

Low-Frequency Noise Identification for the Dark Matter Experiment SuperCDMS using Machine Learning

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und
Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von
Sukeerthi Dharani

Hamburg
2024

Gutachter/innen der Dissertation:	Prof. Dr. Gregor Kasieczka Prof. Dr. Belina von Krosigk
Zusammensetzung der Prüfungskommission:	Prof. Dr. Gregor Kasieczka Prof. Dr. Belina von Krosigk Prof. Dr. Dieter Horns Prof. Dr. Gudrid Moortgat-Pick Prof. Dr. Manuel Meyer
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Dieter Horns
Datum der Disputation:	24.02.2025
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang J. Parak
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Markus Drescher
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

The SuperCDMS collaboration has developed detectors to probe potential interactions between Dark Matter (DM) particles and semiconductor crystals at cryogenic temperatures. During the final phase of the Cryogenic Dark Matter Search low ionization threshold experiment (CDMSlite), vibrations originating from the cooling system were observed. These vibrations generated broadband low-frequency noise (LFN), which frequently triggered the detector. These LFN events were challenging to remove due to their variable pulse shapes, which often resembled low-energy signals produced by particle interactions. In the final analysis, a strict event selection criterion was applied to the dataset to remove the LFN events. However, this increased the energy threshold and lowered the signal efficiency, reducing the experiment's sensitivity to low-mass DM interactions.

In this work, a new LFN selection criterion using machine learning is presented. This approach aims to improve the signal efficiency while maintaining the signal purity. Two neural network architectures are considered: a Convolutional Neural Network (CNN) and a CNN combined with Long Short-Term Memory (LSTM). The CDMSlite detector has four phonon signal channels; therefore, LFN identification is probed using individual channels, the sum of the phonon channels, and multi-channel configurations. The LFN-background samples for training are data-driven. A novel method for signal sample generation using pulse shapes and regression is developed. Both this method and standard phonon signal templates are used for generating signal samples. Among the tested channel configurations and neural network architectures, the multi-channel configuration with a CNN architecture performs the best in terms of signal efficiency and purity. This network is used on the CDMSlite Run 3 data to remove LFN events.

The final spectrum reveals additional events compared to the spectrum obtained with the previous LFN selection criterion. Exclusion limits on the spin-independent DM-nucleon cross-section are calculated for DM masses between $1 \text{ GeV}/c^2$ and $10 \text{ GeV}/c^2$. The new results are consistent with the previous analysis, with an improvement in sensitivity up to 17.5% in the DM mass range of $2.4 \text{ GeV}/c^2$ to $5 \text{ GeV}/c^2$. The gain in signal efficiency due to the ML-based LFN cut is limited by the application of a fiducial volume selection criterion. Therefore, case studies are conducted excluding the fiducial volume selection criterion to investigate its impact, which demonstrates the potential of the ML-based LFN cut in the low-mass DM parameter space. The sensitivities are extended down to DM masses of $0.8 \text{ GeV}/c^2$, with improvements in sensitivity of up to three orders of magnitude below $1.2 \text{ GeV}/c^2$.

Zusammenfassung

Die SuperCDMS-Kollaboration hat Detektoren entwickelt, um mögliche Wechselwirkungen zwischen Teilchen der Dunklen Materie (DM) und Halbleiterkristallen bei kryogenen Temperaturen zu untersuchen. Während der Endphase des Experiments *Cryogenic Dark Matter Search low ionization threshold* (CDMSlite) wurden Vibrationen beobachtet, die vom Kühlsystem ausgingen. Diese Vibrationen erzeugten breitbandiges niederfrequentes Rauschen (LFN¹), das den Detektor häufig auslöste. Diese LFN-Ereignisse waren aufgrund ihrer variablen Pulsformen, die oft niederenergetischen Signalen aus Teilchenwechselwirkungen ähnelten, schwer zu entfernen. In der abschließenden Analyse wurde ein strenges Kriterium für die Ereignisauswahl auf den Datensatz angewendet, um die LFN-Ereignisse zu entfernen. Dadurch wurde jedoch die Energieschwelle erhöht und die Signaleffizienz bei niedrigen Energien gesenkt, was die Empfindlichkeit des Experiments für Wechselwirkungen leichter DM-Teilchen verringerte.

In dieser Arbeit wurde ein neues LFN-Auswahlkriterium mithilfe von maschinellem Lernen entwickelt. Dieser Ansatz zielt darauf ab, die Signaleffizienz zu verbessern und gleichzeitig die Signalreinheit zu erhalten. Es werden zwei neuronale Netzwerkarchitekturen betrachtet: ein *Convolutional Neural Network* (CNN) und ein CNN kombiniert mit *Long Short-Term Memory* (LSTM). Der CDMSlite-Detektor verfügt über vier Phonon-Signalkanäle; daher wird die LFN-Identifizierung anhand einzelner Kanäle, der Phonon-Kanalsumme und von Mehrkanal-Konfigurationen untersucht. Die LFN-Hintergrundproben für das Training basieren auf Daten. Es wird eine neuartige Methode zur Erzeugung von Signalproben unter Verwendung von Pulsformen und Regression entwickelt. Sowohl diese Methode als auch Standard-Phonon-Signalvorlagen, sogenannte Templates, werden zur Erzeugung von Signalproben verwendet. Unter den getesteten Kanalkonfigurationen und neuronalen Netzwerkarchitekturen schneidet die Mehrkanalkonfiguration mit einer CNN-Architektur in Bezug auf Signaleffizienz und -reinheit am besten ab. Dieses Netzwerk wird auf die Daten von CDMSlite Run 3 angewendet, um LFN-Ereignisse zu entfernen.

Das resultierende Spektrum zeigt zusätzliche Ereignisse im Vergleich zu dem Spektrum, das mit dem vorherigen LFN-Auswahlkriterium erhalten wurde. Ausschlussgrenzen werden für den spinunabhängigen DM-Nukleonen-Wirkungsquerschnitt für DM-Massen zwischen $1 \text{ GeV}/c^2$ und $10 \text{ GeV}/c^2$ bestimmt. Die neuen Ergebnisse stimmen mit der vorherigen Analyse überein, wobei die Empfindlichkeit im DM-Massenbereich von $2,4 \text{ GeV}/c^2$ bis $5 \text{ GeV}/c^2$ um bis zu 17,5% verbessert wurde. Die Verbesserung der Signaleffizienz, die sich aus der Verwendung des ML-basierten LFN-Auswahlkriteriums ergibt, wird durch die Anwendung eines Radius-Auswahlkriteriums begrenzt. Daher werden Fallstudien ohne das Radius-Auswahlkriteriums durchgeführt, um dessen Auswirkungen zu untersuchen, was das Potenzial des ML-basierten LFN-Schnitts im Parameterraum der leichten DM-Teilchen demonstriert. Die Empfindlichkeiten werden bis zu DM-Massen von $0,8 \text{ GeV}/c^2$ ausgedehnt, mit Verbesserungen der Empfindlichkeit um bis zu drei Größenordnungen unter $1.2 \text{ GeV}/c^2$.

¹Aus dem Englischen: Low Frequency Noise

Personal Contribution

The experimental data used in this analysis is from the CDMSlite experiment, a part of the SuperCDMS collaboration. The detector, experimental design, and operation were carried out by numerous members of the collaboration. The experimental run used for this analysis ended in 2015. This project builds on the analysis framework developed for the CDMSlite dataset by the SuperCDMS collaboration.

- Chapter 2 and Chapter 3: These chapters are collaborative efforts.
- Chapter 4: The machine learning (ML) models were developed entirely by me. Signal sample generation for training the network in single- and PT-channel analysis involved adding phonon templates to baseline noise. The phonon templates were developed by the collaboration, while the baseline noise was taken from the experimental data. The simulation using pulse shapes and a regressor was my original contribution, and I independently developed and implemented it for multi-channel analysis. I used TensorFlow and Scikit-learn to develop LFN models for low-frequency noise identification and the regressor for simulation, respectively. The development of various channel configurations and the corresponding neural networks were carried out entirely by me.
- Chapter 5: Analysis cuts for the final spectrum and signal efficiencies were developed by the collaboration. I used the SuperCDMS tools, to produce samples for signal efficiency calculations and limit-setting. The sensitivity studies I performed to select the final channel configuration were based on background modeling from the previous analysis developed by the collaboration.
- Chapter 6: Similar to Chapter 5, the analysis for case studies was developed by the collaboration. However, the extension of this ML-based analysis with case studies was carried out by me.

Acknowledgments

My first and foremost gratitude goes to my supervisor, Belina von Krosigk. Thank you for offering me this interesting topic and guiding me throughout the thesis. You have been a great teacher and a mentor, consistently valuing my work. Your expertise and scientific rigor have shaped my approach to research.

Gregor Kasieczka, thank you for agreeing to be my first reviewer. Although our meetings were sparse, your feedback helped me a lot to improve my work.

I am grateful to the SuperCDMS collaboration for giving me the opportunity to work on the experiment and for providing a friendly atmosphere. I thank everybody I collaborated with over the three years. I'm excited to continue after this thesis and to be a part of developing the main experiment at SNOLAB. Special thanks to Amy, Andy, Anthony, Mike, Rik, Stefan, Wolfgang, and Ziqing. I learned so much from you all.

Many thanks to Matthew Wilson, Belina von Krosigk, and Francesco Toschi for providing feedback on this written work, which improved significantly from your comments.

I began my thesis with a scintillator project with Torben Ferber's group. It was a great learning experience, and I appreciate the knowledge I gained during that time.

To my colleagues Matt and Francesco, thank you for always being supportive and lending an ear whenever I needed it.

- Francesco, talking through my ideas with you helped me get back on track when I was too absorbed in one of the many branches of this thesis. Your reminders to wrap things up really helped me to get over the finish line. I'm thankful for your friendship and support, which made the final stressful months easier.

- Matt, the limit-setting part of this thesis owes so much to your patience and expertise. Towards the end of my thesis, you spent so much time reading and discussing details of my results, even after relocating. Thank you for all your insightful questions and for showing genuine interest in my work. Can't go without saying that our time as office-mates was fun!

To our team mates, thank you for providing a friendly atmosphere. Sasha, your "motivational" words gave me perspective when I needed it most. Emanuele and Eleanor, thank you for guiding me with CDMSlite analysis. You both were always available to answer my questions and point me toward the right resources, which I highly appreciated.

Jenny, you've been a mentor to me, not just during my Ph.D. time but since my bachelor thesis days. Your steady presence and guidance have been invaluable throughout my academic career. Thank you!

Vera and Melih, it was so much fun doing my thesis the same time as you. Thank you for the coffee breaks, writing sessions, pep talks, karaoke, food and walks.

Finally, my deepest gratitude goes to Vikas. Thank you for your unwavering support during all these years. Without you, none of this would have been possible. Thank you for encouraging me to apply for that internship at DESY in 2017, which ultimately set me on the path to a career in particle physics.

