude $\hat{A}$ as:

$$\hat{A} = [\vec{v}^\dagger \mathbf{V}^{-1} \vec{v}]^{-1} \vec{v}^\dagger \mathbf{V}^{-1} \cdot \vec{s}. \quad (4.3)$$

In Eq. (4.2), we have transformed the signal, s , and template, v , into vectors, absorbing the summation in Eq. (4.1) into the matrix notation. From Table 4.2, a phonon hybrid waveform will contain $26.2145 \text{ ms} \cdot 39.0625 \text{ kHz} + 1.6384 \text{ ms} \cdot 625 \text{ kHz} + 24.5761 \text{ ms} \cdot 39.0625 \text{ kHz} = 3008$ data points. So $\vec{s}$ and $\vec{v}$ are column vectors of dimension 3008×1 . The covariance matrix, $\mathbf{V}$, captures the correlations between each of these 3008 time bins in the frequency domain and has a dimension of 3008×3008 . Also note that the elements of $\mathbf{V}$ are complex in nature. Fourier coefficients have a mirror symmetry about the zero frequency which means that half the coefficients are simply the reflection of the other half about the zero frequency. Therefore, the actual number of independent components for an FFT of length N is $N/2$. Thus the matrix, $\mathbf{V}$, has 1505 independent components in the frequency domain. The calculation of this matrix is non-trivial and discussed in more detail

in Ref. [1].

Since Eq. (4.3) is derived in the frequency domain, $\vec{s}$ and $\vec{v}$ are naturally the Fourier transformed signal and templates. But as discussed earlier, performing a Fourier transform of a non-uniformly sampled waveform is non-trivial and computationally expensive. Page explored alternative ways to perform this transform. One such avenue was “frequency pasting”. In this technique, the OP pulse was used to reconstruct frequencies above 19.531 kHz, while the DS pulse was used to reconstruct the low frequency components (≤ 19.531 kHz) with higher precision. This technique failed due to power aliasing which causes any non-zero frequency components larger than the Nyquist frequency to produce false peaks at frequencies lower than the Nyquist frequency.

Further, it was noted that the prime advantage of performing a frequency domain fit is lost with the hybrid readout scheme due to the covariance matrix, $\mathbf{V}$, being non-diagonal. The operation of constructing a Fourier transform for non-uniformly sampled data adds additional time and space complexity to our computational problem. Therefore, it was concluded that the HOF ought to be performed in the time domain saving us the trouble of producing Fourier transforms. Then, the summation over frequencies in Eq. (4.1) is replaced by a summation over time bins. Minimizing this time domain χ^2 gives

$$\hat{A} = [\vec{v}^T \mathbf{V}^{-1} \vec{v}]^{-1} \vec{v}^T \mathbf{V}^{-1} \cdot \vec{s}, \quad (4.4)$$

where we have notably changed the complex conjugation, $\dagger$, to a transpose, T . This is because in the time domain, the elements of $\vec{v}$, $\vec{s}$ and $\mathbf{V}$ are all real numbers. The corresponding matrix notation of the signal and template contains the time domain versions and the covariance matrix $\mathbf{V}$ is also constructed in the time domain. The remaining derivation and expression for $\hat{A}$ still hold with these simple replacements.

Although we have now surpassed the complexity of non-uniform Fourier transforms, we are still faced with the complexity of inverting a 3008×3008 matrix. Per computational theory, the time taken to invert a matrix of dimension $n \times n$ increases as n^3 . Cholesky decomposition [119] is a popular technique for matrix inversion which was adopted into the HOF framework.

The HOF was able to demonstrate as good a performance as the simple 1×1 OF described in Ch. 2 at the cost of computational power. Further, Page estimated that to achieve this level of accuracy, the covariance matrix computation requires additional noise data acquisition resulting in a ~ 6.5 -fold increase in our estimated data volume, which is not optimal. These are the two major limitations of the HOF.

The third point brings us back to the missing time delay term in Eq. (4.1). The time delay estimation is simplified in the frequency domain. Consider the 1×1 OF amplitude in Eq. (2.10).

If we define the normalized Optimal Filter as

$$\phi_j = \frac{v_j}{J_j} / \sum_k \frac{|v_k|^2}{J_k}, \quad (4.5)$$

then the amplitude, $\hat{A}$, can be re-written as a function of time delay

$$\hat{A}(t_0) = \sum_j e^{i\omega_j \hat{t}_0} \cdot \phi_j \cdot s_j. \quad (4.6)$$

Eq. (4.6) presents the amplitude as simply the inverse discrete Fourier transform of the optimal filtered waveform. The χ^2 can also be similarly re-written as:

$$\chi^2 = \sum_k \frac{|s_k|^2}{J_k} - \hat{A}^2(t_0) \cdot \sum_k \frac{|v_k|^2}{J_k}. \quad (4.7)$$

Thus, the amplitude as a function of time can be computed and a scan over the time values can be performed. The maximum amplitude naturally minimizes the χ^2 in Eq. (4.7) and the corresponding best fit time delay, $\hat{t}_0$, is the value that maximizes the amplitude function, $\hat{A}(t_0)$.

In the HOF, we lose this advantage of easy mathematical manipulation to obtain $\hat{t}_0$ due to the fit being performed in the time domain. This means that one ought to extract the time delay first from the OP waveform due to its superior time resolution and perform an additional step to align the signal and the template before performing the HOF. This would be perfectly fine, but the additional complexity added by the matrix inversion renders the full process inefficient and time consuming.

The HOF, although working as well as the 1×1 OF, thus requires more computational power and requires additional disk space and memory allocation. Further, the time delay component is not directly obtained in this formalism.

4.3.2 2-speed Optimal Filter

Given the limitations of performing a HOF, it becomes necessary to investigate alternative approaches to fit a hybrid waveform. It is important to acknowledge at this point that the mathematically accurate method of fitting a hybrid pulse is by minimizing the χ^2 in Eq. (4.2). Any alternate approach is only an approximation to this true solution. The goal, however, is to find a close enough approximation that performs as well as fitting the fully sampled pulse while saving on computational power and data volume. Further, the techniques discussed in this chapter assume that the pulses can be modeled with a single template, which is not true in general. Performing

a multi-template fit requires advanced reconstruction techniques which will be discussed in later chapters.

The 1×1 OF works reasonably well in estimating the pulse amplitude. The DS pulse discussed in Sec. 4.2 contains predominantly energy information. This estimate is however closely tied to a proper time delay estimate as is evident from the discussion in Sec. 4.3.1. The templates are usually defined at the trigger point, i.e. the point where the threshold is crossed in the trigger. This number can usually be offset by a few bins to the left or right and could mean that the start of the pulse is off by a bit. The time delay is a measure of this time offset. In the cartoon in Fig. 4.5, the red curve shows the true fit from extracting the right time delay via optimal filtering. When the time delay is estimated poorly, e.g. with insufficient time resolution, the most common outcome is that the fitter attempts to align the rising edge of the pulse with the template which might end up underestimating or overestimating the amplitudes, thus limiting our reconstruction capability.

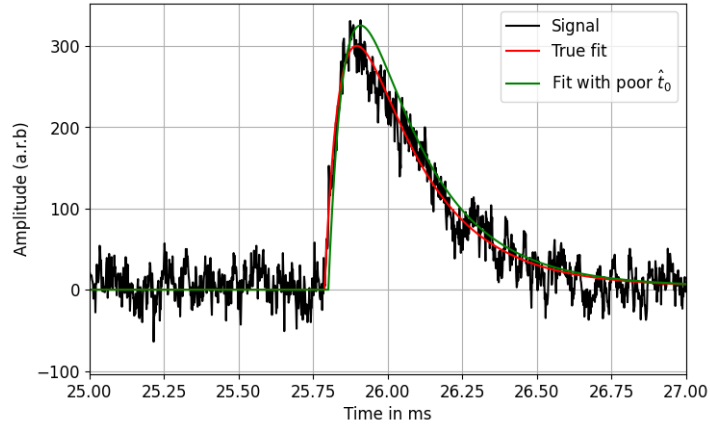


Figure 4.5: A cartoon comparing the true OF fit with a poor time delay fit.

Therefore, in order to estimate the amplitude from the DS pulse (which we will call DSamps from now on), we require a time delay estimate. For hybrid data, the 16-fold enhanced time resolution of the OP region allows us to extract a reasonable time delay estimate from it, similar to the HOF. The OP pulse, however, has poor frequency resolution and can only provide a limited, perhaps even worse energy/amplitude resolution compared to the DS pulse. If the DS pulse has energy information and the OP pulse has time information with some energy information, is there some empirical method that allows us to get the best of the two?

The 2-speed Optimal Filter (2SOF) was developed precisely to answer this question. As the name suggests, the 2SOF performs optimal filtering on waveforms sampled at two different speeds.

One of the waveforms in question is the DS pulse and the other is the OP pulse. As a preliminary step, the OP pulse is subjected to optimal filtering. In order to do this, the necessary quantities in Eq. (2.10) must be first calculated for the OP pulse. The templates are constructed for the fully sampled waveform and contain 32768 data points for a 52 ms phonon pulse. The OP region spans bins 1024 to 2047, a length of 1024 data points. This region is extracted from the fully sampled template to obtain an OP template. The noise data is sampled at the full frequency even with hybrid readout. This is meant to help us reconstruct our noise perfectly in the regions of interest without any power aliasing. The OP region of the noise is extracted the same way as the template. An OP PSD is constructed with the extracted OP trace. This PSD looks similar to the OP PSD in Fig. 4.4 in terms of frequency span. The optimal filtering process produces an amplitude for the OP pulse, called the OPamp, and a time delay, $\hat{t}_0$.

The next step in the processing chain is running an OF on the DS pulse. In order to obtain useful results from this fit the pulse must be aligned well with the template as discussed earlier. Typically, this is achieved via the inclusion of a time delay parameter in the χ^2 , but we've already discussed why this is not feasible for the DS waveform. Therefore, prior to optimal filtering, the fully sampled template is shifted by the time delay estimated from the OP fit, $\hat{t}_0$, so that the templates and pulse align. A cartoon of the time shifting procedure is shown in Fig. 4.6.

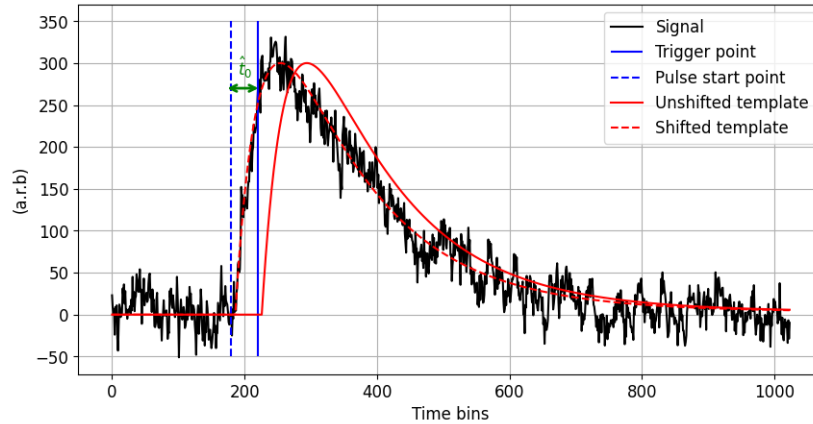


Figure 4.6: A plot showing the effect of template shifting in optimal filtering. The template is usually aligned such that the rising edge begins at the trigger point of the pulse. The time delay estimate, $\hat{t}_0$, is a measure of the difference in start point between the template and the signal. In order to perform an optimal filtering of the DS pulse, the OP pulse is first used to extract $\hat{t}_0$ with which the fully sampled template is shifted and then down-sampled.

Once shifted, the template is downsampled by averaging every 16 bins to obtain the DS template. Similarly, the fully sampled noise trace is downsampled to obtain the downsampled noise trace from which a “DS PSD” is constructed (see Fig. 4.4). Figure 4.7 compares the DS and OP templates to the fully sampled one.

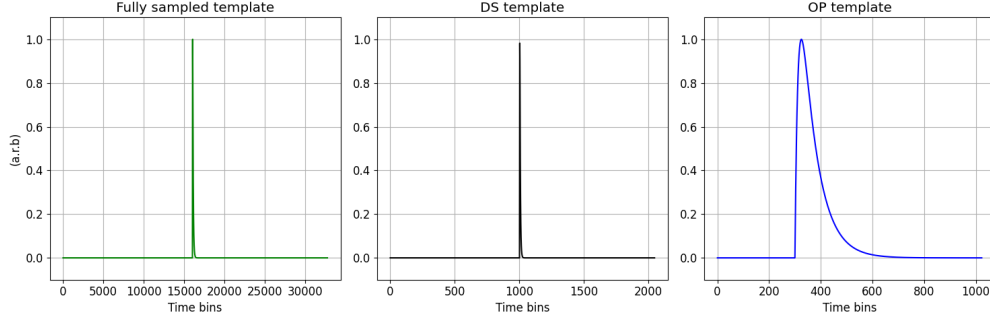


Figure 4.7: A comparison of the fully sampled, downsampled and on-pulse templates used in the 2SOF algorithm. The latter two templates are obtained via downsampling and chopping off tails, respectively, of the fully sampled template as described earlier. The key difference is seen in the x axes of the plots. The fully sampled template has a length of 32768 time bins, the DS template has a length of 2048 time bins and the OP template has a length of 1024 time bins. Note that the OP template is simply a zoomed in version of the fully sampled template.

The optimal filtering of the DS pulse yields an amplitude, which we will call DSamp. The time delay produced in this fit is zero due to the pulse being aligned closely with the template. Even if there is a small offset, the DS filtering does not have the capability to resolve features less than $25.6 \mu\text{s}$.

4.3.2.1 Wrap-around effect

A simple caveat needs to be considered before jumping to the next step in the 2-speed OF. Although the fall time of the pulse is expected to keep the pulse-like part of the signal in the OP region, SuperCDMS has seen cases in the past where this assumption does not hold. A well known case in the collaboration is from a prototype Ge detector constructed for SuperCDMS SNOLAB. During preliminary testing, this detector showed fall times up to 3 orders of magnitude larger than the expected pulse shape. Although this detector qualifies as an anomaly with potential fabrication flaws, it is generally good to anticipate and prepare for scenarios where the falling tail of the pulse is not fully enclosed within the OP window. An example pulse is shown in Fig. 4.8. The tail of the

pulse gets abruptly cut off in such cases, when hybrid sampling is applied.

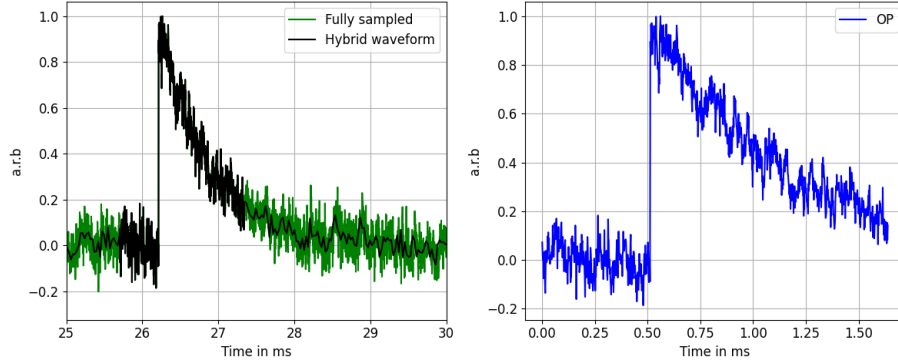


Figure 4.8: An example of a hybrid pulse with a long fall time. The long fall time makes it impossible to capture the full pulse within the OP window.

Thus, the OP pulse need not begin and end at “zero”. In optimal filtering, one works in the frequency domain, effectively assuming that the signal is periodic, in which case its FFT transforms under a time shift t_0 as:

$$s(\omega, t_0) = e^{-i\omega t_0} \cdot s(\omega, t_0 = 0). \quad (4.8)$$

In the 2-speed OF, we shift templates and not the signal by making the same implicit assumption. Note that it is more convenient to shift templates since they are effectively noise-free with tails that go to zero. The alternate approach of fixing the template and shifting the pulse, although in principle possible, could mix up the non-zero noise tails introducing unwanted correlations. But when the template itself does not end in zero, the shifting procedure “wraps” the template around in the time domain for every phase shift in the frequency domain. Thus, Eq. (4.8) does not hold for templates where the tails are cut off by downsampling. This means that the OP optimal filtering with the built-in assumption of Eq. (4.8) cannot estimate the time delay in the frequency domain accurately for such pulses. A demonstration of this wrap-around effect is shown in Fig. 4.9.

The solution to this problem is to perform an extended time delay search, removing the assumption in Eq. (4.8) from the OF. We accomplish this by producing an extended OP template by adding more bins on either end of the template. We generally do not expect our SuperCDMS pulses to be delayed by more than 100 time bins. Pulses that do go beyond that, even though rare, are removed from analysis via cuts. Therefore, we extend the template by, say, 50 bins on either side. Although, the tail of the template may still be cut off with this small extension it no longer affects the time

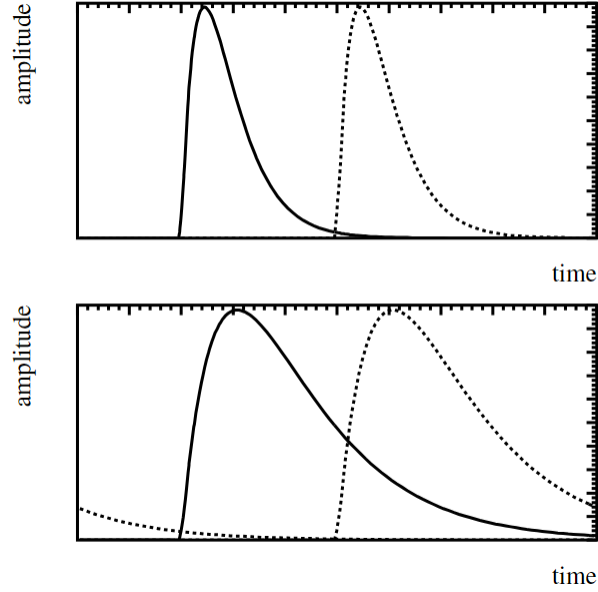


Figure 4.9: A demonstration of the “wrap-around” effect in time shifted pulses. The solid line shows an example pulse and the dashed line shows the same pulse shifted by a certain amount. In the top plot, the pulse begins and ends in zero preventing the wrapping around of the waveform when time-shifted. In the bottom plot, the time domain shift of the pulse wraps the non-zero, right tail of the pulse to the left creating the wrap-around effect. Figure taken from Ref. [1].

delay estimation since we are no longer performing phase rotations in the frequency domain. This also has the consequence that the convenient form in Eq. (4.7) does not hold, i.e., the minimum χ^2 is no longer guaranteed to be produced by the maximum amplitude. A series of sub-templates are generated from this extended OP template as shown in Fig. 4.10.

Each generated sub-template has the same length as the OP template, which is chosen to be 1024 in the figure. The first sub-template selects the first 1024 bins, starting from bin 0 to bin 1023. The second sub-template begins from the 1st bin and goes to the 1024th bin, skipping the zeroth bin. The third sub-template starts from the 2nd bin, the fourth sub-template from the 3rd bin and so on till the last bin is included once. Thus, by extending the template by 100 bins, we obtain 101 sub-templates, each delayed from the prior sub-template by one time bin (or 1.6 μ s).

We perform a manual scan over all sub-templates to identify the one delayed by the same amount as the signal. Remember that this is completely different from the regular 1×1 OF which uses the inverse FFT to obtain the amplitude as a function of time delay (see Eq. (4.6)). Here, we

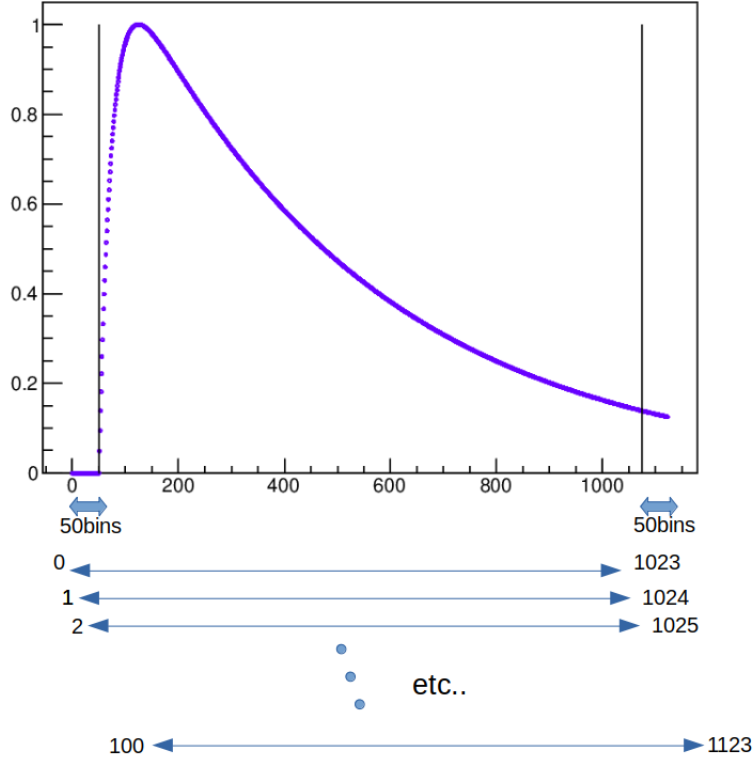


Figure 4.10: A cartoon demonstration of generating sub-templates from an extended OP template. The OP template is marked by the black solid lines. This template is extended by 50 bins on each side. A series of sub-templates are generated from the extended template.

compute the χ^2 for each sub-template and identify the template which produces the minimum χ^2 . The χ^2 here is defined as

$$\chi^2(t_0) = \frac{\sum_i |s_i - A \cdot v_i(t_0)|^2}{J_i}, \quad (4.9)$$

where we have absorbed the phase shift into the FFT of the template, v , since the relationship in Eq. (4.8) does not hold in general. As an example, say you have a number of sieves with different sized holes that filters signals. You drop your signal through each sieve to figure out which sieve will let your full signal pass through. The manual scanning procedure to find the optimal time delay is equivalent to this sieve example with each sieve being a sub-template and the hole size being a

different time delay. This manual delay scan procedure succeeds in estimating time delays for the most generic pulses without making particular assumptions about the underlying templates.

4.3.2.2 Combining amplitudes

Thus, we have an OPamp and $\hat{t}_0$ from the OP pulse fit and a DSamp from the DS pulse fit. The simple 1×1 OF does not account for the frequency correlations between the OP and DS pulses. The HOF, on the other hand, does take these correlations into account via a covariance matrix at the cost of computational power. For the 2SOF to be truly beneficial, one has to take the same correlations into account but more efficiently than the HOF. So far we have obtained two amplitudes. A simple scatter plot of the two reconstructed amplitudes, such as Fig. 4.11, reveals that they are strongly correlated. The question then becomes whether we can exploit these correlations to improve our final energy estimator.

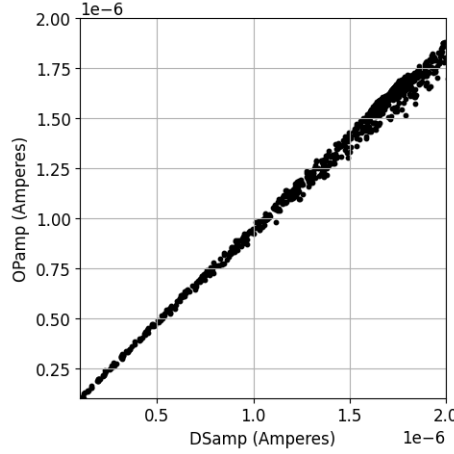


Figure 4.11: A scatter plot between the estimated amplitudes from the DS pulse, DSamp, and the OP pulse, OPamp. The two quantities are clearly correlated from fitting the same pulse. However, notice that they are not perfectly correlated which indicates that there is scope for improvement from considering a combination of these two quantities.

Consider the true amplitude of the pulse, A_0 . Let $\hat{A}_0$ be the best estimate of A_0 that we could obtain from the hybrid sampled waveform. For the same pulse, let us also denote DSamp by A_{DS} and the OPamp by A_{OP} . A_{DS} and A_{OP} represent the best possible estimates of A_0 that we could obtain from the DS and OP pulses, respectively. The difference between the best possible estimate, $\hat{A}_0$ and A_{DS} or A_{OP} gives an estimate of the error in each of our estimate. One could construct a χ^2 such that minimizing this error returns the best fit $\hat{A}_0$. We construct the following χ^2 :

$$\chi_{DS-OP}^2 = \begin{bmatrix} A_{DS} - \hat{A}_0 & A_{OP} - \hat{A}_0 \end{bmatrix} \cdot \mathbf{V}_{DS-OP}^{-1} \cdot \begin{bmatrix} A_{DS} - \hat{A}_0 \\ A_{OP} - \hat{A}_0 \end{bmatrix}, \quad (4.10)$$

where $\mathbf{V}_{DS-OP}^{-1}$ is a covariance matrix that captures the correlations between A_{DS} and A_{OP} . In this approach, rather than explicitly accounting for correlations between frequencies introduced by the downsampling, we instead capture the correlations between the amplitudes determined from the DS and OP pulses. This approach throws away some information but is computationally much simpler than the HOF. Thus, weighting the amplitudes by the respective variances and covariances allows us to correct for their resolution broadening effects. The matrix $\mathbf{V}_{DS-OP}^{-1}$ is always a 2×2 matrix which is a lot easier to invert compared to the 3008×3008 matrix in the HOF. The dimensions of $\mathbf{V}_{DS-OP}^{-1}$ are independent of the tracelength which makes it even more advantageous and efficient than the covariance matrix in the HOF. Further, estimation of this covariance matrix does not require additional data collection, but rather it can be estimated from toy model simulations which will be discussed in the next section.

The best fit amplitude, $\hat{A}_0$, is obtained by minimizing the χ^2 in Eq. (4.10). In order to simplify this minimization by avoiding further matrix algebra, we define the elements of $\mathbf{V}_{DS-OP}^{-1}$ and rewrite Eq. (4.10) in terms of these defined elements. We define the elements of the diagonal matrix, V_{DS-OP}^{-1} as

$$V_{DS-OP}^{-1} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}. \quad (4.11)$$

Minimizing Eq. (4.10) with respect to $\hat{A}_0$, $\frac{d\chi^2}{d\hat{A}_0} = 0$, we get,

$$\hat{A}_0 = \frac{(a+b)A_{DS} + (b+c)A_{OP}}{a+2b+c}. \quad (4.12)$$

The above clean equation allows us to directly evaluate the best fit estimator, which we call the combined amplitude, $\hat{A}_0$. $\hat{A}_0$ has been found to improve upon A_{DS} and A_{OP} as you will see in the next sub-section. We have thus decomposed the problem of evaluating large covariance matrices into a simple semi-empirical solution of a covariance weighted sum between the DS and OP amplitudes. By taking the correlations into account we are able to correct for the frequency correlations introduced by hybrid sampling. This method is also computationally efficient and keeps our data volume for SuperCDMS SNOLAB at bay without sacrificing on energy resolution and data acquisition threshold.

The overall flowchart of the algorithm is shown in Fig. 4.12 in the next page.

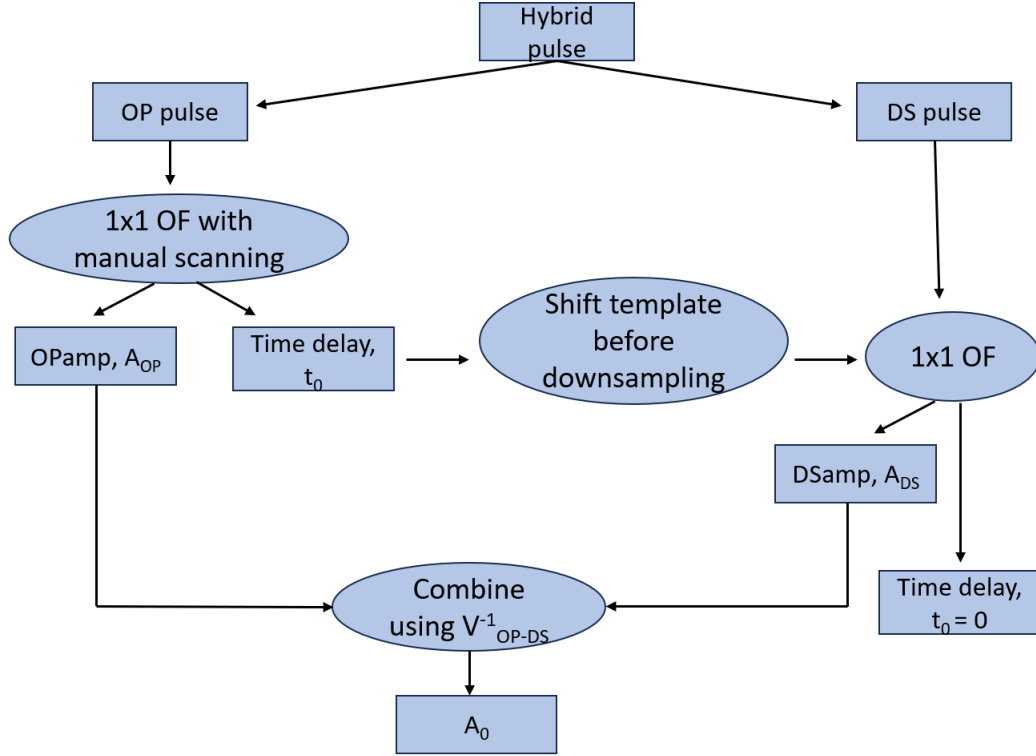


Figure 4.12: A flow chart outlining the steps in the 2-speed OF algorithm. The algorithm begins by extracting an OP and DS pulse from a given hybrid waveform. The 1×1 OF with manual scanning is first applied to the OP pulse to produce an amplitude, A_{OP} , and a best fit time delay, $\hat{t}_0$. $\hat{t}_0$ is used to shift the fully sampled template to align it with the DS pulse which is then downsampled to match the appropriate sample rate. The DS pulse is then optimal filtered to produce a best fit amplitude, A_{DS} . The time delay from the DSOF will be zero due to the pulse being closely aligned with the template. A_{DS} and A_{OP} are correlated due to frequency correlations between the DS and OP pulses. We capture this correlation in a covariance matrix which is then used to obtain the best fit estimator, $\hat{A}_0$.

4.3.2.3 Algorithm validation with toy models

In order to determine the efficiency of the 2SOF, toy models of varying complexity were generated and processed with the 2SOF. Pulses in SuperCDMS are a superposition of some level of noise and a signal with a rising and falling edge. We often approximate the latter via our templates which could be derived from the data itself or produced through some analytical functions. A discussion of some commonly used analytical signal templates will be provided in the description of the 2-speed Minimizer in Sec. 4.3.3.

The noise, to first order, can be assumed to be Gaussian with a mean zero and some standard deviation, σ . The average noise PSD represents the mean variance of the noise at each frequency. One could randomize this variation by sampling from a normal distribution whose mean is zero and variance equals the noise PSD at that frequency. Thus, we could construct a smeared FFT coefficient at frequency f as

$$\tilde{F}(f) = N(\mu, \sigma(f)) + iN(\mu, \sigma(f)), \quad (4.13)$$

where N is a Gaussian distribution whose mean μ will be zero. Two random numbers are generated at each frequency: one constitutes the real part while the other constitutes the imaginary part of the complex FFT coefficient at that frequency. The procedure is repeated to obtain a series of complex coefficients over the full frequency range. This is the full randomized FFT. Each of the randomized FFT produced in this manner corresponds to a noise trace and thus constitutes a randomly sampled noise event from an averaged PSD. If the average PSD is generated using real noise traces, then it represents the true noise of our detector environment. The randomly generated FFT is then inverted by performing an Inverse Fast Fourier transform which converts the complex frequency coefficients to a corresponding time domain noise trace. This procedure is quite effective in capturing the features at each frequency into a corresponding noise trace. In principle, the average PSD could be sampled an infinite number of times subject to the efficiency of the random number generator in use. This method, thus, is a Monte Carlo simulation of noise. Example noise traces generated in this manner are shown in Fig. 4.13. If sufficient number of traces are generated to compute an average noise PSD, we reproduce the original noise PSD which was used to generate the traces in the first place. This is shown in Fig. 4.14.

A “signal” is generated as the sum of the amplitude-scaled template and the simulated noise. The templates are normalized to their maximum height so that the amplitude used in scaling represents the height of the pulse. The simulated pulse is given by

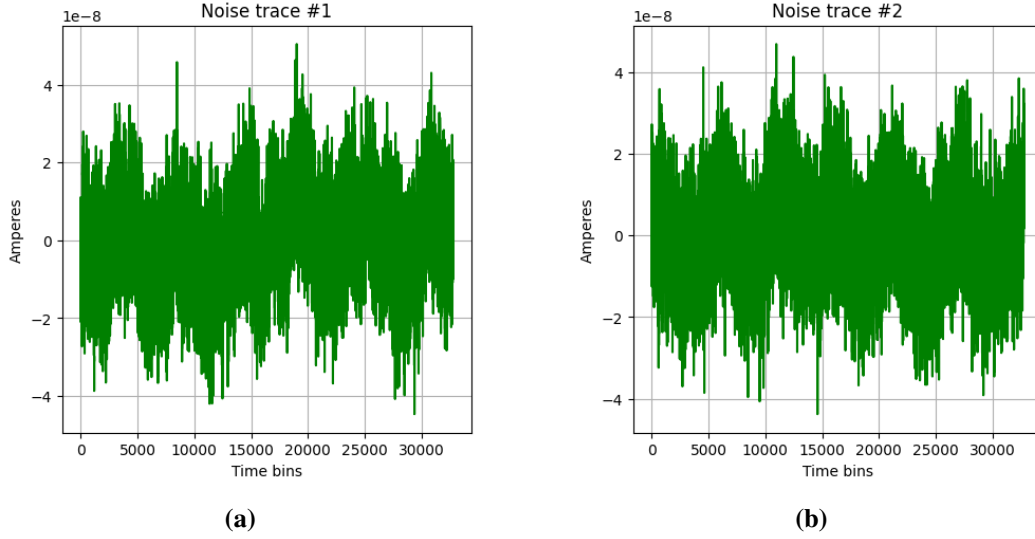


Figure 4.13: Example noise traces generated via Monte Carlo sampling. The difference between the two traces is purely random (see Eq. (4.13)).