Contents

1. Introduction to Dark Matter	1
1.1. Cosmological evidence	1
1.1.1. Rotation curves of galaxies	2
1.1.2. Cosmic Microwave Background (CMB)	3
1.1.3. Gravitational lensing: The Bullet Cluster	5
1.1.4. Modified gravity	7
1.2. Dark matter candidates	7
1.2.1. Weakly Interacting Massive Particle (WIMP)	7
1.2.2. Light Dark Matter (LDM)	9
1.2.3. Bosonic dark matter	10
1.2.4. Other dark matter candidates	10
1.3. Direct dark matter searches	11
1.3.1. Current WIMP search status	12
2. The SuperCDMS Experiment	14
2.1. A brief history of SuperCDMS	14
2.2. Particle detection	15
2.2.1. Semiconductors detectors	15
2.2.2. Detector types	17
2.2.3. Phonon measurement	19
2.2.4. Ionization measurement	23
2.3. SuperCDMS at Soudan	24
2.4. SuperCDMS at SNOLAB	27
3. CDMSlite Run 3	30
3.1. Description of the experiment	30
3.2. Data processing	32
3.3. Energy reconstruction	37
3.3.1. Energy scale	38
3.3.2. Energy calibration	38
3.3.3. Energy Corrections	38
3.3.4. Energy resolution model	41
3.4. Data-quality cuts	42
3.4.1. Live-time cuts	42
3.4.2. Pulse shape cuts	43
3.4.3. Low-frequency noise removal	44
3.4.4. Other efficiency cuts	47

3.4.5.	Fiducial volume cut	48
3.5.	Live-time and efficiencies	49
3.5.1.	Live-time estimate	50
3.5.2.	Trigger efficiency	50
3.5.3.	Data-quality cuts efficiencies	51
3.5.4.	Fiducial volume efficiency	51
3.5.5.	Combining efficiencies	51
4.	Low-Frequency Noise Identification using Machine Learning	53
4.1.	Sample preparation	53
4.1.1.	LFN-background samples	53
4.1.2.	Signal samples	55
4.1.3.	Pre-processing	58
4.2.	Simulation Using Pulse shapes and Regression (SUPR)	60
4.2.1.	K-shell events	60
4.2.2.	Fitting traces	62
4.2.3.	Using the fit parameters for the traces generation	65
4.2.4.	Generation of new traces	65
4.3.	ML-based classification approach	66
4.3.1.	Convolutional neural network (CNN)	67
4.3.2.	Long Short-Term Memory (LSTM)	69
4.3.3.	Model training	69
4.3.4.	Loss function	70
4.3.5.	Metrics	70
4.3.6.	Class activation maps	72
4.4.	LFN classification using CNN	73
4.4.1.	Single-channel	74
4.4.2.	PT-channel	81
4.4.3.	Multi-channel	87
4.4.4.	Hyperparameter optimized CNNs	91
4.5.	LFN classification using CNN-LSTM	96
4.5.1.	Single-channel	97
4.5.2.	PT-channel	98
4.5.3.	Multi-channel	98
5.	CDMSlite Run 3 Dark Matter Search	101
5.1.	Limit setting method	101
5.1.1.	Concept of parameter space exclusion	101
5.1.2.	Maximum gap method	102
5.1.3.	Optimum Interval (OI)	102
5.2.	Model Selection for final analysis	103
5.2.1.	Signal efficiency	103
5.2.2.	Expected energy spectra	104
5.2.3.	Sensitivity projections	107

5.3.	Analysis results	107
5.3.1.	DM signal efficiency	107
5.3.2.	Application of the new LFN cut to CDMSlite Run 3 data	110
5.3.3.	Observed energy spectra	112
5.4.	DM exclusion limit with OI	113
5.4.1.	Systematic uncertainties	113
5.5.	Investigation of the additional observed events	114
5.6.	DM exclusion limit with OI without mis-reconstructed events	116
6.	Case Studies	118
6.1.	Impact of radial cut	118
6.2.	Placing a stringent cut	120
6.2.1.	With radial cut	121
6.2.2.	Without radial cut	123
6.3.	In the absence of observed events	127
7.	Conclusion and Outlook	129
7.1.	Summary	129
7.2.	Outlook	130
	Appendices	133
A.	Hyperparameter Optimization	134
A.1.	CNN architecture	134
A.1.1.	Single-channel	134
A.1.2.	PT-channel	135
A.1.3.	Multi-channel	135
A.2.	CNN-LSTM architecture	137
A.2.1.	Single-channel	137
A.2.2.	PT-channel	138
A.2.3.	Multi-channel	138
B.	Energy Mis-reconstruction	140
C.	Supplementary Information on the Impact of Radial Cut	142
	Bibliography	146
	Acronyms and Abbreviations	158

List of Tables

3.1. Baseline resolution of CDMSlite Run 3	42
3.2. Energy resolution for the ^{70}Ge activation peaks	42
3.3. Best-fit parameters of the energy resolution model	42
4.1. Model summary of the CNN classifier model	74
4.2. Hyperparameter-optimized parameters for the CNN architecture	92
4.3. Hyperparameter-optimized parameters for the CNN-LSTM architecture	97
5.1. Number of CDMSlite Run 3 events passing each selection criterion	111
5.2. Energy and network predictions for the newly identified events	115
B.1. Comparison of assigned and estimated energy for mis-reconstructed events	140

List of Figures

1.1. Rotation curves of various spiral galaxies as a function of distance	3
1.2. Temperature map of the CMB	4
1.3. CMB power spectrum measured by the Planck experiment	5
1.4. Image of the Bullet Cluster showing the stellar components, the X-ray-emitting plasma, and the mass distribution from gravitational lensing. . .	6
1.5. WIMP event rate as a function of recoil energy for Ge target	9
1.6. SI WIMP-nucleon interaction parameter space for direct detection experiments over WIMP mass	13
2.1. Illustration of a dark matter particle χ scattering off an atomic nucleus in a detector	16
2.2. Photograph and schematic of a SuperCDMS Soudan iZIP detector	18
2.3. Photograph and schematic of a SuperCDMS SNOLAB iZIP detector . . .	18
2.4. Photograph and schematic of a SuperCDMS Soudan HV detector	19
2.5. Illustration of the resistance vs. temperature curve for a TES	20
2.6. Schematic of quasiparticle trapping in a QET	21
2.7. Schematic of the TES circuit	22
2.8. Schematic of the SuperCDMS Soudan charge readout circuit	24
2.9. Schematic of the SuperCDMS Soudan experiment	25
2.10. Schematic of the SuperCDMS SNOLAB experiment	28
2.11. Projected sensitivities of SuperCDMS SNOLAB experiment to spin-independent (SI) WIMP-nucleon scattering	29
3.1. Schematic of the CDMSlite detector	30
3.2. Schematic of the HV circuit	31
3.3. Histogram showing the correlation between the trigger rate and the detector voltage	32
3.4. Raw traces from the CDMSlite detector phonon channels PA, PB, PC, and PD	33
3.5. CDMSlite Run3 OF templates	36
3.6. Power Spectral Densities (PSDs) of randomly chosen noise events	36
3.7. Example OF fit with fit results	37
3.8. Histogram comparing the energy distribution before and after the energy corrections	40
3.9. Standard OF template with glitch and LFN templates	44
3.10. Heatmap illustrating the cryocooler cycle variable over the run period . .	45
3.11. LFN templates used for pulse shape cuts	46

3.12. Calculated bias voltage distribution across a 2D plane of vertical position against radial position	49
3.13. Events passing and failing the radial cut shown on energy- ξ plane	50
3.14. Hardware trigger efficiency, quality cuts efficiency, and fiducial volume cut efficiency for CDMSlite Run 3 analysis	52
4.1. PT channel amplitude and energy distributions of the LFN events	54
4.2. Raw traces of example LFN events	54
4.3. Examples of raw traces that did not pass the standard deviation cut	56
4.4. Examples of raw traces that passed the standard deviation cut	57
4.5. Pulse templates of individual phonon channels and the PT channel	57
4.6. Schematic of signal samples generation for single- and PT-channel	58
4.7. Comparison of pre-processed raw traces with unprocessed raw traces	59
4.8. CDMSlite Run 3 event spectrum with K-events highlighted	61
4.9. Partition plot with the K-shell events	62
4.10. Raw traces fitted using an exponential fit function consisting of one rise time and one fall time constants	63
4.11. Raw traces fitted using an exponential fit function consisting of two rise time and two fall time constants	64
4.12. Trace generation using the SUPR method	66
4.13. Schematic of signal samples generation for multi-channel	67
4.14. Illustration of the confusion matrix	71
4.15. Schematic of the CNN architecture	73
4.16. Training performance of single-channel CNN classifier	75
4.17. ROC curve and signal probability for a test sample using single-channel CNN classifier	76
4.18. Confusion matrix for a test sample using single-channel CNN classifier	76
4.19. Signal efficiency and purity of single-channel CNN classifier	77
4.20. Raw traces of example false negatives from single-channel CNN classifier	79
4.21. Activation maps of example false negative traces from single-channel CNN classifier	79
4.22. Raw traces of example false positives from single-channel CNN classifier	79
4.23. Activation maps of example false positive traces from single-channel CNN classifier	80
4.24. Signal efficiency and purity of single-channel CNN classifier for different data processing methods	80
4.25. Training performance of PT-channel CNN classifier	82
4.26. ROC curve and signal probability for a test sample using PT-channel CNN classifier	82
4.27. Confusion matrix for a test sample using PT-channel CNN classifier	83
4.28. Signal efficiency and purity of PT-channel CNN classifier	84
4.29. Raw traces of example false negatives from PT-channel CNN classifier	85

4.30. Activation maps of example false negative traces from PT-channel CNN classifier	85
4.31. Raw traces of example false positives from PT-channel CNN classifier . .	85
4.32. Activation maps of example false positive traces from PT-channel CNN classifier	86
4.33. Signal efficiency and purity of PT-channel CNN classifier for different data processing methods	86
4.34. Training performance of multi-channel CNN classifier	88
4.35. ROC curve and signal probability for a test sample using multi-channel CNN classifier	88
4.36. Confusion matrix for a test sample using multi-channel CNN classifier . .	89
4.37. Signal efficiency and purity of multi-channel CNN classifier	90
4.38. Raw traces of example false negatives from multi-channel CNN classifier .	90
4.39. Raw traces of example false positives from multi-channel CNN classifier .	90
4.40. Signal efficiency and purity of multi-channel CNN classifier for different data processing methods	91
4.41. Performance of the CNN compared to hyperparameter-optimized CNN for single-channel classification	93
4.42. Performance of the CNN compared to hyperparameter-optimized CNN for PT-channel classification	94
4.43. Performance of the CNN compared to hyperparameter-optimized CNN for multi-channel classification	95
4.44. Schematic of the CNN architecture	96
4.45. Performance of hyperparameter-optimized CNN compared to hyperparameterized CNN-LSTM for single-channel classification	98
4.46. Performance of hyperparameter-optimized CNN compared to hyperparameterized CNN-LSTM for PT-channel classification	99
4.47. Performance of hyperparameter-optimized CNN compared to hyperparameterized CNN-LSTM for multi-channel classification	99
5.1. Quality cut efficiency as a function of energy using the three ML-models .	104
5.2. DM signal efficiency as a function of energy using the three ML-models .	105
5.3. The CDMSlite Run 3 best-fit background models with the energy spectrum	106
5.4. Monte-Carlo toy experiments for the sensitivity studies.	106
5.5. Sensitivity projections using the three ML-models	107
5.6. CDMSlite Run 3 radial cut efficiency	109
5.7. Final DM signal efficiency using the previous and new analysis	110
5.8. Final energy spectra of CDMSlite Run 3 using the previous and new analysis	112
5.9. 90 % C.L. OI limits on the SI WIMP-nucleon cross-section using the previous and new analysis without removing the mis-reconstructed events . .	113
5.10. Systematic uncertainty on the 90 % C.L. OI limits on the SI WIMP-nucleon cross-section using the new analysis	114
5.11. Raw traces of the newly observed events	115

5.12. Activation maps of the new observed events	116
5.13. 90 % C.L. OI limits on the SI WIMP-nucleon cross-section using the previous and new analysis with the mis-reconstructed events removed	117
6.1. Comparison of the CDMSlite Run 3 energy spectrum for the analysis with bifurcated LFN cut and analysis with ML-based LFN cut.	119
6.2. Effect of the radial cut on the 90 % C.L. OI-based sensitivities on the spin-independent WIMP-nucleon cross-section for CDMSlite Run 3	120
6.3. Final signal efficiency of analysis using the ML-based LFN cut with a signal probability threshold at 0.99	122
6.4. Comparison of final signal efficiency of the analysis using ML-based LFN cut with signal probability threshold of 0.99 with other methods	122
6.5. Final energy spectrum of the analysis using ML-based LFN cut with signal probability threshold of 0.99, compared with other methods	123
6.6. DM sensitivities for the analysis using ML-based LFN cut with signal probability threshold of 0.99, compared with other methods	124
6.7. Final signal efficiency of the analysis using ML-based LFN cut with a signal probability threshold of 0.99, compared with other methods. The radial cut was not applied	125
6.8. Final energy spectrum of the analysis using ML-based LFN cut with signal probability threshold of 0.99, compared with other methods. The radial cut was not applied.	126
6.9. DM sensitivities for the analysis using ML-based LFN cut with signal probability threshold of 0.99, compared with other methods. The radial cut was not applied	126
6.10. 90 % C.L. OI-based sensitivities on the spin-independent WIMP-nucleon cross-section for CDMSlite Run 3 with the assumption of no observed events	128
A.1. Hyperparameter optimization for CNN architecture with the single-channel configuration	135
A.2. Hyperparameter optimization for CNN architecture with the PT-channel configuration	136
A.3. Hyperparameter optimization for CNN architecture with the multi-channel configuration	136
A.4. Hyperparameter optimization for CNN-LSTM architecture with the single-channel configuration	137
A.5. Hyperparameter optimization for CNN-LSTM architecture with the PT-channel configuration	138
A.6. Hyperparameter optimization for CNN-LSTM architecture with the multi-channel configuration	139
B.1. Raw traces of the additional events observed that have mis-reconstructed energies.	141
B.2. Relation between OF amplitudes in ADC units and energy in keV_{ee} units	141

C.1. Energy spectrum of the CDMSlite Run 3 events below 0.08 keV_{ee}	142
C.2. Raw traces of additional events observed without the radial cut and with a signal probability threshold of 0.9	143
C.3. Raw traces of additional events observed without the radial cut and with a signal probability threshold of 0.99	145

1. Introduction to Dark Matter

The quest to understand the universe — its composition, origins, and underlying principles — has driven human curiosity for centuries. The Standard Model (SM) describes the known elementary particles, and the forces associated with them. However, several key questions remain unanswered, including (but not limited to) the matter-antimatter asymmetry, the strong CP problem, neutrino masses, dark matter, and dark energy. Observations suggest that dark matter constitutes approximately 26% of the universe's total energy content, while visible matter accounts for only about 5%. This thesis is a part of the extensive efforts to uncover the nature of dark matter. In this chapter, an introduction to dark matter is provided. Section 1.1 discusses the cosmological evidence, followed by Sec. 1.2, which describes potential dark matter candidates. Finally, Sec. 1.3 outlines direct dark matter search experiments and their results.

1.1. Cosmological evidence

There were several hints of Dark Matter (DM) uncovered in the 20th century, but a significant observation was made by Fritz Zwicky in 1933[1, 2]. While studying the redshift of galaxies in the Coma cluster, Zwicky applied the virial theorem to estimate the mass of the cluster. The time-averaged kinetic energy of a system consisting of N particles is given by

$$\langle T \rangle = \frac{1}{2} \sum_{k=1}^N \langle \mathbf{F}_k \cdot \mathbf{r}_k \rangle, \quad (1.1)$$

where $\mathbf{F}_k$ is the force acting on the k^{th} particle, and $\mathbf{r}_k$ is the position of the k^{th} particle. If the forces acting on particles are from a potential energy V that depends on the distance between particles r such that $V(r) \propto r^\alpha$, the kinetic energy can be written as

$$\langle T \rangle = \frac{\alpha}{2} \langle V \rangle. \quad (1.2)$$

For a galaxy cluster held by gravity, $\alpha = -1$, the kinetic energy becomes

$$\langle T \rangle = -\frac{1}{2} \langle V \rangle. \quad (1.3)$$

With the gravitational potential energy that depends on mass, the virial theorem implies that the average kinetic energy of particles in the system is higher for a system with more mass. Zwicky applied this principle to galaxies in the Coma cluster and observed that the mass of the galaxies found using luminosity could not account for the observed high

velocities. Thus, the Coma cluster has to contain a significant amount of non-luminous matter, or dark matter, which constitutes most of its mass.

1.1.1. Rotation curves of galaxies

Rotation curves of galaxies are among the most widely discussed evidence for dark matter, derived from observations of the rotational dynamics of galaxies. The time-averaged gravitational potential energy of an object with mass m in a galaxy at a distance r from the galactic center is given by:

$$\langle V \rangle = -\frac{GmM(r)}{r}, \quad (1.4)$$

where G is the gravitational constant, and $M(r)$ is the mass enclosed within radius r . With an orbital velocity v , the time-averaged kinetic energy of an object is given by

$$\langle T \rangle = \frac{1}{2}mv^2. \quad (1.5)$$

Applying Eq. 1.4 and Eq. 1.5 in the virial theorem in Eq. 1.3, the orbital velocity of the object is derived as

$$v = \sqrt{\frac{GM(r)}{r}}. \quad (1.6)$$

Galaxies consist of stars and interstellar gas orbiting the galactic center, and their orbital velocities are determined by Eq. 1.6, assuming circular orbits. As we move farther from the galactic center, the density of stars and interstellar gas decreases. Consequently, the orbital velocity is expected to decline as $1/\sqrt{r}$, as per the relation in Eq. 1.6, if visible matter accounts for the entire mass distribution in a galaxy.

In 1970, Vera Rubin and Kent Ford measured the velocity profile of ionized hydrogen in the Andromeda Galaxy (M31) [4]. Using the Doppler shift of $H\alpha$ emission lines from molecular hydrogen, they determined the velocities and observed a flat rotation curve at large radii. Instead of the expected $1/\sqrt{r}$ decline, the velocities remained constant. Throughout the 1970s, they cataloged rotation curves for spiral galaxies of various sizes, consistently observing this flat velocity behavior at large distances [5, 3]. Figure 1.1 shows the rotation curves for various spiral galaxies by Rubin *et. al.*

Aside from alternative theories, which are discussed in Sec. 1.1.4, there must be additional mass beyond what is visible to explain this behavior. The constant velocity profiles are consistent with a spherically symmetric distribution of non-luminous matter surrounding the galaxy, referred to as the dark matter halo.

Even in the Milky Way, estimates have been made for the local dark matter density. For a spherical halo, the density is estimated to be between 0.2 and $0.4 \text{ GeV}/c^2/\text{cm}^3$ [6]. The commonly used value in the dark matter community is $0.3 \text{ GeV}/c^2/\text{cm}^3$, which

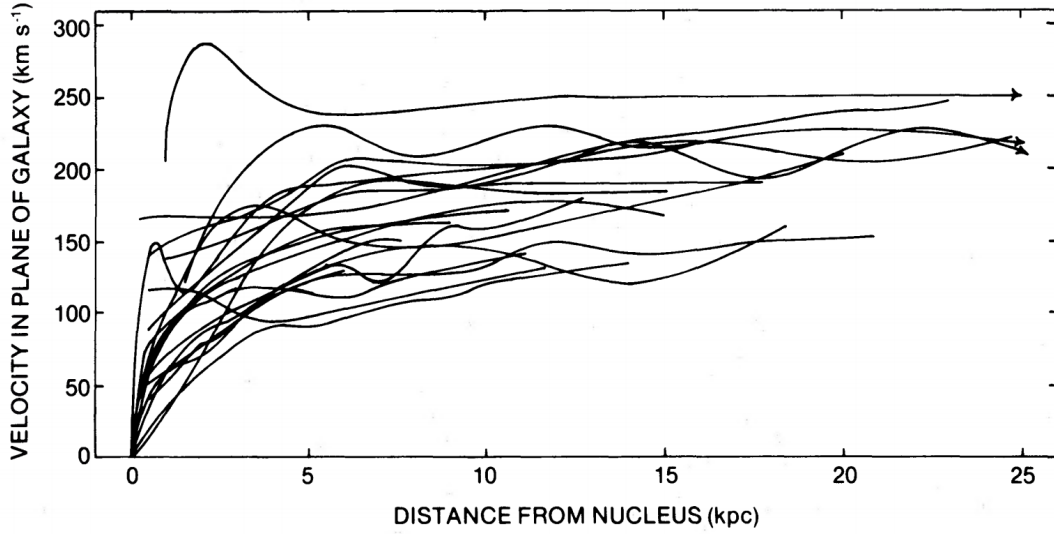


Figure 1.1.: Rotation curves of various spiral galaxies as a function of distance, showing the velocity remains relatively constant for objects far from the galactic center. Figure taken from Ref. [3].

is also adopted in this work.

1.1.2. Cosmic Microwave Background (CMB)

The Cosmic Microwave Background (CMB) is the primordial blackbody radiation of the universe and it provides one of the strongest evidence for the existence of dark matter. Before the epoch of recombination when the universe was cooled enough for electrons and protons to combine and form neutral atoms, they were strongly coupled to photons. As the universe cooled, matter and radiation thermally decoupled, thus the photons freely travel through space. These photons are still observable today at 2.7 K [7]. Thus, the CMB provides a snapshot of the early universe just after recombination. Initially, CMB observations exhibited a nearly perfect blackbody spectrum without anisotropies to the first order [8]. However, anisotropies were later discovered in both temperature and polarization, on the order of 10^{-5} and 10^{-6} , respectively [9].

The primordial fluid consisted of baryonic matter and radiation during recombination. Dark matter was largely decoupled from the primordial fluid and only interacted gravitationally. The gravity of dark matter caused the fluid to compress into over-dense pockets, that in-turn created under-dense regions. The fluid pressure provided a force that moved the fluid out of the over-dense regions, leading to oscillating compressions and decompressions of the fluid. This phenomenon is known as Baryon Acoustic Oscillations [11, 12]. As radiation escaped from the baryonic matter at the end of recombination, the oscillation of the fluid density stopped. The leftover baryonic matter “froze” preserving

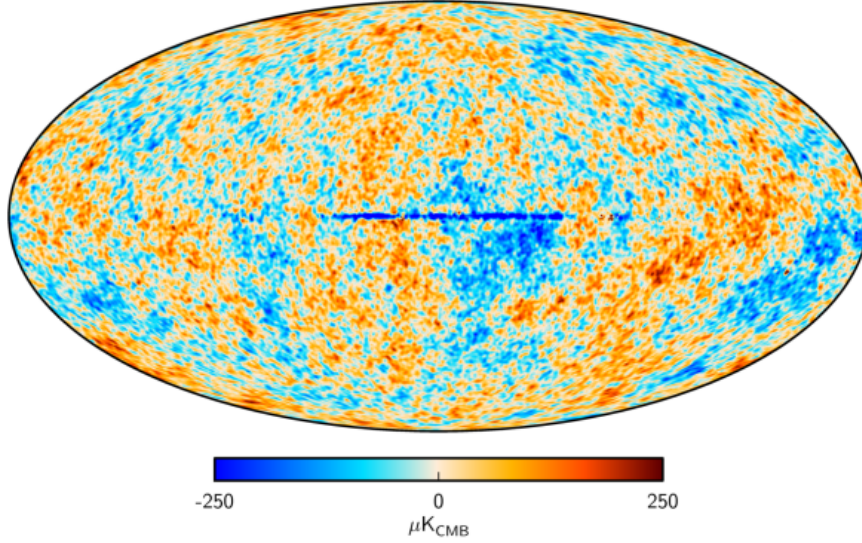


Figure 1.2.: Temperature map of the CMB showing the temperature fluctuations from the inhomogeneities in the photon-baryon fluid in the early universe. The information was derived from the joint baseline analysis of Planck, WMAP, and 408 MHz observations. Figure taken from Ref. [10].

the high and low-density regions. The photons escaping from high-density regions would have a lower temperature, while those escaping from low-density regions would have a higher temperature. This resulted in the fluctuations observed in the CMB spectrum. These frozen high-density regions would later become large-scale structures seen in the universe. As the inhomogeneities in the photon-baryon fluid happened before photon decoupling, the CMB captures the temperature fluctuations, as shown in Fig. 1.2.

The universe is composed of matter, radiation, and dark energy, each characterized by a density parameter Ω_i , which is given by

$$\Omega = \sum \Omega_i = \Omega_R + \Omega_M + \Omega_\Lambda, \quad (1.7)$$

where Ω_R , Ω_M , and Ω_Λ are densities of radiation, matter, and dark energy, respectively. The matter density is composed of ordinary baryonic matter and dark matter such that $\Omega_M = \Omega_b + \Omega_{DM}$.