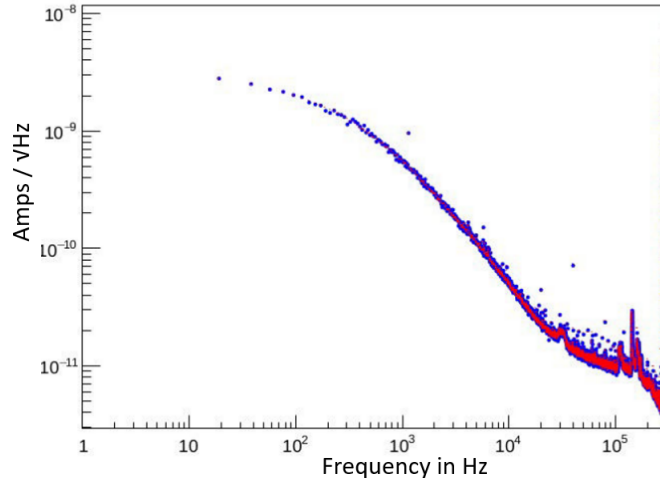


Figure 4.14: A comparison of the averaged PSD between simulated noise samples and original noise samples. The blue PSD is obtained by calculating the PSD of each Monte Carlo sampled noise trace and computing their mean. The red curve is the PSD that was used to generate the Monte Carlo samples. The average PSD of the generated samples matches the PSD they were originally generated from.

$$S_{sim}(t|A) = m(t) + A \cdot v(t). \quad (4.14)$$

In Eq. (4.14), $m(t)$ is the simulated noise which introduces the randomness. A is the amplitude of the simulated pulses and $v(t)$ is the template used to generate the pulses. A cartoon demonstration of Eq. (4.14) is shown in Fig. 4.15.

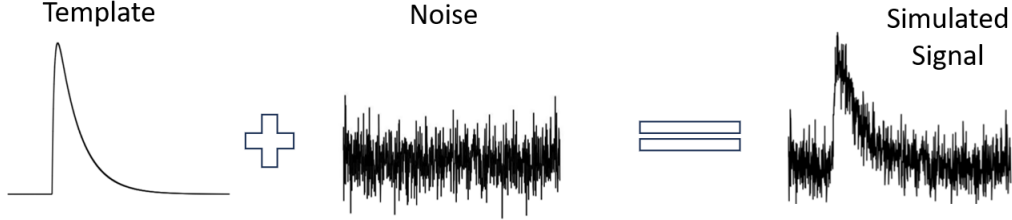


Figure 4.15: A cartoon demonstration of simulating a signal from a noise trace and a pre-defined template.

The amplitude in Eq. (4.14) could be a constant in which case the simulated pulses would be termed mono-energetic. In this case the distribution of OF amplitudes is a Gaussian centered at the magnitude of A . The smearing is caused by the uncertainty in measurement introduced by the noise. One could also construct a set of pulses, each with a different amplitude by randomizing A and sampling from a distribution. The most commonly used distribution is uniform. In this case, the estimated amplitudes will follow a distribution similar to the underlying distribution of A smeared by the reconstruction resolution.

The more efficient way to evaluate an algorithm is to use a mono-energetic sample. The width of the Gaussian represents the ability of the algorithm to correctly place the amplitude close to the true A which is given by the mean. A better algorithm produces a thinner peak exhibiting excellent signal discrimination power. Performing a similar analysis on uniformly sampled A requires a few additional steps to “Gaussian-ize” the uniform spectrum by estimating the difference between the true and fitted amplitude of each event. The “error” constructed in this manner will be Gaussian and can be used to perform a similar assessment. In this thesis, we stick to Gaussian features.

So far we have only considered the case of signals which start exactly at the same point as the input template. The 2SOF has been designed to be robust to cases of pulses with a non-zero time offset. For a true evaluation of this criterion, we must generate pulses which also contain a time offset relative to the template used in the processing. To do this we modify Eq. (4.14) in the following way:

$$S_{sim}(t|A, t_0) = m(t) + A \cdot v(t, t_0). \quad (4.15)$$

In Eq. (4.15), the signal, $S_{sim}(t)$ becomes a function of time delay by means of a time-shifted template. The delay parameter, t_0 , is sampled randomly from a uniform distribution. Good pulses typically only deviate from the trigger point by small amounts, of the order of 10's of μs . In the study documented in this thesis, we sample t_0 in the range of $[-25.6 \mu s, 25.6 \mu s]$.

The simulated fully sampled traces are then “hybridized” by downsampling their tails by a factor of 16. About 10000 events with true amplitude $A = 0.003 \mu A$ are simulated in this manner and processed with the 2SOF. The covariance matrix is calculated directly from A_{DS} and A_{OP} for these events. Four different output quantities are studied – A_{DS} , A_{OP} , the 1×1 OF on the full trace (denoted simply by “ 1×1 OF”) and the combined estimator, A_0 . The spectra are compared in Fig. 4.16, but however it is difficult to gauge which estimator performs better simply from this plot.

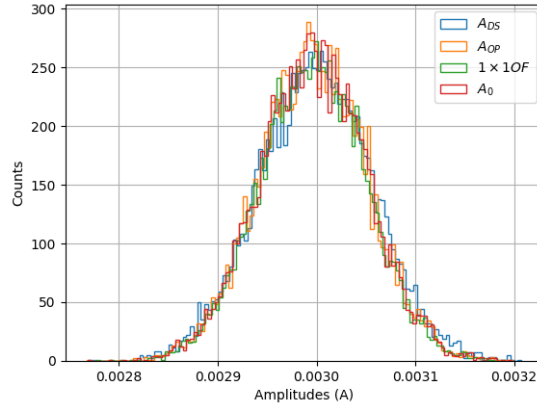


Figure 4.16: A comparison of the energy spectra from various estimators for mono-energetic toy pulses.

I fit each of the distributions in Fig. 4.16 with a Gaussian of the form

$$h \cdot \exp\left(-\frac{(\hat{A} - \mu)^2}{2\sigma^2}\right),$$

where h is a normalization constant which accounts for the height of the distribution in units of bin count, μ is the mean of the distribution and σ is the standard deviation or the width of the Gaussian. The results of this fit are reported in Table 4.4.

The widths obtained from a Gaussian fit to the spectra in Fig. 4.16 are converted to a percentage

Estimator	% resolution on the estimate
A_{DS}	$1.92 \pm 0.02 \%$
A_{OP}	$1.83 \pm 0.02 \%$
1×1 OF	$1.79 \pm 0.02 \%$
$\hat{A}_0$	$1.81 \pm 0.02\%$

Table 4.4: A comparison of reconstruction resolution between various energy estimators for a mono-energetic toy spectrum.

using, $\% \text{ resolution} = \frac{\sigma}{\mu} \times 100$. The numbers reported in the table are based on observations from running multiple trials of fitting with multiple sets of toy pulses. The results are a bit counter-intuitive. Given that the DS pulse contains predominantly energy information, it is surprising that A_{DS} has the worst resolution of all the estimators. There are two competing effects at play here. First is that the downsampling produces a low-pass filtering effect which enhances the signal-to-noise ratio. This factors in as the dominant effect at low energies. The second effect is the loss of information from downsampling, i.e., the factor of 16 reduction in amount of information in the pulse. At the lowest energies, which is our region of interest in a dark matter analysis, the low pass filtering effect of the DSamps dominates making it a better estimator than OPamp. This is the reason DSamps gives a better baseline resolution for real data as you will see in Sec. 4.3.2.4. You will also see this to be true in the case of toy pulses generated for a study of a special algorithm described in the next two chapters. At higher energies, we do not benefit from any low pass filtering of the pulse due to the large signal-to-noise ratio. The dominating effect here then is the loss of information. Thus, the pulses used in this study have higher amplitudes which causes a relative broadening of A_{DS} with respect to A_{OP} .

The OP pulse is also relatively insensitive to low frequency noise due to its frequency resolution of 610 Hz. Thus, which of the DS and OP pulses gives the better estimate is dependent on the pulse energy, pulse shape and the ambient noise. At a given moment, it is not straightforward to pick one of the two as the better one, however the combination is guaranteed to be the better energy estimate on average since it combines both OP and DS amplitudes, disregarding their individual performance at a given energy. The combined amplitude, $\hat{A}_0$ is within 9% of the amplitude estimated from the fully sampled waveform. This demonstrates that this technique provides us the benefits that we hope to gain from accounting for frequency correlations.

Before jumping to the next topic to present results from applying the 2SOF to a real data set, I would like to also show the distribution of estimated time delays here. In Fig. 4.17, I compare the

estimated time delay from the OP pulse and the fully sampled pulse. Since the delays are uniformly drawn, the estimated distribution is also uniform. We see the advantage of time delay estimation from the OP pulse here – it is equivalent to the estimate from the fully sampled pulse.

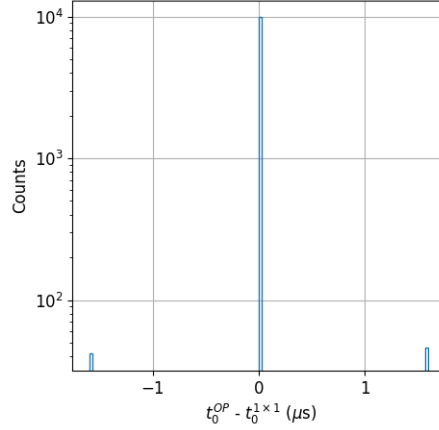


Figure 4.17: The distribution of difference in time delays estimated from the OP pulse and the fully sampled trace. Most events have the time delay estimated to perfect accuracy using the OP pulse. The handful of events showing deviations are numerically within one time bin difference. This corresponds to the time resolution of the OP and fully sampled pulses ($1.6 \mu\text{s}$).

4.3.2.4 Algorithm validation with real data

Here, I present results from applying the 2SOF to real data. The data used was taken during the testing of a prototype SuperCDMS device at CUTE. The device was a gram-scale photo-detector with a single readout channel, called PD2. The original data was collected in fully sampled mode and was later “hybridized” for my study by hand. An example pulse for this detector and the corresponding hybrid waveform are shown in Fig. 4.18.

A real energy spectrum from a detector is not mono-energetic but spans many values. A cartoon image of a SuperCDMS energy spectrum is shown in Fig. 4.19. Although the energy span and the particular features mostly vary from detector to detector, certain features remain common across all our devices. The first thing to notice is the “noise blob” at the left-most end of the spectrum. These are the lowest energy pulses or even noise which exceeded the trigger threshold. The pulses here, if present, are nearly invisible by eye but sensitive to optimal filtering. The name “noise blob” comes from the dominating presence of events triggering on random noise in the blob which diminish the

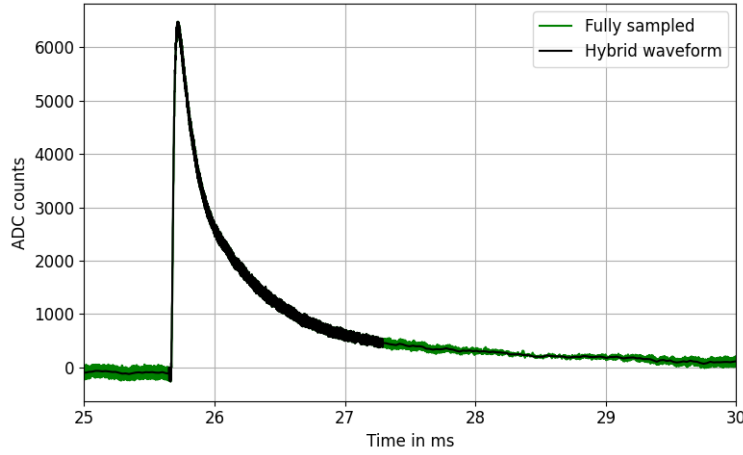


Figure 4.18: An example pulse from the PD2 detector compared to the hybridized pulse.

performance of our OFs. The ability of an OF to distinguish noise from a signal gives an estimate of its resolution.

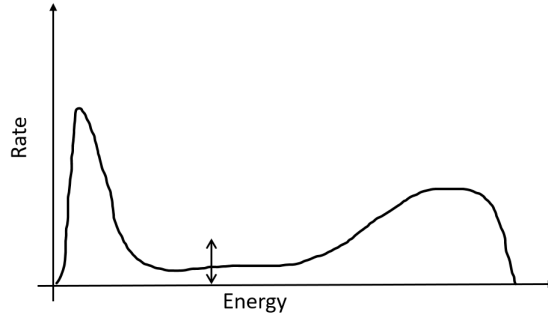


Figure 4.19: A cartoon showing a typical SuperCDMS energy spectrum. Three features stand out: (a) the left-most noise blob, (b) the central arrow representing a peak-like or a step-like feature, and the right-most saturation peak. The saturation peak shown here is a little exaggerated and is only that prominent in calibration data sets where the detector is directly exposed to a radioactive source.

A second common feature is the peak at the right-most end of the spectrum. The pulses in this broad, rising feature are almost certainly saturated, meaning that they are energetic enough to push the TES out of transition for a very short period of time before it cools back down. The extent of saturation determines the pulse shape going from slower fall time all the way to a characteristic flat top due to which nearly every OF we have fails to reconstruct their true shape and returns large χ^2

values for the fits. Thus, everything larger than a certain energy gets pushed into this right broad peak. Note that this is a false peak for which it is not possible to extract a true energy without complicated analyses. Such analyses are yet to be developed.

Other incidental features are often present in the spectra. These features could be peaks or steps depending on the detector material and are commonly used in detector calibration. The device whose pulses were used in this study had a spectrum with two calibration features. The two peaks were produced by inserting an ^{55}Fe source inside the shielding material to expose the detectors to X-rays. These X-rays are not penetrating for a full-sized SuperCDMS detector but work quite well for these small gram-scale devices. The X-rays come in two energies corresponding to the K_α and K_β energy levels of the source. The energies of these don't matter for this study since our goal is not detector calibration, but algorithm validation for which units do not matter. We will utilize these peaks to judge the efficiency of the 2SOF.

The hybridized pulses from this detector are passed through the 2SOF, following the full series of steps shown in Fig. 4.12. A slight difference from the toy model analysis is that the covariance matrix here is not estimated from the data, but by generating about 1500 Monte Carlo pulses, passing them through the 2SOF and estimating the covariance between A_{DS} and A_{OP} . Real noise pulses are used for this purpose and bootstrapped to generate enough samples in case there aren't enough noise events.

In Fig. 4.20, I compare the distribution of χ^2 as a function of energy in units of current. The minimum χ^2 value is of the order of magnitude of the number of samples. Since the fully sampled waveform has 32768 time samples, the χ^2 curve for this begins at roughly this value, while the OP and DS are an order of magnitude lower in terms of χ^2 . The upward trend of the curve is due to changing pulse shape with energy which is not accounted for by a fixed template. The curvature is larger for the OP and DS pulses due to insufficient number of samples to capture the true pulse shape. The double spots present in all 3 curves are the two ^{55}Fe peaks which we will study.

A comparison of the ^{55}Fe peak distribution obtained from this fit is shown in Fig. 4.21. The first thing you notice is that the peaks are all centered at different means. The OP peak measures lower energies due to the full pulse tail not being enclosed within the OP window. The DS pulse measures larger amplitudes due to the low pass filtering effect. The fully sampled pulse estimates lie in between these two estimators. The combined amplitude closely resembles the trend of the DS amplitude.

Table 4.5, compares the percentage resolution of various energy estimators. We see results roughly in line with the toy simulations. The combined amplitude, A_0 , is within 1% of the 1×1 OF result considering error bars. The OPamp spectrum fails at the highest energies due to the OP

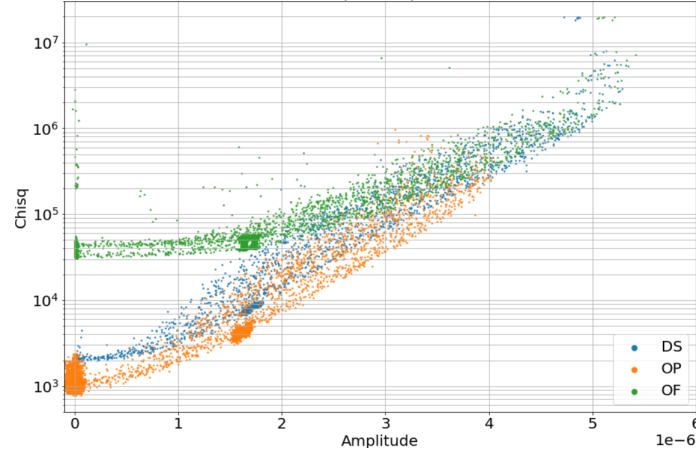


Figure 4.20: A plot showing the χ^2 vs amplitude plane in PD2. The y-axis is unitless while the x-axis is in units of amperes. The green points, labelled as OF in the legend, show the 1×1 OF fit performed on the fully sampled waveform.

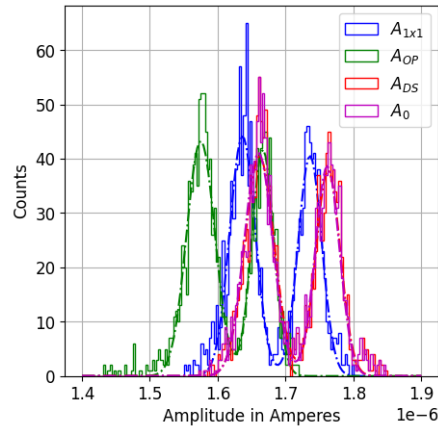


Figure 4.21: A comparison of the two ^{55}Fe peaks obtained from various estimators. Solid lines show the fitted Gaussians used to extract the resolution.

pulse not enclosing the full pulse, an effect which gets worse with wider pulse shapes at higher energies. The DS spectrum is immune to this effect although its resolution could benefit from the OP spectrum.

The final, reconstructed energy spectrum is shown in Fig. 4.22. The primary conclusion here is that the estimator A_0 functions well in general irrespective of energy and is comparable to the 1×1 OF estimator, while the estimates from the DS or OP pulses often face shortcomings depending on

Estimator	Peak 1, % resolution	Peak 2, % resolution
A_{DS}	$1.31 \pm 0.05 \%$	$1.03 \pm 0.04 \%$
A_{OP}	$1.32 \pm 0.05 \%$	$0.97 \pm 0.04 \%$
1×1 OF	$1.16 \pm 0.05 \%$	$1.01 \pm 0.05 \%$
A_0	$1.27 \pm 0.05 \%$	$1.01 \pm 0.04 \%$

Table 4.5: A comparison ^{55}Fe peak resolutions from various energy estimators.

the pulse window or noise.

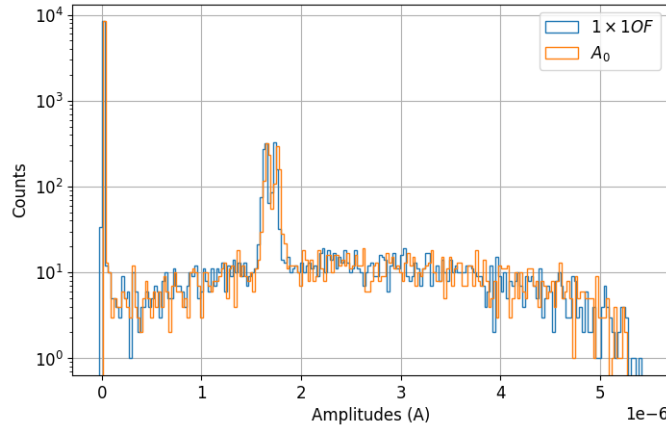


Figure 4.22: The complete energy spectrum for PD2 detector. Both the 1×1 OF and the final A_0 are shown for comparison. The x-axis is in units of Amperes and the y-axis is unitless

The calibration peaks we studied are clearly subject to saturation effects. At the lowest energies we can estimate a baseline resolution which is free of saturation that better represents the efficiency of the 2SOF. The baseline resolution is calculated from a set of randomly collected events with no pulses, that are representative of the ambient noise in the detector. Fitting a template to such events will tell us how well the algorithm is able to distinguish a signal pulse from random fluctuations. Table 4.6 shows the baseline resolution determined from various estimators. Remember that this quantity represents the ability of the estimator or algorithm to correctly identify signal from noise. A lower baseline resolution means a better ability to distinguish signals. From the table, it is clear that the A_0 is the best possible estimate of baseline resolution from a hybrid pulse. The table also demonstrates that the DS pulse outperforms the OP pulse by about 39% at the lowest energies.

Before concluding this discussion, I have one last point to make regarding the time delay of

Estimator	Baseline resolution
A_{DS}	4.49 nA
A_{OP}	6.24 nA
1×1 OF	4.12 nA
A_0	4.48 nA

Table 4.6: A comparison baseline resolutions from various energy estimators.

the DS pulse. We noted earlier that this quantity is zero for the toy pulses and is theoretically expected to be so. However, this only holds true for the case where the pulse shape matches the data reasonably well and if the trace is not from a noise event. The OF is biased to signals and always tries to find the most signal-like region of a trace even if there isn't a signal. When the OF fit is poor, like in the noise blob, the estimated delay from the OP pulse is also poor which means that the trace alignment is not perfect. In addition, the downsampling procedure might remove features that the OP fit decided was a pulse. Therefore, in the noise blob the time delay from the DS pulse could be non-zero depending on the trace.

4.3.3 2-speed Minimizer

Although the 2-speed OF produces reasonable results, it has certain limitations. Some of these limitations are also applicable to the 1×1 OF. The covariance matrix used in the combination procedure is pulse-shape dependent. This is because it is generated using toy pulses which have a fixed pulse shape. In real data, the pulse shape varies as a function of energy which is also a reason why the combined amplitudes perform much better relative to the 1×1 OF fits for the toy pulses. In general, the 1×1 OF also does not take pulse shape variation into account, but this effect is doubled in the 2SOF – once via the optimal filtering process itself and then again via the covariance matrix.

In order to reduce the effect of pulse shape in the optimal filtering process, we make this a free parameter in our fit. We achieve this by replacing the predefined template in the χ^2 in Eq. (2.9) with an analytical function whose parameters determine the shape of the pulse. Once the template is replaced, we can no longer analytically minimize this χ^2 and must instead do this computationally with the help of a minimizer.

4.3.3.1 Pulse shape functions

SuperCDMS pulses have an exponentially rising edge and an exponentially falling tail. The rise is defined by a “rise time”, τ_r , and the fall by one or more “fall times”, τ_f . The following set of 3 models are the most commonly used:

$$\begin{aligned} f_{1FT}(t) &= h \cdot \left(e^{-\frac{t-t_d}{\tau_f}} - e^{-\frac{t-t_d}{\tau_r}} \right), \\ f_{2FT}(t) &= h_1 \cdot e^{-\frac{t-t_d}{\tau_{f1}}} + h_2 \cdot e^{-\frac{t-t_d}{\tau_{f2}}} - (h_1 + h_2) e^{-\frac{t-t_d}{\tau_r}}, \\ f_{3FT}(t) &= h_1 \cdot e^{-\frac{t-t_d}{\tau_{f1}}} + h_2 \cdot e^{-\frac{t-t_d}{\tau_{f2}}} + h_3 \cdot e^{-\frac{t-t_d}{\tau_{f3}}} - (h_1 + h_2 + h_3) e^{-\frac{t-t_d}{\tau_r}}. \end{aligned} \quad (4.16)$$

In Eq. (4.16), each model progressively adds a fall time which tunes the falling tail of the pulse. The models are labelled by the number of fall times with the subscripts 1FT, 2FT and 3FT corresponding to 1 fall time, 2 fall times and 3 fall times, respectively. Each fall time is weighted by a parameter h which controls its relative contribution to the overall pulse shape. The time shift of the pulse is accounted for in the functional form via a delay parameter, t_d . Note that $f(t) \equiv 0$ for $t < t_d$ for all templates.

The question of which model is accurate is an age-old one. The most general one is the 3 fall time model since the other two models are regained from this model by setting the appropriate weights to zero. The physicality of a certain model is often tested via other methods which study the properties of the TES and relate them to a specific fall time. The three fall time model by definition will fit the data better. This is expected since the more you fine tune a given shape the better it fits a generic pulse.

The disadvantage here with a more sophisticated model is the consequent increase in the complexity of the fit. In addition to the rise and fall times, we must fit for the time delays and pulse amplitudes. At the minimum, if we fix the weight of the components and the rise time, even though not ideal, this becomes a three parameter fit, which doesn't sound too bad. But at worst, this is a 9 parameter fit for the three fall time model. Each fully sampled waveform is 32768 samples and therefore, any minimizer would require a lot of computational power to perform this fit. This is why optimal filtering is performed analytically whenever possible and not computationally.

But hybrid data opens a new door in achieving this with a reasonable amount of computational power. This is because the DS pulse and OP pulse have a much lesser number of data points, an order of magnitude lower than the fully sampled waveform.

4.3.3.2 Algorithm

The 2-speed Minimizer (2SM) performs the optimal filter computationally by minimizing the following χ^2 :

$$\chi^2 = \sum_i \frac{|s_i - A \cdot v_i|^2}{J_i}, \quad (4.17)$$

where we have absorbed the time delay into the definition of the functional template, v_i . v_i is the discrete Fourier transform of one of the functions in Eq. (4.16). We perform this minimization in the frequency domain, similar to the analytical optimal filter since we want to exploit the diagonality of the frequency domain covariance matrix. Further, the inclusion of time delay as a free parameter in the fit means that we do not have to manually shift the templates to line up with the DS pulse, allowing us to simplify over overall code structure.

The “2-speed” in the name of the filter comes from us applying this to the DS and OP pulses separately, which are sampled at two different speeds. The 2SOF simply produced two correlated amplitudes which we combined via a covariance weighted sum. Here, depending on the number of parameters you choose to fix, we have multiple correlated parameters whose correlation information could be used to obtain a combined overall pulse shape and energy.

Since the dimensions of the covariance matrix depend on the number of free parameters in the fit, there is no general analytical solution which is the equivalent of Eq. (4.12). There is thus a second step of minimization here to obtain the combined parameters.

Consider a three parameter fit with the amplitude (A), the time delay (t_d) and a fall time (τ_f) corresponding to the single fall time model in Eq. (4.16), fixing the other two parameters τ_r and h . We obtain 2 sets of estimated for these parameters – one from the OP pulse and one from the DS pulse. We then construct a covariance matrix which captures the correlations between the 6 fitted parameters, V_{DS-OP} . The best fits for each of the three parameters in our model are obtained by computationally minimizing the following expression:

$$\chi_{cov}^2 = \begin{bmatrix} A^{DS} - \hat{A}^0, & A^{OP} - \hat{A}^0, & t_d^{DS} - \hat{t}_d^0, & t_d^{OP} - \hat{t}_d^0, & t_f^{DS} - \hat{t}_f^0, & t_f^{OP} - \hat{t}_f^0 \end{bmatrix} \cdot V_{DS-OP}^{-1} \cdot \begin{bmatrix} A^{DS} - \hat{A}^0 \\ A^{OP} - \hat{A}^0 \\ t_d^{DS} - \hat{t}_d^0 \\ t_d^{OP} - \hat{t}_d^0 \\ t_f^{DS} - \hat{t}_f^0 \\ t_f^{OP} - \hat{t}_f^0 \end{bmatrix}. \quad (4.18)$$

In Eq. (4.18), the superscripts *DS* and *OP* denote estimates from the DS and OP pulses, respectively. The superscript 0 denotes the best fit estimator that we can expect from the hybrid pulse. The covariance matrix V_{DS-OP}^{-1} is a 6×6 dimensional matrix computed from simulations using toy pulses. The dimensionality of this matrix increases with the number of free parameters but it still is much simpler to invert compared to the HOF. The overall work flow of the 2SM for this example case is shown in Fig. 4.23.

4.3.3.3 Algorithm validation with toy models

A validation of the 2SM was performed using toy models similar to Sec. 4.3.2.3. The noise was generated the same way but the templates used for pulse generation were produced differently. The templates or shapes were created using the functional forms in Eq. (4.16) so that we are better able to control the input to understand the efficiency of the algorithm.

Here, I report results from a 3-parameter fit performed with the 2SM, following the same workflow described in the previous sub-sub-section. The rise time and weight were fixed to values of $25 \mu s$ and 0.9, respectively. The pulses were generated with fall times of $50 \mu s$ and time delay of zero. The results from performing this fit are shown in Fig. 4.24. The distributions are fit with Gaussians and the results of the fit are reported in Table 4.7.

The results confirm that the timing information from the DS pulse is poorly resolved and hence the minimizer cannot efficiently detect the time delay and fall time to the same extent as the OP pulse. The OP pulse is able to capture the full trace in this toy model and therefore presents better amplitude estimates. The combined best fit estimator for all three parameters provides the best outcome by correctly taking the correlations into account.

The primary disadvantage of this formalism is the increase in computational time compared to the 2-speed OF, the high degree of degeneracy of the parameter space which worsens with

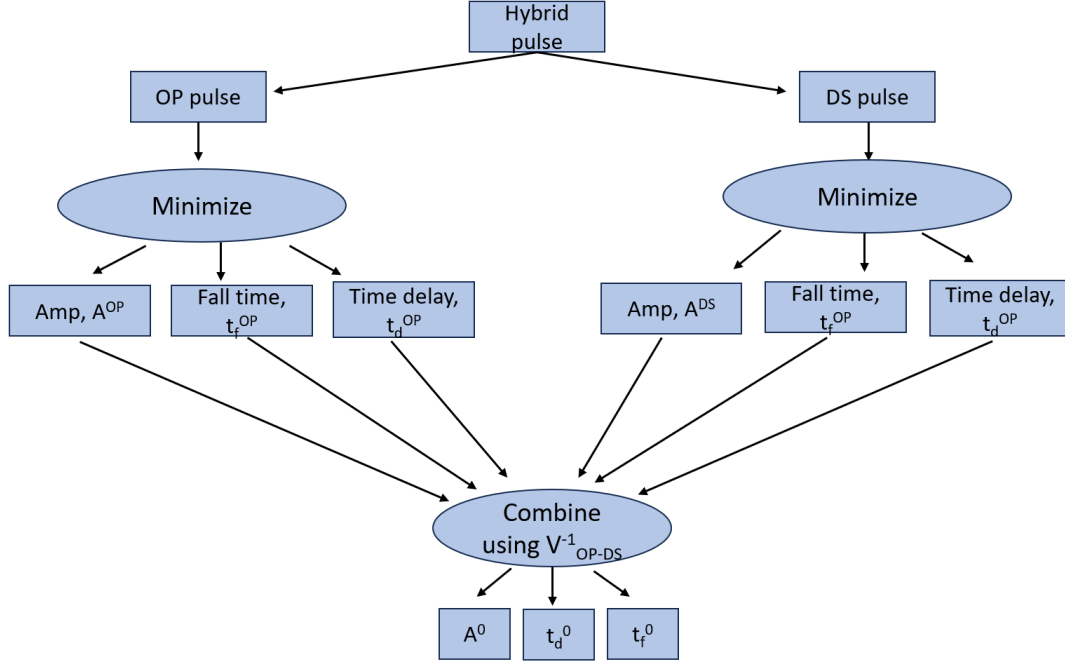


Figure 4.23: A flowchart showing the workflow of the 2-speed Minimizer algorithm with a three parameter fit. The three parameters being fit for are the pulse amplitude, time delay and pulse fall time. The algorithm begins similarly to the 2SOF by taking a hybrid trace and extracting the OP and DS pulses. The pulses are minimized using the TMinuit class in ROOT [120]. The minimization yields 2 sets each of the amplitude, the time delay and the fall time which are correlated with each other. A covariance matrix is constructed by simulation for these 6 parameters and used to combine them to obtain the best fit amplitude (A_0), time delay (t_d^0) and fall time (t_f^0).

Estimator	% resolution on amplitude	% resolution on fall time	Resolution on time delay
DS	2 %	4.41 %	0.58 μ s
OP	1.82 %	3.23 %	0.22 μ s
Combined	1.80%	3.20 %	0.22 μ s

Table 4.7: A comparison of percentage error between various estimators for a mono-energetic toy spectrum. The time delay is centered about zero and therefore only the width of the distribution is reported.

4.3. EVENT RECONSTRUCTION WITH HYBRID DATA

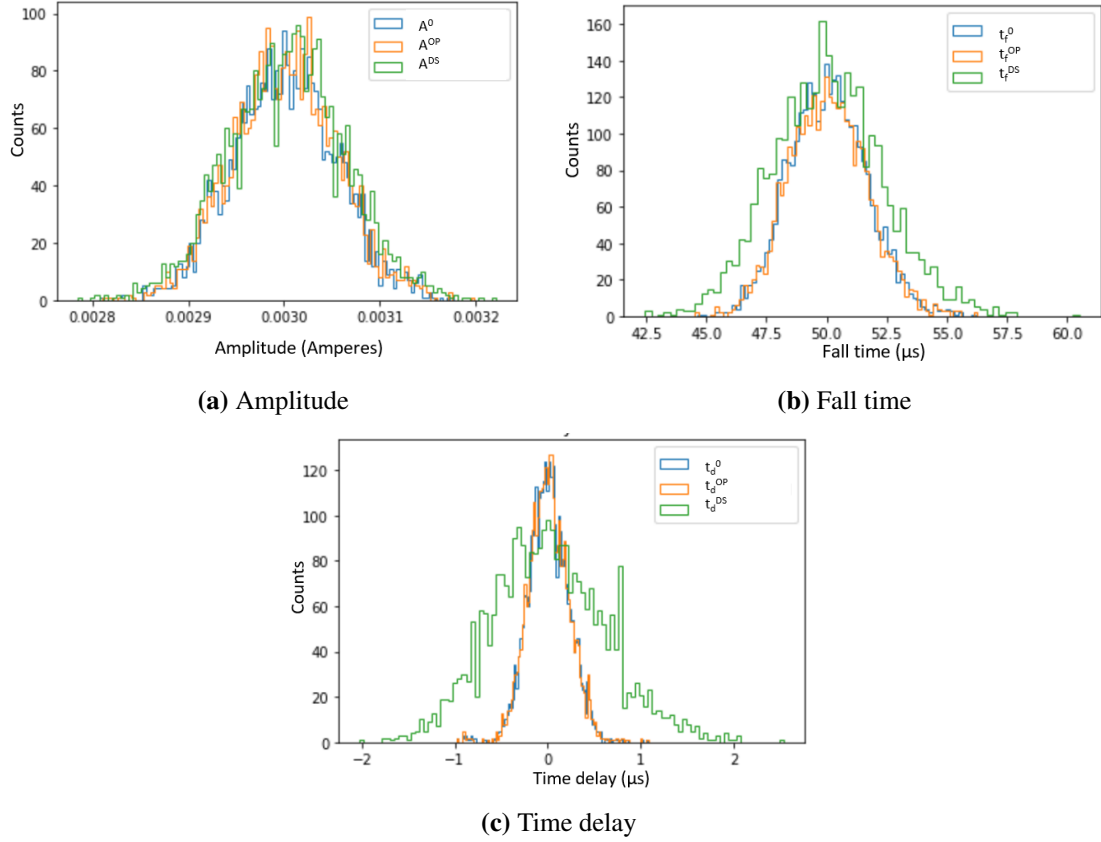


Figure 4.24: Distribution of fit parameters from applying the 2SM to toy pulses.

more parameters, and non-convergence of the minimizer. The first is dependent on the number of parameters and to some extent cannot be controlled. The last two points go hand-in-hand. The more degenerate your parameter space, the more difficult for the minimizer to find the true minimum of the χ^2 . Some common work-arounds for this problem are to constrain your parameters either by adding constraint terms which penalize the cost function or by defining search boundaries for the minimization routine. Both these solutions are only efficient if some information about the parameters is known a priori. This is in principle possible for the time delay and the pulse shape parameters, but careful tuning is required.

Thus the 2SM, although overcoming some limitations of the 2SOF and capable of generalization to the fully sampled pulse, introduces new challenges in optimization which need careful studies. There is a fundamental trade-off between the dimensionality of your χ^2 , computational

power and fit efficiency, which limits its usage.

4.4 Conclusion and future directions

In this chapter, we discussed some of the data acquisition challenges we expect to face at SuperCDMS SNOLAB. In order to overcome these challenges, the experiment plans to take data in hybrid format by downsampling the pre-pulse and post-pulse regions by a factor of 16. This makes our waveforms non-uniformly sampled in time, thus introducing significant consequences for event reconstruction.

A former collaborator had attempted to resolve this problem by developing a Hybrid Optimal Filter which correctly accounts for the frequency correlations that arise between the downsampled and fully sampled regions of the pulse. This algorithm proved to be computationally expensive and increased our net required data volume by a factor of ~ 6.5 .

The 2-speed OF and 2-speed minimizer are two algorithms with better computational efficiency and lower data needs that were developed as part of this thesis to reconstruct hybrid pulses. This chapter discussed the algorithms in great detail and presented some results which showcased the ability of the algorithm to perform reasonably well compared to the fully sampled pulse without compromising on computational efficiency. Both algorithms have certain limitations including a need for toy model pulses to capture frequency correlations. The covariance matrix used in both algorithms are pulse shape dependent which needs to be carefully considered in interpreting the results. The 2SM has the additional disadvantage of using a minimizer which occasionally results in non-convergence.

The 2-speed OF has been implemented into the SuperCDMS data processing software, CDMS-Bats, as part of a larger effort to ready the software to handle hybrid data once the main experiment is commissioned. The effort was done in collaboration by 4 people led by the UBC group including myself and our post-doc. Over a time scale of months, the 2-speed OF and other algorithms such as pulse integrals were modified.

Further work was done as part of this thesis to develop a multi-channel, multi-template fit for hybrid data which takes into account correlated noise between detector channels. This will be discussed briefly in the next chapter.

The work done in this chapter is not the end. This is still an active area of research in the collaboration with new graduate students working on creative ways to reconstruct hybrid data. A top contender is the Missing Data Infilling OF which fills in replacements for the data points we lose from downsampling in a statistically consistent manner allowing us to run an OF with some

4.4. CONCLUSION AND FUTURE DIRECTIONS

constraint terms. Preliminary results from this algorithm have shown to be at par with the 2SOF results. More information on this will likely be covered in theses by other graduate students in the near future.

Chapter 5

$N \times M$ reconstruction and machine learning

In the previous chapter, we discussed two novel reconstruction techniques to address our reconstruction needs at SuperCDMS SNOLAB. In this chapter, I discuss an advanced optimal filter that overcomes the limitations of our commonly used estimators. I will also present a “2-speed” version of this advanced algorithm which may be applied to hybrid data from SuperCDMS SNOLAB. The contents of this chapter and the next will roughly follow the outline of Ref. [1], but will present more details than what is discussed in the paper.

5.1 Multi-channel reconstruction

SuperCDMS detectors are segmented into readout channels with the channel geometry optimized to preserve position information in the data (see Fig. 2.3). An event occurs when an incoming particle collides with the detector producing an electron or nuclear recoil in the crystal which generates phonons and charges that propagate through it. Thus, for every event, a certain amount of energy is collected in every readout channel. The amount of energy collected in each channel is directly related to the location of the primary interaction in the crystal. The total energy transferred to the crystal from the interaction is the sum total of energy collected across all readout channels. Estimation of this total energy presents some challenges.

The approach adopted historically in SuperCDMS has been to calculate a sum of the traces from each channel. For a given event, the pulse in each channel is delayed from the trigger point by a certain amount dependent on the event location. The pulses are first lined up at the trigger

point using the time delay estimates so that the rising edges are appropriately aligned and then summed such that the i^{th} data point in the time domain is given by:

$$s_{total}(t_i) = \sum_c w_c \cdot s_c(t_i). \quad (5.1)$$

In Eq. (5.1), every time domain data point is a weighted sum over the corresponding data points in each readout channel as indicated by the index c . The weight w_c for channel c is otherwise called the relative calibration. It is a number that represents the inherent energy collection capability of each channel relative to an arbitrary chosen channel. The weights equalize the relative contribution of every channel to the total sum accounting for potential inequalities arising from design features. The summed pulse obtained this way is then fit with the 1×1 OF defined in Eq. (2.10). By summing up information from every channel we produce an averaging effect that diminishes the position dependence in the summed trace, s_{total} , thereby reducing resolution broadening from position uncertainties. Although this strategy has been commonly employed in SuperCDMS over the years, it is only moderately successful in addressing position effects.

The 1×2 OF extracts the position information into a second amplitude which is sensitive to prompt phonons. If a second template can capture position information better, then can we do even better with more than two templates? Adding multiple templates increases the complexity of our χ^2 , but at the same time allows us to fine-tune the shape of our model to produce optimal fits. Each template is then some empirical form that captures varying degrees of position and energy information depending on how they are constructed. For any one channel, c , an M -template fit is performed by minimizing the following frequency domain χ^2 :

$$\chi_c^2 = \sum_i \frac{|s_{i,c} - \sum_{j=0}^{M-1} a_{j,c} \cdot v_{i,j,c}|^2}{J_{i,c}}. \quad (5.2)$$

It is important to note here that we allow the M templates to be uniquely defined for each channel. This is the simplest way to accommodate a multi-template fit. Minimizing Eq. (5.2) separately for each channel produces a set of M empirical amplitudes, each of which contains some measure of the energy collected in that channel. The relative amounts of energy collect in different channels contains valuable position information. The quantity of physical interest is not the individual channel amplitude but the energy and position, which relate in a complicated way to the set of amplitudes. The χ^2 for an N -channel fit which considers an M -template model is the sum of χ^2 s of the individual channels:

$$\chi^2 = \sum_{c=0}^{N-1} \chi_c^2 = \sum_{c=0}^{N-1} \sum_i \frac{|s_{i,c} - \sum_{j=0}^{M-1} a_{j,c} \cdot v_{i,j,c}|^2}{J_{i,c}}. \quad (5.3)$$

Unlike the strategy of minimizing Eq. (2.9) for s_{total} , Eq. (5.3) simultaneously fits all N pulses for each event simultaneously, thus preserving the position information in the amplitudes, $a_{j,c}$. Each amplitude also depends on the total energy of the event. The nature of this relation is however non-trivial to determine. A significant portion of this chapter will be dedicated to methods for deriving the true energy and position of an event encoded in these amplitudes.

Although Eq. (5.3) is a type of $N \times M$ filter, the description is incomplete without considering correlated noise. Correlated noise falls into two broad categories for SuperCDMS. The first category includes noise sources that simultaneously affect all channels in a single detector, while the second includes noise sources that affect multiple detectors at the same time. Discussing a multi-detector fit is beyond the scope of this thesis and therefore we will stick to correlated noise between channels on the same detector.

Correlated noise between channels can have many sources. For example vibrations from the environment, such as the cryogenic system surrounding the cryostat, can couple in via vibrational modes. Since vibrations affect the detector as a whole, they could manifest as coupled noise peaks in the PSDs of all channels. A second and dominant source of correlated noise is the readout electronics since channels on the same side of the detector in SuperCDMS SNOLAB share a common return line. Correlated noise has the undesirable effect of broadening energy resolution. Eq. (5.3), can only be used under the assumption that correlated noise between channels is negligible which turns out to be a false assumption in our situation.

The accurate prescription for performing a multi-channel, multi-template fit is by accounting for frequency domain correlations between two channels. This is given by the following equation:

$$\chi^2 = \sum_{c_1=0}^{N-1} \sum_{c_2=0}^{N-1} \sum_i \sum_j \left(s_{i,c_1} - \sum_{k=0}^{M-1} a_{k,c_1} \cdot v_{i,k,c_1} \right)^* \cdot \Upsilon_{i,j,c_1,c_2}^{-1} \cdot \left(s_{j,c_2} - \sum_{k=0}^{M-1} a_{k,c_2} \cdot v_{j,k,c_2} \right), \quad (5.4)$$

where i, j are frequency indices and c_1, c_2 are channel indices. In Eq. (5.4), we have introduced the general cross-channel, frequency domain covariance matrix, Υ . Υ is a 4-D matrix of dimensions $N \times N \times L \times L$, where L is the number of frequency bins. The matrix is composed of multiple 3-D matrices that capture the correlations between any two channels at any two frequencies. The diagonal matrix blocks of Υ are composed of covariance matrices between two channels at the same frequency, while the off-diagonal blocks contain covariances matrices of the two channels

between any two pairs of frequencies. This is further complicated by the complex nature of this matrix which needs to be handled with certain special considerations.

The total covariance matrix of two complex random numbers, $A = a_1 + ia_2$ and $B = b_1 + ib_2$, is represented by a combination of ten independent¹, real covariances:

$$\begin{bmatrix} \text{cov}(a_1, a_1) & \text{cov}(a_1, a_2) & \text{cov}(a_1, b_1) & \text{cov}(a_1, b_2) \\ \text{cov}(a_1, a_2) & \text{cov}(a_2, a_2) & \text{cov}(a_2, b_1) & \text{cov}(a_2, b_2) \\ \text{cov}(b_1, a_1) & \text{cov}(b_1, a_2) & \text{cov}(b_1, b_1) & \text{cov}(b_1, b_2) \\ \text{cov}(b_2, a_1) & \text{cov}(b_2, a_2) & \text{cov}(b_2, b_1) & \text{cov}(b_2, b_2) \end{bmatrix}. \quad (5.5)$$

In the matrix above, the 2×2 diagonal blocks are autocovariance matrices of the respective complex random variables, while the off-diagonal blocks are the cross-covariance matrices between the two complex variables. The second order moment, i.e. total covariance, of two complex random vectors can be completely determined from the complex-valued, 2×2 covariance matrix, C and pseudocovariance matrix, $\tilde{C}$, of the two variables [121]. Both of these are complex quantities with the definitions:

$$\begin{aligned} C(A, B) &= E[(A - E[A])(B - E[B])^*], \text{ and} \\ \tilde{C}(A, B) &= E[(A - E[A])(B - E[B])] = C(A, B^*), \end{aligned} \quad (5.6)$$

where $*$ represents complex conjugation. The total covariance matrix in Eq. (5.5) can be described completely in terms of the matrices C and $\tilde{C}$. If the two random complex vectors are completely uncorrelated, both their mutual covariance and pseudo-covariance matrices are zero. Further, a complex variable which is symmetric under phase rotation is called circular and such variables have zero pseudo-covariance. Such a complex variable is called “proper”. This is true for most of our fully sampled waveforms. This allows us to simplify the structure of our 4D covariance matrix.

We also go one more step and make the assumption that our noise is stationary, i.e., the statistical properties of the noise are time-invariant. In this scenario, the correlation between two different frequencies no longer exists in the frequency domain as we saw in the case of the simple 1×1 OF. In the 1×1 case, this allowed us to use a simple PSD in Eq. (2.9) that characterizes the noise for each frequency in a given channel. For the $N \times M$, the equivalent assumption simplifies our 4D covariance matrix into a 3D matrix which is composed of the covariance matrices between two channels, c_1 and c_2 , at the same frequency f . Eq. (5.4) then simplifies to:

¹Since the matrix is Hermitian the lower triangular section is identical to the upper triangle.

$$\chi^2 = \sum_{c_1=0}^{N-1} \sum_{c_2=0}^{N-1} \sum_f \left(s_{f,c_1} - \sum_{k=0}^{M-1} a_{k,c_1} \cdot v_{f,k,c_1} \right)^* \cdot \mathcal{V}_{f,c_1,c_2}^{-1} \cdot \left(s_{f,c_2} - \sum_{k=0}^{M-1} a_{k,c_2} \cdot v_{f,k,c_2} \right). \quad (5.7)$$

Eq. (5.7) is much easier to deal with. Here, $\mathcal{V}$ is a complex 3D matrix of dimension $N \times N \times L$. This equation represents the frequency domain χ^2 for a multi-channel, multi-template fit with stationary noise.

In the next section, we will rewrite Eq. (5.7) in matrix form and minimize it with respect to the amplitudes, a , to analytically obtain the weights or contribution of each template to the total pulse shape.

5.2 The N×M filter

To further simplify our derivation, we consider the χ^2 at a given frequency, f . Re-writing Eq. (5.4) in matrix form, we obtain,

$$\chi_f^2 = \left(\vec{S}_f - T_f \cdot \vec{A} \right)^\dagger \cdot \mathcal{V}_f^{-1} \cdot \left(\vec{S}_f - T_f \cdot \vec{A} \right). \quad (5.8)$$

Here, we have absorbed all the summations into the matrix notation such that the matrix multiplication returns the full double sum over channels at frequency f . The total χ^2 in Eq. (5.7) can be calculated as, $\chi^2 = \sum_f \chi_f^2$. In Eq. (5.8), $\vec{S}_f$ is an N -dimensional vector, each of whose elements are the complex coefficients of the Fourier transform at frequency f of the signal/trace in each channel for an event. Similarly, $\vec{A}$ is an NM -dimensional vector constructed from the amplitudes, a , and T is an $N \times NM$ matrix of complex Fourier coefficients of the templates, v . The covariance matrix, $\mathcal{V}$ is an $N \times N$ complex matrix, whose elements capture the correlations between the N channels at a single frequency.

Working with complex matrices is difficult especially when one has to perform inversions which end up being precision limited. The inverse of a complex matrix is difficult to derive computationally and there has always been a preference to perform such inversions by representing complex matrices in real form [122, 123] owing to the computational simplicity of this technique. The most common technique used in signal processing is to represent a complex matrix in a composite real form such as in Ref. [122]. The basic idea behind such representations is to represent a simple complex number with a real matrix whose elements correspond to the real and imaginary parts of the number. We adopt this technique to represent all our complex matrices and vectors as

real quantities in terms of the real ($\Re$) and imaginary ($\Im$) parts of their respective Fourier coefficients as follows:

$$\begin{aligned}
 & \bullet \vec{S} = \begin{bmatrix} \Re(S_0) \\ \Im(S_0) \\ \Re(S_1) \\ \Im(S_1) \\ \vdots \\ \Re(S_{N-1}) \\ \Im(S_{N-1}) \end{bmatrix} = 2N \times 1 \text{ dimensional vector,} \\
 & \bullet T = \begin{bmatrix} \Re(T_{00}) & \cdots & \Re(T_{0(M-1)}) & & & \\ \Im(T_{00}) & \cdots & \Im(T_{0(M-1)}) & & & \\ & & & \ddots & & \\ & & & \ddots & & \\ & & & & \Re(T_{(N-1)0}) & \cdots \Re(T_{(N-1)(M-1)}) \\ & & & & \Im(T_{(N-1)0}) & \cdots \Im(T_{(N-1)(M-1)}) \end{bmatrix} \\
 & \quad = 2N \times NM \text{ dimensional matrix,} \\
 & \bullet \mathcal{V} = \begin{bmatrix} \langle \Re(\mathcal{F}_0) \cdot \Re(\mathcal{F}_0) \rangle & \langle \Re(\mathcal{F}_0) \cdot \Im(\mathcal{F}_0) \rangle & \langle \Re(\mathcal{F}_0) \cdot \Re(\mathcal{F}_1) \rangle & \langle \Re(\mathcal{F}_0) \cdot \Im(\mathcal{F}_1) \rangle & \cdots \\ \langle \Im(\mathcal{F}_0) \cdot \Re(\mathcal{F}_0) \rangle & \langle \Im(\mathcal{F}_0) \cdot \Im(\mathcal{F}_0) \rangle & \langle \Im(\mathcal{F}_0) \cdot \Re(\mathcal{F}_1) \rangle & \langle \Im(\mathcal{F}_0) \cdot \Im(\mathcal{F}_1) \rangle & \cdots \\ \langle \Re(\mathcal{F}_1) \cdot \Re(\mathcal{F}_0) \rangle & \langle \Re(\mathcal{F}_1) \cdot \Im(\mathcal{F}_0) \rangle & \ddots & \ddots & \\ \langle \Im(\mathcal{F}_1) \cdot \Re(\mathcal{F}_0) \rangle & \langle \Im(\mathcal{F}_1) \cdot \Im(\mathcal{F}_0) \rangle & \ddots & \ddots & \\ \vdots & \vdots & \ddots & \ddots & \end{bmatrix} = 2N \times \\
 & \quad 2N \text{ dimensional matrix, where } \mathcal{F}_i \text{ are the mean Fourier coefficients of the } i^{th} \text{ channel at a} \\
 & \quad \text{given frequency, and} \\
 & \bullet \vec{A} = \begin{bmatrix} a_{10} \\ a_{11} \\ \vdots \\ a_{NM} \end{bmatrix} = NM \times 1 \text{ dimensional vector.}
 \end{aligned}$$

Some features of the above representation are given below:

- $\vec{S}$, T and $\mathcal{V}$ have undergone a transformation from complex to real as $\mathbb{C}^N \rightarrow \mathbb{R}^{2N}$, while $\vec{A}$ has always been a real vector representing weights or amplitude of each template.
- The matrix T represents the Fourier coefficients of the M templates for every channel. In the representation shown above, we assume that the template shapes of one channel are uncorrelated with the shapes of another channel. This produces the block diagonal form of T .
- The covariance matrix $\mathcal{V}$ is Hermitian in its real and complex representations. This is a common property of covariance matrices.

This representation now allows us to use all the same machinery that is available for manipulating real matrices. In this new representation, our χ^2 becomes a real quantity computed as:

$$\chi_f^2 = \left(\vec{S}_f - T_f \cdot \vec{A} \right)^T \cdot \mathcal{V}_f^{-1} \cdot \left(\vec{S}_f - T_f \cdot \vec{A} \right), \quad (5.9)$$

where I've replaced the Hermitian conjugate with a simple transpose and the total χ^2 is

$$\chi^2 = \sum_f \left(\vec{S}_f - T_f \cdot \vec{A} \right)^T \cdot \mathcal{V}_f^{-1} \cdot \left(\vec{S}_f - T_f \cdot \vec{A} \right). \quad (5.10)$$

Minimizing Eq. (5.10) with respect $\vec{A}$, we obtain A as:

$$\vec{A} = \left(\sum_f T_f^T \mathcal{V}_f^{-1} T_f \right)^{-1} \cdot \sum_f \left(T_f^T \mathcal{V}_f^{-1} \vec{S}_f \right). \quad (5.11)$$

The full derivation for Eq. (5.11) is shown in Appendix A. Eq. (5.11) is extremely convenient since it allows us to draw parallels between the N×M and the simple 1×1 OF. Similar to the “Optimal Filter” quantity defined in Eq. (4.5), we can define a normalized “N×M” filter quantity that optimally filters the signal from multiple channels simultaneously:

$$\theta_f = \left(\sum_f T_f^T \mathcal{V}_f^{-1} T_f \right)^{-1} \cdot T_f^T \mathcal{V}_f^{-1}. \quad (5.12)$$

Here, θ_f is an NM×2N filter matrix. Then, Eq. (5.11) has an alternate form:

$$\vec{A} = \sum_f \theta_f \vec{S}_f. \quad (5.13)$$

So far in this discussion, I have ignored the time delay quantity in the N×M filter. The formulae discussed in this chapter so far are all derived under the assumption that the time delay is zero or, in other words, the signal is perfectly aligned with the templates used in the N×M filter. Use of Eq. (5.11) thus requires us to estimate the time delay as a prior step (e.g. using the 1×1 OF or the 2-template OF), then align the templates with the data. Including the time delay modifies Eq. (5.10) as:

$$\chi^2 = \sum_f \left(\vec{S}_f - R_f \cdot T_f \cdot \vec{A} \right)^T \cdot \mathcal{V}_f^{-1} \cdot \left(\vec{S}_f - R_f \cdot T_f \cdot \vec{A} \right), \quad (5.14)$$

where the matrix R is a rotation matrix which captures the phase rotations in frequency domain due to time shifts in the time domain. If $\vec{t}$ is a vector of time delays, one per channel such that $\vec{t} = [t_0, t_1, \dots, t_{N-1}]$, then the matrix, R_f , in terms of the phase, $\phi_f(t) = 2\pi\omega_f t$, has the following form:

$$R_f(\vec{t}) = R_f(t_0, t_1, \dots, t_{N-1})$$

$$= \begin{bmatrix} \cos[\phi_f(t_0)] & \sin[\phi_f(t_0)] & & & & & 0 \\ -\sin[\phi_f(t_0)] & \cos[\phi_f(t_0)] & & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ & & & & \cos[\phi_f(t_{N-1})] & \sin[\phi_f(t_{N-1})] \\ & & & & -\sin[\phi_f(t_{N-1})] & \cos[\phi_f(t_{N-1})] \\ 0 & & & & & & \end{bmatrix}. \quad (5.15)$$

Although Eq. (5.14) can be analytically minimized with respect to the amplitude to obtain the amplitude as a function of R_f , it is non-trivial to determine the elements of R_f analytically. Computationally solving this involves fitting for NM + N parameters including all amplitudes and time delays which is at least a 24-parameter fit with a single template for a 12-channel SuperCDMS SNOLAB detector. This parameter space is also degenerate and collinear which further complicates the computational feasibility of this minimization. More work is necessary to determine an optimal method which allows time delay estimation from the N×M filter directly. This is not an area of active interest in the collaboration since a closely related problem can be analytically solved to obtain the time delays directly.

The 1×M OF is obtained by minimizing Eq. (5.2) directly. Including the time delay for channel c , t_c , this equation looks like:

$$\chi_c^2 = \sum_j \frac{|s_{j,c} - \sum_{k=0}^{M-1} a_{k,c} \cdot v_{j,k,c} e^{i\omega_j t_c}|^2}{J_{j,c}}. \quad (5.16)$$

Here, we fit one channel at a time, disregarding any contributions from cross-channel correlations. The $1 \times M$ OF can be analytically solved to obtain the amplitudes as a function of time delay. In this case, there is a straightforward method which allows us to calculate the optimal time delays with ease. By using the same templates in the $1 \times M$ filter as the $N \times M$ filter, we ensure a direct utilizability of the time delays estimated from the former to the latter. I will not delve into the details of this derivation here, but will point interested readers to Ref. [124].

The current implementation of this $N \times M$ code allows the user to select between the 1×1 , the 1×2 and the $1 \times M$ OFs to estimate time delays as a preliminary step in $N \times M$ fitting.

5.3 The $N \times M$ processing pipeline

Although we have explained how the filter itself works, we have not discussed how one ought to interpret the outputs of this filter. We are also yet to discuss the nature of the M templates to be used in the filtering which will determine how we interpret our amplitudes. Our primary goal in reconstruction is to determine the event energy from the amplitude and shape of a pulse. We are also interested in the location of the event for the purpose of correcting position effects as well as identifying potential sources of spurious noise. The templates ought to be constructed such that they capture the relevant position and energy information in the output amplitudes. There is also the question of how many templates to use since M can be as large as we like in Eq. (5.7). Further, the exact relationship that maps the amplitudes to a physical event energy and position is not known a priori. In situations like this where the underlying relationships are difficult to determine analytically, empirical approaches are often sought as a replacement. The growth of machine learning over the last decade has led to the development of several tools that could be adapted into the realm of physics. Here, I present the use of unsupervised machine learning for template generation and supervised learning to obtain empirical mappings of the amplitudes to a physical energy and position.

5.3.1 Template generation with unsupervised learning

The templates used in the $N \times M$ filter must be able to optimally characterize the shape of a pulse via a weighted sum of the templates. The best fit weights/amplitudes are, of course, derived from the filter itself (see Eq. 5.11). The templates must be reasonably smooth so that high frequency

noise does not leak into the shape. Such high frequency noise in the templates could often bias the output of the filter. The templates are also normalized such that the maximum height of the template is one.

I have discussed in earlier chapters the close relationship between template shape and event position in the detector. The 1×2 OF carries information about event location in the residual amplitudes made possible by the residual templates which represent the prompt phonon signal. The templates utilized by the $N \times M$ filter must be able to characterize position information better than the 1×2 templates. By doing so, the $N \times M$ can produce better energy estimates by opening new avenues for reducing position related uncertainties.

An additional factor for consideration is the number of templates to be used. In principle, Eq. (5.7) can accommodate an arbitrary number of templates. A larger number of templates naturally equals a finely tuned fit that suits the most generic pulse shape at the cost of computational power. Moreover, the larger the number of templates, the less room we leave for reasonable physical interpretations of the templates. Therefore, a template selection procedure must be devised to pick the M best templates that together represent the average pulse shape across the detector and deviations from it.

Principle Component Analysis (PCA) is a linear feature reduction technique that has been extensively used since the 20th century, but has more recently been classified as an “unsupervised” learning approach in machine learning [125–127]. PCA provides an elegant mathematical framework which allows one to map high-dimensional datasets into a new vector space populated by uncorrelated, orthonormal components derived from the data itself. This factorization technique has found applications in two broad categories: (1) feature identification, and, (2) feature reduction. Both these properties are what we hope to achieve in our template generation task. Feature identification will allow us to re-map our data (a set of pulses) in terms of its critical features while feature reduction will aid us in restricting the number of templates.

Say you have n observations of p features in your data set. Your data can then be expressed as an $n \times p$ matrix, X , each of whose rows is one observation or measurement of all p features. The p features span a p -dimensional vector space. PCA is simply a process of finding a new set of unit vectors that define a rotated vector space where the data set can be projected such that the most meaningful features are easily identified. The new basis vectors are calculated as some linear combination of the original basis vectors. “Meaningful information” refers to the statistical variance in the data set. Typically, the data set is centered to have means of zero by subtracting the mean of each feature/column so that the variance represents an excess or deviation from zero. The projection of the k^{th} measurement in the new rotated vector space is:

$$\vec{Y}^{(k)} = \vec{X}^{(k)} \cdot W, \quad (5.17)$$

where $\vec{Y}$ is the new p -dimensional vector, W is a matrix each of whose columns are basis vectors of the new vector space and $\vec{X}$ is the old p -dimensional vector². The most insightful result of PCA is that the columns of matrix W are sorted such that the column $j = 0$ has the most information/variance, $j = 1$ has the second-most, $j = 2$ has the third-most and so on. Since each column is independent and orthonormal to the other columns of the matrix, removing a few columns has minimal impact on the net information in $\vec{Y}$. This means that we can remove enough columns of W , i.e. neglect unimportant features in the data, to reduce the dimensionality of $\vec{Y}$.

The new basis vectors are calculated by maximizing the variance of the data. Consider the first column vector, $\vec{w}_1$, of the matrix W . Following the convention in Ref. [127], if S is a $p \times p$ matrix that represents the variance in our data matrix, X , then, $\text{Var}(X) = X^T X = S$. Considering only the first feature, the variance of the rotated matrix, Y , is $\text{Var}(Y) = \text{Var}(X \cdot w_1) = w_1^T X^T X w_1 = w_1^T S w_1$. Therefore, maximizing the variance is equivalent to maximizing this quadratic form. Additionally, since basis vectors are unitary, we impose the condition that $w^T w = 1$. Thus, we maximize the following expression:

$$w_1^T S w_1 - \lambda (w_1^T w_1 - 1) = 0, \quad (5.18)$$

where λ is a Lagrange multiplier. Taking the first derivative of Eq. (5.18) with respect to w_1 gives $S w_1 = \lambda_1 w_1$. This equation reveals that w_1 is in fact an eigenvector of the covariance matrix, S , with λ_1 being the corresponding eigenvalue. The first basis vector, w_1 , which points in the direction of maximum variance in the data, is normalized to be a unit vector and has information proportional to the magnitude of its eigenvalue relative to the other eigenvectors. To find the subsequent basis vectors, the sum of all previous $\vec{X}_i$ are projected to the new basis and subtracted from X as

$$X_{new} = X - \sum_{k=0}^i X w_k \cdot w_k^T. \quad (5.19)$$

Then, the eigenvector which maximizes the variance of the new subtracted data matrix, X_{new} , gives the next weight vector. It can be shown that this iterative approach ultimately yields the eigenvectors of the initial covariance matrix, S [125]. Further, any two vectors in the new vector space are uncorrelated since $\text{cov}(Y_i, Y_j) = w_i^T S w_j = \lambda w_i^T w_j = 0$.

²It is a single row vector of the matrix, X .

For the purpose of template generation, we apply PCA to a collection of time-aligned traces. In the time domain, every time bin has a certain level of information about the pulse shape. Each time bin in a trace thus acts as one of p features we seek to study, with p being equal to the tracelength. Each trace is one of n samples. The traces are baseline subtracted to ensure that the signal-like portion in each trace represents an excess over zero. The traces are arranged into an $n \times p$ dimensional matrix, Z . The covariance matrix is calculated as $Z^T Z$ and eigenvalues and vectors of this covariance matrix are estimated via singular value decomposition [128]. This process is shown in the schematic in Fig. 5.1. The eigenvectors are called PCs of the data and each PC points in the direction of subsequent importance in the data. The “importance” or magnitude of each direction is determined by the eigenvalues with the vector with the largest eigenvalue being the most significant PC, PC0. PC1 is the second most important direction, PC2 is the third-most and so on. Every subsequent component captures information that was unaccounted for by all the preceding components. Furthermore, every new direction vector is produced from a combination of information from all the samples and therefore the loss of information from dropping features of less importance is minimal. One can arrive at a decision on how many components to drop by looking at a scree plot of the eigenvalues [129]. A scree plot visualizes the “importance” of each PC by plotting the eigenvalues in increasing order of component number. Since the eigenvalues get smaller for every subsequent component by construction in PCA, the plot typically has a falling trend. An often used rule of the thumb is that if the proportion of eigenvalue in the component is less than 0.1, then that component and subsequent ones do not add significant information to the population. This can of course be modified to specific situations and the cutoff value is often determined by studying the scree plot and comparing it to the actual PCs to determine the PCs with relevant information. In our case, subsequent PCs capture more and more noise information with most of the signal information captured in the first few PCs. This allows us to significantly cut down the number of PCs under consideration from order of 10,000 to under 5, with the exact number depending on the noise environment and detector properties. An example scree plot is shown in Fig. 5.2 to provide a visual intuition for what this curve looks like.

M templates are selected using a combination of scree plots and visual comparison of PCs. For N channels, the $N \times M$ filter outputs NM amplitudes. Since an even mix of traces across the detector are used in PC generation, the PCs represent variation in pulse shape across the detector. Thus, the amplitudes contain a combination of energy and position information. The exact analytical relationship between these amplitudes and the true physical parameters is unknown. In such cases, empirical methods serve as reasonable approximations. In high dimensional mappings, machine learning techniques have proven to be successful as empirical models. I use supervised learning

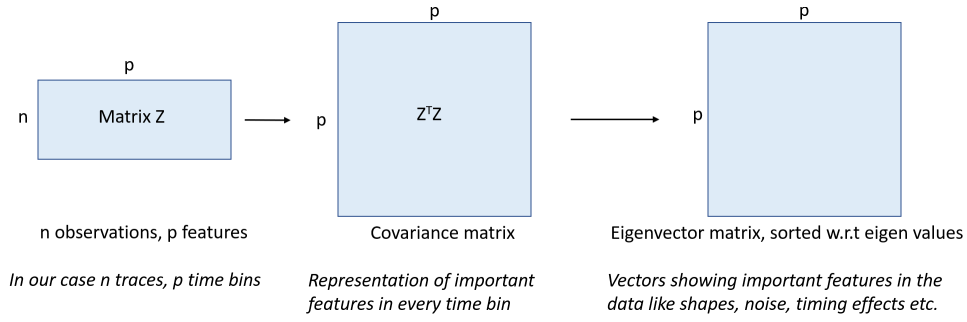


Figure 5.1: A schematic showing the construction of Principal Components in PCA. The matrix, Z , consists of n baseline-subtracted traces with p features each. The covariance matrix constructed from Z is decomposed to estimate the Principal Components of the data.

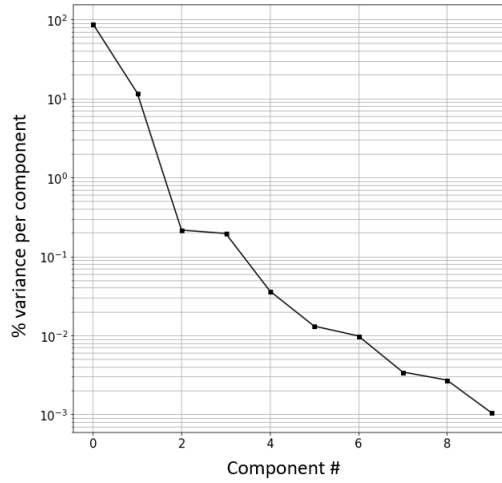


Figure 5.2: An example scree plot for the first ten PCs. The plot shows the proportion of each eigenvalue with respect to the sum of eigenvalues in percentage. This example scree plot was produced from an older SuperCDMS data set. The traces in this data set had 4096 features and thus, have 4096 PCs. The y-axis is normalized such that the sum over all PCs equals 100%.

techniques to map our NM amplitudes to an event energy and physical position on the detector.

5.3.2 Reconstructing physics with supervised learning

Supervised learning algorithms learn mathematical relationships between a set of input features, X , and a set of output labels, Y [130, 131]. If the output labels are discrete, then the model serves

as a classifier that classifies an observable into a discrete category based on its input features. On the other hand, if the output labels are continuous, then we have a regression model that learns an empirical relationship between the high dimensional inputs and the outputs.

We seek an algorithm that maps our NM amplitudes to an energy and position, thus performing a regression task. Most supervised learning algorithms are data hungry and often require a large training data set. Part of the data is kept aside for training the model and another part is kept aside for tuning model parameters and independent validation. A third portion is used as a “test” data set on which a completely trained and tuned model is tested for independent assessment of model performance. Simple regression techniques such as linear or polynomial regression are difficult to apply given the lack of knowledge about the exact mapping function and the high dimensionality of our input features. Further, rare event search experiments, as the name suggests, produce vast data sets; however candidate or pulse-like events are rare enough that we cannot reach the target of $\geq O(10^6)$ to apply deep learning. However, sophisticated decision trees can outperform other methods in regression tasks for smaller data sets with larger number of features [132]. Therefore, we use decision trees to map the NM amplitudes to meaningful physical output labels.