The fluctuations observed in the CMB are converted into a power spectrum using spherical harmonics, resulting in the CMB power spectrum shown in Fig. 1.3. The locations and relative heights of the spectrum's peaks and troughs reflect the oscillatory behavior of the baryon-photon fluid at the time of decoupling. The relative heights and the spacing between the peaks are used to estimate Ω_b , Ω_M , and Ω_Λ [14]. Using the best-fit of the data, the current estimates of the densities of baryonic and dark matter are $\Omega_b \approx 0.0495$

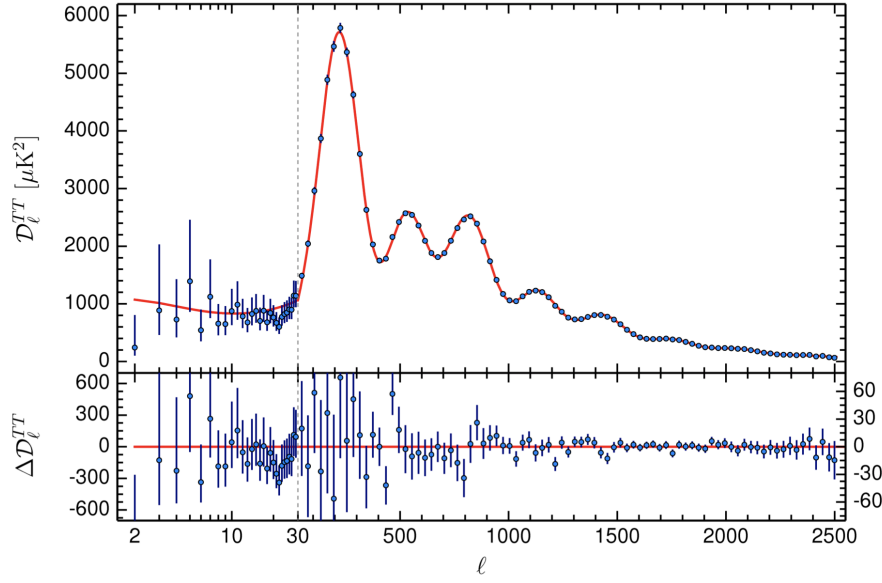


Figure 1.3.: The CMB power spectrum measured by the Planck experiment. The temperature fluctuation are plotted as a function of the multipole number ℓ . The relative peak heights and positions of the peaks are used to infer the energy contents of the universe. Figure taken from Ref. [13].

and $\Omega_{DM} \approx 0.264$ [15]. This implies that most of the matter content in the Universe is non-baryonic, and makes up roughly 85% of all matter. The anisotropies in the CMB and observations of large-scale structures suggest that the non-baryonic matter must be cold (non-relativistic). If the non-baryonic matter was relativistic, it would have resulted a structure formation paradigm inconsistent with our cosmological observations [16].

1.1.3. Gravitational lensing: The Bullet Cluster

General Relativity predicts that light follows the curvature of spacetime, bending in the presence of mass [17]. In other words, in the presence of a mass in the line of sight, the mass acts like a lens and deflects light. This effect is known as gravitational lensing. In the simplest example, a point-like lens of mass M that causes the deflection of light by an angle α is given by

$$\alpha = \frac{4GM}{rc^2}, \quad (1.8)$$

where G is the gravitational constant, r is the distance between the light source and the lens in the plane perpendicular to the observer, and c is the speed of light. By observing the distortions caused gravitationally by the lens, the mass of the lens can be estimated without knowing its composition [18]. Extreme lensing examples are arcs, multiple images, and ring images (a ring image occurs when the background object, lens, and the observed are all aligned). However, the majority of the lensing is subtle and referred

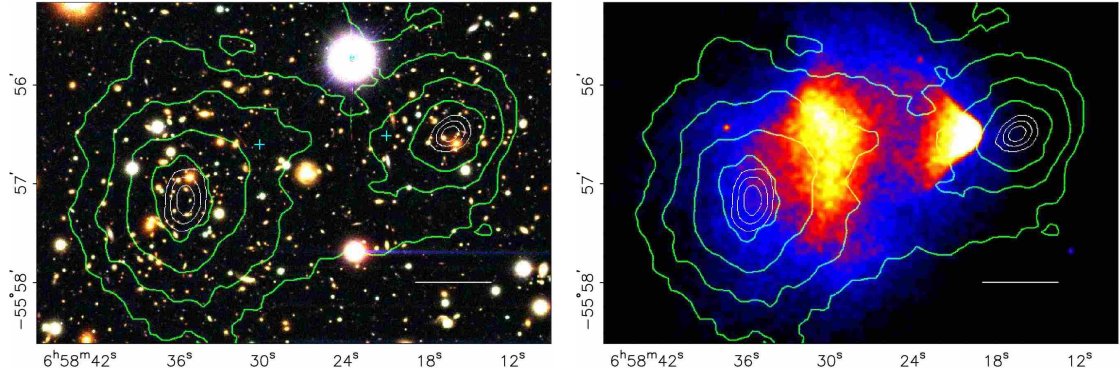


Figure 1.4.: Image of the Bullet Cluster showing the stellar components (left) and the X-ray-emitting plasma (right). The green contours represent the mass distribution determined using gravitational lensing. The center of the mass distribution is offset from the dense hot plasma region, showing the separation between dark matter and baryonic matter. Figure taken from Ref. [20].

to as weak lensing, where images of the background objects are only slightly distorted [19]. Also, weak lensing is observed on a large scale with a collection of background objects rather than individual objects. One of the examples of gravitation lensing that provides evidence of the existence of dark matter is the Bullet Cluster (1E 065756) [20], as shown in Fig. 1.4. It demonstrates the different properties of baryonic matter and dark matter in the collision of two clusters of galaxies.

The Bullet Cluster is composed of stellar components such as stars and galaxies, as well as X-rays-emitting hot plasma. The stellar components were observed by the Hubble Space Telescope, and the hot plasma was observed with the Chandra X-ray Observatory. The distribution of total mass was derived using gravitational lensing. It was seen that the hot plasma, which accounts for most of the cluster’s visible mass (ordinary matter), consists of hot gas whose interactions are slowed down due to electromagnetic interactions. The mass distributions using lensing show that the center of mass does not align with the plasma. If only baryonic matter were present, the mass distribution would follow the distribution of the interstellar plasma. However, it is observed that there are two large regions of non-luminous matter, which accounts for the majority of the cluster’s mass. This provides evidence of the existence of dark matter. The large spatial separation between the centers of total mass and baryonic mass in the Bullet Cluster is 8σ [20]. Additionally, it was observed that the dark matter components passed through each other with minimal interaction, indicating that dark matter is mostly collisionless. From this observation, the self-interaction cross-section of dark matter per unit mass was estimated to be less than $1.25 \text{ cm}^2/\text{g}$ [21].

This phenomenon has also been observed in the merging cluster MACS J0025.4-122 [22]. These observations also challenge modified Newtonian dynamics theories,

which propose that gravity behaves differently at large scales.

1.1.4. Modified gravity

The observed behavior of galaxy rotation curves and the mass-to-luminosity discrepancy has led to alternative theories, including modifications to gravity on large scales. The Modified Newtonian Dynamics (MOND) theory was introduced by Mordehai Milgrom in 1983 [23, 2]. MOND proposes that $F = ma$ holds under Newton's laws, but not in the regime of very low accelerations ($a \ll \sim 10^{-10} \text{ m/s}^2$). Another theory based on modified gravity is the AQUAdratic Lagrangian (AQUAL) theory, which modifies the Newtonian gravitational Lagrangian. AQUAL was later extended to include relativity, resulting in Relativistic AQUAL (RAQUAL) [24]. While modified gravity theories like MOND and RAQUAL don't account for the lensing, another possible theory to the lensing problem is the Tensor-Vector-Scalar (TeVeS) gravity proposed by Jacob Berkstein in 2004 [25, 26].

1.2. Dark matter candidates

The study of bullet clusters implied that dark matter is non-interacting electromagnetically and non-baryonic. Observations of CMB indicate that dark matter is non-relativistic, and constitutes approximately 85% of all matter. This section focuses on Weakly Interacting Massive Particles (WIMP) in detail, while other particle candidates, including Light Dark Matter (LDM), dark photons, axions, and Axion-Like Particles (ALPs), are briefly reviewed. For an in-depth discussion of these candidates, refer to Refs. [27, 28, 29].

1.2.1. Weakly Interacting Massive Particle (WIMP)

The Weakly Interacting Massive Particles (WIMPs) have long been considered a strongly motivated and leading candidate for dark matter. During the radiation-dominated era in the early universe, when temperatures were much higher than the WIMP mass, WIMPs were in thermal equilibrium. The equilibrium was maintained by annihilation with their antiparticles into lighter particles and the reverse process for creation [30].

As the universe expanded and cooled, the temperature dropped below the WIMP mass. At this stage, the annihilation rate of WIMPs fell below the Hubble expansion rate, halting the creation and annihilation processes [30, 31]. This phenomenon is known as thermal freeze-out. The remaining WIMPs with their number density fixed, constitute the current relic abundance that can be observed using terrestrial detectors.

To account for the amount of dark matter in the universe, the velocity averaged annihilation cross-section is estimated to be $\sim 10^{-26} \text{ cm}^3/\text{s}$ for particles with a mass in the range of 10s to 100s GeV/c^2 [32]. The cross-section is roughly at the scale of electro-weak force. WIMPs are also predicted by supersymmetry (SUSY), which addresses the hierarchy problem in particle physics, making WIMPs a compelling

candidate.

Consider a WIMP with mass m_χ that scatters off of a nucleus inside a target material, resulting in recoil energy E_r . The differential event rate for this WIMP-nucleus scattering is given by [33]

$$\frac{dR}{dE_r} = N_T \frac{\rho_\chi}{m_\chi} \frac{m_T \sigma_0}{2\mu_T^2} F^2(E_r) \mathcal{I}, \quad (1.9)$$

where N_T is number of target nuclei, m_T is the mass of the target nuclei, ρ_χ is the local WIMP matter density, μ_n is the WIMP-nucleus reduced mass given by $\mu_T = m_\chi m_T / (m_\chi + m_T)$, σ_0 is the point-like WIMP-nucleus cross-section, and $F(E_r)$ is the nucleus form factor. The function $\mathcal{I}$ denotes WIMP halo model which is given by

$$\mathcal{I} = \int_{v_{\min}}^{v_{\max}} \frac{f(\mathbf{v}, \mathbf{v}_E)}{v} d\mathbf{v}, \quad (1.10)$$

where $f(\mathbf{v})$ describes the dark matter velocity distribution that depends on $\mathbf{v}$ the velocity of the WIMPs in the halo's frame, and $\mathbf{v}_E$ the average velocity of the Earth with respect to the halo. The total WIMP-nucleus cross-section σ_0 in Eq. 1.9 consist of Spin-Independent (SI) cross-section σ_0^{SI} and Spin-Dependant (SD) cross-section σ_0^{SD} , and the Eq. 1.9 can be written as

$$\frac{dR}{dE_r} = N_T \frac{m_T \rho_\chi}{2m_\chi \mu_T^2} \left[\sigma_0^{\text{SI}} F_{\text{SI}}^2(E_r) + \sigma_0^{\text{SD}} F_{\text{SD}}^2(E_r) \right] \int_{v_{\min}}^{v_{\max}} \frac{f(\mathbf{v}, \mathbf{v}_E)}{v} d\mathbf{v}. \quad (1.11)$$

Assuming Maxwellian distribution boosted to the lab frame of the Earth [34],

$$f(\mathbf{v}, \mathbf{v}_E) \propto \exp(-|\mathbf{v} + \mathbf{v}_E|^2 / v_0^2), \quad (1.12)$$

where v_0 is the Galactic circular velocity. The minimum WIMP speed $v_{\min}$ in Eq. 1.11 is given by $\sqrt{m_T E_r / 2\mu_T^2}$ [33]. Thus, the $v_{\min}$ scales with the mass of the nucleus, so heavier nuclei require higher WIMP velocity to produce a recoil of energy E_r . The $v_{\max}$ in Eq. 1.11 is determined by the escape velocity of the galaxy which is $v_{\text{esc}} \approx 544 \text{ km/s}$ [35]. Comparing and combining results from experiments with various target materials requires isolating the WIMP-nucleon cross-section $\sigma_{n,0}^{\text{SI}}$ from the WIMP-nucleus cross-section σ_0^{SI} such that the SI cross-section for any target nucleus containing A nucleons is given by

$$\sigma_0^{\text{SI}} = \sigma_{n,0}^{\text{SI}} \frac{\mu_T^2}{\mu_n^2} A^2, \quad (1.13)$$

where μ_n^2 is the WIMP-nucleon reduced mass [36]. Spin-dependant (SD) interactions require both the dark matter particle and the nucleus to have non-zero spin. The majority of isotopes in materials like germanium and silicon have zero spin, which limits their sensitivity to SD interactions. For example, only 8% of germanium nuclei and 5% of silicon nuclei are spin-sensitive [37]. In contrast, spin-independent (SI) cross-section

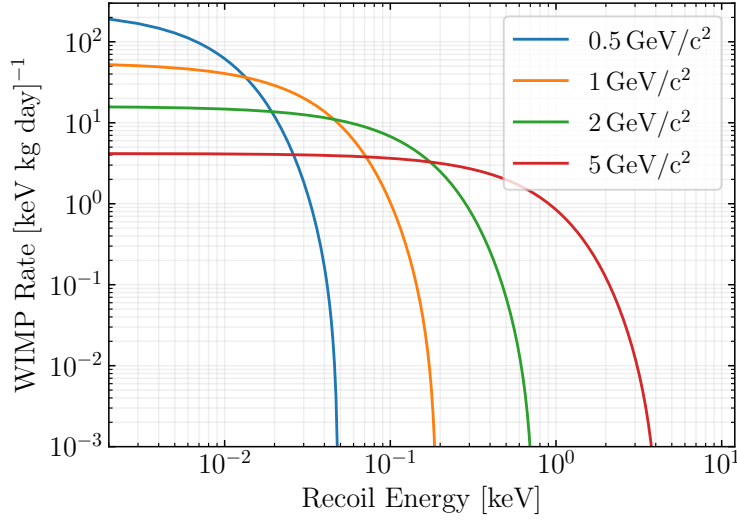


Figure 1.5.: Differential event rate as a function of recoil energy for WIMP recoils on a Ge target with WIMP-nucleon spin-independent cross section of 10^{-41} cm^2 for various WIMP masses.

scales with the square of the atomic mass (A^2) as seen in Eq. 1.13, favoring heavy nuclei as targets. However, heavy nuclei targets have limitations such as high energy thresholds which limit the sensitivity to low-mass WIMPs, making them more suited for exploring high-mass WIMPs. In this thesis work, only spin-independent cross-sections are discussed. Differential event rate as a function of recoil energy for a Ge target with WIMP-nucleon spin-independent cross section of 10^{-41} cm^2 for various WIMP masses is shown in Fig. 1.5.