Decision trees are popular classifier algorithms that learn a sequence of yes or no statements from the training data [133]. The smallest working unit of a decision tree is called a decision stump. A decision stump simply decides if a given feature satisfies a certain condition and places the measurement into one of the possible output categories. A sequence of decision stumps, like the one shown in Fig. 5.3, form a full tree. Each stump consists of one input node (node A in Fig. 5.3) and at least two output nodes (nodes B and C in Fig. 5.3), connected by edges. The terminal nodes of the tree where no further branching occur are called leaf nodes, while the initial node of a tree is called the root node. The depth of a given node is the number of edges from that node to the root node. The height of a decision tree is the distance of the root node from the deepest leaf node, or in other words, it represents the maximum depth. The maximum depth of the tree is one of the hyperparameters which can be controlled to prevent over-fitting. Similarly, the minimum number of samples in a node can be preset to restrict further splitting of the tree.

The condition for splitting at each node is learned by “greedy recursive splitting” in most algorithms. This technique attempts several possible ways to split the feature space and picks the split that gives the best cost. Once the first split is determined, the data is divided into groups and then recursively split again based on the next best possibility determined by the cost. This process is continued until some user-defined condition for termination is achieved. The cost of each split is assessed via a cost function which can be information gain for classifiers or Mean Squared Error (MSE) for regressors. New models are estimated by adjusting the splitting rules such that we

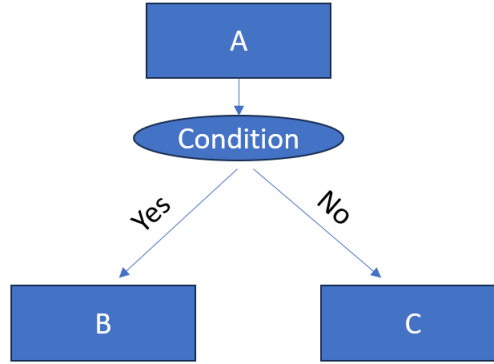


Figure 5.3: A schematic of a binary decision stump which is the smallest decision making unit of a binary decision tree. Decision stumps can be non-binary meaning that more than two outcomes could exist.

minimize (maximize) the MSE (information gain) in the prediction. In decision tree regression, a tree learns the optimal way to split the feature space into regions such that each region, represented by a leaf node, gives the best possible prediction of the output with its mode or mean for any new point that falls in that region of the feature space.

Figure 5.4 shows an example of decision tree regression. The blue points represent data with a single input feature, X , which maps to a single output label, Y . The X - Y plane represents the two-dimensional feature space for this simple case. The decision tree learns the optimal decision boundaries marked in red, effectively partitioning this feature space. The green line marks the best prediction for each region of the feature space which is usually the mode of Y for all X in that region in the training data. When we seek the value of Y for a new input, X_{new} , the decision tree determines which of the three regions the new input belongs to and returns the best prediction of Y in that region. Although the curve in the plot is discontinuous, the same principle can be applied to a continuous feature space.

In our case we have an NM -dimensional feature space populated by the $N \times M$ amplitudes and therefore the decision tree learns how to split the data meaningfully in these NM dimensions. For our complicated and continuous feature space, it is obvious that a single decision tree may not fare well. Ensemble methods utilize multiple decision trees, each of which on its own is a “weak predictor”, but will jointly produce one strong prediction. In such ensemble methods, each successive tree is trained to predict the error correction necessary to fix the predecessor’s predictions. In simple terms, the following sequence is followed:

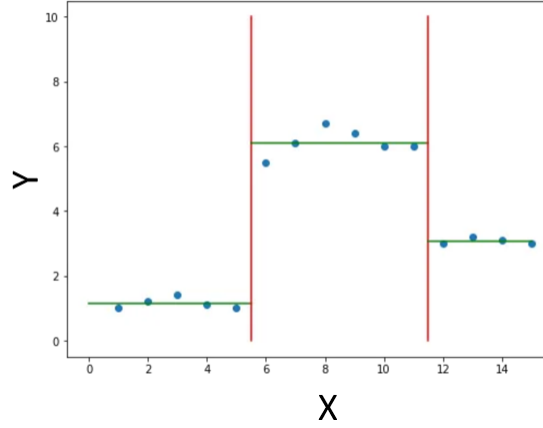


Figure 5.4: A schematic of decision boundaries in a decision tree regressor. Figure taken from [134].

- First tree trains on the label Y , which is the true value of the output.
- First tree's predictions: $\hat{Y}$.
- Second tree trains on the errors of the first tree's predictions: $y_1 = Y - \hat{Y}$.
- Combined prediction of first and second trees, $\hat{Y}_1 = \hat{Y} + y_1$.
- Third tree trains on the combined error from the first and second trees: $y_2 = Y - \hat{Y}_1$.
- Combined prediction of second and third trees, $\hat{Y}_2 = \hat{Y}_1 + y_2 = \hat{Y} + y_1 + y_2$.
- Continue to last tree.

In this scheme, every subsequent tree corrects for mistakes not accounted for by all its predecessors. Thus, higher trees predict smaller and smaller corrections to the prediction of the first tree which trained directly on the output labels. This technique is called boosting. Many algorithms subtract a gradient of a loss function similar to gradient descent instead of subtracting the predictions directly. Such ensemble trees are termed Gradient Boosted Decision Trees (GBDTs) [132, 135]. We specifically use the LightGBM regressor due to its excellent performance with relatively small data sizes [132, 136].

The NM amplitudes obtained from the N×M filter serve as our input features. In order to predict the true energy, we require events whose true energies are known. A common way to

obtain this is via simulations that take into account the full set of physical processes that occur in our crystal and associated electronics in the readout chain. Simulated data can be labeled in both energy and position and can produce large training samples. However, in the case of SuperCDMS, simulations are known to not fully represent our real data sets. Therefore the possibility of training on simulated data and applying the trained model to real data is still placed far away in the future. Another potential training data set is of course real data itself. While it is possible to obtain energy-labeled events from real data sets in the form of calibration peaks, it is impossible to produce position-labeled real events from the detector. Although we would be limited in terms of training data size when using real data, we benefit from our training data exactly resembling our test data sets for which we seek a prediction provided that both training and test data sets are acquired under the same detector conditions. Therefore, if using simulated data, we train on both energy and position labels while we train on energy labels only if we use real data sets.

A caveat to remember is that to first order a weighted sum of the NM amplitudes will give a preliminary energy estimator. Since each amplitude contains energy and position information, this preliminary estimator can correct for linear resolution-broadening effects arising from position dependence of our pulse shape. Note that this is made possible by our specific construction of templates that capture shape variation across the whole detector. This linear estimator is defined as

$$E_{lin-est} = \sum_{i=0}^{NM-1} A_i \cdot \alpha_i, \quad (5.20)$$

where A_i is the i^{th} amplitude and α_i is an optimized weight for that component. These weights can be obtained using our training data. For example, if our training data comes from a calibration peak, then the optimal weights are those that minimize the width of the Gaussian peak. Mathematically, this is equivalent to minimizing the following expression:

$$\chi^2 = \sum_j \left| E_{true}^j - \sum_i (A_i^j \cdot \alpha_i) \right|^2, \quad (5.21)$$

where E_{true}^j is the true energy for event, j . The above expression has an analytical solution that allows us to directly calculate the weights:

$$\alpha_i = \frac{\sum_j A_i^j \cdot E_{true}^j}{\sum_j (A_i^j)^2}. \quad (5.22)$$

The estimator in Eq. (5.20) only takes into account the linear effects of position by weighting

each template component in each channel optimally. In reality, the energy resolution as a function of true energy is expected to be:

$$\begin{aligned}\sigma &= \sqrt{\sigma_{bs}^2 + \sigma_{fano}^2 + \sigma_{pos}^2} \\ &= \sqrt{\sigma_{bs}^2 + BE + (AE)^2},\end{aligned}\tag{5.23}$$

where σ_{bs} is the baseline resolution of the detector which is independent of energy, σ_{fano} is the random variation in the number of charges produced in each interaction and σ_{pos} is the uncertainty arising from our imperfect knowledge of event position in the detector. The Fano term scales as the square root of energy while the position term scales linearly with energy. The weights, α , are energy dependent owing to this linear dependence of the position uncertainty on energy, $\alpha(E) = \alpha_0 + \beta(E)$, where α_0 is an arbitrary constant and $\beta(E)$ captures the energy dependence of α . The optimal weights thus must be calculated at each energy. Since this is not practically feasible in real data sets due to the limited availability of events with known true energy across the spectrum, the weights will be calculated at calibration peaks and linearly interpolated to points in between the calibrated peaks. This however introduces the need for multiple calibration peaks in the data set at different energies for weight estimation and limits the application of this technique to data sets from our Ge detectors. As discussed in Ch. 3, Ge produces multiple electron capture peaks at different energies thus satisfying our condition that multiple peaks be present in the data. For Si detectors, which lack any internal activation lines, alternate techniques must be devised to estimate these weights.

While $E_{lin-est}$ accounts for first order effects of the broadening, σ_{pos} , the true energy is some non-linear combination of the NM amplitudes. The true functional form of these non-linear dependencies is unknown but correcting for these is essential to minimizing σ_{pos} .

When training on real data sets, calibration peak events have the same energy labels, i.e. all events in the Gaussian peak map to a delta function. Thus, training directly on energy labels with a GBDT will map all events to the same value that equals the true energy label which is of course a result of the way in which we have framed the problem. To understand this, I remind the reader that GBDTs work by partitioning the feature space into regions. GBDTs, or most machine learning models for that matter, do not extrapolate well. When all our training data maps to the same output label, the GBDT cannot extrapolate beyond the values defined by this output label since what the GBDT reports is the most common value of the label in that region of the feature space. Therefore,

5.3. THE $N \times M$ PROCESSING PIPELINE

instead of directly mapping to true energy, we map the amplitudes to a correction factor:

$$BDT\text{Gain} = \frac{E_{true} - E_{lin-est}}{E_{lin-est}}. \quad (5.24)$$

This correction factor accounts for any broadening effects not captured by $E_{lin-est}$. The GBDT effectively is then being trained to identify deviations of the linear energy estimator $E_{lin-est}$, relying on the linear estimator for the overall extrapolation in energy while using the GBDT to account for higher order position dependencies. The final energy estimator is then given by:

$$E_{N \times M} = E_{lin-est} \cdot (1 + BDT\text{Gain}). \quad (5.25)$$

It is important to point out that similar position corrections have been applied to SuperCDMS data sets in the past [71] but such corrections have never considered non-linear dependencies of amplitudes on position.

Here we have built a full reconstruction pipeline for SuperCDMS data. This pipeline seamlessly integrates the mathematical framework of the $N \times M$ filter with the supporting machine learning techniques I've described in this chapter. This pipeline is shown in Fig. 5.5. The pipeline starts with template generation from the data using PCA. The generated templates are used to process the data with the $N \times M$ filter. The filter outputs NM amplitudes from which a linear energy estimator is calculated using Eq. (5.20). The NM amplitudes are then mapped to a meaningful energy and position using a GBDT. Only energy labels, in the form of BDTGain, are considered for real data sets due to lack of position-labelled data. But in the scenario where position labels do become available, they can be reconstructed directly from the NM amplitudes using the GBDT. This result along with the non-linear position dependence of the amplitudes on true position will be illustrated using simulated data in the next chapter.

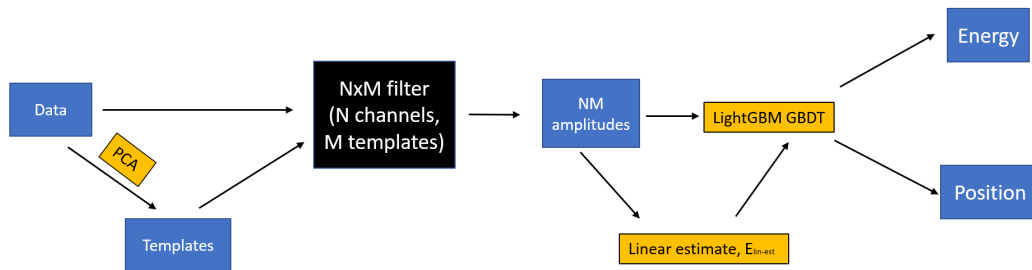


Figure 5.5: The reconstruction pipeline for the $N \times M$ filter.

The $N \times M$ filter very clearly outperforms all other conventional reconstruction techniques which will also be confirmed in the next chapter where I show results from applying this reconstruction pipeline to different data sets.

5.4 $N \times M$ filter for hybrid data

In this section, we jump back to the case of hybrid data. The $N \times M$ filter discussed in Sec. 5.2 assumes that the data is uniformly sampled. This allows us to effortlessly calculate DFTs of all the quantities necessary for the $N \times M$ filter. With non-uniformly sampled data, we are faced with the same issues which were described in Ch. 4. The main disadvantage arises from having to use Eq. (5.4) now that frequency correlations become a significant factor. This equation is analytically and computationally difficult to solve. Therefore, in this section I explore the idea of a “2-Speed $N \times M$ filter” which is a semi-empirical approximation to running the $N \times M$ filter on fully-sampled data.

5.4.1 2-Speed $N \times M$ filter

The basic idea behind the 2-Speed $N \times M$ (2SN $\times$ M) filter is similar to other two-speed algorithms where we process the DS and OP pulse separately (see Sec. 4.3.2 and Sec. 4.3.3) and then combine the two sets of amplitudes taking into account the correlations between them. There are a few subtle differences in the 2SN $\times$ M algorithm compared to the others.

The 2SOF in Sec. 4.3.2 applies the optimal filter to the OP trace to get the time delay using which the templates are aligned with the signal. Thus, the optimal filtering of the DS trace gives us zero time delay. For the $N \times M$ filter, it is not possible to obtain the time delay without expending significant computational power due to the complicated form of Eq. (5.14). Remember that the $N \times M$ filter in its current implementation following Eq. 5.11 is only accurate for perfectly (or near perfectly) aligned pulses. For the 2SN $\times$ M filter, we skip the time delay step and directly utilize the delay estimated from another algorithm like the 1×1 OF for which a two-speed version capable of processing hybrid data exists. The OP trace is processed with one of these algorithms as a first step to give a time delay with which the $N \times M$ filter employs its usual routine of shifting pulses.

A second difference is that now we have 2NM amplitudes in place of 2 amplitudes, NM each for the DS and OP traces. All the 2NM amplitudes are correlated to each other, but not perfectly. Typical correlation coefficients between the DS and OP amplitudes range from $\sim 70\%$ - 90% which tells us that we could benefit from combining them. An example plot showing the correlation between DS and OP amplitudes is shown in Fig. 5.6.

Since there are not simply two estimated amplitudes like in the case of 2SOF, but 2NM, there is

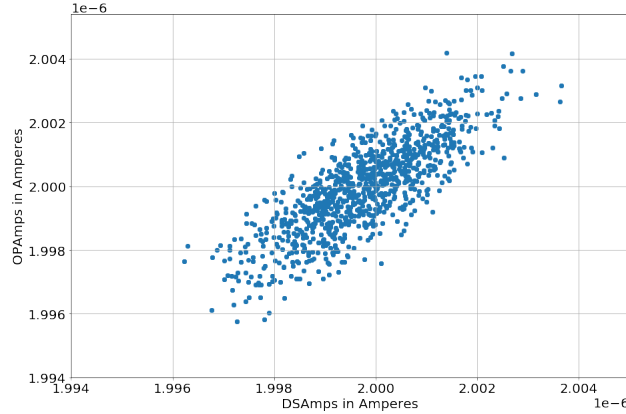


Figure 5.6: A plot of OPamps vs DSamps for one channel from a 2×1 fit (2 channels, 1 template) for simulated toy data. Here, we see that the two amplitudes are strongly correlated. The correlation coefficient was measured to be $\sim 80\%$.

no simplified analytical expression to combine the amplitudes directly without matrix algebra. The correlations between amplitudes can be described by a $2NM \times 2NM$ covariance matrix. Although it is possible to simulate toy pulses with correlated noise (see Sec. 5.4.2), it is not easy to capture pulse shape variation in toy pulses as was discussed in Ch. 4. Since pulse shape difference is significant to the performance of the $N \times M$ filter, this cannot be neglected in the covariance matrix calculation. Therefore, the covariance matrix is computed directly from the amplitudes of real pulses for the $2SN \times M$.

A final point of difference is in the combination procedure. The procedure is a two step process. The first involves reducing the number of amplitudes from $2NM$ dimensions to NM dimensions. At this stage we become equivalent to the usual $N \times M$ filter which allows us to use Eq. (5.20) to combine them. The first step requires minimizing another χ^2 expression:

$$\chi^2 = (A - M \cdot P)^T \mathbf{V}^{-1} (A - M \cdot P), \quad (5.26)$$

where A is a $2NM$ -dimensional vector of the $N \times M$ amplitudes of the OP and DS traces and P is an NM -dimensional vector of true $N \times M$ amplitudes, one per channel per template. The matrix, M , is a $2NM \times NM$ mapping matrix whose elements are zeros and ones which maps P to appropriate dimensions. $\mathbf{V}$ is the $2NM \times 2NM$ covariance matrix discussed earlier. This equation has a well known analytical solution which has been presented a few times before in this thesis, which I show below in terms of our current variables of interest:

$$P = [M^T \mathbf{V}^{-1} M]^{-1} M^T \mathbf{V}^{-1} A. \quad (5.27)$$

The amplitudes, P_i , are then combined to obtain a combined estimator:

$$E_{2SN \times M} = \sum_i P_i \cdot \alpha_i. \quad (5.28)$$

Here, we do not go beyond a simple linear estimate for sake of simplicity and will present results from applying the 2SN×M to a toy model. The full flowchart for the 2SN×M algorithm is shown in Fig. 5.7.

5.4.2 Algorithm validation

The 2SN×M algorithm was validated using toy model data sets. The toy pulses were constructed as a sum of energy scaled templates with noise traces. The noise traces were generated from the averaged Cross Spectral Density (CSD) instead of the simple PSD to ensure that correlated noise effects are present in the simulated pulses. In this example, I chose a CSD measured from the CUTE facility at SNOLAB. The CSD matrix represents the mean covariance of noise at each frequency normalized by the the sampling frequency. The covariance matrix is the matrix $\mathcal{V}$ in the N×M filter that captures correlations between channels at the same frequency. The covariance matrix can equivalently be converted to a correlation matrix using $\rho_{ijf} = \frac{\mathcal{V}_{ijf}}{\sqrt{\mathcal{V}_{iif} \mathcal{V}_{jff}}}$ for easy interpretation, where f is a frequency index. Examples of correlations vs. frequency are shown in Fig. 5.8. The correlations are clearly non-zero and vary as a function of frequency. This plot elucidates the need to account for correlated noise in our optimal filtering process.

The square root of the covariance matrix gives the corresponding standard deviation of noise. Assuming Gaussian noise, we generate a smeared DFT coefficient at a given frequency f as:

$$G_f = \mu + \sqrt{cov_f} \cdot \vec{\mathcal{N}}(0, 1), \quad (5.29)$$

where cov_f is a 2N×2N covariance matrix between all channels at frequency f , $\mu = 0$ and $\vec{\mathcal{N}}$ is a vector of random numbers drawn from a uniform distribution. The square root of the covariance matrix is calculated using eigendecomposition [137]. The real and imaginary DFT coefficients generated in this manner are grouped appropriately into complex numbers to obtain a set of correlated complex coefficients for each channel. The inverse DFT of these complex coefficients gives correlated noise traces for every channel. Every such iteration produces a different set of correlated

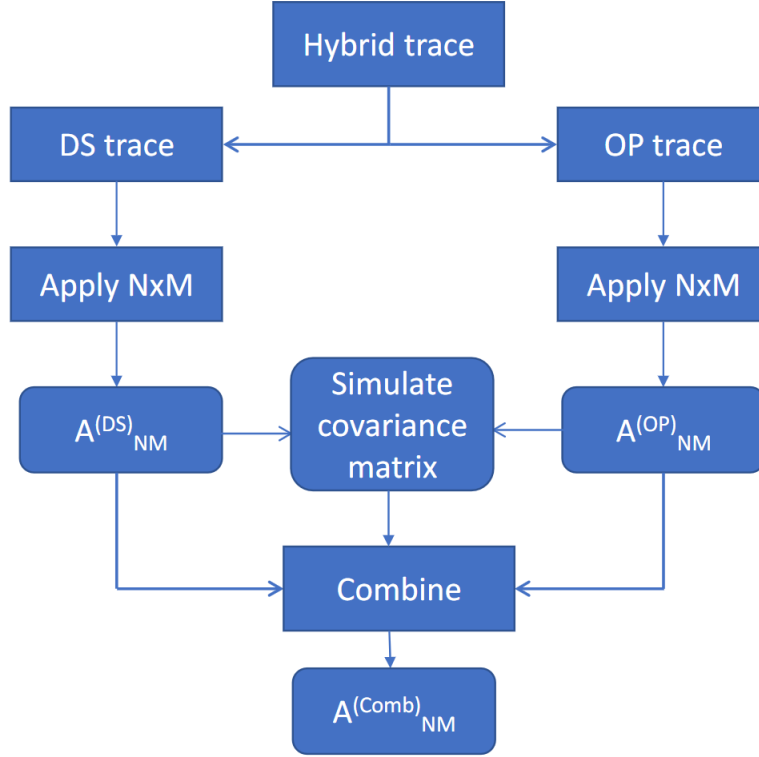


Figure 5.7: A flowchart showing the workflow of the 2-Speed $N \times M$ algorithm. The algorithm begins the same way as the 2SOF or the 2SM by extracting a DS and OP trace from the hybrid trace. Both the DS and OP traces are passed through the $N \times M$ filter. Each of them will produce NM amplitudes that contain energy and position information. These amplitudes are correlated with each other and therefore a covariance matrix is constructed from them. The covariance matrix is used to combine the DS and OP amplitudes using Eq. (5.27) to give NM amplitudes, A^{comb}_{NM} , equivalent to those obtained from applying the $N \times M$ filter to a fully sampled waveform. As a further step, these amplitudes may be combined to return a single energy estimator.

noise events thus acting as a Monte Carlo sampling technique.

About 3000 Monte Carlo samples were generated in this manner. The results reported in this section come from mono-energetic samples with a total amplitude of $1\mu A$ and no time delay, i.e. the rising edge of every pulse starts at the same location in the trace. Pulses generated in this manner are then hybridized by downsampling their tails and processed with the 2SN $\times$ M. A few different versions were tested by progressively increasing the number of channels— 1×1 , 2×1 ,

5.4. $N \times M$ FILTER FOR HYBRID DATA

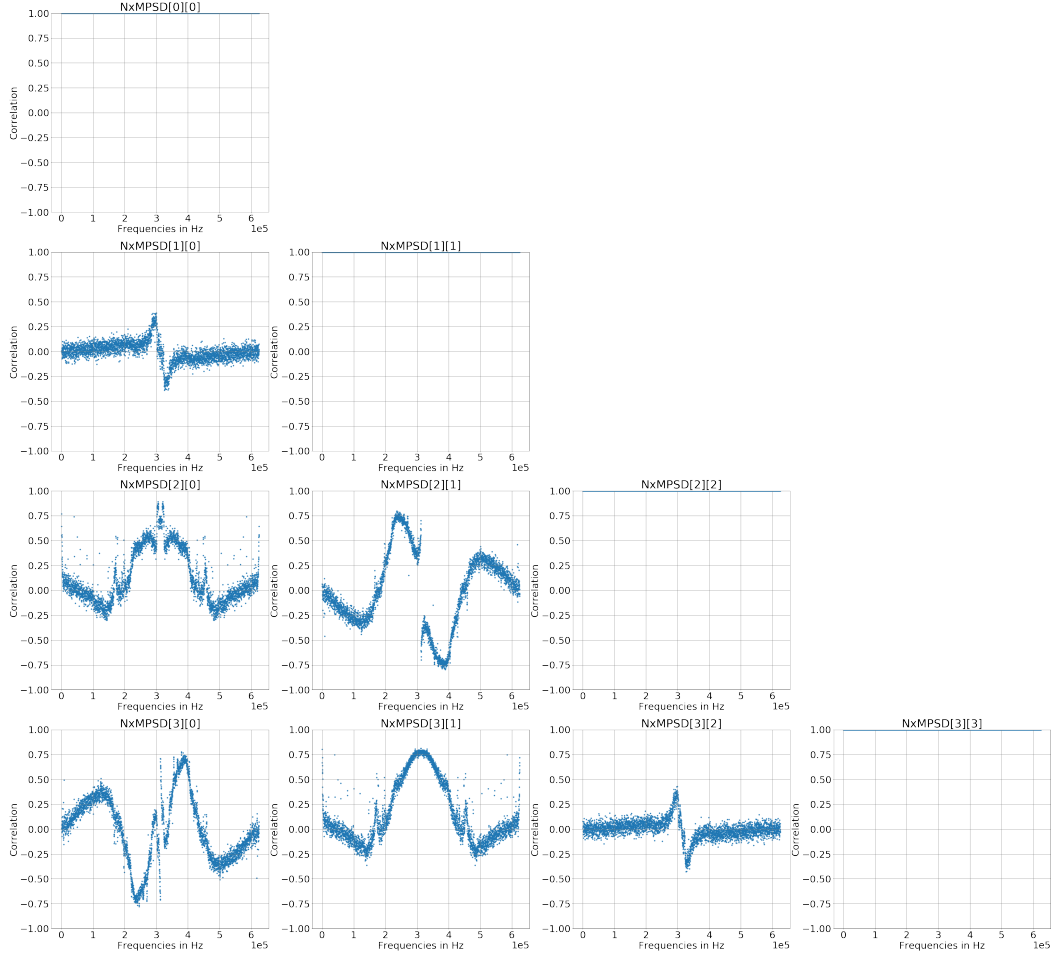


Figure 5.8: A plot of the correlations vs. frequency for 2 channels. The real and imaginary parts are separated in this representation as discussed earlier in this chapter to give a 4×4 symmetric matrix. Hence, only the lower triangle of the matrix is shown.

3×1 , 4×1 and 5×1 . We only process a single template for simplicity. Since the model generated is too simplistic and does not in fact have any pulse shape variation we do not benefit from adding additional templates.

Since the events generated were mono-energetic, we fit Gaussians to the peak and compare the widths of the fitted Gaussians. A comparison was made with the simple 1×1 OF and the $N \times M$ OF applied to fully sampled data to understand the resulting loss/gain from downsampling traces. The $N \times M$ result reported does not use GBDT or other sophisticated corrections and is simply equal to the amplitude. The results are tabulated in Table 5.1.

5.4. $N \times M$ FILTER FOR HYBRID DATA

Test case	Algorithm	Gaussian Width
1 channel	1×1	0.617 ± 0.020 nA
	$N \times M$	0.613 ± 0.020 nA
	A^{DS}	0.626 ± 0.015 nA
	A^{OP}	0.714 ± 0.026 nA
	$A^{2SN \times M}$	0.623 ± 0.019 nA
2 channels	1×1	1.195 ± 0.04 nA
	$N \times M$	1.194 ± 0.034 nA
	A^{DS}	1.204 ± 0.045 nA
	A^{OP}	1.414 ± 0.030 nA
	$A^{2SN \times M}$	1.198 ± 0.045 nA
3 channels	1×1	1.551 ± 0.060 nA
	$N \times M$	1.497 ± 0.054 nA
	A^{DS}	1.549 ± 0.050 nA
	A^{OP}	1.815 ± 0.057 nA
	$A^{2SN \times M}$	1.536 ± 0.043 nA
4 channels	1×1	3.210 ± 0.143 nA
	$N \times M$	2.985 ± 0.100 nA
	A^{DS}	3.094 ± 0.104 nA
	A^{OP}	3.399 ± 0.113 nA
	$A^{2SN \times M}$	3.094 ± 0.115 nA
5 channels	1×1	3.705 ± 0.143 nA
	$N \times M$	3.667 ± 0.123 nA
	A^{DS}	3.667 ± 0.121 nA
	A^{OP}	4.086 ± 0.133 nA
	$A^{2SN \times M}$	3.640 ± 0.083 nA

Table 5.1: A comparison of widths of energy distributions of simulated toy pulses. 1×1 refers to results from the single channel, single template OF which does not take any correlated noise into account. $N \times M$ refers to the combined $N \times M$ output. A^{DS} , A^{OP} refer to results obtained from applying the $N \times M$ filter to the DS and OP pulses, respectively. $A^{2SN \times M}$ refers to the final combined output of the $2SN \times M$ filter. The 1σ statistical error bars on the standard deviation are also shown.

Considering the statistical uncertainties on the Gaussian widths shown in the table, all the estimators seem more or less consistent with each other owing to the simplicity of the toy simulation being used. However, by comparing the mean widths certain trends can be studied. In all cases, the $N \times M$ energy estimator applied to the fully sampled data consistently produces a smaller mean

width compared to the simple 1×1 OF. This improvement is purely from better handling of correlated noise by the former. A^{DS} is consistently narrower in its distribution than A^{OP} . In the previous chapter, I mentioned that which of DS or OP pulse gives the better energy estimator is dependent on the event energy and shape of the pulse with a note that at the lowest energies DS outperforms OP due to a dominance of the low pass filtering effect from downsampling. We clearly see this in action here with the DS outperforming OP estimates at a low event amplitude of $1 \mu\text{A}$. The combined amplitude always provides a peak that is narrower than either DS or OP on its own. And finally an interesting behavior is observed in A^{DS} . Since A^{DS} takes into account correlated noise one may expect this to give better mean peak widths compared to the 1×1 OF. The interesting point is that this is only true if the loss of information due to downsampling is smaller than the degradation from correlated noise in the system. These two effects compete with each other and depending on which effect dominates A^{DS} maybe better or worse than 1×1 OF. For example, in Table 5.1, for $N < 3$, A^{DS} has a larger mean width compared to the 1×1 which roughly equalizes at $N = 3$ and then consistently improves for a larger number of channels. The trends seen here may hold for real data however a study of how the $2\text{SN} \times \text{M}$ performs on real data was not performed as part of this thesis. Further, as mentioned before, considering the error bars in this simplistic case, one cannot conclusively say that any one of these estimators is absolutely better than the other. Based purely on the numbers in Table 5.1 the DS amplitudes are just as good as the combined estimator from the $2\text{SN} \times \text{M}$ and could potentially save us some computational time and power in data processing.

5.5 Summary

In this chapter, the $N \times M$ filter was introduced and a path for how to apply this filter in practice was laid down. A combination of machine learning techniques and the mathematical framework of the $N \times M$ filter can give us outstanding performance with this novel filter. This will be demonstrated in the next chapter where we apply the $N \times M$ processing pipeline to a couple of different types of data sets from SuperCDMS. A brief discussion of the 2-Speed $N \times M$ filter was also provided to motivate some core ideas about how one could adapt this technique to hybrid data.

Gravitational wave detectors such as LIGO often encounter sources of correlated noise similar to SuperCDMS, such as vibrational couplings across channels. The amplitude and shape of such time series pulses in gravitational wave detectors are fit in a similar fashion and could benefit from a similar analysis. This technique is thus adaptable to applications outside SuperCDMS where correlated data exist.

Chapter 6

Application of $N \times M$ filter to SuperCDMS data

The $N \times M$ filter described in the previous chapter was first validated on toy models that were simple combinations of a template and noise simulated from an average noise PSD. Such models are too simplistic to produce noticeable differences in the final energy spectra between simple algorithms such as the 1×1 OF and complex ones like the $N \times M$ OF. Although these models incorporated correlated noise via the mechanism described in Sec. 5.4.1, position dependence of the pulse shape can only be generated via detector Monte Carlo simulations which use packages like GEANT4 [138] capable of simulating the passage of particles through matter.

This chapter will present results from the application of the $N \times M$ processing pipeline to three different data sets. I will first delve into data obtained from GEANT4 simulations of the SuperCDMS SNOLAB HV detectors. The second part will discuss results from re-analyzing old SuperCDMS data from Soudan and the third part will show results from the most recent testing of SuperCDMS SNOLAB detectors at the CUTE facility which was discussed in Ch. 3.

6.1 Validation with GEANT4 simulations

6.1.1 The data

The data used in this study was generated using the simulation framework developed and maintained by the SuperCDMS collaboration. The framework consists of two parts – (a) SourceSim, which simulates particle interactions in the crystal, and, (b) DetectorSim, which integrates the out-

put of SourceSim with detector response and electronics effects to produce “raw data” in the format it is received and analyzed by humans. An overview of this structure is provided in the cartoon in Fig. 6.1.

SuperCDMS Simulation

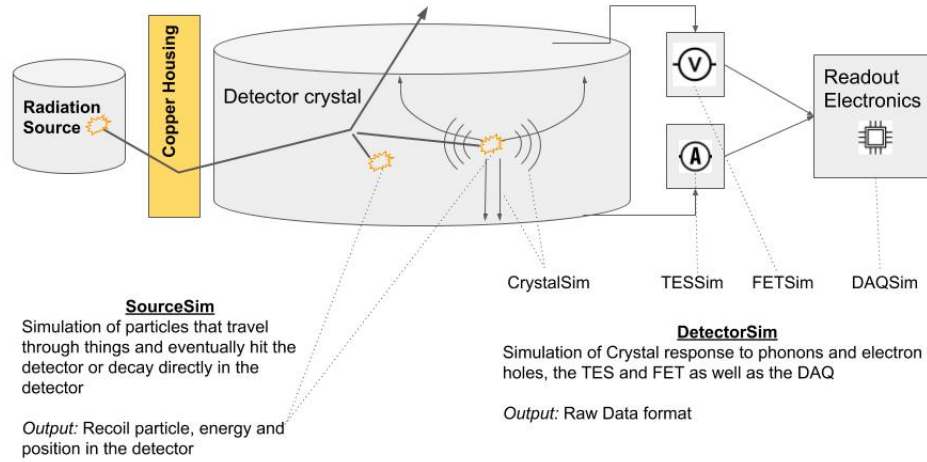


Figure 6.1: A schematic showing the workflow of the SuperCDMS detector simulations framework. Figure taken from internal SuperCDMS simulation documentation.

GEANT4 is integrated into the SourceSim part of the simulation which uses it to generate particle hits. The particles are then propagated through the crystal using CrystalSim and the outputs fed through the readout simulation that includes TESSim for phonons and FETSim for charges. Finally, the output goes through DAQSim which converts the simulated traces into ADC counts, adds noise and produces the raw data format expected by SuperCDMS processing software. The package provides an extensive set of options to tune the settings for each facility, taking into account the relevant backgrounds and geometries.