1.2.2. Light Dark Matter (LDM)

Light Dark Matter (LDM) are dark matter candidates with masses below $2 \text{ GeV}/c^2$. For WIMPs with masses below $\sim 2 \text{ GeV}/c^2$ interacting via SM weak mediators, the annihilation cross-section is too small to account for the observed relic abundance. This is called the Lee-Weinberg limit [38]. The LDM models introduce a new “dark sector” of particles that mediates the interactions between LDM and SM. The new force mediators associated with LDM allow for a relic abundance consistent with observations despite their smaller masses. Unlike WIMPs, which primarily undergo nuclear scattering, LDM is expected to predominantly interact via electron scattering.

1.2.3. Bosonic dark matter

Dark Photons

Dark photons arise in a hidden dark sector under a new $U(1)$ gauge group and kinetically mix with Standard Model photons [39]. They act as mediators between DM and ordinary matter. They can interact with detectors via electromagnetic cross-sections proportional to the mixing parameter. Experiments like SuperCDMS search for dark photons through their absorption signatures in recoil spectra.

Axions and Axion-Like Particles (ALPs)

Axions are low-mass particles ($\leq \text{meV}/c^2$) originally proposed to resolve the strong CP problem in quantum chromodynamics (QCD) [40, 41]. Axion-Like Particles (ALPs) are general pseudo-Nambu-Goldstone bosons arising from the spontaneous breaking of a global $U(1)$ symmetry [42, 43]. The axions and ALPs provide low-mass DM candidates that interact weakly with photons that could be detected via their electromagnetic coupling.

1.2.4. Other dark matter candidates

Thus far, the candidates explored within the SuperCDMS collaboration were described. However, a wide variety of alternative dark matter candidates have been proposed, each motivated by different theoretical frameworks and experimental observations. Some of these candidates are briefly outlined below:

Sterile Neutrinos Sterile neutrinos are right-handed chirality neutrinos that interact only gravitationally and not via the electroweak force (thus, sterile) [44]. They are massive fermions that can mix with Standard Model neutrinos via the Higgs mechanism [45, 46]. They could be detected indirectly through their electroweak decay emitting X-rays or by mixing with a standard neutrino, scattering off of an electron [46, 47].

WIMPzillas WIMPzillas are supermassive DM candidates with masses between 10^{12} and $10^{16} \text{ GeV}/c^2$. Unlike standard WIMPs, they were hypothesized to not be in thermal equilibrium during freeze-out, but the production is explained by a freeze-in mechanism. These non-thermal DM are motivated by unexplained ultra-high-energy cosmic rays, which may result from the decay of the WIMPZilla [48, 49].

Primordial Black Holes (PBHs) PBHs are non-particle DM candidates formed in the early Universe, before Big Bang nucleosynthesis [50]. They remain a viable candidate, particularly after LIGO/Virgo detections of black hole mergers, which reignited interest in PBHs as a potential DM candidate [51]. The current constraints on the mass range are detailed in Ref. [52, 53].

MACHOs MACHOs include non-luminous baryonic objects like brown dwarfs, planets, and black holes [2]. The microlensing constraints (e.g., EROS-2 [54]) limit their contribution to less than 8% of the halo mass.

1.3. Direct dark matter searches

Before introducing the direct dark matter detection methods, a brief overview of two other primary methods for detecting dark matter particles is outlined below.

Indirect detection: This involves searching for the products of dark matter particle self-annihilation or decay. The products are expected to be SM particle-antiparticle or high-energy gamma rays. Experiments look for potential excess of decay products near massive objects such as black holes and stars. The Fermi Gamma-ray Space Telescope have searched for gamma-ray excesses [55]. PAMELA [56] and AMS-02 [57] have searched for positron excesses.

Production mechanisms: Collider and beam dump experiments may produce dark matter particles alongside standard model particles. Due to their rare interactions with SM particles, it can only be detected as missing mass. This approach has been explored in the Large Hadron Collider (LHC) with CMS and ATLAS experiments [58] and the SuperKEKB electron-positron collider with BelleII experiments [59].

The **direct detection methods** look for signatures of dark matter particles from the galactic halo interacting with detector materials. These direct dark matter interactions are characterized by small energy depositions occurring at low rates, so detectors with low background, high exposure and sensitivity are required. The interaction of DM with the detector material induces low-energy recoils, which can be detected using various techniques, including ionization, heat, and scintillation.

Solid state detectors: The detectors are operated at cryogenic temperatures and measure phonon, scintillation and ionization signals. They can search for DM with masses from sub-GeV to several 10s of GeV/ c^2 . Their ability to measure both phonon and charge signals provides discriminative power to distinguish between electron and nuclear recoils. Apart from WIMP searches, their sensitivity to small signals enables them to explore other dark matter candidates such as dark photons and ALPs.

SuperCDMS utilizes Ge and Si crystals to search for DM interactions. The measurements are done using TES and High-Electron-Mobility Transistors (HEMTs) for phonon and ionization signals, respectively. Details of the experiment are provided in the next chapter. EDELWEISS [60, 61] used ultra-pure Ge detectors with neutron transmutation doped (NTD) thermometers for phonon measurements. Cryogenic Rare Event Search using Superconducting Thermometers (CRESST) [62, 63] uses CaWO_4 crystals instrumented with silicon-on-sapphire wafers and W-TESs for measuring scintillation light and phonons.

Noble liquid detectors Noble liquid detectors measure scintillation light produced when the DM particles interact with the liquid target, such as xenon or argon. There are two common types of detectors: single-phase and dual-phase time projection chambers (TPCs). Single-phase detectors measure only scintillation light, using pulse shape discrimination to distinguish nuclear and electron recoils based on the timing distribution of emitted photons. The dual-phase TPCs add an electric field to drift liberated electrons into a gaseous phase, where they produce secondary scintillation. Experiments such as XENON [64, 65], PandaX [66, 67], and LZ (LUX-ZEPLIN) [68] use a xenon target, while experiments such as DEAP [69], and DarkSide [70] use liquid argon to search for DM. Xenon-based detectors set competitive constraints on WIMP-nucleon cross-sections due to xenon’s high atomic mass. Lower atomic mass noble liquids like helium are promising for detecting low-mass dark matter.

Bubble chambers and Charge-coupled devices (CCDs) Superheated bubble chambers detect particle interactions by observing bubble formations in a superheated fluid target. Energy depositions create bubbles, which are tracked using cameras. The PICO experiment [71, 72] uses chlorofluorocarbons as the target to search for interactions. The Charge-Coupled Devices (CCDs) can detect small energy deposits from particle interactions in the bulk of CCDs and are well-suited to detect low-mass DM interactions. The SENSEI [73] and DAMIC [74] experiments use CCDs to search WIMP, LDMs, and dark photons.

1.3.1. Current WIMP search status

Over the last several decades, numerous experiments have searched for WIMP-nucleon interactions, excluding a significant portion of the parameter space. Fig. 1.6 shows the current status of the WIMP parameter space for WIMP-nucleon interactions. Noble liquid detectors have excluded most of the mass region above $10 \text{ GeV}/c^2$. Their cross-sections are getting closer to the “neutrino fog”, where experiments will face irreducible neutrino backgrounds. For masses below $10 \text{ GeV}/c^2$, solid-state detectors have excluded parts of the parameter space, with ongoing and future experiments aiming to probe even lower cross-sections. Also, this figure highlights the growing focus on low-mass dark matter searches.

With the information about dark matter explained in this chapter, Chapter 2 details the SuperCDMS experiment and its detection principles. The CDMSlite Run 3 analysis, which forms the basis of this work, is outlined in Chapter 3. The CDMSlite Run 3 experiment faced issues with the supplementary cryocooler, leading to vibration-induced Low-Frequency Noise (LFN). In Chapter 4, a machine learning-based method to remove these LFN events is detailed. This new method for removing LFN events was applied to the CDMSlite Run 3 dataset. The results are presented in Chapter 5. Chapter 6 explores various case studies based on the newly developed method to remove the LFN events. Finally, Chapter 7 concludes this written work with a summary and an outlook.

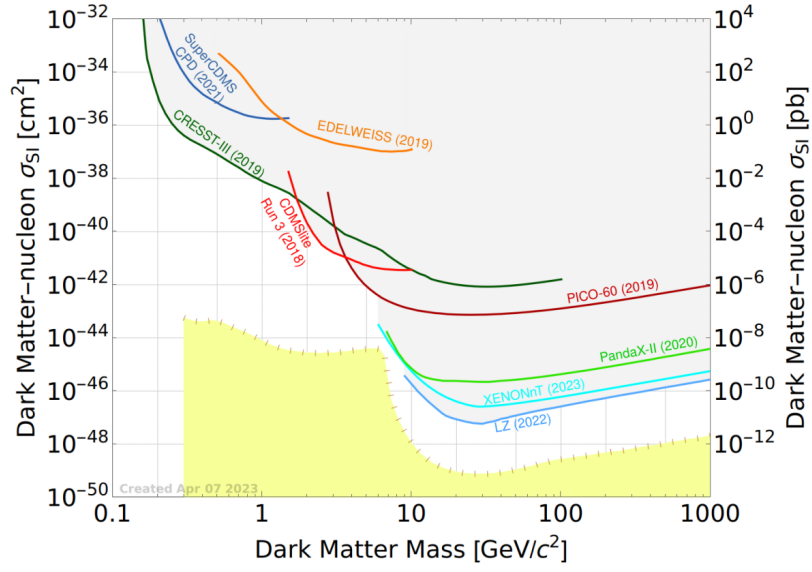


Figure 1.6.: SI WIMP-nucleon interaction parameter space for direct detection experiments over WIMP mass. The current results of various DM search experiments are shown by the colored curves [61, 75, 63, 76, 72, 67, 65, 68]. The excluded parameter space is shown in shaded-grey, and the neutrino fog for the Xe target is shown in shaded-yellow [77]. The SuperCDMS Limit Plotter was used to make this figure [78].

2. The SuperCDMS Experiment

The Super Cryogenic Dark Matter Search (SuperCDMS) experiment uses semiconductor crystal detectors to search for DM interactions with the detector material. The silicon (Si) and germanium (Ge) semiconductor crystals used in SuperCDMS are kept at cryogenic temperatures and are configured to measure either ionization or phonons or in some cases both. This chapter begins with a short summary of the origins of SuperCDMS in Sec. 2.1, followed by the detection principle in Sec. 2.2. Section 2.3 and Section 2.4 detail the SuperCDMS experiment at Soudan and SNOLAB, respectively.