The data used for the study described in this chapter comprised 10,000 events simulated for a SuperCDMS SNOLAB HV detector. The events were mono-energetic and emulated a full 12-channel readout in Ge. The underlying model was set to generate WIMP signals undergoing electron recoils with recoil energy of 1 keV before NTL amplification and no additional backgrounds producing events. A voltage of 100 V was also included to simulate the NTL effect in the crystal.

Noise was added artificially from an average PSD. Note that the noise was generated simply from a PSD and not a cross-PSD meaning that the effect of correlated noise is minimal in the simulated sample. In addition to the 10,000 signal-like events that were simulated, about 2500 noise traces were generated separately to be able to produce noise PSDs and cross-PSDs for data processing and fitting.

Since the simulation includes the response of the TES and accounts for the propagation of phonons and charges through the crystal, pulse shape variation was a significant factor in the final pulses. This allowed us to test the effectiveness of the position correction techniques with machine learning. Further, since the data are labeled in both energy and position, we can reconstruct position via supervised machine learning. Fig. 6.2 shows the generated pulses that were used in the template generation process.

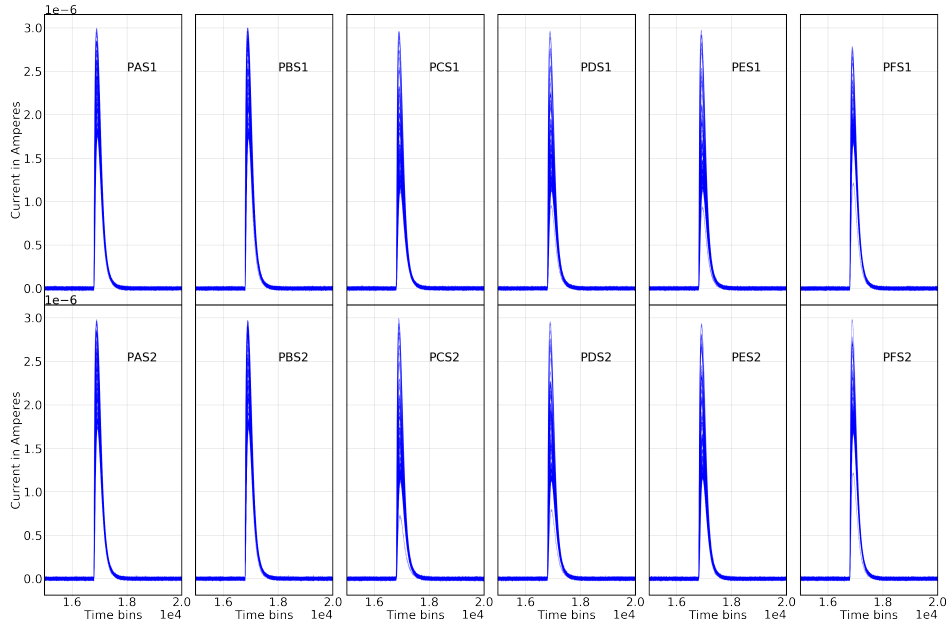


Figure 6.2: A selection of 100 simulated mono-energetic DMC events for a SNOLAB HV detector at 100 V for all 12 channels.

6.1.2 Template generation

The pulses generated were used to create templates with PCA. A simple amplitude selection was applied to remove the lowest amplitude pulses in each channel and the most saturated ones. The former cut off was placed at $0.5 \mu\text{A}$, while the latter cut at $3 \mu\text{A}$. This cut is meant to produce

higher quality templates. PCA can be sensitive to outliers and the goal behind applying PCA to our raw pulses is to capture position variation in the pulse shapes. Low energy pulses with low signal to noise ratio will amplify the noise contribution in PCA while high energy pulses with saturation tend to have non-peaky or rounded tops. Such features will be captured by the PCs thus biasing the templates and potentially increasing the need for more templates to fully characterize a pulse. The final selection of pulses is shown in Fig. 6.2. This choice would obviously mean that saturated pulses with flat/rounded tops would not be fit perfectly by the filter. A potential option in the future is to construct a saturation template that may be able to better account for the special shape of these pulses.

Due to the large cumulative number of pulses across channels, only the first 500 pulses from each channel were used in the template generation. This still comes out to be 6000 training pulses for the algorithm. The pulses were aligned such that the data point corresponding to 0.5% of the maximum height of the pulse shows up at the trigger point. Only the first 10 PCs were calculated for computational simplicity and due to most of the information being contained in the first few components. The PCs for this data set are shown in Fig. 6.3.

The zeroth component looks similar to typical single fall time templates used by SuperCDMS. This represents the mean pulse shape across the detector. All subsequent components represent variations from this mean shape. The dominant factor causing pulse shape variation across the detector is position dependence and therefore, the $N \times M$ filter will capture position dependence in its amplitudes via the weights of each of these templates. Since the input pulses are aligned with respect to the trigger point the first two components have a smooth rise. As sub-dominant effects become dominant in higher components, the imperfections in the pulse alignment are picked up by the PCA. This results in small kinks in the direction opposite to the rise of the pulse such as in component 2. This effect is more difficult to visually identify in components with higher noise.

Each PC is independent and orthogonal to all other PCs. The scree plot in Fig. 6.4 tells us that nearly 95% of the variance is captured by component zero. Component 1 adds about 2%, component 2 adds 0.5%, component 3 adds 0.2% and so on. All higher order components add negligible information. A quick comparison with Fig. 6.3 shows that higher components are mostly noisy. Therefore, we select the first 3 PCs as templates to the $N \times M$ discarding the others due to the independent and orthogonal nature of the PCs. Each PC is then normalized such that the maximum height is one. This is done to ensure that the $N \times M$ amplitudes represent absolute weights of each template and not relative weights with respect to the scale of the respective PC. This way the contribution of each PC can be studied by comparing the $N \times M$ amplitudes directly. With the templates prepared, we proceed to do a 12×3 fit to the simulated events.

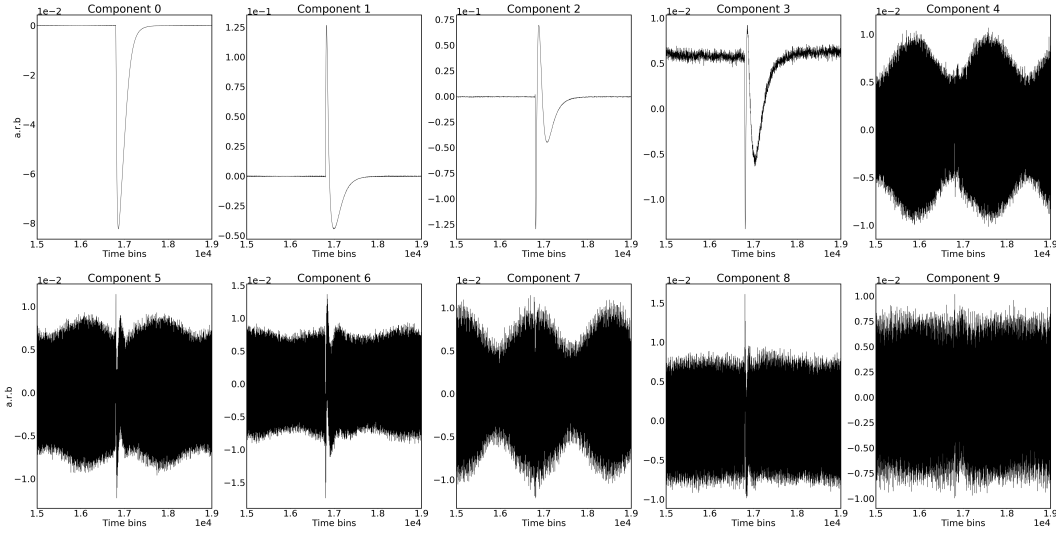


Figure 6.3: First ten Principal Components for simulated pulses. The y-axes are unit-less and denoted by a.r.b for “arbitrary”. The pulse-like nature diminishes with each component owing to the capture of smaller and smaller effects in each subsequent component in PCA. Noise becomes the dominant effect at component 4 and beyond. Although small pulse-like portions are present in between the noise in these higher components, a comparison with the scree plot in Fig. 6.4 shows that the contribution of the component itself is small and consequently the contribution of the peak-like portion of higher order components (if any) are negligible.

6.1.3 Energy reconstruction

The 12×3 fit produces 36 output amplitudes, each carrying varying levels of energy and position information. The amplitudes are non-linear functions of true position as shown in Fig. 6.5. In the figure, I show only the zeroth component for the outer and inner channels where the difference is most obvious. The simplest idea behind position dependence of energy is that events that occur closer to a channel deposit more energy in that channel compared to channels that are farthest from it. PAS1 is the outermost ring channel and PFS1 is the central ring channel in a HV detector (see Fig. 2.3 for channel geometry of the detector). Events that occur in PFS1 deposit the most energy in this channel and the least energy in the farthest channel from PFS1, PAS1. In the figure, the amplitudes are strongly correlated with the X and Y positions of the events. The exact nature of the relationship is complicated and highly non-linear. This translates to a non-linear dependence of the energy on the event location. The GBDT technique described in the previous chapter or other non-linear position corrections are thus an absolute necessity for obtaining the best energy

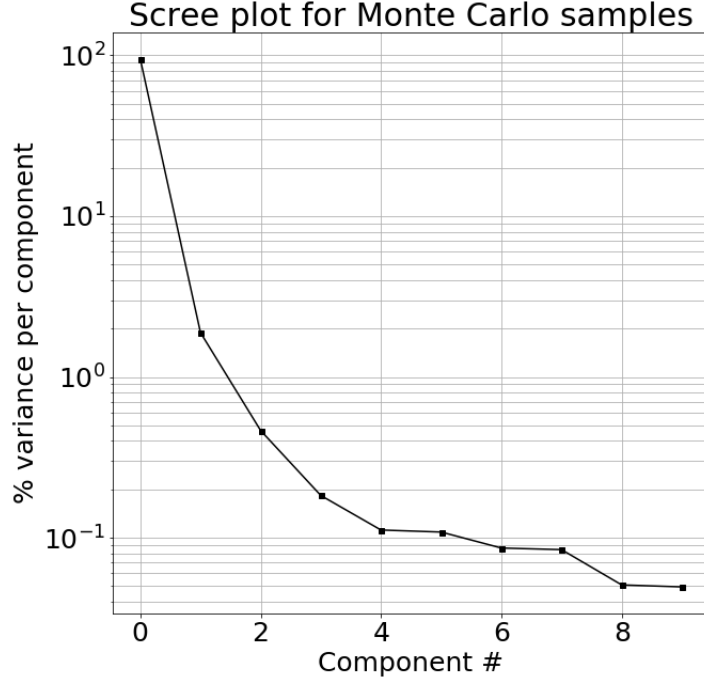


Figure 6.4: Scree plot showing the amount of information/variance in the first ten PCs plotted against the component number. The y-axis is normalized such that the sum of all components (including higher order terms beyond the 10 shown) is 100%.

resolution from SuperCDMS data.

The Z position warrants a brief discussion. Figure 6.5 indicates that the dependence of Z on the output of the $N \times M$ filter is negligible or non-existent. This implies that reconstructing Z-position for this data is near impossible. This is because the data is simulated for an HV detector which relies on phonon signal amplification in the presence of an external electric field. The electric field is applied on the vertical (Z) axis of the detector, drifting the charges in this direction and consequently producing additional phonons rapidly in this direction. The reader may recall that most of the position information is brought by the prompt phonons, but the quick production of secondary phonons in the Z direction washes out the Z-dependence of the energy. In the past, SuperCDMS has utilized parameters such as the difference in energy between the two sides of the detector or the difference in phonon arrival times captured in the time delay as empirical parameters that correlate with the Z position. None of these parameters appear to correlate with the true Z labels in this particular simulated data set, foreboding a difficult situation for fiducialization in the future. However, during the testing of SuperCDMS detector tower at CUTE, it was noted by

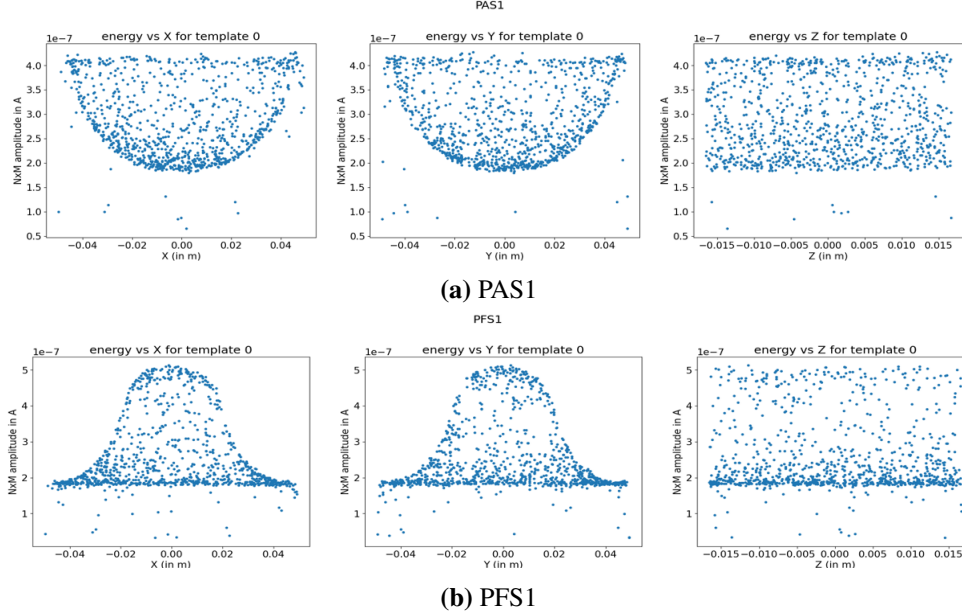


Figure 6.5: The distribution of fitted template amplitudes as a function of the true event position in the detector. Fig. 6.5a and Fig. 6.5b plot the zeroth amplitude component as a function of true X, Y and Z positions for channels PAS1 (outer ring) and PFS1 (center channel), respectively. Events that deposit most energy in the bulk of the detector, i.e. at the center, correspondingly produces the least energy deposition on the outermost edges of the detector and vice versa. The data lacks Z-sensitivity due to the large and quick production of phonons in the z direction which washes out much of the Z-sensitivity in energy.

Shubham Pandey that some residual dependence on Z exists in the energy and time delay parameter, possibly from the non-uniform electric field close to the surface of the detector. The construction of a Z parameter that takes this into account is in the works as part of the analysis effort for data from CUTE as of the time of writing this thesis.

Since the energy correlates with X and Y positions, correcting for the effects of these will enhance the resolution of the spectrum. We achieve this by first computing $E_{lin-est}$ (see Eq. (5.20)) which corrects for linear effects of position and then a second correction is applied by GBDT to address non-linear effects. The weights for the 36 amplitudes were computed using Eq. (5.22). Of the 10,000 events, 90% were used in training the GBDT and the remaining 10% used for independent validation. Note that we do not attempt to create a third independent sample since the goal here is not to judge the generalizability of this model, but to find an empirical model which

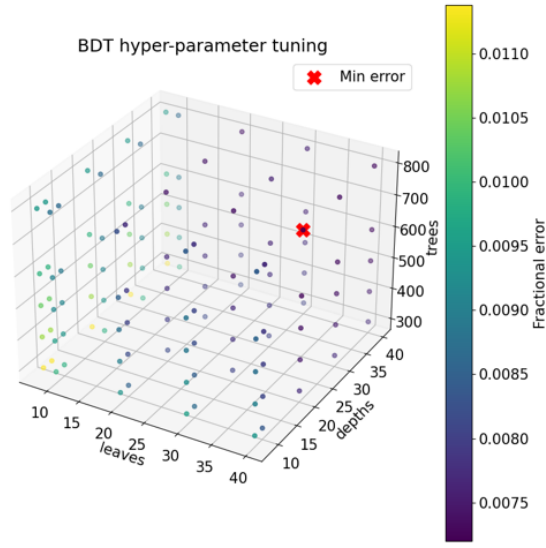
estimates a correction factor for this particular data set. Several parameters were tuned to prevent overtraining and underfitting, namely the number of trees, the depth, the number of leaves and the learning rate. The first three were picked via a grid search which maximizes the resolution of the reconstructed spectrum. An example of the grid search is shown in Fig. 6.6. As the maximum depth goes beyond ten, the GBDT continues to produce improved resolution on the training data, but the corresponding validation data ceases to improve beyond twenty leaves irrespective of the depth. This is a classic sign of overtraining. The model was tuned to give the best energy predictions for the validation data set.

The training was performed with the LightGBM GBDT. The 36 amplitudes served as the input, while the BDTGain defined in Eq. (5.24) was the main energy-related label¹ and X, Y, Z were the position labels. The package used is meant to only handle the mapping of the inputs to a single output and therefore scikit-learn’s [139] MultiOutputRegressor was used to serve as a “meta-estimator” which functions by fitting one regressor per target label. This choice simply presents a slightly more elegant algorithm structure over training four separate regressors manually. The discussion of the position predictions is found in the next sub-section.

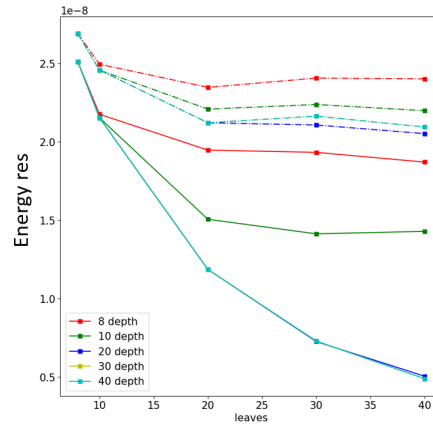
The contour map of the predicted BDTGain is plotted in the true X and Y plane of the detector in Fig. 6.7. The contour lines roughly follow the channel map of the detector shown in Fig. 2.3 when viewed from the top or bottom. This points to the GBDT learning positional information from the amplitudes and mapping this appropriately into the energy correction term, BDTGain.

The final GBDT-corrected $N \times M$ spectrum is compared with the spectrum before correction and with the simple 1×1 OF reconstruction of the same spectrum. The 1×1 OF spectrum was not corrected for any position effects. The comparison is shown in Table 6.1. For this model simulated without any correlated noise, the 1×1 OF is similar to the linear $N \times M$ estimator. The numbers in the table also point to only minimal linear resolution broadening since $E_{lin-est}$ is equivalent to 1×1 OF. A noticeable improvement comes from BDTGain correction – about a two-fold enhancement in energy resolution. This improvement means that use of the $N \times M$ filter, properly accounting for position effects, has potential to significantly enhance the experiment’s sensitivity. This is also a strong encouragement to attempt this with real-world data sets where the noise is not as perfect as in the simulations, and assess how well the technique performs. The plot of the reconstructed energy spectrum compared to the other estimators in the table is shown in Fig. 6.8. The resolutions reported in Table 6.1 are obtained by fitting a Gaussian to the reconstructed peak and then dividing by the mean of the peak to get a value in percentage. The fitted Gaussians are overlaid in Fig. 6.8.

¹Labels refer to the output of the machine learning model in this context.



(a)



(b)

Figure 6.6: A plot showing the result of tuning the GBDT hyperparameters. (6.6a) shows the constructed grid that spans the number of trees, leaves and the tree depth with ‘x’ marking the chosen point. The color is the fractional error in the prediction of the validation energy. (6.6b) shows the dependence of the number of leaves and depth for the training (solid line) and validation (dashed line) data. The y-axis is the spectral energy resolution for the respective data and configuration.

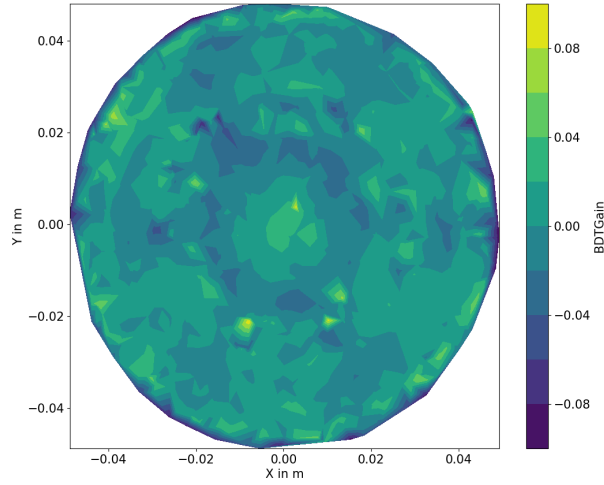


Figure 6.7: A contour plot of the BDTGain in the horizontal plane of the detector. The contour lines map out channel boundaries in the detector (see Fig. 2.3) when viewed along the vertical axis.

Estimator	Peak width in %
1×1 OF	1.93%
$N \times M$ OF without position correction ($E_{lin-est}$)	1.92%
$N \times M$ OF with <i>BDTGain</i> position correction ($E_{N \times M}$)	0.92%

Table 6.1: A comparison of energy resolution between different energy estimators for the validation data set which is independent from the training data.

6.1.4 Position reconstruction

Reconstruction of all three coordinates of the event location – X, Y and Z – was attempted even though the scope for reconstructing Z for this data set was poor. The reconstructed positions are plotted against the true positions in Fig. 6.9a and the final distributions are shown in Fig. 6.9b and Fig. 6.9c. Clearly, the GBDT is able to reconstruct X and Y position nearly perfectly while the Z position is poorly reconstructed. Although the predicted Z seems to correlate with true Z, albeit poorly, the GBDT fails to learn the right Z from the amplitudes as evidenced by the incorrect distribution of the prediction. This is expected from the lack of correlation between the amplitudes and true Z position as discussed earlier in the chapter. X and Y positions are reconstructed to within 0.52 mm and 0.55 mm, respectively². The Z position, on the contrary, is predicted within 5.49 mm.

²These numbers are the 1σ errors in the GBDT prediction.

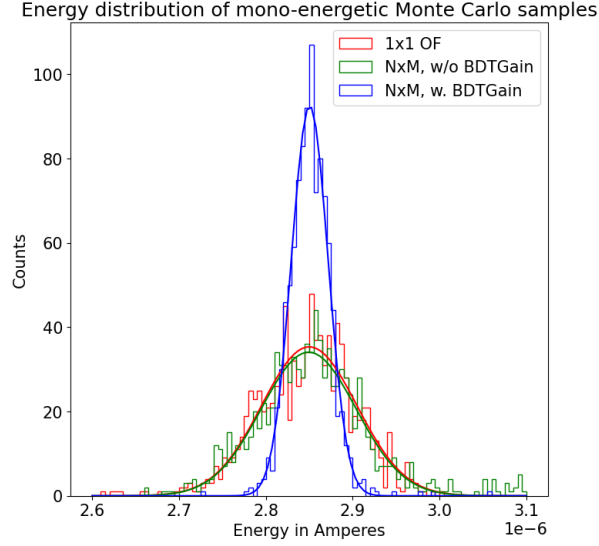


Figure 6.8: A comparison of the reconstructed mono-energetic spectra of simulated events for a SNOLAB HV detector operated at 100 V. The events shown here are from an independent test data set.

The limitation in Z reconstruction with GBDT could be due to imperfections in the simulations that do not fully capture the detector response. An example is the electric field near the surface of the detector, where field lines become non-uniform. This non-uniform electric field can create reduced amplification for near surface events and establish some Z -dependence in the energy. The simulated data used in this study was generated without accounting for this effect and could be the reason for reduced Z -information in the $N \times M$ amplitudes. It is very likely that residual Z -dependencies exist in real data. This study is thus hinting at two major points of caution to the reader: (a) the results here need to be read with a grain of salt when comparing to real data, and, (b) the SuperCDMS simulation package requires upgrades to serve as a direct analog to real data, some of which are already underway since this particular study.

Here, it is important to emphasize that position reconstruction was made possible by the availability of position labels from the simulation truth. Similar information is not available in the case of real data sets due to the unavailability of calibration sources that produce events at known specific positions. Furthermore, SuperCDMS simulations are still limited in terms of producing perfect simulations that capture every physical aspect of a detector in an operational environment perfectly. Since this ability is still under development, training on simulated data and applying the trained model to real data will likely not yield best results. For most machine learning models,

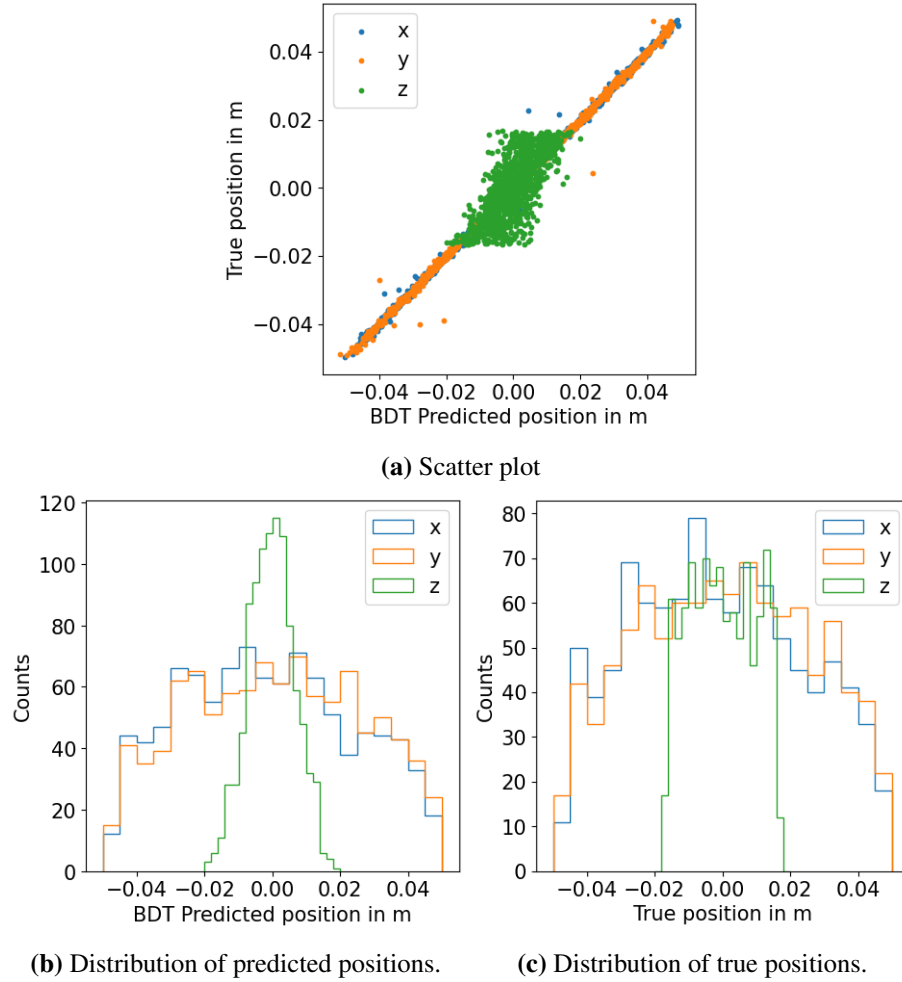


Figure 6.9: A comparison of the GBDT predicted position and the true position. 6.9a directly compares the true and predicted position as a scatter plot. 6.9b and 6.9c show the expected and predicted distributions of the event position. An assessment of all three plots shows that the GBDT is able to reconstruct X and Y positions very well but not the Z position.

the training and test data sets must come from similar distributions, i.e. be produced under similar conditions from a physics perspective. Since the performance of the simulation has not been sufficiently validated against real data, we do not attempt to train the model to be applied to real data on a simulated data set. Therefore, as you will see in the following sections, direct position reconstruction is not studied in the case of real data sets.

6.2 Validation with CDMSLite data

In this sub-section, I discuss the major results obtained from applying the $N \times M$ filter to past data from CDMSLite.

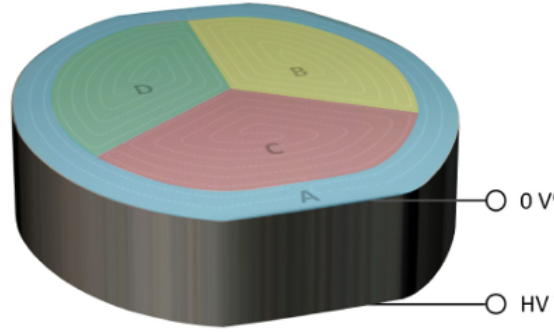
6.2.1 CDMSLite experiment

CDMSLite was a small-scale experiment operated towards the end of the period when we operated an earlier generation of SuperCDMS detectors in Soudan, MN [70, 71]. It consisted of a single tower payload with three iZIP detectors, two of which were used only as vetos. The third detector was grounded on one side and biased on the other side to a large negative voltage. The goal behind this mini-experiment was to assess the advantage of the NTL effect and its impact on dark matter sensitivity. An amplified phonon signal allows us to probe lower dark matter masses than would have been possible otherwise by amplifying the low energy depositions to levels measurable by the detector. The schematic of the CDMSLite detector biasing and the full circuitry is shown in Fig. 6.10.

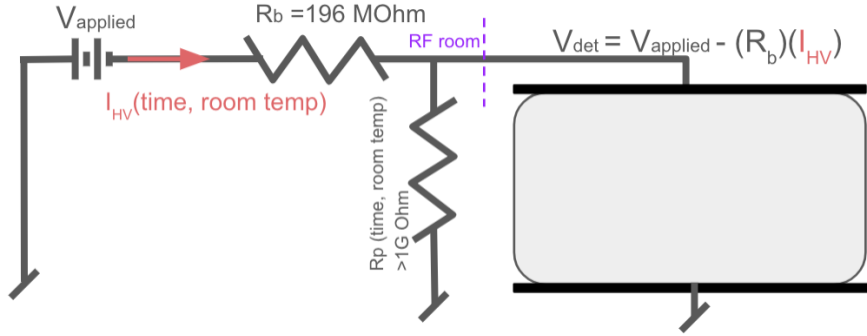
The experiment spanned 3 successful runs which demonstrated superior sensitivity for spin-independent nuclear recoiling dark matter below $10 \text{ GeV}/c^2$ (see Fig. 6.11). In the schema shown in Fig. 6.10, half the detector was grounded and thus one side (four out of the eight channels) was read out. In addition to phonon signals, charge signals were also collected in this setup and this provided additional position and data quality information necessary to perform event selection. This biasing scheme, novel at the time for SuperCDMS, effectively lowered the threshold, with the lowest measured energy at about 56 eV in electron-equivalent units³.

In this section, I present results from re-analyzing the data from CDMSLite Run 3. The primary goal behind this re-analysis effort was to study the effectiveness of the $N \times M$ filter in producing enhanced energy resolution and understand the effect of this enhanced resolution on the dark matter sensitivity. This is also meant to serve as a proof of principle execution of the full $N \times M$ filter pipeline on a real high quality data set.

³See Sec. 2.1 for more information on the unit.



(a) A schematic showing biasing of a Soudan-style iZIP detector in the CDMSLite scheme. Figure taken from [70].



(b) A schematic showing the external circuitry supplying the bias voltage to the detector. Figure taken from CDMSLite internal documentation.

Figure 6.10: A schematic representation of the CDMSLite detector biasing scheme.

6.2.2 Template generation

The templates were generated following the same method described in the previous chapter and section. Many of the standardized data quality cuts used in the published analysis were adopted into the template generation procedure to obtain a sample of “clean” representative pulses of the sort that dark matter might produce. A basic cut to select only events that triggered on the detector was performed and of that we only select events that are “physics type”, i.e. events which were collected in dark matter data mode. The χ^2 quantity from the non-stationary filter was used as a parameter which is sensitive to “bad events” which include, but are not limited to, pile-ups (events with multiple signals in one trace) and saturated pulses (events which have slower fall times or rounded/flat peaks). A χ^2 cut was performed in the following manner. The χ^2 quantity

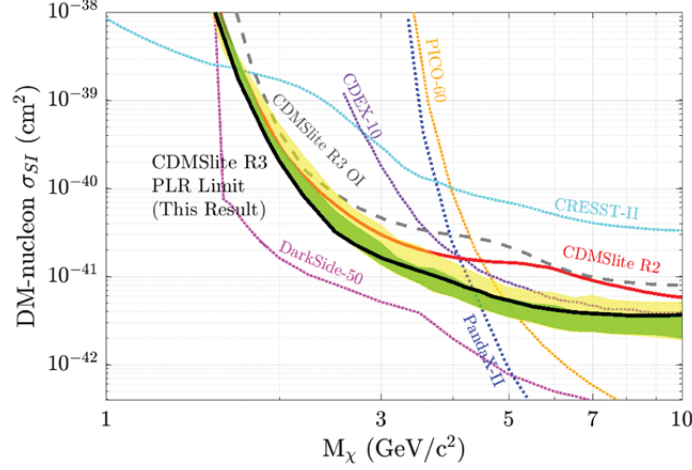


Figure 6.11: 90% profile likelihood ratio confidence limit for nuclear-recoiling dark matter performed with CDMSlite Run 3 data. Limits from CDMSlite Run 2 is also shown in the plot. Figure taken from [71].

was grouped in energy and the mean χ^2 in each group found. The cut was a polynomial interpolation between these predefined points at 3.5σ deviation from the mean fit polynomial. A glitch cut was also performed to remove glitch events originating from the electronics. Glitches usually are a kind of correlated noise which exists across detectors and triggers on multiple detectors. The number of phonon triggers for each event was used as a means to identify these events. A standard deviation cut was performed to remove events with a large pre-pulse baseline standard deviation, often originating from a secondary pulse in the baseline. Finally, an amplitude selection was placed in the broad range of 1 to 12 μA (0.47 to 5.7 keV_{ee}). I also further impose a restriction on the minimum pulse height by only allowing events with at least 1 μA height in all channels to be selected. The final selected pulses are shown in Fig. 6.12 and consists of 1960 pulses per channel.

The first 10 PCs of these pulses were calculated using the same techniques described in the earlier section, the first four of which are shown in Fig. 6.13. The corresponding scree plot is shown in Fig. 5.2. I do not show the other components since they were ultimately not used in the processing. The selection was made based on the scree plot which shows that components higher than 3 contribute less than 0.1% variance per component. The PCs are normalized such that the maximum height is unity.