2.1. A brief history of SuperCDMS

SuperCDMS has a history dating back to 1996, beginning with the predecessor Cryogenic Dark Matter Search (CDMS) experiment, followed by CDMS-II, the SuperCDMS experiment at Soudan, and the upcoming SuperCDMS experiment at SNOLAB. With each generation, detection volumes have increased, the experiments have moved deeper underground for better shielding against cosmic rays, and detector technologies have advanced. This section briefly outlines the progression of these generations; for a more detailed history, refer to Ref. [79].

CDMS: The CDMS experiment was conducted at the Stanford Underground Facility (SUF) at Stanford University. The experiment used 1-cm-thick, 7.6-cm-diameter germanium and silicon detectors in 1998 [80, 81]. The shallow underground facility provided 16 meter water equivalent (m.w.e.) overburden, which suppresses the background from cosmic radiation by a factor of 5. Despite this reduction, the dominant background remained the hadronic component of cosmic radiation [82].

CDMS-II: The second phase, CDMS-II, was relocated to the Soudan Underground Laboratory (SUL) in Minnesota, USA. It is situated 780 m underground in a decommissioned iron mine. With an overburden of 2090 m.w.e., the muon flux was reduced by a factor of 10^4 [82]. Between 2003 and 2008, the experiment deployed 19 germanium detectors (239 g each) and 11 silicon detectors (106 g each), maintaining the same 1-cm thickness and 7.6-cm radius as the original design [83]. The analysis using the Ge detectors set leading limits (at that time) on the spin-independent WIMP-nucleon cross-section for WIMP masses above $70 \text{ GeV}/c^2$ [84].

SuperCDMS Soudan: The subsequent generation, SuperCDMS, utilized 2.5-cm-thick, 7.6-cm-diameter interleaved Z-sensitive Ionization and Phonon (iZIP) detectors.

There were five towers, each consisting of 3 germanium detectors with a total payload of ~ 9 kg [85]. An analysis of 577 kg-days of data resulted in a spectrum consistent with background expectations while setting leading limits (at that time) on WIMP-nucleon cross-section in the mass range of 4 and 6 GeV/c² [86]. Sec. 2.3 describes the experiment facility.

Between 2012 and 2015, two of the detectors were applied a high bias voltage to probe low-mass WIMP signatures in an experimental mode called the CDMS low ionization threshold experiment (CDMSlite) [87]. The data taken with this configuration presented leading spin-independent dark matter interaction limits for WIMP masses ranging from 1.6 to 5.5 GeV/c². This thesis work was developed using the dataset of the final run [76] in this configuration, and Chapter 3 discusses the experimental conditions and the analysis in detail.

SuperCDMS SNOLAB The newest generation, SuperCDMS SNOLAB, is under construction at SNOLAB in an active nickel mine in Sudbury, Ontario, Canada. With an overburden of 6010 m.w.e., the facility offers 2.5-3 times the shielding against cosmic rays compared to Soudan. The experiment features upgraded detectors with increased volume, operating at colder temperatures, and in different measurement configurations [88]. Further details of this experiment are provided in Sec. 2.4.

2.2. Particle detection

This section discusses how particles interact within semiconductor materials, the types of detectors used in SuperCDMS, the measurement of charge, and phonon signals.

2.2.1. Semiconductors detectors

Semiconductors are widely used in particle detectors due to their unique electronic properties. In solid-state materials, the discrete energy levels transform into energy bands: the valence band where electrons are bound to the atomic nucleus, and the conduction band where they are free to move throughout the material. The energy difference between the valence and the conduction band is referred to as the band gap. The band gap is an important quantity in determining conductivity: if the band gap is large it leads to reduced conductivity and vice versa.

In semiconductors at cryogenic temperatures, electrons predominantly occupy the valence band. To excite an electron into the conduction band, energy at least equal to the band gap (E_g) is required. This excitation creates a “hole” in the lattice, which is a quasiparticle with a net positive charge. The electron and the hole together are referred to as the electron-hole pair or e^-h^+ pair. When an electron moves from the valence to the conduction band, the corresponding hole moves from the conduction band to the valence band.

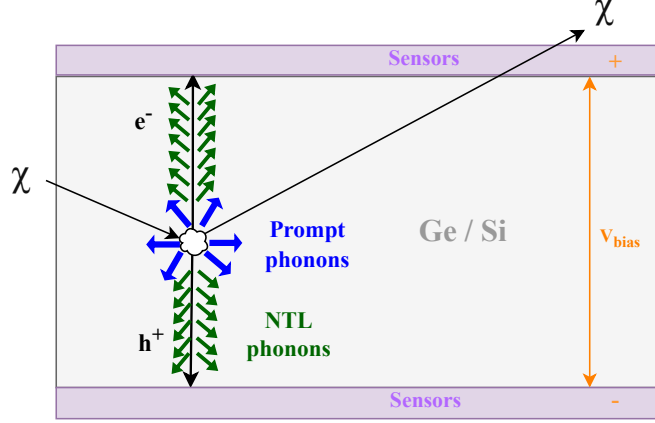


Figure 2.1.: Illustration of a dark matter particle χ scattering off an atomic nucleus in a detector. The interaction produces prompt phonons and liberates a certain number of e^-h^+ pair. With an application of bias voltage V_{bias} across the detector, the e^-h^+ pairs drift towards the electrodes on the detector surface, denoted by their polarity (+/-). Apart from the prompt phonons, drifting charges generate additional phonons along the way via the Neganov-Trofimov-Luke (NTL) effect, referred to as NTL phonons. Sensors on the detector's surface measure the energy deposited by these phonons.

When a particle interacts in a semiconductor, it can recoil against an atomic nucleus or a electron, depositing energy. The scattering process can liberate e^-h^+ pairs, depending on the energy and the recoil type. The interaction also produces phonons in the crystal called prompt phonons. When an electric field is applied across the detector, the e^-h^+ pairs drift to the respective electrodes on the opposite sides of the detector and can be measured by the sensors on the detector surface. When the charge carriers drift across the detector, they reach terminal velocity due to interactions with atoms along the way and the additional energy gained from the electric field is dissipated into the crystal lattice as heat in the form of phonons (lattice vibrations). This phenomenon is known as Neganov-Trofimov-Luke (NTL) effect [89, 90]. The phonons produced from the NTL effect are referred to as NTL phonons. A schematic of a particle interaction in the detector is shown in Fig. 2.1.

Thermal phonons are background thermal fluctuations due to the temperature of the semiconductor target. At crystal temperatures in the range of mK, the thermal phonons are negligible. The phonons generated by the interactions are athermal phonons which have energies higher than thermal phonons. Unlike thermal phonons, the generated athermal phonons have enough energy to be absorbed by the sensors on the surface of the detectors, and will be further discussed in Sec. 2.2.3.

2.2.2. Detector types

SuperCDMS uses silicon and germanium crystal detectors in two distinct configurations: Inter-leaved Z-sensitive Ionization and Phonon (iZIP) detectors and high voltage (HV) detectors. The iZIP detectors measure both phonon and charge signals. This allows for discrimination between electron recoils (ER) and nuclear recoils (NR), which is crucial for effective background rejection. In contrast, the HV detectors amplify low-energy signal events by applying a high bias voltage via the NTL effect but lack charge channels. The measurements from the phonon and charge channels are detailed in Sec. 2.2.3 and Sec. 2.2.4, respectively. With a low-energy threshold, HV detectors enable low-mass DM searches. The two configurations are designed to address different scientific objectives, such as low-mass WIMP searches or accurate background discrimination. In this section, an overview of the detector configuration types based on Refs. [88, 79] is presented.

Inter-Leaved Z-Sensitive Ionization and Phonon (iZIP) detector: In this type of detector, the charge channels are interleaved with the phonon channels. Typically, a small bias voltage of a few volts is applied to collect the charge carriers. NR events produce fewer electron-hole e^-h^+ pairs for a given recoil energy than ER events of the same recoil energy. Consequently, measuring both the charge and phonon signals enables discrimination between ER and NR events. This event-by-event ER/NR discrimination is crucial, as most background events are γ and β induced ER events, and WIMPs are expected to interact predominantly via NR.

For the SuperCDMS Soudan experiment, each detector had four phonon channels and two charge channels on each side. Phonon channels consisted of an outer ring and three wedge-shaped inner channels. The top and the bottom phonon channels were rotated by 60° for better position information. Figure 2.2 shows a photograph of the SuperCDMS Soudan iZIP detector and a schematic of the channel layout. In contrast, the SuperCDMS SNOLAB detectors feature six phonon channels on each side, arranged as an inner core, four wedge-shaped channels, and an outer ring. An “outer” ionization channel overlaps with the outermost phonon ring, while an “inner” ionization channel overlaps with the remaining phonon channels. The wedge-shaped phonon channels in the bottom are rotated by 45° degrees relative to the top. The outer ring channels enhance event identification at higher radii. A photograph of the SuperCDMS SNOLAB iZIP detector and a schematic of the channel layout is shown in Fig. 2.3.

CDMSlite: iZIP with high bias voltage: During the SuperCDMS Soudan experiment, two iZIP detectors were operated with a high bias voltage between 70-75 V, utilizing NTL amplification to achieve a low energy threshold. The bias was applied to one side, with the other side grounded, where the phonon signals were read out. This approach enabled a low-mass WIMP search but did not allow for ER/NR discrimination. The success of this method motivated the design of the dedicated high voltage (HV) detectors for the SuperCDMS SNOLAB experiment.

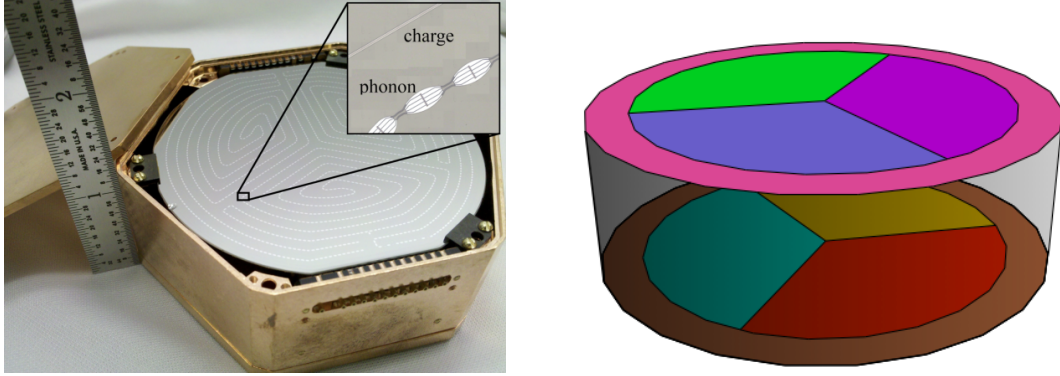


Figure 2.2.: A photograph of a SuperCDMS Soudan iZIP detector in its copper housing is shown on the left, and a schematic of its channel layout on the right, with phonon channels represented in varying colors. The inset displays a zoomed-in view of the detector surface, showing the phonon and charge channels. Figures are taken from Refs. [79, 91].

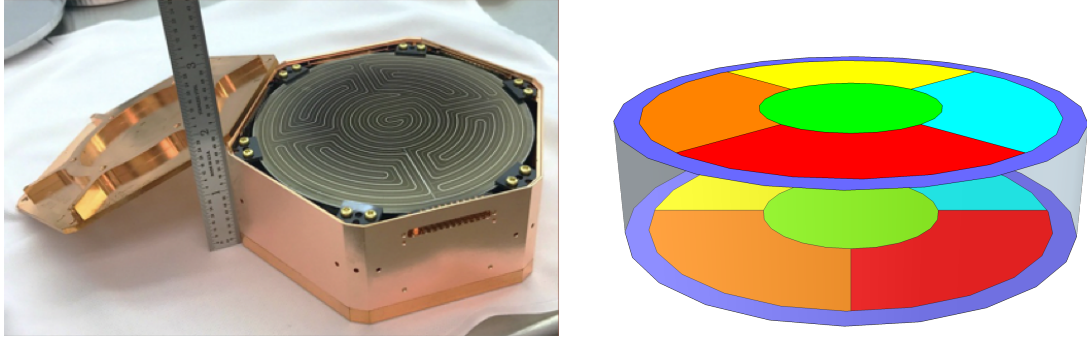


Figure 2.3.: A photograph of a SuperCDMS SNOLAB iZIP detector in its copper housing is shown on the left, and a schematic of its channel layout on the right, with phonon channels represented in varying colors.