We perform a 4×4 fit and obtain 16 amplitudes as the output. The amplitudes were first corrected for all the same effects that were accounted for in the published analysis. This is to ensure

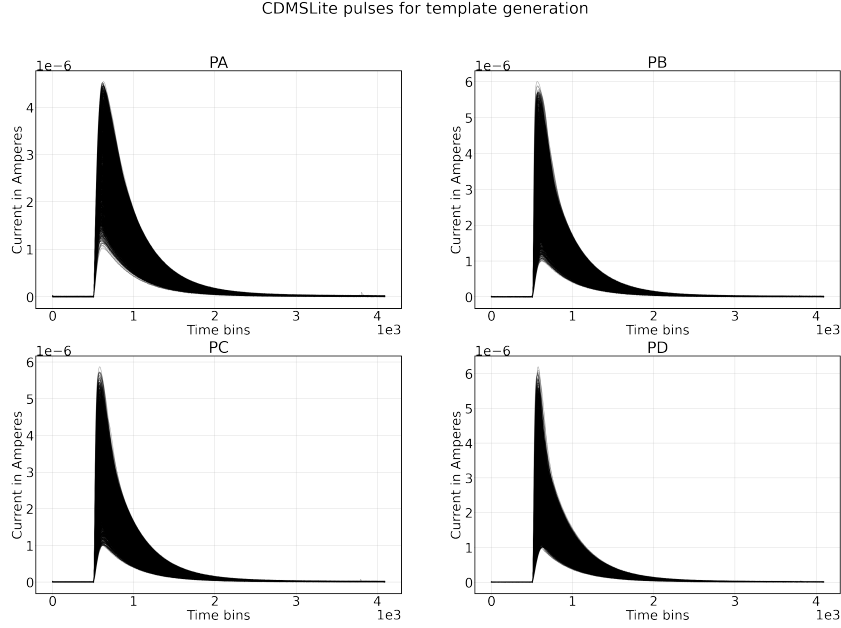


Figure 6.12: The final selection of pulses used in template generation for the four channels that were read out in the CDMSLite detector.

that the results in this study are ultimately comparable to the published results. More details to follow in the next sub-section.

6.2.3 Energy reconstruction

CDMSLite operated Ge detectors and therefore observed a few internal electron capture peaks in the energy spectrum whose true energies are known. These calibration peaks, namely the K, L and M shell lines, serve as calibration points for the spectrum as discussed in Ch. 3. For Run 3, the detectors were “activated” with a ^{252}Cf neutron source. The neutrons produce an increased fraction of ^{71}Ge whose decay via electron capture produces the peaks. The calibration peaks and their branching ratios were described in Ch. 3 and also used in the calibration of the SuperCDMS SNOLAB Ge HV detectors at CUTE. The calibration events for this study were selected using the same overall procedure followed in the original analysis which plotted the amplitude vs. current for candidate events in which the distribution of events in the peak is clearly visible. A straight line fit to this plot followed by a cut at certain number of standard deviations ensures that most of the peak is selected. The number of standard deviations was selected visually in a manner that incorporates most events.

6.2. VALIDATION WITH CDMSLITE DATA

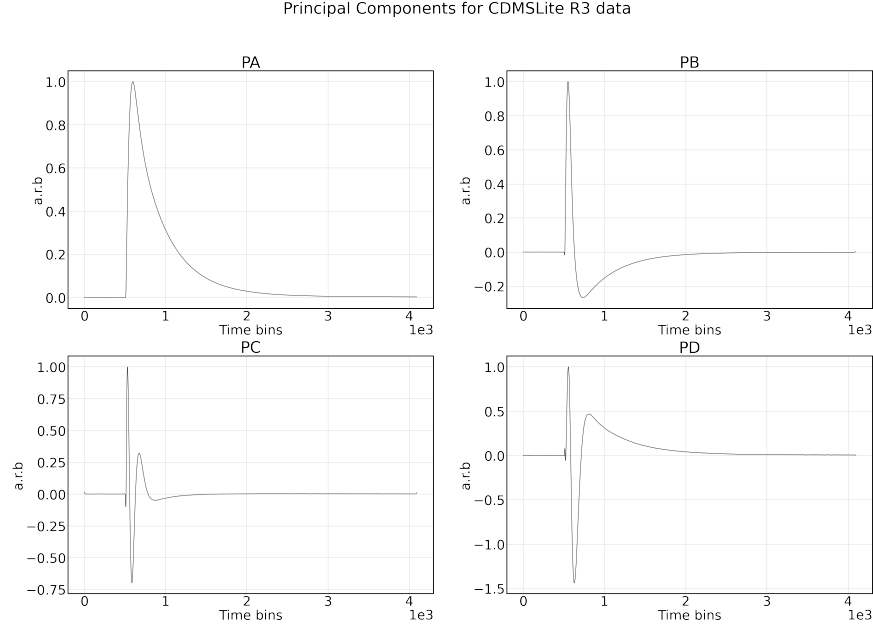


Figure 6.13: The first four PCs derived from CDMSLite data normalized to a maximum height of 1.

The full machinery of cuts employed in the published analysis were rerun for this study. This includes 9 live time cuts⁴ and 9 other data quality cuts. These cuts together removed bad events like those with pile-ups, saturation, glitches, noise and events that are not dark matter candidate events like those which triggered on multiple detectors. In addition the cuts also removed series taken during periods of high noise over the course of the five-month data taking period in 2015. More information can be found in the published paper Ref. [71] and the thesis Ref. [76] and in the internal summary documentation for the analysis in Ref. [112].

During CDMSLite Run 3, several environmental factors impacted the measured energy and noise of the detector. While most of them simply required appropriate cuts for removal, others required simple corrections to the energy estimator to “fix” the issues. The same corrections were applied in sequence prior to interpreting the $N \times M$ energy estimator.

The circuit diagram in Fig. 6.10b shows the biasing mechanism in CDMSLite mode. A constant voltage must be supplied across the detector over the course of the run and, ideally, no current must be dumped across the detector. However, a large parasitic resistance to the ground, R_p , enabled small currents (of the order of ~ 100 nA) to be dumped across the detector causing the de-

⁴Cuts which reduce the livetime of the detector

tector voltage to drift over the course of the run. This drift in voltage results in a drift in measured energy, thus worsening our resolution. It was found that the measured phonon energy decreases with the measured leakage current I_{HV} (Fig. 6.10b). A similar correlation was observed in the zeroth component of the NxM amplitude. Higher amplitude components sometimes seem to correlate the opposite way to the leakage current. A correction term depending on the leakage current was used to correct for this drift for each event as follows:

$$A_{hv-corr} = \frac{A_{uncorrected}}{1 - \frac{I_{hv} R_b}{V_{norm} + \epsilon}}. \quad (6.1)$$

Here, V_{norm} is the nominal bias voltage of 75 V and ϵ is the ionization potential for a single e^-/h^+ pair (~ 3 V for Ge). This correction is applied to each N×M amplitude.

A second factor that varied over the course of the run was the detector temperature. The variation in temperature effectively shifts the bias point of the TES and thus changes the amount of energy collected. Although CDMSLite was operated at a mean base temperature of 50 mK, the temperature was found to drift over the course of the run between 45 mK and 60 mK. The measured, current-corrected phonon amplitudes correlated positively with the measured base temperature for most energy estimators. So a line was fit to the energy-temperature distribution and used to correct this dependence. For the N×M, the correlation was not clearly visible in the plots but a fit returned positive and negative slopes on a channel by channel basis. The corrected amplitude is obtained as:

$$A_{C2} = A_{hv-corr} \cdot [1 + (T_{base} - \text{mean}(T_{base})) \cdot C_1]. \quad (6.2)$$

The above correction is empirical. The factor C_1 is estimated by fitting a straight line to the amplitude - temperature curve and $\text{mean}(T_{base})$ is the mean base temperature of 50 mK.

The amplitudes corrected for these environmental effects are, at this stage, combined to obtain $E_{lin-est}$ as in Eq. (5.20). The weights, α , are computed using Eq. (5.22) applied on the K shell, L shell and M shell peaks separately. The amplitudes of the events within 5σ deviation of each peak are combined using the weight at that peak. The weights for the events outside the 5σ region of a calibration peak are computed using an iterative piece-wise linear interpolation method since this is the simplest first order approximation to any function. This means that a preliminary estimate of the energy of each event is computed by summing all the amplitudes with a weight of one. These preliminary energy sums are then used to linearly interpolate the weights from the calibration peaks to other energies. Events with energy outside the calibration range (0.16 keV - 10.37 keV)

use weights that are equal to the weights at the left-most or right-most peak, i.e. the M shell or the K shell depending on which peak is closest. Note that due to the limited availability of points to compute these weights before linear interpolation, it introduces systematic uncertainties in the combined energies.

It was found that using all events in the training sample worked fine for the K shell peak due to the larger number of available events but caused the peaks to be artificially narrow at the L and M shells due to overtraining. Therefore, in order to do a fair assessment of the final estimators at lower energies, the L and M shell events were sampled repeatedly about 500 times via bootstrapping. Each sample was split into a training and test data set with weights calculated separately for each. For the K shell, the first 300 events were kept aside as completely independent test samples and bootstrapped 500 times to obtain a fair estimate of the resolution. $E_{lin-est}$ was computed for each test data set and converted from amplitudes to energy units using a conversion factor computed for each calibration peak. The calibration factor for events outside the 5σ region of a peak are interpolated in the same manner as the weights. I denote this combined $N \times M$ estimator in energy units as E_{C2} .

The final step was a position correction. The $N \times M$ filter was conceived and implemented only recently. At the time of the original analysis, the residual amplitudes from the 1×2 OF were used as a parameter that correlated with position. The energy estimator used in the original analysis showed a correlation with these residual amplitudes which was corrected by another linear empirical correction with the same form as in Eq. (6.3). The form is shown below:

$$E_{C3} = E_{C2} \cdot [1 + (A_{res} - \text{mean}(A_{res})) \cdot C_3]. \quad (6.3)$$

Here, both C_3 and $\text{mean}(A_{res})$ were estimated from a straight line fit to the energy vs residual amplitude curve. I compute this correction to obtain an energy estimator from the $N \times M$ fit that has identical corrections to the published estimator. The combined $N \times M$ energy, E_{C2} , is already corrected for linear position effects by tuning the weights. Thus, this estimator only correlates by a small amount with the residual amplitudes and is calculated only for legacy reasons. Table 6.2 shows the values of all the calibration factors described in this section as calculated from the data. Further the full sequence of corrections described in this section is shown in Fig. 6.14.

Our primary mode of position correction for this data is using the GBDT. A GBDT which takes the 16 amplitudes as input is trained on the K shell peak. The availability of data was limited, and therefore augmentation was performed to triple the data size. SuperCDMS detectors have symmetry about the vertical axis. The channel schematic for the CDMSLite detector in Fig. 6.10a

C ₁ (in K ⁻¹)				mean(T _{base}) in K
PA	PB	PC	PD	0.05 for all channels
-1.064e-09	1.272e-08	-2.131e-08	-2.986e-08	
C ₂		C ₃	mean(A _{res})	
(in keV/Amperes)		(in Ampere ⁻¹)	(in Amperes)	
127706999.058		-1155.031	-3.213e-06	

Table 6.2: A list of correction factors to correct the $N \times M$ energy estimator for CDMSLite Run 3 data. PA, PB, PC and PD are the names of the channels which were read out on the detector. The values are calculated from the data itself (except the mean base temperature which was fixed from the nominal operational temperature) and applied to the estimator. The linear position correction factors are only shown here for legacy reasons.

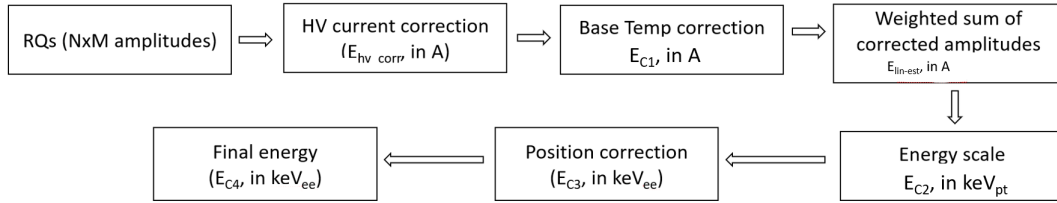


Figure 6.14: The full sequence of corrections applied to the $N \times M$ energy estimator to account for environmental and event position effects. The conversion from total phonon (pt) to electron equivalent (ee) units was discussed in detail in Ch. 2. The position correction shown in the sequence is performed using GBDT.

shows three inner channel and one outer ring. We use the rotational symmetry of the detector to rotate the data about the vertical axis of the detector. This triples the data size due to three possible rotational configurations. Including rotated versions of each training event essentially builds the rotational symmetry of the detector into the training set, effectively reducing the number of free parameters. Unfortunately, due to the unavailability of the second side of the detector, further rotations were made impossible which would have otherwise sextupled the data size.

Here, we only train on energy labels since we do not know the true positions of events from real data sets. The model is trained to predict the BDTGain described in the previous chapter. 96% of the augmented K-shell data was used in the training and about 50 data points (roughly 4%) were saved for independent validation and hyperparameter tuning. The L and M shell points also serve as independent test data sets to validate the model on. The final BDTGain corrected estimator is compared to the published estimator and the slow amplitude from the 1×2 OF. The latter two estimators do not account for correlated noise in the data and are only corrected for linear position

effects using the 1×2 OF residual/fast amplitudes.

The K shell peak was fit with a Gaussian to extract the resolution from its width in electron-equivalent units. The number of events available to determine weights for the L and M shell peak is an order of magnitude lower than the K shell peak which makes it difficult to keep a sufficient number of independent test events. To overcome this limitation, events from the respective peaks are bootstrapped 500 times, each time keeping aside a different subset of events for training and independent validation. This gives a sense of the statistical error in the resolution estimate. A resolution is obtained for all 500 bootstrapped data sets and histogrammed to show the distribution of resolutions due to the uncertainty in the weights. This distribution is fit with a Gaussian and the mean value of the fitted Gaussian is used as an estimate of the resolution for that calibration peak obtained from this estimator. This exercise is repeated for the non-stationary filter and the 1×2 OF to compare the equivalent quantities. The results are tabulated in Table. 6.3. Plots of the K peak and the distribution of L and M resolutions overlaid with the fits are shown in Fig. 6.15.

Estimator	K shell σ in eV_{ee}	L shell σ in eV_{ee}	M shell σ in eV_{ee}
1×2 OF	130.31 ± 0.68	37.72 ± 0.17	12.48 ± 0.04
NF	130.37 ± 0.67	35.67 ± 0.15	12.71 ± 0.04
$N \times M$ OF with BDTGain	49.68 ± 0.18	34.62 ± 0.17	12.15 ± 0.06

Table 6.3: Comparison of energy resolution of the $N \times M$ filter with other energy estimators in SuperCDMS along with the 1σ statistical uncertainty from the Gaussian fit. NF stands for the non-stationary filter. The distribution of the resolutions from 500 bootstrapped samples are used to compare the K, L and M shell resolutions.

Here, we see a factor of 2.5 improvement in the resolution of the K shell peak over other energy estimators owing to the enhanced position correction offered by the BDT. The contribution of the position uncertainty decreases linearly with energy for SuperCDMS detectors (see Eq. (5.23)). Therefore, the largest position effect is felt at the highest energy and is minimal at lower energies. At lower energies the factor which drives the resolution is the noise which is represented by the baseline resolution of the detector. Accounting for the correlated noise improves the baseline resolution and therefore most of the improvement at the lowest energies is from this effect and not the BDTGain correction. This was also ascertained by calculating the $N \times M$ resolution without BDTGain correction using only a simple linear correction shown in Eq. 6.3. At the M shell peak the BDTGain corrected estimator is equivalent to the linear corrected $N \times M$ energy. The amount of enhancement from BDTGain correction decreases as a function of energy, but nevertheless the

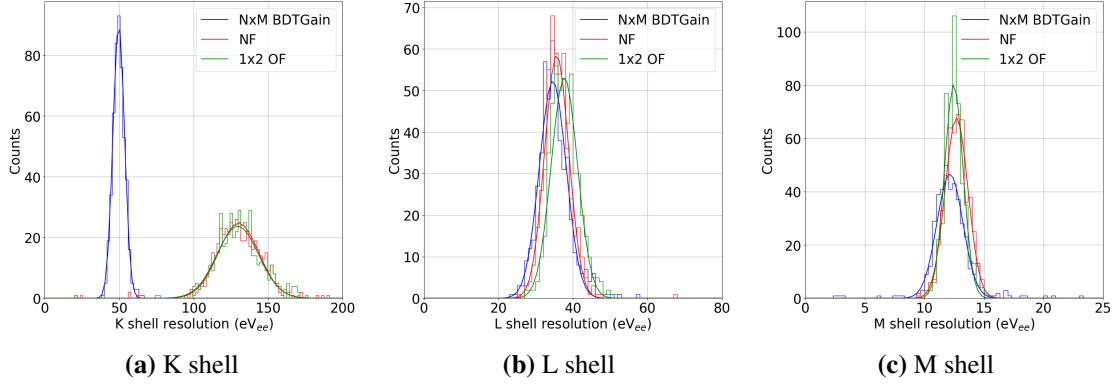


Figure 6.15: A comparison of the resolution of calibration peaks from a CDMSLite Run3 data. The distribution of resolutions from 500 bootstrapped samples for BDTGain-corrected $N \times M$ energy, the non-stationary filter (NF) energy and the two-template filter (1×2 OF) energy are shown for the K, L and M shell peaks.

$N \times M$ filter as a whole offers better resolution at all three peaks.

The baseline resolution requires some additional discussion. The first important point is that the baseline must be calculated from amplitudes that were extracted without any time shifting of the pulses. This is to ensure that we do not bias ourselves by forcing the optimal filter to search for peak-like structures in the data. Secondly, we only use the first amplitude component from all channels, setting the higher order amplitudes to zero. The higher order components, again, fine-tune the fit towards signal-like structures and will bias our estimates. The baseline estimate simply represents the OF’s ability to distinguish signal from noise at the simplest level. Therefore, even though a 4×4 fit was done for this study, the final 3 amplitude components were set to zero to estimate the baseline resolution.

Since the weights are calculated as a function of energy and the “true energy” of the baseline peak is 0 keV, it is not clear how to estimate the weights at the baseline. There is a mathematical degeneracy in the method used here for the other peaks (see Eq. (5.22)) which when applied at $E=0$ sets all the weights to zero as a trivial solution. We do not yet have a way to directly compute these weights from the baseline peak and leave that as a task for further studies. For the purpose of this thesis, we produce an approximation to demonstrate a feel for the numbers and use it to compute a resolution model. We use a set of weights estimated from the M shell peak which is the closest to the baseline peak by inherently making the assumption that position effects below this energy are negligible. This is of course not entirely true and thus the baseline resolution computed in this approximation does not show much improvement. CDMSLite Run3 was divided into two sub-runs

due to change in the environmental noise mid-run. The published analysis computed baselines separately for each sub-run and extracted a limit that took into account the correlation between the resolution parameters between the two sub-runs. This study follows the same technique and computes the baseline resolution and a resolution model for each sub-run. The baseline calculation uses only the first component of the 4×4 fit, without any time shifting. The average values for each sub-run are shown in Table 6.4. The published baseline resolution in the original analysis was calculated from the simple 1×1 OF. I recalculate these numbers in my analysis by fitting Gaussians to the final baseline peak. The baseline was calculated for each individual data set and then averaged over the period of the whole sub-run.

Estimator	σ_{bs} from R3a	σ_{bs} from R3b
N×M	$12.27 \pm 2.42 \text{ eV}_{ee}$	$16.24 \pm 1.01 \text{ eV}_{ee}$
1×1	$10.62 \pm 2.02 \text{ eV}_{ee}$	$13.1 \pm 0.71 \text{ eV}_{ee}$

Table 6.4: A comparison of baseline resolution for N×M and 1×1 OFs for each sub-run in CDMSLite. We expect the N×M baseline resolution to improve from better treatment of correlated noise. However, the observation is contrary to our expectations in the table due to limitations in estimating $\vec{\alpha}$ for zero energy events.

The baseline resolution we observe is unfortunately worse than what was computed in the original analysis despite our expectation that it would improve with better treatment of correlated noise. This is pointing to the insufficiency of our technique of applying the M shell weights to the baseline peak.

The usual SuperCDMS resolution model defined in Eq. (5.23) was used to model the resolution of the NF and 1×2 filters. We note that application of the BDT logarithmically suppresses the resolution at high energies and therefore, an empirical model is constructed to model the resolution for the N×M filter. The model is defined as:

$$\sigma(E) = a \ln(bE + c), \quad (6.4)$$

where a, b and c are empirical parameters with units of keV, keV^{-1} and none, respectively. These parameters will be estimated from the data. The data used is the resolution at each calibration peak along with the baseline peak. The resulting model with the 3σ statistical uncertainty on the fit is shown in Fig. 6.16. The NF and 1×2 models are also shown for comparison. The N×M filter improves energy resolution by reducing effects of correlated noise and position dependence. While the former is a constant irrespective of energy, the latter is a linear function of energy. Therefore,

the amount of improvement gained from the combined $N \times M$ filter and GBDT technique decreases as we go down in energy. At the lowest energies, if the amount of correlated noise in the detector is low, then the improvement from using the $N \times M$ filter is also correspondingly small. This is what we see in Fig. 6.16. However, for the new detectors that were tested at CUTE, the effect of correlated noise is larger and therefore we expect a bigger improvement from the technique described in this chapter. The improvement may also be dependent on the amount of training samples available to calculate the weights, α , which turned out to be our limiting factor in the data sets that were studied in this chapter.

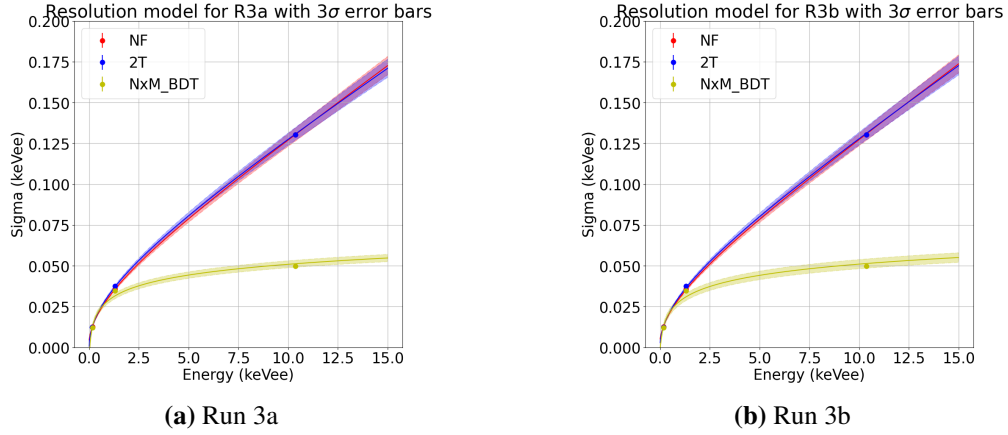


Figure 6.16: A comparison of resolution models from various energy estimators for CDMSLite Run3. NF refers to the non-stationary filter, 2T refers to the 1×2 OF and $N \times M$.BDT refers to the BDTGain corrected $N \times M$ model.

The values of the parameters a, b and c are shown in Table. 6.5. The empirical modeling limits our ability to meaningfully interpret the various uncertainty components described in Eq. (5.23) such as the Fano term and the position uncertainty. However, the model still suffices as a way to predict resolution at a given energy in this range.

Run	a (keV)	b (keV) $^{-1}$	c
3a	9.53×10^{-3}	20.91	0.314
3b	10.00×10^{-3}	16.41	0.856

Table 6.5: Estimates of the parameters in Eq. (6.4) for CDMSLite data.

6.2.4 Impact on dark matter sensitivity

A quick study was performed to understand the effect of an improvement in resolution on the dark matter sensitivity. The resolution values shown in Table 6.6 were chosen to study the change in dark matter sensitivity. The values reflect resolutions from the K, and M shell peak, if all events were used in estimating the weights. Thus, the numbers in the table are somewhat artificially narrowed for the N×M since the calculated resolution is estimated with the same events used to tune the estimator. They are however a perfect example of an improved resolution model.

Estimator	K shell σ in eV _{ee}	L shell σ in eV _{ee}	M shell σ in eV _{ee}
NF	116.85 ± 8.32	34.61 ± 1.60	14.51 ± 0.92
N×M OF with BDTGain	49.52 ± 3.10	29.20 ± 1.45	11.27 ± 0.50

Table 6.6: A table of resolution values chosen for demonstrating the effect of resolution on dark matter sensitivity. The two models differ only modestly in resolution at low energies, but show two-fold reduction at 10.37 keV. The difference in models roughly capture the improvement in resolution from performing a GBDT correction.

The dark matter limits were calculated using the same background and uncertainty assumptions as the published Run 3 analysis. The analysis incorporated the following background models:

- Electron capture peaks from ^{71}Ge , ^{68}Ga , ^{65}Zn , ^{55}Fe ,
- β decay of tritium,
- Compton scatter from various sources in the experimental setup, and,
- surface events from the decay of ^{210}Pb .

I will only provide a brief overview of the background models since this information is available in great detail via other references such as Ref. [70, 71, 76]. The electron capture peaks were modeled as Gaussians centered at a peak value taken from Ref. [140] and the amplitudes of the peaks were taken from Ref. [141]. The tritium background was modeled from the non-relativistic β -decay theory [142] as

$$f_{\text{tritium}}(E_{KE}) = C \sqrt{E_{KE}^2 + 2E_{KE}m_e c^2} \cdot (Q - E_{KE})^2 \cdot (E_{KE} + m_e c^2) \cdot F(Z, E_{KE}). \quad (6.5)$$

Here, C is a normalization constant, $F(Z, E_{KE})$ is the non-relativistic approximation to the Fermi function, Q is the endpoint energy of the spectrum, m_e is the electron mass and E_{KE} is the kinetic

energy of the electron with $f(E_{KE})$ denoting the distribution of this energy. The Compton background produces steps at distinct energies (see Ch. 3). The location and heights of the steps in the CDMSLite detectors were determined from GEANT4 simulations and modeled as a sum of error functions. Finally, the surface backgrounds were generated from GEANT4 simulations, modeling the decay of ^{210}Pb produced from radon exposure. Three locations were considered in the simulations, namely, the surface of the crystal itself, the copper on the top of the detector (the lid) and the cylindrical detector housing. Strong systematic uncertainties were noted from the variability in the electric field across the detector and the location of the fiducial volume cut. These uncertainties altered the shape of the surface background spectrum. This effect was taken into account via “morphing parameters” which changed the shape of the spectra based on fits to the data. The mean shape, ρ_{med} , and the 1σ uncertainty on the shapes were determined from the simulations and used to mark the upper (ρ_{up}) and lower (ρ_{low}) bounds of the shape. The shape of the final background spectrum was allowed to vary between these two bounds, controlled by the morphing parameter m as:

$$\rho(E, m) = \begin{cases} \rho_{med}(E) + m \cdot (\rho_{up}(E) - \rho_{med}(E)), & m > 0 \\ \rho_{med}(E) + m \cdot (\rho_{med}(E) - \rho_{low}(E)), & m < 0 \\ \rho_{med}, & m = 0. \end{cases} \quad (6.6)$$

The dark matter signal is modeled with Eq. (1.6) smeared by the resolution and cut efficiencies. In total, a 20-parameter profile likelihood fit was performed to the data in the published analysis. However, since this is a short sensitivity study, I attempt this on simulated data sampled from the sum of backgrounds and a signal smeared by the resolution obtained from our resolution model and scaled by cut efficiencies and live time. An example of a generated data set is shown in Fig. 6.17.

Since the same cuts and data were used, the cut efficiencies and live times were kept the same as the original analysis. The author acknowledges that some of the simulated background models such as the surface and tritium backgrounds depend on the resolution used in the simulation, although these effects are likely to be small since these distributions are much broader than the energy resolution. Since we do not regenerate these simulated background spectra with our alternate model, the results shown here should not be interpreted as a formal dark matter limit and must be read with adequate caution in mind.

The simulated data was fit with a model with 20 parameters which included the number of events from each background components – one each (six in total) for Compton, Tritium, Ge, Ga,

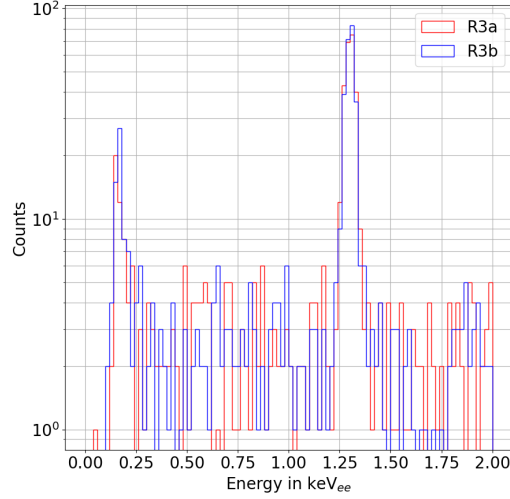


Figure 6.17: Example spectra for Run 3a and Run3b generated from our background and signal models for CDMSLite Run3, smeared by the resolution and scaled by cut efficiency and livetime.

Zn, Fe, three morphing parameters with constraints (one per surface background), three efficiency constraint parameters, six resolution parameters with constraints (three per sub-run), a constraint term for the Lindhard k term which determines the yield (see Eq. (2.7)) and the dark matter interaction cross section σ_χ . The data is fit with the total likelihood function which has three components [71, 76]:

$$\mathcal{L} = \mathcal{L}_{\text{poisson}} \times \mathcal{L}_{\text{energy-distr}} \times \mathcal{L}_{\text{constraints}}. \quad (6.7)$$

The components of the above likelihood function after taking the logarithm and dropping constant terms which are irrelevant to the fit are:

$$\begin{aligned} \ln(\mathcal{L}_{\text{poisson}}) + \ln(\mathcal{L}_{\text{energy-distr}}) = & - \left[v_\chi + \sum_b v_b + \sum_{sb} v_{sb} \right] + \\ & \sum_{i=1}^N \ln \left[v_\chi f_\chi(E_i, \vec{n}) + \sum_b v_b f_b(E_i, \vec{n}) + \sum_{sb} v_{sb} f_{sb}(E_i, m_{sb}, \vec{n}) \right], \end{aligned} \quad (6.8)$$

and the constraint term comprises Gaussian priors as follows:

$$\begin{aligned} \ln(\mathcal{L}_{\text{constraint}}) = & -\frac{(k - \mu_k)^2}{2\sigma_k^2} - \frac{1}{2} \sum_{i,j}^3 (e_i - \mu_{e_i}) \mathbf{E}_{ij}^{-1} (e_j - \mu_{e_j}) - \frac{1}{2} \sum_{i,j}^6 (r_i - \mu_{r_i}) \mathbf{R}_{ij}^{-1} (r_j - \mu_{r_j}) \\ & - \frac{1}{2} \sum_{i,j}^3 (m_i - \mu_{m_i}) \mathbf{M}_{ij}^{-1} (m_j - \mu_{m_j}). \end{aligned} \quad (6.9)$$

In Eq. (6.8), the subscripts χ , b and sb denote dark matter/signal components, backgrounds and surface backgrounds, respectively. The v terms represent the Poisson counts of the number of events from that component. The probability density function of each component is represented by f which is a function of the energy, E_i , of each event i and $\vec{n}$ which is a vector of the cut efficiencies and resolution parameters. m_{sb} is the morphing parameter for the corresponding surface background. In Eq. (6.9), k , e , r and m are the Lindhard parameters, efficiency parameters, resolution parameters and morphing parameters, respectively. μ denotes the mean values used to constrain the respective terms. σ_k , $\mathbf{E}$, $\mathbf{R}$ and $\mathbf{M}$ denote the standard deviation and covariance matrices between the respective parameters. The Lindhard parameter, k , was constrained to a mean of 0.157 and a deviation of 0.05. Although the published paper simply constrains the efficiency with a Gaussian prior, the code base used in this study instead implements it as a morphing of the shape of the efficiency curve with 1σ bounds, similar to the shape of the surface background curve. The other two matrices and mean values for resolution and morphing were derived from the data with the estimated resolution model parameters serving as μ_r and $\mu_m = 0$ for the mean surface background curve, ρ_{med} .

Five data sets were generated from the signal + background model similar to what is shown in Fig. 6.17. Ideally, more data sets are preferred for a sensitivity study, but limitations in computational resources prevented the use of a larger number of data sets. The full limit was only calculated for a randomly chosen data set. The uncertainty in the limit for a small region of dark matter masses was computed to see if any improvement seen in this specific data set is statistically significant or not. The limit from an arbitrarily chosen simulated set is shown in Fig. 6.18. The limit from the data set generated with $N \times M$ resolution parameters is compared to the limit obtained from a dataset generated with the non-stationary filter resolution parameters. Note here that the only difference between the two data sets is the underlying resolution model and, of course, the randomization in the sampling.

The limits were generated by scanning over a range of dark matter masses. For each dark matter mass the likelihood function is fit with and without the cross-section as a free parameter. If

$\hat{\theta}$ are the nuisance parameters calculated with the cross section constrained to a fixed value, σ , and $\hat{\hat{\theta}}$ are the nuisance parameters calculated with the cross section floating ($\hat{\sigma}$ is the estimated cross section), then the test statistic, q , is defined as:

$$q = \begin{cases} -2\ln(\lambda_q), & \hat{\sigma} < \sigma \\ 0, & \hat{\sigma} > \sigma \end{cases} \quad (6.10)$$

where

$$\lambda_q = \frac{\mathcal{L}(\sigma, \hat{\theta})}{\mathcal{L}(\hat{\sigma}, \hat{\hat{\theta}})}. \quad (6.11)$$

The test statistic is calculated for each dark matter mass and over a range of cross sections. The p-value represents the probability that the signal hypothesis is true. So a p-value close to 0.1 implies a rejection at 90% confidence level. The p-value is the likelihood of obtaining a q as small as is observed and is given by:

$$p(q, \sigma_\chi) = \int_{q_{test,obs}}^{\infty} f(q_{test}, \sigma_\chi) dq_{test}. \quad (6.12)$$

This would ideally need to be simulated over a large number of datasets, but it was shown that CDMSLite data satisfies Wilks' theorem [143] which says that in the asymptotic limit, this $f(q_{test}, \sigma_\chi)$ maybe approximated as a sum of a delta function and a χ^2 function [71, 76]. Under this assumption, it becomes formally easier to compute the p-value. The limit is drawn at the lowest cross section which gives only a 10% probability of being a signal-like event, i.e., the null hypothesis is rejected at the 90% Confidence Limit (CL). The limit was calculated at select points to reduce the use of computational resources. The curve is then interpolated with a 7th order polynomial via a least squares fit to obtain the smooth curve shown in Fig. 6.18. In the regions where the two curves deviate, an approximate uncertainty was calculated from fitting five different data sets, computing a limit from each of them, smoothing with a second order polynomial fit and then computing the standard deviation of the curves from the mean value at each test mass. This exercise was only performed in one of the deviating regions, once again due to limited computational time and resources.

The key takeaway from this plot is that the limits obtained from the $N \times M$ energy estimator is equivalent to the NF estimator. An apparent difference between the two estimators is observed below 0.8 GeV/c² and 2 – 6 GeV/c². However, considering the uncertainty of the green curve between 4 and 6 GeV/c², it is likely that the limit from the red curve likely has a similar uncertainty.

This uncertainty likely spans outwards to higher and lower masses. This means that limit from the improved resolution model, which is primarily driven by the improvement at 10 keV, is equivalent to that obtained from the NF resolution model. If there is an improvement, it is in the lower mass range of the curve since the $N \times M$ limit applied to the real CDMSLite R3 data set showed a small improvement in this region. It is possible that proper accounting of the systematic uncertainties and regeneration of the simulated background models with the improved resolution may produce noticeable differences in the limits, but this was not explored in this study since this is only meant to serve as a proof of principle and to provide a rough feel for what one ought to expect in real-world datasets. I also note here that although the sensitivity enhancement is small here, since CDMSLite did not have a lot of correlated noise, we do expect that SNOLAB-style detectors will benefit more from the $N \times M$ especially at lower energies, due to better treatment of correlated noise.

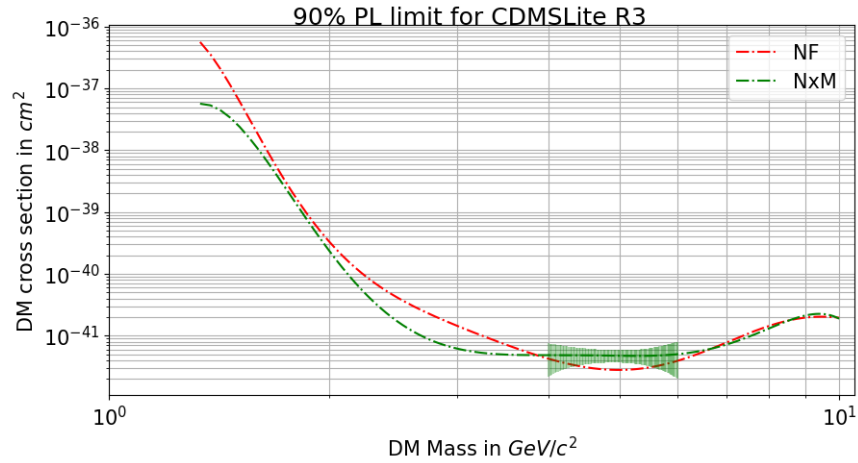


Figure 6.18: An example limit derived from a randomly chosen Monte Carlo data set simulated from the summed background and signal model. The red and green curves show the limits for data sets generated using the non-stationary and $N \times M$ filter resolution model parameters, respectively. The shaded green region shows the 1σ uncertainty in that mass range for the $N \times M$ limit computed from five different data sets.

6.3 Validation with data from the CUTE facility

Here, I present results from applying the $N \times M$ filter to a more recent dataset from CUTE. The data set is from a SNOLAB HV detector and was obtained during the tower testing effort which spanned the period of October 2023 - March 2024.

6.3.1 Data preparation

The data used for this study came from a Ge detector (T3Z1) in the stack operated at a bias voltage of ± 50 V. The detector was activated via controlled exposure to a radioactive neutron source several times from November - December 2023. The neutrons emitted from the source produce ^{71}Ge in the crystal which decay via electron capture (see Ch. 3 for more information). About 65 hours of calibration data collected during December was used in this study. The rate of ^{71}Ge K shell electron capture varied between 0.010 and 0.015 Hz depending on how soon after activation the data was collected.

The events were fit with both the simple 1×1 OF and the $N \times M$ filter. The templates for the two filters were generated from the raw traces in the K shell region with a few basic data quality cuts applied. This included selecting only those events which triggered on that detector, and a χ^2 cut. The template generation is an iterative process where a preliminary assumption about the pulse shape is used to generate a model to process the data as soon as it is acquired. This data is analyzed further offline and then used to generate more representative templates that align well with the data. The same set of events were used to generate templates for the 1×1 OF and the $N \times M$ filter. A simple, flat χ^2 cut was devised from the preliminary processing, tuned visually based on the distribution of points in Fig. 6.19. The cut only removes events that have large χ^2 and are visually distinct in the χ^2 -amplitude plane. Since this χ^2 is measured from fitting only approximate models that are not entirely representative of the pulse shape in the data set, the loose stringency of this cut ensures that only truly bad events are removed without any fine tuned selections.

The pulses were shifted such that the point of 0.5% of the height of the pulse is aligned at the trigger point of the data. The final aligned pulses are shown in Fig. 6.20. The $N \times M$ templates were computed with PCA with the PCs normalized. PCA was performed on the first 200 events from each channels, thus on a training data set of 2400 pulses. As mentioned before, the higher order PCs pick up more noise than others in spite of applying a low pass filter. Therefore, the noisiness of the higher order PCs were smoothed over with a piece-wise Savitzky–Golay filter which used a first-order polynomial⁵. The Discrete Fourier Transform of the smoothed template was compared to the non-smoothed version to ensure that the frequency domain components of the smoothed templates did not acquire unwanted or unusual artifacts from the smoothing process. The final templates are shown in Fig. 6.21c, the first ten PCs are shown in Fig. 6.21a and the corresponding scree plot used in the selection is shown in Fig. 6.21b.

From a combined assessment of the PCs and the scree plot, four templates were chosen and a

⁵The smoothing was performed by Musaab Al Bakry, a graduate student at UBC.

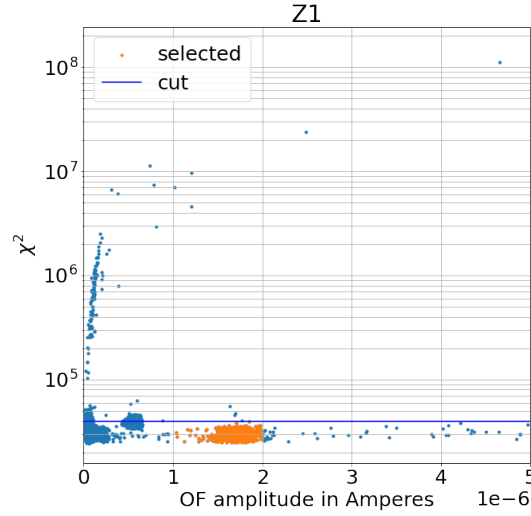


Figure 6.19: The χ^2 vs OF amplitude curve for detector T3Z1 from the SuperCDMS SNOLAB tower tested at CUTE. The blue line shows the cut which was roughly set by eye. The orange points show the final selected events in the K shell peak.

12×4 fit was performed on the data.

6.3.2 Reconstructing the Ge spectrum

About 3001 K shell events were selected from the 1×1 OF spectrum by placing an amplitude cut in the range $25 - 30 \mu\text{A}$. Similarly, the L and M shell peaks were selected with amplitude cuts in the range $3 - 4 \mu\text{A}$ and $0.3 - 0.6 \mu\text{A}$, respectively. The selected events were then used to estimate the weights, α . The L and M shell cuts selected 317 and 123 events for each peak, respectively. The limited number of events has the consequence that the estimated weights from the L and M shells are overtuned to these samples. In order to produce a fair assessment of the technique, the events in the peak are once again bootstrapped similar to CDMSLite. 500 samples were generated from each calibration peak. The bootstrapped dataset was split into training and test samples, the weights were calculated for each bootstrapped training set and resolutions computed for each test set. The same procedure as CDMSLite is followed where events outside the 5σ region of the peaks use interpolated weights while those within use the same weights as the peaks. Further, events lower in energy than the M shell and higher in energy than the K shell use the weights of the nearest peaks. The results are shown in Fig. 6.22. The distribution of the resolutions at each calibration peak is once again fit with a Gaussian, the mean of which is quoted as the resolution in Table 3.7. We find

6.3. VALIDATION WITH DATA FROM THE CUTE FACILITY

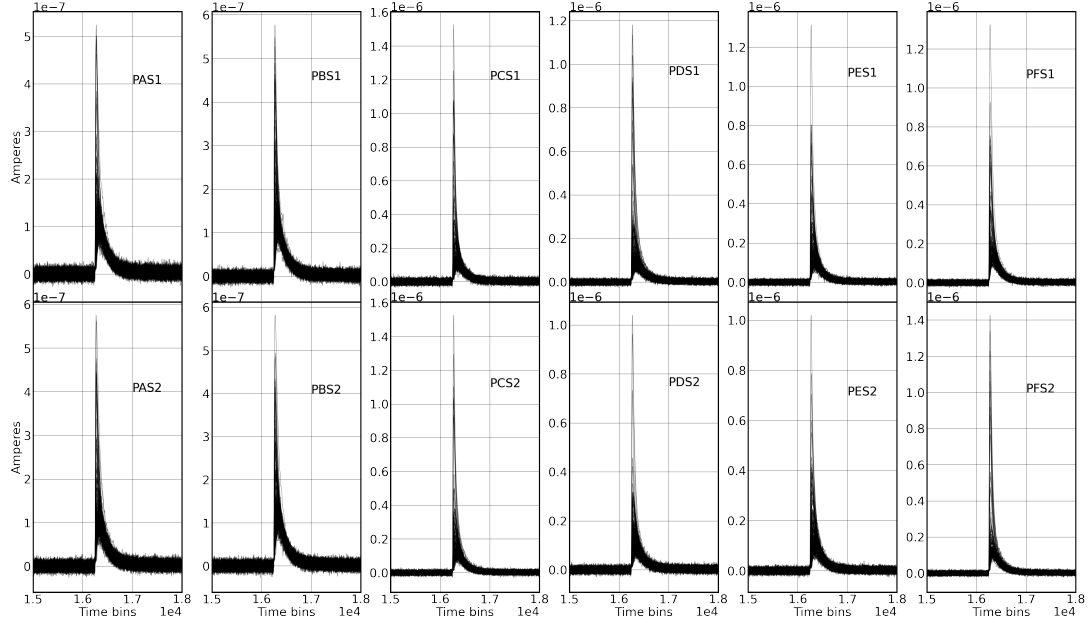


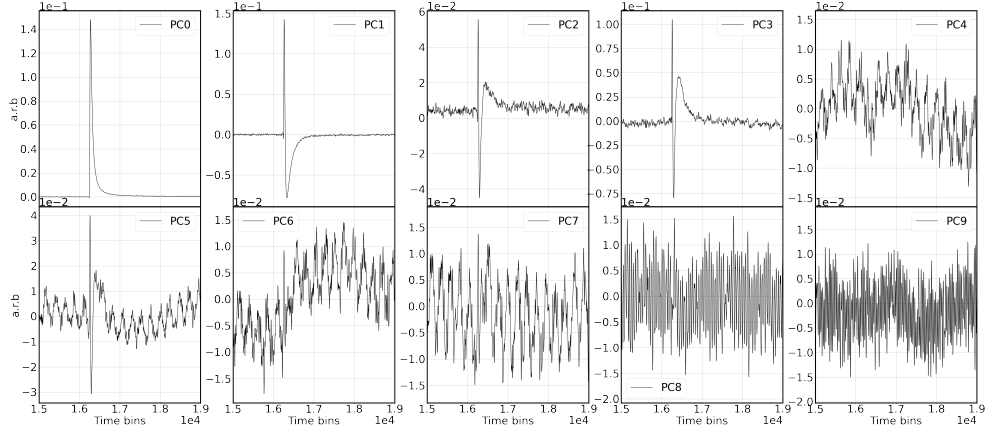
Figure 6.20: The final selection of pulses used in template generation for T3Z1, a Super-CDMS SNOLAB HV detector tested at CUTE. The x-axis shows time bins, while the y-axis shows the pulse height in amperes. The subplots are labeled by channel names as PAS1, PBS1, etc.

a significant improvement over the values estimated from the 1×1 OF, which was also discussed in Ch. 3. The full results are tabulated in Table. 3.7. Most of this can be attributed to the handling of correlated noise in the $N \times M$ filter and the correction of position effects achieved by weighting the contribution of each amplitude component.

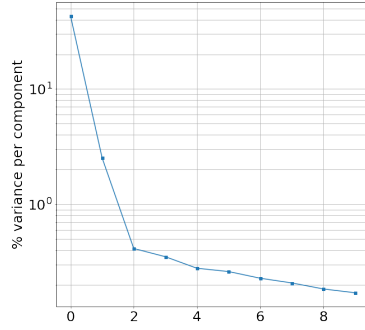
The baseline resolution was also calculated by refitting the data with the $N \times M$ filter but without time shifting of the pulses. The amplitude of the baseline events were then combined with the weights from the M shell peak to obtain a baseline resolution of about 5 eV_{ee} which is modestly better than CDMSLite. Note that the baseline quoted here is from when multiple detectors were operated simultaneously which caused excess noise to couple in via the readout electronics. As discussed in Ch. 3, this is an issue that is actively being tackled by the collaboration. The resolution of this and other detectors is expected to improve once many of the factors that have an impact on the resolution such as the temperature of the still stage of the fridge and the electronics noise are resolved.

Although the author would have liked to perform a study of non-linear position correction with

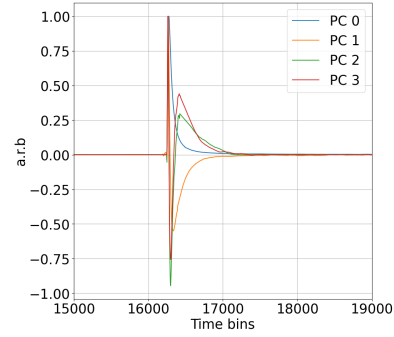
6.3. VALIDATION WITH DATA FROM THE CUTE FACILITY



(a) Principal components for T3Z1 detector. The x-axes are in time bins while the y-axes have arbitrary normalization from the PCA. The first four components have the most pulse-like information.



(b) Scree plot for T3Z1 detector. The x-axis is component number, while the y-axis is the % of variance in each component.



(c) Normalized templates used in the $N \times M$ processing. The first four PCs were chosen based on the scree plot and visual assessment of the PCs.

Figure 6.21: This figure shows (a) the first 10 PCs of the data set along with (b) the scree plot and (c) the final normalized and smoothened templates.

the BDTGain for this dataset, there were severe limitations in data size. For a 48-dimensional input feature space corresponding to the 48 $N \times M$ amplitudes, a larger number of events is necessary to train the decision tree. This means that the training data size of 2900 or so (keeping aside, say 100 independent test events) would need to be increased significantly. We hoped to achieve this by data augmentation similar to what was done in the CDMSLite data where the rotational symmetries of the detector was exploited to rotate the input features along the vertical axis of the detector. The availability of additional channels would have allowed us to sextuple the data

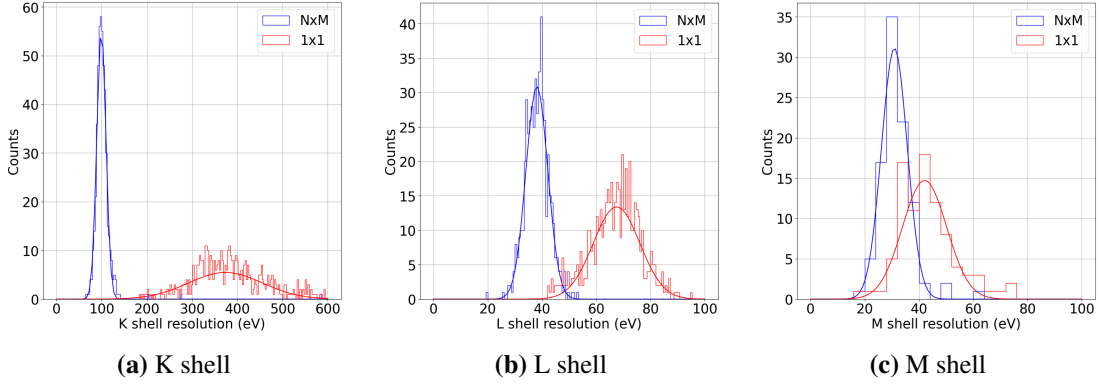


Figure 6.22: A comparison of calibration peaks for T3Z1 detector from CUTE tower testing. The distribution of resolutions from 500 bootstrapped samples for BDTGain-corrected $N \times M$ energy and the 1×1 OF energy estimators are shown for the K, L and M shell peaks.

size through rotations. However, it was found that sources of asymmetries existed in the detector, such as a couple of channels with relatively excess noise that made them distinctive. Such noise seems to impact the reconstructed energy in a manner that made swapping or rotating detector channels impossible. Therefore, augmentation could not be performed for this data which is quite unfortunate. This brings to attention an important limitation of the $N \times M$ pipeline developed in this thesis – a strong need for large data sets.

6.3.3 Effect of missing channels in the detector

A third study was performed to assess the impact of a missing channel on the resolution of the detector. It was noted during tower testing that a few detectors often had channels that were rendered non-operational due to a variety of reasons. Some examples of these situations were reported in Ch. 3. The net result of a missing channel is that the data acquired will miss the energy component from that channel. In addition, depending on the location of the channel, varying degrees of information about the position of the event may be lost. This will have some effect on the final energy resolution that we reconstruct. The original goal of this study was (a) to assess the impact of a missing channel on the energy resolution of the activation peaks, and, (b) to study the usefulness of the GBDT in handling a missing channel, perhaps even compensating for the missing positional information. The second part of this study ultimately could not be carried out due to the limited data size available to train the GBDT. Therefore, this study is only an assessment of how much the

resolution changes when a channel is non-functional. Note that we use the term “missing channel” to denote such non-functional channels; they are not in fact physically absent.

The study was performed on a detector with all twelve channels fully functional. This means that we are able to make a direct comparison between the scenario where a single channel is missing and none are missing. For this reason, we pick T3Z1 from Tower 3 (the same detector studied earlier), since it is also our most investigated detector with a full set of functional channels. We only consider the K shell peak since this peak offered sufficient number of events without the risk of overtraining. First, the resolution for the K shell peak was calculated using all twelve channels in the manner described earlier. Then, a single channel is dropped, i.e., its amplitudes ignored. The remaining 44 amplitudes are used to calculate 44 new weights which are then used to combine the energies using Eq. (5.20). The procedure is repeated by adding the previous channel back and dropping a new channel. This is continued until all channels are dropped to ensure that the effect of losing channels of different types are captured in the study.

We broadly define three categories of channels based on shape for this purpose:

- Outer channels – The four outer rings; PAS1, PAS2, PBS1 and PBS2
- Inner channels – The two central circles; PFS1 and PFS2, and,
- Sector channels – The six angular channels surrounding the inner channels on all sides; PCS1, PDS1, PES1, PCS2, PDS2 and PES2.

The combined amplitudes are converted to energy units and then fit with Gaussians. The width of the Gaussian is used as a measure of the resolution of the peaks in each scenario. The results from this study are summarized in Table 6.7.

The conclusions drawn in the list below are purely based on the K shell resolution. A few general observations can be made from the above table:

- On average, the least effect on resolution results when missing an outer channel. This is likely because the outer channels often collect the least energy and offer the least positional information due to their circular geometry.
- On average, the loss of an inner channel has a medium effect on the energy resolution.
- On average, the loss of a sector channel has the most effect on energy resolution. This is because the geometrical arrangement of the sector channels have been designed such that positional information is easier to pull out from the data in these channels. Therefore, the

# channels	Missing channel type	Missing channel name	K shell σ (in eV_{ee})
12	N/A	N/A	96.982 ± 1.37
11	PFS1	Central	98.07 ± 1.37
11	PFS2	Central	101.20 ± 1.40
11	PAS1	Outer	98.34 ± 1.35
11	PAS2	Outer	99.15 ± 1.40
11	PBS1	Outer	97.31 ± 1.40
11	PBS2	Outer	97.71 ± 1.36
11	PCS1	Sector	98.15 ± 1.35
11	PCS2	Sector	104.24 ± 1.46
11	PDS1	Sector	103.72 ± 1.41
11	PDS2	Sector	110.42 ± 1.55
11	PES1	Sector	99.34 ± 1.39
11	PES2	Sector	106.52 ± 1.51

Table 6.7: A comparison of activation peak resolutions when various types of channels are “missing” in the detector. The 1σ errors are also shown.

loss of one of these channels contributes significantly to a loss of energy resolution from increased position uncertainty.

- The loss of a channel on side 2 appears to have a larger effect on the energy resolution than the loss of a side 1 channel in this detector. This is pointing to somewhat of a biased collection of energy in these detectors or perhaps a biased trigger mechanism that causes more events to trigger closer to this side on average, thus driving the loss of resolution from missing a channel on this side.

The linear weight method used in the $N \times M$ allows us to “retrain” on the data to produce new optimal weights to account for missing channels. This is guaranteed to give better resolutions than simply using the weights from all twelve channels. For example, the $N \times M$ resolution with channel PFS2 missing, when calculated with the weights generated for twelve channels give a resolution of 144.38 ± 2.48 eV. But regenerating the weights after dropping PFS2 gives a resolution of 101.20 ± 1.40 eV, which is $\sim 43\%$ better than the previous estimate.

This study does not investigate the loss of more than a single channel since in the case of our science grade detectors, there is only ever one missing channel as seen from the CUTE tower testing. This study posits a fortunate scenario for the SuperCDMS collaboration. Although a few

detectors are missing a single channel, in the scenario where the missing channels are not recovered, the observed loss in reconstructed energy resolution will be minimal because the missing channels are always an outer channel (specifically, either PAS1 or PAS2).

6.3.4 Reconstructing the Si spectrum

The goal of this section is not to provide an example of a Si spectrum reconstructed with the $N \times M$ filter, but rather to provide some intuition for how one might go about performing such a fit. Reconstructing the Si spectrum is challenging even with simple OFs, and the $N \times M$ filter adds additional complexities on top of this. This is non-trivial and is well beyond the scope of this thesis to perform despite some availability of data from the CUTE tower testing effort. I leave this section in this chapter as a concluding remark and to lay out some ideas for future analyzers to follow when trying to reconstruct Si data.

The whole of this chapter focused on reconstructing the energy spectrum of Ge detectors. This is due to a couple of factors. Firstly, the Ge detectors produce multiple calibration points which can be used at various important stages of the $N \times M$ processing pipeline, including weight calculation and GBDT training. Secondly, the calibration points provided by Ge are peaks which are easily modeled by Gaussians. This provides a trivial condition to estimate the weights by simply finding those weights that minimize the width of the peak.

As discussed before, Si detectors do not produce any internal activation lines and therefore their primary mode of calibration is via Compton steps. These are step-like features in the spectrum which are modeled with error functions to account for the broadening effect of the detector resolution. Therefore, rather than fitting an energy the goal here is to optimize the shape such that the step is the sharpest. For Ge, the weights were calculated by effectively modeling the peak as a delta function by setting the true energy to a constant. For Si, there is no simple analytical form that allows an easy weight calculation. Rather, the weights ought to be calculated as those numbers which produce the sharpest step function when combining the amplitudes. If A stands for the $N \times M$ amplitudes and α are the weights, then the error function is defined as:

$$f(A, \alpha, \sigma; \mu, H) = \sum_j H_j \cdot \operatorname{erf} \left(\frac{\sum_i \alpha_{i,j} A_{i,j} - \mu_j}{\sigma_j} \right), \quad (6.13)$$

where μ , σ , H are the mean location, width and height of the steps and the subscript j loops over the number of steps. μ is known since it corresponds to the atomic energy level of the Si atom. H is entirely detector and source dependent and need to be calculated with careful simulations. This function would then have to be numerically fit to obtain α and the width of the step, σ with an

($NM + 1$)-parameter unbinned likelihood fit.

One can also imagine selecting events with energy right around the rising edge of the step (say a 1σ region from the mean) and then finding those weights which bring these energies close to the mean. This makes the rising edge steeper but has the risk of over-fitting if event selection is not performed appropriately. Yet another approach could be partitioning/binning the spectrum around the region of the step and fitting a Gaussian to each bin with a likelihood estimator. Then the final shape of the step will be given by the sum of all the Gaussians. The systematic variation of the weights as a function of energy in the region of the step could then be used to interpolate the weights to other regions of the spectrum. Performing a Si spectrum reconstruction with the $N \times M$ filter could, thus, easily be a thesis in its own right.

Further, attempting to find non-linear position corrections to the energy is a whole other beast. One could imagine selecting $E_{lin-est}$ within a certain range of the mean (perhaps 3σ again) and then construct a BDTGain that brings these events close to the mean. This will severely be limited in terms of the number of events and once again risks overfitting. But the author hopes that future students with an interest in pursuing Si spectrum reconstruction will use some of the ideas proposed in this thesis as inspiration for their work.

6.4 Summary and future directions

In this chapter, we discussed the application of the $N \times M$ filter and GBDT position corrections on a variety of datasets from simulations to real data. A noticeable improvement in energy resolution was demonstrated along with the opportunity to reconstruct position in the scenario where position-labeled data may be available. A brief discussion of potential ways to reconstruct Si spectra with the $N \times M$ was also included in this chapter.

A few limitations of the technique presented in this chapter and the previous one include the requirement for multiple independent calibration peaks, large datasets and of course computational resources. It was found during the data processing effort for CUTE data that the $N \times M$ processing is both memory and time intensive requiring about 4 GB of memory per detector and a couple of hours to process a single data file (usually containing about 1000 events). This is extremely slow for the purpose of processing 24 detectors in parallel and needs further work to improve.

In order to improve the performance on small data sizes, the author proposes exploring transfer learning [144] as a potential alternative. Transfer learning is a technique by which a model trained to perform a closely related task may be adapted to your own model. This is usually used in the context of neural networks where earlier layers of the network learn more generic features and are

adaptable to other models. The earlier layer can be cut and pasted to the model you wish to train, by freezing those layers, i.e., the model only learns the parameters in the other layers using your data. For example, one could imagine a scenario where we train the model on a large number of simulated data, which in theory is possible to obtain, and then adapt parts of this trained model to be integrated into a model to be trained on real data. The primary advantage of this technique is that part of the layers are already trained, and therefore fewer data points are needed to train the subsequent layers.

Chapter 7

Conclusions

In this thesis, I presented two frontiers of the SuperCDMS SNOLAB experiment. The first is the detector testing and characterization effort of the pre-commissioning campaign at CUTE to benchmark detector performance. The second is the active software R&D campaign to develop event reconstruction algorithms for (a) complicated data formats and, (b) deweighting correlated noise and correcting position dependence in the detectors.

The analysis of data from the tower testing campaign is ongoing and therefore some of the preliminary results shown in this thesis will change with time. However, the key points of success remain the same – the functionality of the detectors, our ability to control the detector state, to reconstruct the data and to calibrate the detectors at the lowest energies even in the absence of discernible peaks in the Si detectors. The ability to hold voltages up to and beyond 100 V in these detectors implies that they are capable of achieving amplifications up to and beyond $\times 27$ (34) for Si (Ge) and an equivalent lowering of thresholds in the nuclear recoil energy scale. An exciting piece of information is the potential over-amplification in Si which is not proportional to voltage. Although the origin of the effect is yet to be understood, this would allow the Si detectors to achieve lower thresholds than expected at the applied bias voltage, thus furthering the sensitivity of the experiment.

The baseline resolution is an indication of the strength of our reconstruction techniques in being able to tell signal from noise and is also a fundamental characteristic of our detectors. The Si detectors show baseline resolutions commensurate with their design, however the Ge detectors showed resolution numbers that are further from expectation. We believe that this is due to unexpected noise coupling in from the readout electronics and the surroundings at CUTE. Active work is being done to resolve some of these issues that are expected to affect the main experiment as

well.

Development of novel reconstruction techniques to fit non-uniformly sampled data were also developed in this thesis. These techniques will be critical to SuperCDMS's ability to process advanced data formats and in allowing the DAQ to achieve the thresholds it was designed for.

Due to the larger size of the detectors and differences in the readout electronics design, SuperCDMS SNOLAB detectors will show enhanced position dependence and correlated noise in the data. The $N \times M$ filter reconstruction studied in this thesis will bring significant enhancement in the energy resolution of the detectors. The validation of this novel filter and the machine-learning based analysis pipeline presented in this thesis successfully demonstrated two-fold enhancement in the reconstructed energy resolution of previous data and four-fold enhancement in the new detectors to be deployed at SNOLAB. The new methods developed here thus stand to outperform all other energy reconstruction techniques currently in use by the collaboration. Limitations of the presented analysis pipeline were also discussed along with potential pathways for improvement. This includes the difficulty in interpreting a baseline resolution, and the need for large training data sets and multiple peak-like features of known energies in the spectrum. Alternate techniques to overcome these limitations will be explored by the SuperCDMS collaboration.

In conclusion, this thesis pieces together different aspects of the SuperCDMS SNOLAB experiment and lays down promising directions for future research which will be essential to the experiment's success. The techniques developed in this thesis show exceptional improvement in event reconstruction. Preliminary characterization of the detectors also demonstrate a bright future for SuperCDMS!

Although supersymmetric WIMPs, as predicted by the "WIMP miracle", have not been detected, WIMPs at the GeV and TeV mass scales have not yet been completely ruled out. Vast regions of the parameter space defined by the dark matter mass and cross-section still remain viable, with many experiments directing their searches towards these previously unexplored regions both below and above the GeV scale. There is general consensus within the community that dark matter does not have to be one kind of particle but rather could be composed of many different particles. Searches have continued to focus on all dark matter mass scales rather than a small restrictive search window for this reason. At the highest dark matter masses, searching with colliders for BSM physics at the TeV scale is the most prominent of such efforts. BSM searches probe production channels for dark matter via Higgs interactions or heavy supersymmetric WIMP production from particle collisions. Such searches are critical to studying the physics at high energy scales where new physical phenomena are likely to emerge.

Most direct detection experiments, on the other hand, are beginning to push for low mass sensi-

tivity by lowering their detection thresholds. With the next generation of experiments projected to reach the neutrino fog at the 100s of GeV mass scale, investments in detector development are increasingly directed towards building low threshold devices whether by using lighter target materials or doping to reduce the minimum excitation energy. The low mass parameter space is particularly interesting because candidates such as axions and sterile neutrinos have other implications within the Standard Model. The development of low threshold technology, especially for solid-state detectors, has found parallel applications in quantum computing with devices quickly approaching sub-eV energy resolutions and with it the ability to detect single electrons. Although the feasibility of scaling liquid noble technology quickly starts approaching cost and background limitations for dark matter searches, such experiments will have sensitivity to neutrino physics phenomena such as coherent electron neutrino nucleus scattering. Despite higher thresholds, liquid noble technology has, in the past, demonstrated sensitivity to low mass dark matter candidates using the Migdal effect¹. Experiments can also pursue searches in the small cross-section region beyond the neutrino fog, but this requires directional sensitivity and large exposures to discriminate neutrino events from potential signals.

Thus, both categories of experiments are marching towards a low-mass dark matter direct detection program, but it remains to be seen which one would be the front-runner. At this point of time, it certainly appears that solid-state devices possess an advantage due to low target masses and small bandgaps. But to conclude with an open outlook, whatever experiment ultimately produces a successful detection, there is scope for multiple technologies to co-exist and make complementary measurements to further study the properties of dark matter!

¹A yet to be confirmed phenomenon that postulates that a nuclear recoil can be associated with atomic ionization, producing detectable charge signals from an otherwise undetectable nuclear recoil.