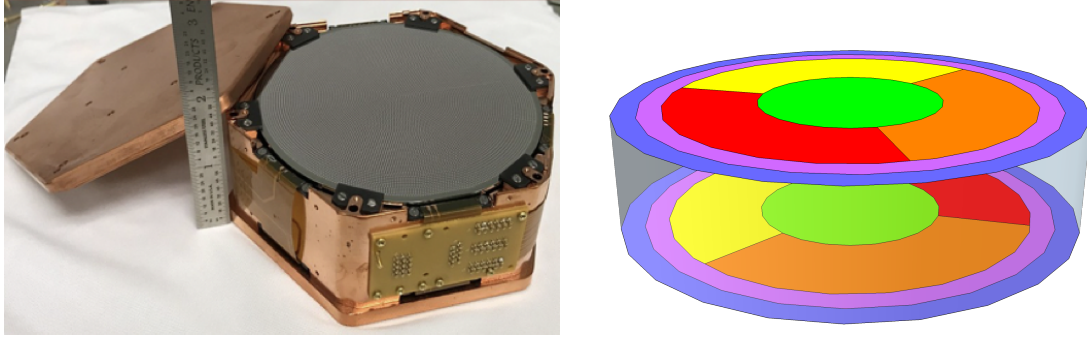


Figure 2.4.: A photograph of a SuperCDMS SNOLAB HV detector in its copper housing is shown on the left, and a schematic of its channel layout on the right, with phonon channels represented in varying colors.

High Voltage (HV) detector: This detector was specifically developed to operate with a high bias voltage, amplifying the phonon signal via the NTL effect. The detector is designed to be operated at a symmetrical bias voltage of ~ 100 V for better field uniformity. The high bias voltage allows for the detection of interactions that deposit small amounts of energy, enabling sensitivity to sub-GeV WIMP interactions. However, due to the absence of charge channels, this detector configuration does not provide ER/NR discrimination. Each side of the detector has 12 phonon channels, with six phonon channels on each side, consisting of an inner channel surrounded by three wedge-shaped channels and two outer ring channels. This arrangement is designed to improve the rejection of events near the detector’s edge. Fig. 2.4 shows a photograph of the SuperCDMS SNOLAB HV detector and a schematic of the channel layout. Additionally, the wedge-shaped phonon channels on the bottom surface are rotated by 60° relative to those on the top side. With a significant fraction of the detector surface covered by phonon sensors ($\sim 35\%$), this detector configuration has enhanced phonon absorption.

2.2.3. Phonon measurement

Phonons are measured using Quasiparticle-Trap-Assisted Electrothermal-Feedback Transition Edge Sensors (QETs) on the detector surfaces [92] and each phonon channel consists of QETs. These sensors consist of photo-lithographically patterned aluminum (Al) fins, which are connected to narrow Transition Edge Sensor (TES) made of tungsten (W). The aluminum fins act as phonon collectors. The TES consists of a superconducting material (W in this case), whose resistance increases with temperature, particularly exhibiting a steep rise near its critical temperature, T_c . Below the critical temperature, the material is superconducting with zero resistance and above the critical temperature, it transitions to a normal state with finite resistance [93, 94]. Fig. 2.5 shows the resistance vs. temperature curve for a TES. It is seen that the transition forms a narrow band where small temperature changes result in significant changes in the resistance, making the TES highly sensitive to minor temperature changes.

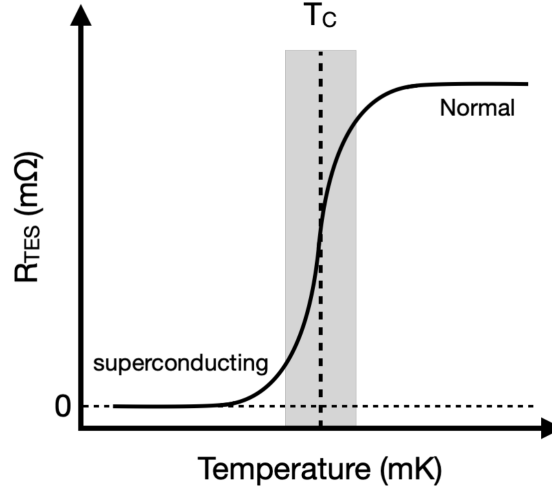


Figure 2.5.: An illustration of the resistance vs. temperature curve for a TES showing the transition region T_c , where resistance changes steeply with temperature. The TES operates in this region to detect small temperature variations. Below T_c , the TES is superconducting with zero resistance, while above T_c , it transitions to the normal resistive state. Figure taken from Ref. [79].

The setup is cooled to temperatures where both Al and W are superconducting. Most phonons produced in an interaction bounce off the surface of the detector several times before reaching the Al fins on the surface of the detectors. As the phonons reach the Al fins, they may either be absorbed into the fin or reflect back into the substrate. This probability of absorption is characterized by the phonon collection efficiency, which depends on factors such as the design of the sensor and the sensor coverage on the detector surface. If the absorbed phonons have energy greater than twice the superconducting gap of aluminum, $2\Delta_{\text{Al}} \approx 340 \mu\text{eV}$, they can break Cooper pairs in the superconducting Al and create quasiparticles. These quasiparticles diffuse through the Al fins. Absorbed phonons in Al can generate additional phonons and break Cooper pairs if they have enough energy. The overlap of Al and W forms a bi-layer region, which reduces phonon reflection. The quasiparticles diffuse through the Al fin to the W TES, where they are trapped and deposit energy. This energy increases the temperature of the TES, which is then measured. An illustration of this process is shown in Fig. 2.6. At cryogenic temperatures, the thermal phonons have energies on the order of $\mathcal{O}(10^{-6})$ eV, which is less than the energy required to break Cooper pairs in both Al and W, and thus they are not detected. In contrast, the athermal phonons generated in an interactions have an average energies on the order of $\mathcal{O}(10^{-4})$ eV in Ge and Si crystals, and thus they can be detected. Apart from particle interactions, vibrations can also produce phonons with sufficient energy to be absorbed by the Al fins. Developing a method to distinguish and remove these vibration-induced events is the major focus of this thesis.

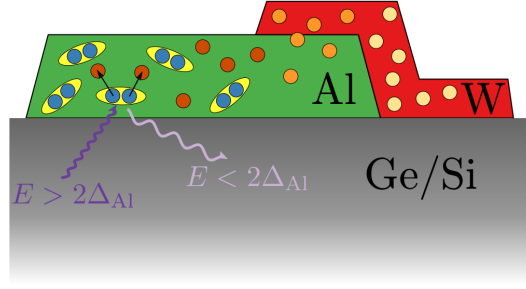


Figure 2.6.: Schematic of quasiparticle trapping in a QET. Phonons from the substrate with energy $E > 2\Delta_{Al}$ (dark purple) enter the superconducting aluminum fins (green) and break cooper pairs (blue circles in yellow oval), creating quasiparticles (solo circles with the gradient of orange indicating the energy). If a phonon has energy $E < 2\Delta_{Al}$ (light purple) cannot break cooper pairs. The quasiparticles diffuse into the tungsten TES (red) depositing their energy. Figure taken from Ref. [95].

A schematic of the TES circuit is shown in Fig. 2.7. The TES is voltage-biased with a fixed current source such that $V_b \approx I_b \times R_{sh}$, where I_b is the applied bias current. To maintain a stable bias voltage, a shunt resistor (R_{sh}) is wired in parallel to the TES. The energy deposited in the TES by the quasiparticles increases the temperature, causing the resistance of the TES (R_{TES}) to increase. This, in turn, changes the current through the TES circuit, resulting in a change in current through the input inductor L_i .

The current change in the TES circuit L_i is observed as a change in the magnetic field through the SQUID array, which is registered by an amplifier as the signal readout (V_0). Superconducting Quantum Interference Devices (SQUIDs) are sensitive magnetometers that measure changes in magnetic flux and output a voltage change [96, 97]. A feedback inductor (L_{fb}) is connected to the circuit, producing a current to maintain constant flux through the SQUID, ensuring stable operation of the SQUID. An amplifier, operated in feedback mode, stabilizes and linearizes the response from the SQUID. The current change in the L_i induces a current in the feedback inductor (L_{fb}), where the relationship between the two inductors is given by the inductor ratio L_i/L_{fb} . The ratio is 10 for SuperCDMS Soudan, and 2.4 for SuperCDMS SNOLAB. This way, by monitoring the current through the TES, the energy from the phonons can be measured. The QET is in thermal contact with the crystal substrate, so the TES is cooled down and goes back to the transition state. The signal is readout at V_0 with a digitization rate of 625 kHz for SuperCDMS Soudan.

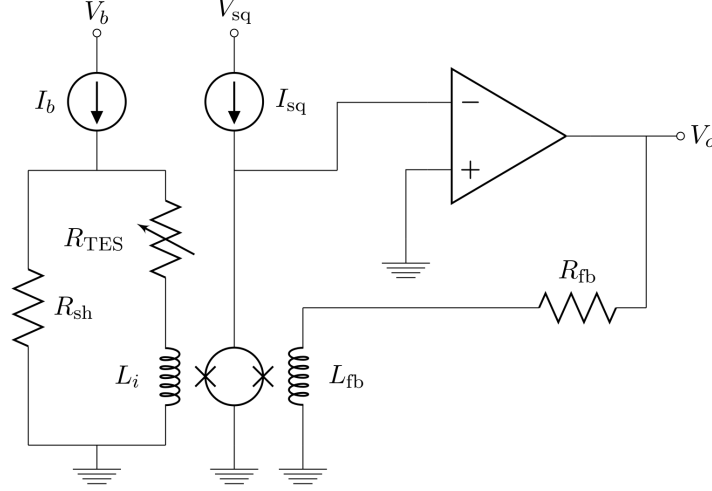


Figure 2.7.: Schematic of the TES circuit where the TES is voltage-biased with a fixed current, as shown on the left. When energy is deposited in the TES, its resistance R_{TES} increases, which leads to a change in the current through the inductor L_i . This change in the current is measured by a SQUID circuit coupled with an amplifier to read out the phonon signal. The figure is taken from Ref. [95].

The total phonon energy E_{pt} from an interaction is a sum of the recoil energy E_r and the energy gained via the NTL effect E_{NTL} .

$$E_{pt} = E_r + E_{NTL}. \quad (2.1)$$

For an e^-h^+ pair drifting across the detector, the energy produced as NTL phonons is equal to $e \cdot V_{bias}$, where e is the elementary charge and V_{bias} is the voltage bias applied across the detector. This is assuming that the charges transverse through the full electric potential and are not trapped in defects or impurities [98]. The total phonon energy can be written as:

$$E_{pt} = E_r + n_{eh} \cdot e \cdot V_{bias}, \quad (2.2)$$

where n_{eh} is the number e^-h^+ pairs produced. This shows how a small energy deposition can be amplified by applying a high bias voltage across the detector, utilizing the NTL effect. This is the motivation for the CDMSlite operation of the iZIP detectors during the SuperCDMS Soudan experiment and the HV detectors for the SuperCDMS SNOLAB experiment.

The number of e^-h^+ pairs n_{eh} produced is proportional to the recoil energy. For a given recoil energy, n_{eh} is a distribution with a mean $\langle n_{eh} \rangle$ and a width determined by the Fano factor parameter [99]. The mean number of e^-h^+ pairs produced for a given

recoil energy E_r is given by

$$\langle n_{e/h} \rangle = Y(E_r) \frac{E_r}{\epsilon_{eh}}, \quad (2.3)$$

where $Y(E_r)$ is the ionization yield and, ϵ_{eh} is the average energy required to produce an e^-h^+ pair. ϵ_{eh} for Si is ~ 3.8 eV and for Ge it is ~ 3.0 eV [88]. The ionization yield $Y(E_r)$ describes how much of the recoil energy is converted into e^-h^+ pair production. For an Electron Recoil (ER), all of the recoil energy is used for producing e^-h^+ pairs, thus $Y(E_r) = 1$. A Nuclear Recoils (NR) event produces fewer e^-h^+ pairs in comparison with an ER event. Thus, the yield for NR events is $Y(E_r) < 1$ and is characterized based on the Lindhard theory [100, 101, 102]. Using Eq. 2.3 and Eq. 2.2, the total phonon energy is written as

$$E_{pt} = E_r \left(1 + Y(E_r) \frac{eV_{bias}}{\epsilon_{eh}} \right). \quad (2.4)$$

2.2.4. Ionization measurement

As this thesis focuses on the CDMSlite dataset, where only the measurements from the phonon channels were recorded and the charge channels were used for applying the high bias voltage, only a brief summary of the ionization measurement is provided here. For a comprehensive account of ionization measurements, refer to Refs. [79, 95].