Bibliography

- [1] Y. Liu, S. M. Oser, and A. Pradeep. Multi-channel, multi-template reconstruction for SuperCDMS. *Manuscript in preparation*, 2024. → pages vi, 101, 107, 130
- [2] A. H. Guth. The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. *Phys. Rev. D*, 23:347–356, 1981. [doi:10.1103/PhysRevD.23.347](https://doi.org/10.1103/PhysRevD.23.347). → page 1
- [3] A. D. Linde. Chaotic Inflation. *Phys. Lett. B*, 129:177–181, 1983. [doi:10.1016/0370-2693\(83\)90837-7](https://doi.org/10.1016/0370-2693(83)90837-7). → page 1
- [4] S. Dodelson. *Modern Cosmology*. Academic Press, Elsevier Science, 2003. ISBN 9780122191411. → pages 1, 3
- [5] Planck Collaboration. Planck 2018 results - VI. Cosmological parameters. *A&A*, 641:A6, 2020. [doi:10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910). → pages 2, 10, 12
- [6] C. L. Bennett et. al. Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *APJS*, 208:20, 2013. [doi:10.1088/0067-0049/208/2/20](https://doi.org/10.1088/0067-0049/208/2/20). → page 2
- [7] NASA. What is the Universe made of? URL https://wmap.gsfc.nasa.gov/universe/uni_matter.html. Accessed on 2024-01-17. → page 2
- [8] D. H. Rogstad and G. S. Shostak. Gross Properties of Five Scd Galaxies as Determined from 21-CENTIMETER Observations. *ApJ*, 176:315–321, 1972. [doi:https://doi.org/10.1086/151636](https://doi.org/10.1086/151636). → page 4
- [9] V. C. Rubin et. al. Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa → Sc. *APJL*, 225:L107–L111, 1978. [doi:10.1086/182804](https://doi.org/10.1086/182804). → page 4
- [10] M. Persic, P. Salucci, and F. Stel. The universal rotation curve of spiral galaxies — I. The dark matter connection. *Monthly Notices of the Royal Astronomical Society*, 281(1):27–47, 07 1996. [doi:10.1093/mnras/278.1.27](https://doi.org/10.1093/mnras/278.1.27). → page 5

- [11] S. Golwala. *Exclusion Limits on the WIMP-Nucleon Elastic-Scattering Cross Section from the Cryogenic Dark Matter Search*. PhD thesis, UC Berkeley, 2000. URL <https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/golwala.pdf>. → pages 4, 9
- [12] A. Arbey and F. Mahmoudi. Dark matter and the early Universe: A review. *Progress in Particle and Nuclear Physics*, 119:103865, 2021. doi:<https://doi.org/10.1016/j.pnpnp.2021.103865>. → pages 4, 15
- [13] F. Zwicky. On the Masses of Nebulae and of Clusters of Nebulae. *APJ*, 1937. doi:[10.1086/143864](https://doi.org/10.1086/143864). → page 4
- [14] D. Clowe et. al. A Direct Empirical Proof of the Existence of Dark Matter. *ApJ*, 648:L109, 2006. doi:[10.1086/508162](https://doi.org/10.1086/508162). → page 6
- [15] T. M. Undagoitia and Ludwig Rauch. Dark matter direct-detection experiments. *Journal of Physics G: Nuclear and Particle Physics*, 43:013001, 2016. doi:[10.1088/0954-3899/43/1/013001](https://doi.org/10.1088/0954-3899/43/1/013001). → pages 6, 15, 16, 17
- [16] Planck Collaboration et. al. Planck 2018 results - V. CMB power spectra and likelihoods. *A&A*, 641:A5, 2020. doi:[10.1051/0004-6361/201936386](https://doi.org/10.1051/0004-6361/201936386). → pages 7, 8
- [17] M. J. Geller and J. P. Huchra. Mapping the Universe. *Science*, 246:897–903, 1989. doi:[10.1126/science.246.4932.897](https://doi.org/10.1126/science.246.4932.897). → page 9
- [18] M. Colless et. al. The 2dF Galaxy Redshift Survey: spectra and redshifts. *Mon. Not. R. Astron. Soc.*, 328:1039–1063, 2001. doi:<https://doi.org/10.1046/j.1365-8711.2001.04902.x>. → page 9
- [19] D. G. York et. al. The Sloan Digital Sky Survey: Technical summary. *Astron. J.*, 120: 1579–1587, 2000. doi:[10.1086/301513](https://doi.org/10.1086/301513). → page 9
- [20] V. Springel, C. Frenk, and S. White. The large-scale structure of the Universe. *Nature*, 440: 1137–1144, 2006. doi:[10.1038/nature04805](https://doi.org/10.1038/nature04805). → pages 9, 12
- [21] N. Suzuki et. al. The Hubble Space Telescope Cluster Supernova Survey: V. Improving the Dark Energy Constraints Above $z > 1$ and Building an Early-Type-Hosted Supernova Sample. *ApJ*, 746:85, 2012. doi:[10.1088/0004-637X/746/1/85](https://doi.org/10.1088/0004-637X/746/1/85). → pages 10, 11
- [22] G. Jungman, M. Kamionkowski, and K. Griest. Supersymmetric dark matter. *Physics Reports*, 267(5):195–373, 1996. doi:[https://doi.org/10.1016/0370-1573\(95\)00058-5](https://doi.org/10.1016/0370-1573(95)00058-5). → pages 11, 12
- [23] G. F. R. Ellis. Issues in the philosophy of cosmology. In *Philosophy of Physics*, Handbook of the Philosophy of Science, pages 1183–1285. North-Holland, Amsterdam, 2007. doi:<https://doi.org/10.1016/B978-044451560-5/50014-2>. → page 12

- [24] S. P. Martin. A Supersymmetry Primer. *Perspectives on Supersymmetry*, pages 1–98, 1998. doi:https://doi.org/10.1142/9789812839657_0001. → page 12
- [25] G. Arcadi, M. Dutra, and P. Ghosh et. al. The waning of the WIMP? a review of models, searches, and constraints. *Eur. Phys. J. C*, 78:203, 2018. doi:<https://doi.org/10.1140/epjc/s10052-018-5662-y>. → pages 12, 13
- [26] L. Roszkowski et. al. WIMP dark matter candidates and searches—current status and future prospects. *Rep. Prog. Phys.*, 81:066201, 2018. doi:[10.1088/1361-6633/aab913](https://doi.org/10.1088/1361-6633/aab913). → page 13
- [27] B. Dasgupta and J. Kopp. Sterile neutrinos. *Physics Reports*, 928:1–63, 2021. doi:<https://doi.org/10.1016/j.physrep.2021.06.002>. → page 13
- [28] S. P. Ahlen and F.T. Avignone et. al. Limits on cold dark matter candidates from an ultralow background germanium spectrometer. *Physics Letters B*, 195(4):603–608, 1987. doi:[https://doi.org/10.1016/0370-2693\(87\)91581-4](https://doi.org/10.1016/0370-2693(87)91581-4). → page 13
- [29] E. W. Kolb, D. J. H. Chung, and A. Riotto. WIMPZILLAS! *arXiv*, 1998. doi:<https://doi.org/10.48550/arXiv.hep-ph/9810361>. → page 14
- [30] Dark sectors 2016 workshop: Community report. *arXiv*, 2016. doi:<https://doi.org/10.48550/arXiv.1608.08632>. → page 14
- [31] G. N. Wojcik and T. J. Rizzo. Forbidden scalar dark matter and dark Higgses. *J. High Energ. Phys.*, 33, 2022. doi:[https://doi.org/10.1007/JHEP04\(2022\)033](https://doi.org/10.1007/JHEP04(2022)033). → page 14
- [32] T. Aralis et. al. Constraints on dark photons and axionlike particles from the SuperCDMS Soudan experiment. *Phys. Rev. D*, 101:052008, 2020. doi:[10.1103/PhysRevD.101.052008](https://doi.org/10.1103/PhysRevD.101.052008). → pages 14, 16, 20
- [33] S. M. Vermeulen, P. Relton, and H. Grote et. al. Direct limits for scalar field dark matter from a gravitational-wave detector. *Nature*, 600:424–428, 2021. doi:<https://doi.org/10.1038/s41586-021-04031-y>. → page 14
- [34] D. J. E. Marsh. Axions and ALPs: a very short introduction. *arXiv*, 2017. doi:<https://doi.org/10.48550/arXiv.1712.03018>. → page 14
- [35] C. Robilliard et. al. No “Light Shining through a Wall”: Results from a Photoregeneration Experiment. *Phys. Rev. Lett.*, 99:190403, 2007. doi:[10.1103/PhysRevLett.99.190403](https://doi.org/10.1103/PhysRevLett.99.190403). → page 14
- [36] B. M. Brubaker et. al. First Results from a Microwave Cavity Axion Search at 24 μeV . *Phys. Rev. Lett.*, 118:061302, Feb 2017. doi:[10.1103/PhysRevLett.118.061302](https://doi.org/10.1103/PhysRevLett.118.061302). → page 14
- [37] Z. Ahmed et. al. Search for Axions with the CDMS experiment. *Phys. Rev. Lett.*, 103:141802, 2009. doi:[10.1103/PhysRevLett.103.141802](https://doi.org/10.1103/PhysRevLett.103.141802). → pages 14, 19

- [38] Axion-like particles and high-energy gamma rays: interconversion revisited. *International Cosmic Ray Conference (ICRC2023) - Dark Matter Physics (DM)*, 2023. [doi:https://doi.org/10.22323/1.444.1383](https://doi.org/10.22323/1.444.1383). → page 14
- [39] K. Petraki and R. R. Volkas. Review of asymmetric dark matter. *International Journal of Modern Physics A*, 28:1330028, 2013. [doi:https://doi.org/10.1142/S0217751X13300287](https://doi.org/10.1142/S0217751X13300287). → page 14
- [40] S. Clesse and J. García-Bellido. Seven hints for primordial black hole dark matter. *Physics of the Dark Universe*, 22:137–146, 2018. [doi:https://doi.org/10.1016/j.dark.2018.08.004](https://doi.org/10.1016/j.dark.2018.08.004). → page 14
- [41] C. Alcock et. al. The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations. *ApJ*, 542:281, 2000. [doi:10.1086/309512](https://doi.org/10.1086/309512). → page 14
- [42] M. Aaboud et. al. Search for dark matter in events with a hadronically decaying vector boson and missing transverse momentum in pp collisions at $\sqrt{s} = 13\text{TeV}$ with the ATLAS detector. *Journal of High Energy Physics*, 2018:180, 2018. [doi:https://doi.org/10.1007/JHEP10\(2018\)180](https://doi.org/10.1007/JHEP10(2018)180). → page 15
- [43] G. Aad et. al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. [doi:https://doi.org/10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). → page 15
- [44] S. Chatrchyan et. al. Observation of a new boson at a mass of 125 gev with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. [doi:https://doi.org/10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). → page 15
- [45] G. Arcadi, A. Djouadi, and M. Raidal. Dark Matter through the Higgs portal. *Physics Reports*, 842:1–180, 2020. [doi:https://doi.org/10.1016/j.physrep.2019.11.003](https://doi.org/10.1016/j.physrep.2019.11.003). → page 15
- [46] R. Abbasi et. al. Search for neutrino lines from dark matter annihilation and decay with IceCube. *Phys. Rev. D*, 108:102004, 2023. [doi:10.1103/PhysRevD.108.102004](https://doi.org/10.1103/PhysRevD.108.102004). → page 15
- [47] K. Abe et. al. Search for Cosmic-Ray Boosted Sub-GeV Dark Matter Using Recoil Protons at Super-Kamiokande. *Phys. Rev. Lett.*, 130:031802, 2023. [doi:10.1103/PhysRevLett.130.031802](https://doi.org/10.1103/PhysRevLett.130.031802). → page 15
- [48] M. Ackermann et. al. Search for gamma-ray spectral lines with the Fermi Large Area Telescope and dark matter implications. *Phys. Rev. D*, 88:082002, 2013. [doi:10.1103/PhysRevD.88.082002](https://doi.org/10.1103/PhysRevD.88.082002). → page 15
- [49] T. Daylan and D. P. Finkbeiner et. al. The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter. *Physics of the Dark Universe*, 12:1–23, 2016. [doi:https://doi.org/10.1016/j.dark.2015.12.005](https://doi.org/10.1016/j.dark.2015.12.005). → page 15

- [50] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue. Evidence for Unresolved γ -Ray Point Sources in the Inner Galaxy. *Phys. Rev. Lett.*, 116:051103, 2016. [doi:10.1103/PhysRevLett.116.051103](https://doi.org/10.1103/PhysRevLett.116.051103). → page 15
- [51] J. D. Lewin and P. F. Smith. Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil. *Astroparticle Physics*, 6(1): 87–112, 1996. [doi:https://doi.org/10.1016/S0927-6505\(96\)00047-3](https://doi.org/10.1016/S0927-6505(96)00047-3). → page 16
- [52] E. Aprile et. al. First Dark Matter Search with Nuclear Recoils from the XENONnT Experiment. *Phys. Rev. Lett.*, 131:041003, 2023. [doi:10.1103/PhysRevLett.131.041003](https://doi.org/10.1103/PhysRevLett.131.041003). → page 16
- [53] P. Agnes et. al. Low-Mass Dark Matter Search with the DarkSide-50 Experiment. *Phys. Rev. Lett.*, 121:081307, 2018. [doi:10.1103/PhysRevLett.121.081307](https://doi.org/10.1103/PhysRevLett.121.081307). → page 16
- [54] G. Angloher et. al. Results on sub-GeV dark matter from a 10 eV threshold CRESST-III silicon detector. *Phys. Rev. D*, 107:122003, 2023. [doi:10.1103/PhysRevD.107.122003](https://doi.org/10.1103/PhysRevD.107.122003). → page 16
- [55] C. Amole et. al. Dark matter search results from the complete exposure of the PICO-60 C₃F₈ bubble chamber. *Phys. Rev. D*, 100:022001, 2019. [doi:10.1103/PhysRevD.100.022001](https://doi.org/10.1103/PhysRevD.100.022001). → page 16
- [56] B. von Krosigk et. al. DELight: a Direct search Experiment for Light dark matter with superfluid helium. *arXiv*, 2022. [doi:https://doi.org/10.48550/arXiv.2209.10950](https://doi.org/10.48550/arXiv.2209.10950). → page 16
- [57] J. Liu, X. Chen, and X. Ji. Current status of direct dark matter detection experiments. *Nature Phys*, 13:212–216, 2017. [doi:https://doi.org/10.1038/nphys4039](https://doi.org/10.1038/nphys4039). → page 16
- [58] R. L. Workman et. al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. [doi:10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097). → page 17
- [59] N. Wang et. al. A cryogenic phonon detector with simultaneous measurement of phonons and ionization. *Physica B: Condensed Matter*, 165-166:3–4, 1990. [doi:https://doi.org/10.1016/S0921-4526\(90\)80851-9](https://doi.org/10.1016/S0921-4526(90)80851-9). Proceedings of the 19th International Conference on Low Temperature Physics. → page 19
- [60] B. A. Young et. al. A pilot dark matter particle search experiment. *AIP Conference Proceedings*, 272:1260–1265, 1992. [doi:https://doi.org/10.1063/1.43548](https://doi.org/10.1063/1.43548). → page 19
- [61] P. D. Barnes et. al. The first cryogenic dark matter experiment. *J Low Temp Phys*, 93: 791–796, 1993. [doi:https://doi.org/10.1007/BF00693514](https://doi.org/10.1007/BF00693514). → page 19
- [62] D. S. Akerib et. al. Installation and commissioning of the CDMSII experiment at Soudan. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators*,

- Spectrometers, Detectors and Associated Equipment*, 520(1):116–119, 2004.
[doi:https://doi.org/10.1016/j.nima.2003.11.256](https://doi.org/10.1016/j.nima.2003.11.256). → page 19
- [63] Z. Ahmed et. al. Search for Weakly Interacting Massive Particles with the First Five-Tower Data from the Cryogenic Dark Matter Search at the Soudan Underground Laboratory. *Phys. Rev. Lett.*, 102:011301, 2009. [doi:10.1103/PhysRevLett.102.011301](https://doi.org/10.1103/PhysRevLett.102.011301). → page 19
- [64] D. S. Akerib et. al. Low-threshold analysis of CDMS shallow-site data. *Phys. Rev. D*, 82:122004, 2010. [doi:10.1103/PhysRevD.82.122004](https://doi.org/10.1103/PhysRevD.82.122004). → page 19
- [65] P. L. Brink et. al. First test runs of a dark-matter detector with interleaved ionization electrodes and phonon sensors for surface-event rejection. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(2):414–416, 2006.
[doi:https://doi.org/10.1016/j.nima.2005.12.026](https://doi.org/10.1016/j.nima.2005.12.026). → page 20
- [66] R. Agnese et. al. Demonstration of surface electron rejection with interleaved germanium detectors for dark matter searches. *Applied Physics Letters*, 103:164105, 2013.
[doi:http://dx.doi.org/10.1063/1.4826093](http://dx.doi.org/10.1063/1.4826093). → page 20
- [67] R. Agnese et. al. Search for Low-Mass Weakly Interacting Massive Particles with SuperCDMS. *Phys. Rev. Lett.*, 112:241302, 2014. [doi:10.1103/PhysRevLett.112.241302](https://doi.org/10.1103/PhysRevLett.112.241302).
→ page 20
- [68] B. Neganov and O. Trofimov. Otkrytye. Izobreteniya. page 215, 1985. → pages 20, 22
- [69] P. Luke. Voltage-assisted calorimetric ionization detector. *J. Appl. Phys.*, 64:6858, 1988.
→ pages 20, 22
- [70] R. Agnese et. al. Low-mass dark matter search with CDMSlite. *Phys. Rev. D*, 97:022002, 2018. [doi:10.1103/PhysRevD.97.022002](https://doi.org/10.1103/PhysRevD.97.022002). → pages 20, 67, 81, 168, 169, 180
- [71] R. Agnese et. al. Search for low-mass dark matter with CDMSlite using a profile likelihood fit. *Phys. Rev. D*, 99:062001, 2019. [doi:10.1103/PhysRevD.99.062001](https://doi.org/10.1103/PhysRevD.99.062001). → pages 20, 37, 67, 81, 148, 168, 170, 172, 180, 182, 184
- [72] D. A. Bauer. Dark matter detection with cryogenic detectors. *Journal of Physics: Conference Series*, 120(4):042002, 2008. [doi:10.1088/1742-6596/120/4/042002](https://doi.org/10.1088/1742-6596/120/4/042002). → page 20
- [73] A. Münster, S. Schönert, and M. Willers. Cryogenic detectors for dark matter search and neutrinoless double beta decay. proceedings of the vienna conference on instrumentation 2016. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 845:387–393, 2017.
[doi:https://doi.org/10.1016/j.nima.2016.06.008](https://doi.org/10.1016/j.nima.2016.06.008). → page 20

- [74] M. Wilson. *A new search for low-mass dark matter and an examination and reduction of the uncertainty due to the photo-electric absorption cross section using a cryogenic Silicon detector with single-charge sensitivity*. PhD thesis, University of Toronto, 2022. URL <https://tspace.library.utoronto.ca/handle/1807/123283>. → page 20
- [75] E. Fascione. *Searching for Electron-Interacting Dark Matter with the Super Cryogenic Dark Matter Search*. PhD thesis, Queen’s University, 2023. URL <https://hdl.handle.net/1974/31999>. → pages 23, 24
- [76] W. Page. *Searching for low-mass dark matter with SuperCDMS Soudan detectors*. PhD thesis, University Of British Columbia, 2019. URL <https://dx.doi.org/10.14288/1.0376351>. → pages 20, 34, 172, 180, 182, 184
- [77] C. Kittel. *Introduction to Solid State Physics*, page 190. John Wiley & Sons, 2005. → page 21
- [78] K. Ramanathan and N. Kurinsky. Ionization yield in silicon for eV-scale electron-recoil processes. *Phys. Rev. D*, 102:063026, 2020. doi:[10.1103/PhysRevD.102.063026](https://doi.org/10.1103/PhysRevD.102.063026). → page 21
- [79] R. C. Alig and S. Bloom. Electron-Hole-Pair Creation Energies in Semiconductors. *Phys. Rev. Lett.*, 35:1522–1525, Dec 1975. doi:[10.1103/PhysRevLett.35.1522](https://doi.org/10.1103/PhysRevLett.35.1522). → page 21
- [80] B. A. Young. Phonon - mediated particle detection : An overview. *Nuclear Physics B - Proceedings Supplements*, 32:127–137, 1993. doi:[https://doi.org/10.1016/0920-5632\(93\)90017-Z](https://doi.org/10.1016/0920-5632(93)90017-Z). → page 21
- [81] B. Sadoulet et. al. Phonon-mediated Detection of Particles. *Report of the Round Table Discussion*, 1987. URL <https://www.osti.gov/servlets/purl/7070325>. → page 21
- [82] P. A. Fernandez Garrillo. *Development of highly resolved and photo-modulated Kelvin probe microscopy techniques for the study of photovoltaic systems*. PhD thesis, Universite Grenoble Alpes, 2018. URL <https://theses.hal.science/tel-01970171/>. → page 22
- [83] M. C. Pyle. *Optimizing the design and analysis of cryogenic semiconductor dark matter detectors for maximum sensitivity*. PhD thesis, Stanford University, 2012. URL <https://doi.org/10.2172/1127926>. → pages 22, 23, 45, 46, 55
- [84] W. Li, J. C. Fenton, Y. Wang, D. W. McComb, and P. A. Warburton. Tunability of the superconductivity of tungsten films grown by focused-ion-beam direct writing. *J. Appl. Phys*, 104:093913, 2008. doi:<https://doi.org/10.1063/1.3013444>. → page 23
- [85] J. Lindhard et. al. Integral equations governing radiation effects. (Notes on atomic collisions, III). *Kgl. Danske Videnskab., Selskab. Mat. Fys. Medd.*, 33, 1963. URL <https://www.osti.gov/biblio/4701226>. → page 25

- [86] J. Lindhard and M. Scharff. Energy Dissipation by Ions in the keV Region. *Phys. Rev.*, 124:128–130, 1961. [doi:10.1103/PhysRev.124.128](https://doi.org/10.1103/PhysRev.124.128). → page 25
- [87] J. I. Collar, A. R. L. Kavner, and C. M. Lewis. Germanium response to sub-keV nuclear recoils: A multipronged experimental characterization. *Phys. Rev. D*, 103:122003, 2021. [doi:10.1103/PhysRevD.103.122003](https://doi.org/10.1103/PhysRevD.103.122003). → page 25
- [88] M. F. Albakry et. al. Ionization yield measurement in a germanium CDMSlite detector using photo-neutron sources. *Phys. Rev. D*, 105:122002, 2022. [doi:10.1103/PhysRevD.105.122002](https://doi.org/10.1103/PhysRevD.105.122002).
- [89] M. F. Albakry et. al. First Measurement of the Nuclear-Recoil Ionization Yield in Silicon at 100 eV. *Phys. Rev. Lett.*, 131:091801, 2023. [doi:10.1103/PhysRevLett.131.091801](https://doi.org/10.1103/PhysRevLett.131.091801). → page 25
- [90] The SuperCDMS Collaboration. SuperCDMS SNOLAB Technical Design Report, 2017. → pages 26, 27, 47, 61, 67, 68, 93
- [91] J.S. Wilson et. al. The level-1 trigger for the SuperCDMS experiment at SNOLAB. *Journal of Instrumentation*, 17:P07010, 2022. [doi:10.1088/1748-0221/17/07/P07010](https://doi.org/10.1088/1748-0221/17/07/P07010). → pages 27, 28
- [92] H. Meyer zu Theenhausen, B. von Krosigk, and J.S. Wilson. Neural-network-based level-1 trigger upgrade for the SuperCDMS experiment at SNOLAB. *Journal of Instrumentation*, 18:P06012, 2023. [doi:10.1088/1748-0221/18/06/P06012](https://doi.org/10.1088/1748-0221/18/06/P06012). → page 28
- [93] B. Cabrera et. al. Prompt phonon signals from particle interactions in Si crystals. *J. Low Temp. Phys.*, 93:365–375, 1993. [doi:https://doi.org/10.1007/BF00693447](https://doi.org/10.1007/BF00693447). → page 33
- [94] R. Underwood. Rejecting outer radius backgrounds in SuperCDMS high voltage dark matter searches, 2017. → page 33
- [95] J. A. Bearden and A. F. Burr. Re-evaluation of X-Ray Atomic Energy Levels. *Rev. Mod. Phys.*, 39:125–142, 1967. [doi:10.1103/RevModPhys.39.125](https://doi.org/10.1103/RevModPhys.39.125). → page 35
- [96] D. Barker. Low Energy Background Spectrum in CDMSlite. *38th International Conference on High Energy Physics (ICHEP 2016)*, 2016. [doi:https://doi.org/10.48550/arXiv.1611.05792](https://doi.org/10.48550/arXiv.1611.05792). → pages 36, 64
- [97] D. Norcini et. al. Precision measurement of Compton scattering in silicon with a skipper CCD for dark matter detection. *Phys. Rev. D*, 106:092001, 2022. [doi:10.1103/PhysRevD.106.092001](https://doi.org/10.1103/PhysRevD.106.092001). → pages 36, 64
- [98] S. Yellin. Finding an upper limit in the presence of an unknown background. *Phys. Rev. D*, 66:032005, 2002. [doi:10.1103/PhysRevD.66.032005](https://doi.org/10.1103/PhysRevD.66.032005). → page 37

BIBLIOGRAPHY

- [99] SuperCDMS collaboration. A Strategy for Low-Mass Dark Matter Searches with Cryogenic Detectors in the SuperCDMS SNOLAB Facility. *arXiv*, 2022. doi:<https://doi.org/10.48550/arXiv.2203.08463>. → pages 37, 38
- [100] H. Zu, W. Dai, and A.T.A.M. de Waele. Development of dilution refrigerators—a review. *Cryogenics*, 121:103390, 2022. doi:<https://doi.org/10.1016/j.cryogenics.2021.103390>. → page 40
- [101] R. Radebaugh. Development of the pulse tube refrigerator as an efficient and reliable cryocooler. *Proc. Institute of Refrigeration*, page 00, 1999-2000. URL <https://corpora.tika.apache.org/base/docs/govdocs1/535/535594.pdf>. → page 40
- [102] W Rau, G Gerbier, P. Camus, K. Dering, A. Cazes, S. Scorza, X. Zhang, and A. Dastgheibi-Fard. CUTE - A Cryogenic Underground Test Facility at SNOLAB. *Journal of Physics: Conference Series*, 1342:012128, 2020. doi:[10.1088/1742-6596/1342/1/012128](https://doi.org/10.1088/1742-6596/1342/1/012128). → page 40
- [103] P. Camus and J. Corbett et. al. CUTE: A Cryogenic Underground TEst facility at SNOLAB. *Front. Phys.*, 2024. doi:[10.3389/fphy.2023.1319879](https://doi.org/10.3389/fphy.2023.1319879). → pages 40, 41
- [104] R. Germond. *Techniques and challenges in low-mass dark matter searches using CDMS style detectors*. PhD thesis, Queen’s University, ON, Canada, 2023. URL <https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/germond.pdf>. → page 40
- [105] F. T. Avignone et. al. Theoretical and experimental investigation of cosmogenic radioisotope production in germanium. *Nucl. Phys. Proc. Suppl.*, 28A:280–285, 1992. doi:[https://doi.org/10.1016/0920-5632\(92\)90184-T](https://doi.org/10.1016/0920-5632(92)90184-T). → page 47
- [106] D. Lal et. al. Cosmic-ray-produced ^{32}Si in nature. *Science*, 131:332–337, 1960. doi:[10.1126/science.131.3397.332](https://doi.org/10.1126/science.131.3397.332). → page 47
- [107] W. Hampel and L. P. Remsberg. Half-life of ^{71}Ge . *Phys. Rev. C*, 31:666–667, 1985. doi:[10.1103/PhysRevC.31.666](https://doi.org/10.1103/PhysRevC.31.666). → page 62
- [108] J. Hall. Electron Capture Probabilities for Selected Isotopes Often Found in Germanium, 2013. URL https://scdms.slac.stanford.edu/cdms_restricted/Soudan/R133/ebook/130422/ElectronCaptureProb.html. → page 62
- [109] K. Ramanathan, A. Kavner, A. E. Chavarria, P. Privitera, D. Amidei, T.-L. Chou, A. Matalon, R. Thomas, J. Estrada, J. Tiffenberg, and J. Molina. Measurement of low energy ionization signals from Compton scattering in a charge-coupled device dark matter detector. *Phys. Rev. D*, 96:042002, 2017. doi:[10.1103/PhysRevD.96.042002](https://doi.org/10.1103/PhysRevD.96.042002). → pages 64, 65, 66

BIBLIOGRAPHY

- [110] J. E. Parks. Compton Scattering with Scintillation Detector (accessed on May 27th 2024), 2024. URL https://wanda.fiu.edu/boeglinw/courses/Modern_lab_manual3/scint_compton.html. → page 65
- [111] S. Das. Investigating Compton steps in SuperCDMS Si HVeV detectors. *XVIII International Conference on Topics in Astroparticle and Underground Physics (TAUP2023) - Dark Matter and its Detection*, 441, 2023. doi:<https://doi.org/10.22323/1.441.0077>. → page 64
- [112] SuperCDMS Collaboration. Summary of CDMSlite Run 3 First Results, 2018. URL https://scdms.slac.stanford.edu/cdms_restricted/dokuwiki/doku.php?id=analysis:r133:cdmslite:run3_summary. → pages 66, 172
- [113] V. P. Chechev and N. K. Kuzmenko. Table of radionuclides (accessed on may 27th 2024). URL http://www.lnhb.fr/nuclides/Ba-133_tables.pdf. → page 66
- [114] A. J. Anderson. Phonon Calibration Update for Reprocessing, 2013. URL https://scdms.slac.stanford.edu/cdms_restricted/Soudan/R133/ebook/130130/index.html. → page 75
- [115] H. J. Landau. Sampling, data transmission, and the Nyquist rate. *Proceedings of the IEEE*, 55(10):1701–1706, 1967. doi:[10.1109/PROC.1967.5962](https://doi.org/10.1109/PROC.1967.5962). → pages 93, 99
- [116] A. Dutt and V. Rokhlin. Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.*, 14:1368–1393, 1993. doi:<https://doi.org/10.1137/0914081>. → page 98
- [117] D. Scargle. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Ap. J.*, 263:835–853, 1982. doi:[10.1086/160554](https://doi.org/10.1086/160554). → page 98
- [118] W. Page. Data Acquisition for SuperCDMS SNOLAB, 2015. URL <http://hdl.handle.net/2429/54536>. → pages 98, 100
- [119] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C++*. Cambridge University Press, 1992. → page 101
- [120] R. Brun and F. Rademakers. An Object Oriented Data Analysis Framework. In *Proceedings AIHENP'96 Workshop, Lausanne*, 1997. doi:[10.5281/zenodo.848818](https://doi.org/10.5281/zenodo.848818). → page 126
- [121] F. D. Neeser and Massey J. L. Proper Complex Random Processes with Applications to Information Theory . *IEEE Transactions on Information Theory*, 39:1293–1302, 1993. doi:[10.1109/18.243446](https://doi.org/10.1109/18.243446). → page 133
- [122] W. Smith and S. Erdman. A note on the inversion of complex matrices. *IEEE Transactions on Automatic Control*, 19:64–64, 1974. doi:[10.1109/TAC.1974.1100466](https://doi.org/10.1109/TAC.1974.1100466). → page 134

- [123] Z. Dai et. al. Complex matrix inversion via real matrix inversions. *arXiv:2208.01239*, 2022. doi:<https://doi.org/10.48550/arXiv.2208.01239>. → page 134
- [124] Y. Liu and S. M. Oser. Analytic solution of the $1 \times M$ optimal filter. 2023. URL https://confluence.slac.stanford.edu/display/CDMS/SNOLAB+Reconstruction?preview=%2F227170247%2F373358279%2Fanalytical1xMOF_v2.0.pdf. → page 138
- [125] K. Pearson. LIII. On lines and planes of closest fit to systems of points in space . *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2: 559–572, 1901. doi:[10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720). → pages 139, 140
- [126] M. Greenacre, P.J.F. Groenen, and T. Hastie et. al. Principal component analysis. *Nat Rev Methods Primers*, 2:100, 2022. doi:<https://doi.org/10.1038/s43586-022-00184-w>.
- [127] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci.*, 13:2065, 2016. doi:[10.1098/rsta.2015.0202](https://doi.org/10.1098/rsta.2015.0202). → pages 139, 140
- [128] G. W. Stewart. On the Early History of the Singular Value Decomposition. *SIAM Review*, 35:551–566, 1993. doi:[10.1137/1035134](https://doi.org/10.1137/1035134). → page 141
- [129] R. Wilcox. Chapter 6 - Some Multivariate Methods. In Rand Wilcox, editor, *Introduction to Robust Estimation and Hypothesis Testing (Third Edition)*, Statistical Modeling and Decision Science, pages 215–289. Academic Press, third edition edition, 2012. ISBN 978-0-12-386983-8. doi:<https://doi.org/10.1016/B978-0-12-386983-8.00006-8>. → page 141
- [130] I. H. Sarker. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.*, 2:160, 2021. doi:<https://doi.org/10.1007/s42979-021-00592-x>. → page 142
- [131] R. Caruana and A. Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, page 161–168. Association for Computing Machinery, 2006. doi:[10.1145/1143844.1143865](https://doi.org/10.1145/1143844.1143865). URL <https://doi.org/10.1145/1143844.1143865>. → page 142
- [132] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf. → pages 143, 145
- [133] S. B. Kotsiantis. Decision trees: a recent overview. *Artif Intell Rev*, 39:261–263, 2013. doi:<https://doi.org/10.1007/s10462-011-9272-4>. → page 143

- [134] A. Prasad. Regression Trees — Decision Tree for Regression — Machine Learning. URL <https://medium.com/analytics-vidhya/regression-trees-decision-tree-for-regression-machine-learning-e4d7525d8047>. → page 145
- [135] T. Chen and C. E. Guestrin. XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016. doi:<https://doi.org/10.1145/2939672.2939785>. → page 145
- [136] LightGBM documentation V4.3.0, 2024. URL <https://lightgbm.readthedocs.io/en/v4.3.0/>. → page 145
- [137] H. Abdi. The Eigen-Decomposition : Eigenvalues and Eigenvectors. 2006. URL <https://api.semanticscholar.org/CorpusID:14829978>. → page 151
- [138] S. Agostinelli et. al. GEANT4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506:250–303, 2003. doi:[https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). → page 156
- [139] F. Pedregosa and G. Varoquaux et. al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. doi:<https://doi.org/10.48550/arXiv.1201.0490>. → page 163
- [140] A. Thompson et. al. X-Ray Data Booklet, 2009. URL <http://xdb.lbl.gov/>. → page 180
- [141] E. Schönfeld. Calculation of fractional electron capture probabilities. *Applied Radiation and Isotopes*, 49:1353–1357, 1998. doi:[https://doi.org/10.1016/S0969-8043\(97\)10073-2](https://doi.org/10.1016/S0969-8043(97)10073-2). → page 180
- [142] K. S. Krane. *Introductory Nuclear Physics*. John Wiley & Sons Inc, 2 edition, 1988. ISBN 978-0471805533. → page 180
- [143] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9:60–62, 1938. doi:[10.1214/aoms/1177732360](https://doi.org/10.1214/aoms/1177732360). → page 184
- [144] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109:43–76, 2021. doi:[10.1109/JPROC.2020.3004555](https://doi.org/10.1109/JPROC.2020.3004555). → page 194

Appendix A

Deriving the N×M Filter amplitudes

Here, I analytically derive the N×M amplitudes by minimizing the χ^2 in Eq. (5.10), copied below for convenience:

$$\chi^2 = \sum_f \left(\vec{S}_f - T_f \cdot \vec{A} \right)^T \cdot \mathcal{V}_f^{-1} \cdot \left(\vec{S}_f - T_f \cdot \vec{A} \right).$$

The above equation can be minimized with respect to the amplitude vector, $\vec{A}$ as follows:

$$\begin{aligned} \frac{d\chi^2}{d\vec{A}} &= \sum_f \frac{d \left[\left(\vec{S}_f - T_f \cdot \vec{A} \right)^T \cdot \mathcal{V}_f^{-1} \cdot \left(\vec{S}_f - T_f \cdot \vec{A} \right) \right]}{d\vec{A}} \\ &= - \sum_f \left[T_f^T \mathcal{V}_f^{-1} \left(\vec{S}_f - T_f \cdot \vec{A} \right) + \left(\vec{S}_f - T_f \cdot \vec{A} \right)^T \mathcal{V}_f^{-1} T_f \right] \\ &= - \sum_f \left[T_f^T \mathcal{V}_f^{-1} \vec{S}_f - T_f^T \mathcal{V}_f^{-1} T_f \cdot \vec{A} + \vec{S}_f^T \mathcal{V}_f^{-1} T_f - \vec{A}^T T_f^T \mathcal{V}_f^{-1} T_f \right]. \end{aligned} \tag{A.1}$$

Using the fact that T and $\mathcal{V}$ are symmetric matrices allows me to simplify Eq. (A.1) to

$$\frac{d\chi^2}{d\vec{A}} = -2 \cdot \sum_f \left[T_f^T \mathcal{V}_f^{-1} \vec{S}_f - T_f^T \mathcal{V}_f^{-1} T_f \cdot \vec{A} \right]. \tag{A.2}$$

The minimum χ^2 is obtained by equating the first derivative to zero. This gives:

$$\begin{aligned}
& -2 \cdot \sum_f \left(T_f^T \mathcal{V}_f^{-1} \vec{S}_f \right) + 2 \cdot \sum_f \left(T_f^T \mathcal{V}_f^{-1} T_f \cdot \vec{A} \right) = 0 \\
\Rightarrow \sum_f \left(T_f^T \mathcal{V}_f^{-1} T_f \cdot \vec{A} \right) &= \sum_f \left(T_f^T \mathcal{V}_f^{-1} \vec{S}_f \right) \\
\Rightarrow \vec{A} &= \sum_f \left(T_f^T \mathcal{V}_f^{-1} T_f \right)^{-1} \sum_f \left(T_f^T \mathcal{V}_f^{-1} \vec{S}_f \right).
\end{aligned} \tag{A.3}$$