A schematic of the charge readout setup is illustrated in Fig. 2.8. In this setup, C_d represents the detector, where drifting charges induce charge. The charges are collected into the feedback capacitor C_f in the feedback loop, creating a voltage spike (V_0). The feedback loop consists of an amplifier, capacitor C_f , and resistor R_f . The capacitor C_f drains through R_f with a time constant $C_f \cdot R_f$. The coupling capacitor (C_c) protects the feedback loop from the biasing source. The bias resistor R_b prevents the induced current from draining into the bias source. C_s represents any stray capacitance.

Charges propagate through the detector on nanosecond time scales. With the digitization rate of 1.25 MHz (for the SuperCDMS Soudan experiment), the rising edge is instantaneous, and the fall time is controlled by the feedback loop. The amplitude, which is the only measurement of the ionization, is proportional to the collected charge or the energy generated. Considering 100% charge collection efficiency, the measured ionization energy E_Q is given by

$$E_Q = n_{eh} \cdot \epsilon_{eh}. \quad (2.5)$$

As previously mentioned, for NR events, only a fraction of the recoil energy E_r is converted to producing e^-h^+ pairs, which is determined by the ionization yield $Y(E_r)$. Consequently, the measured ionization energy is written as

$$E_Q = Y(E_r) \cdot E_r. \quad (2.6)$$

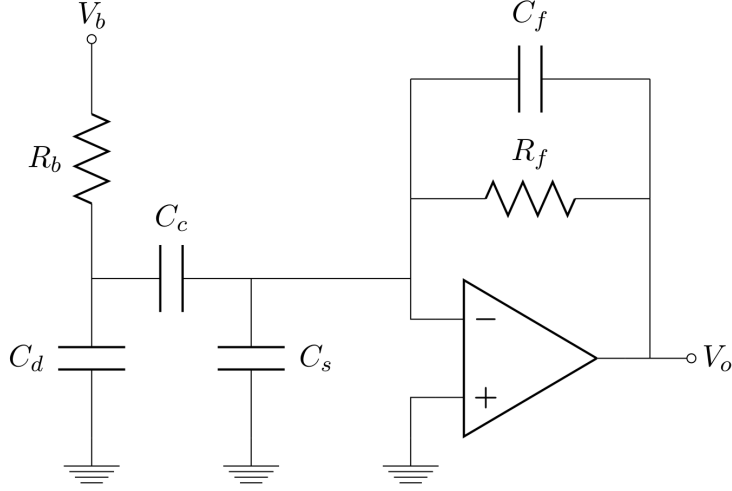


Figure 2.8.: A schematic of the SuperCDMS Soudan charge readout circuit. A bias voltage source V_b is connected to the detector represented by C_d through the biasing resistor R_b . The feedback loop consists of the amplifier, feedback resistor R_f , and feedback capacitor C_f . C_c and C_s are the coupling capacitor, which protects the feedback loop, and stray capacitance in the circuit, receptively. The figure is taken from Ref. [95].

For ER events, $Y(E_r) = 1$, as all of the recoil energy is converted to producing e^-h^+ pairs. Therefore, for E_r events, $E_Q = E_r$.

Yield discrimination: The ionization yield term was introduced to relate the recoil energy with the phonon energy. Rewriting Eq. 2.4 using Eq. 2.6, the ionization yield is written as

$$Y(E_r) = \frac{E_Q}{E_{pt} - \frac{eV_b}{\epsilon_{eh}} E_Q}. \quad (2.7)$$

The values of E_Q and E_{pt} are measured from charge and phonon channels. As the iZIP detectors have both phonon and charge channels, event-by-event ER/NR discrimination is possible, enabling the removal of ER backgrounds. However, for CDMSlite and SuperCDMS SNOLAB HV detectors, where charge channel signals are absent, ER/NR discrimination is not possible.

2.3. SuperCDMS at Soudan

The SuperCDMS Soudan experiment was conducted at the Soudan Underground Laboratory in Minnesota, located 713 m underground (2090 m.w.e.) [85, 86]. Operated between 2012 and 2015, the experiment used five towers of iZIP detectors with a total payload of approximately ~ 9 kg. Each tower consisted of three Ge detectors,

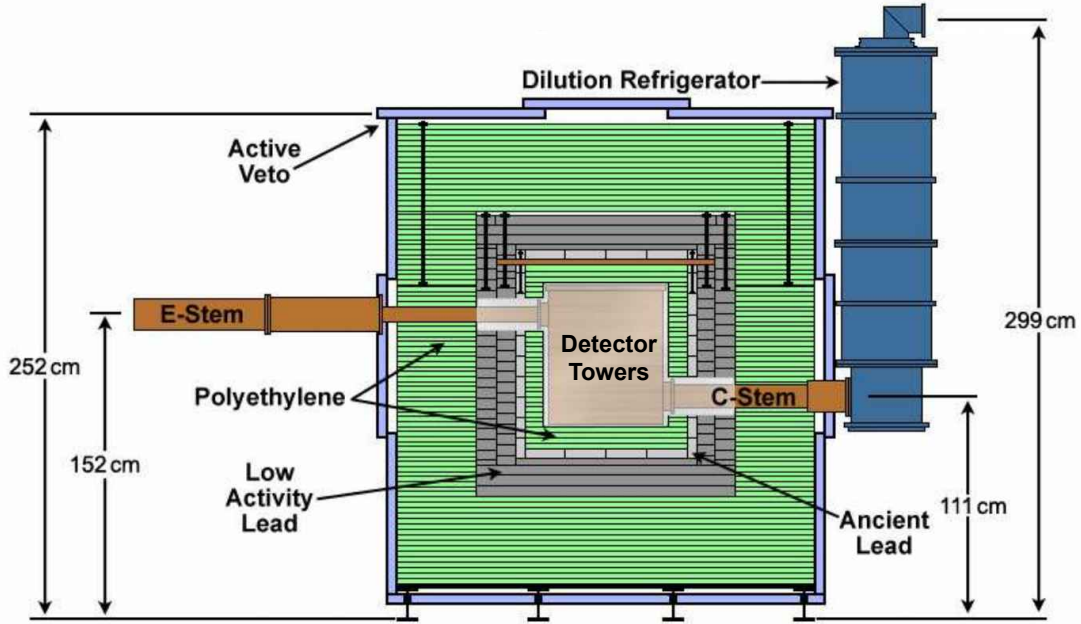


Figure 2.9.: Schematic of the SuperCDMS Soudan experiment, illustrating the shielding and veto setup. The detector towers are housed within the icebox, surrounded by multiple shielding layers, including polyethylene and lead, enclosed within muon veto panels made of plastic scintillator. The detectors are cooled using a dilution refrigerator connected to the icebox via the C-stem. The electronics are connected to the detector towers through the E-stem. Figures courtesy of J. Sander.

each with a diameter of ~ 76 mm, a height of ~ 25 mm, and a mass of ~ 600 g. This section provides an overview of the experimental infrastructure at Soudan, including its shielding system, cryogenic refrigeration setup, and trigger and Data Acquisition (DAQ) systems. Information about the experiment is obtained from Refs. [103, 104] and references therein.

The experimental setup with the shielding layout is shown in Fig. 2.9. A six-stage Kelvinox 400-S ^3He - ^4He dilution refrigerator was used to cool the detectors from room temperature. The stages include 300 K (room temperature), 77 K (nitrogen shield), 4 K (helium bath), 600 mK (still), 50 mK (cold plate), and 10 mK (mixing chamber). The detector towers were housed inside the “icebox” consisting of six copper cans. Each of the thermal stages of the dilution refrigerator was connected to the corresponding copper housing can via the C-stem. On the other side of the C-stem opening in the cans is the E-stem, which connects the detectors to the electronics. The electronics were supplemented with cooling from a Gifford-McMahon cryocooler, which was coupled to

the 4K and 77K stages. The cryocooler had a piston used to expand and contract helium gas, generating vibrations that contribute to Low-Frequency Noise (LFN), the subject of this thesis.

The backgrounds in the SuperCDMS Soudan experiment arose from cavern rock radioactivity, internal radioactive decay from the apparatus, and from cosmic rays. To mitigate these backgrounds, effective shielding was crucial. An active muon veto, consisting of forty overlapping 5 cm thick panels of BC-408 plastic scintillator connected to photomultiplier tubes (PMTs), formed the outer shielding layer, effectively removing events coincident with muons passing through. The outer layer of neutron shielding consisted of 40 cm thick polyethylene bricks. For gamma shielding, two sequential layers of lead, totaling 22.5 cm in thickness, were used, with the inner 4.5 cm made from ancient lead. Lead is typically contaminated with radiogenically produced ^{210}Pb , whose decay produces gamma radiation. The ancient lead within the natural lead shield helped reduce this background. An additional 10 cm thick inner layer of polyethylene was placed between the lead shield and the cryostat icebox, providing protection against neutrons generated in the outer shields. Magnetic field protection was provided by a 0.381 mm thick mu-metal (81% Ni, 19% Fe) shield inside the inner polyethylene layer. The gap between the mu-metal and cryostat was flushed with dry nitrogen to avoid radon (^{222}Rn) contamination. The cryostat that houses the detector tower consists of copper cans that were made of radio-pure copper, providing γ and β shielding.

Even though most of the gammas and neutrons are mitigated by the shielding, the experiment still has backgrounds. The primary source of background after shielding is from the radioactivity of the shielding and other materials used in the apparatus. Additionally, cosmogenic activation produces significant background. Cosmic rays induce spallation, leading to the activation of both detector and apparatus materials during fabrication and handling. For Ge detectors, background are from tritium contamination as well as other isotopes that primarily decay through β -decay or electron capture. Some of these cosmogenic isotopes, which decay via electron capture, are present in the measured energy spectrum. Neutrons from the ^{252}Cf calibration source can be captured by ^{70}Ge , producing ^{71}Ge . With electron capture, ^{71}Ge produces distinct peaks in the energy spectrum [105]. Although these isotopes contribute to an irreducible background, they are useful for calibration. Other sources of background include Compton scattering and radon exposure, which produce surface events.

The trigger system for the detectors required signals to exceed predefined thresholds, which were set for charge and phonon channels. Each detector had two trigger control boards, one for each side. A trigger control boards combined signals from four phonon channels and the two charge channels. Based on a logical OR to determine if the trigger condition was met, the board issues a trigger. When one of the detectors is triggered, the data from all detectors is saved, with the triggered detector's data aligned to the global trigger time bin. The stored event data contained

traces of 4096 time samples, with the bin that issued the trigger set at the 500th time bin.

The SuperCDMS Soudan experiment collected various data types to support WIMP searches and detector calibration. Low-background data, primarily used for the WIMP search, is the main science dataset. For calibration, two radioactive sources were used: ^{133}Ba and ^{252}Cf which are ER and NR sources, respectively. ^{133}Ba decays to ^{133}Cs , producing β and γ radiation, with energy peaks at 356.0 keV, 81.0 keV, 302.8 keV, and 383.8 keV [106]. ^{252}Cf mainly decays via α emissions, but with $\sim 3\%$ probability it undergoes a spontaneous fission process which produces 3-4 neutrons per day in energies of in the range of MeV [107].

2.4. SuperCDMS at SNOLAB

SNOLAB is located in the Creighton nickel mine, operated by Vale mining company in Sudbury, Canada. The lab is located 2 km underground (6010 m.w.e.), which reduces cosmic muon flux by an additional factor of 2.5-3 compared to the Soudan levels, eliminating the need for a muon veto. The entire laboratory is a class-2000 cleanroom standard to minimize intrinsic radiogenic backgrounds.

The schematic of the experiment, along with its shielding layout, is shown in Fig. 2.10. The shielding consists of passive layers: a 60 cm water shield protects the top and sides, while the bottom is shielded by 60 cm of High-Density Polyethylene (HDPE) plates. This protects the experiment from the cavern neutrons. Inside the HDPE layers is a radon purge barrier, constantly flushed with nitrogen, to mitigate atmospheric ^{222}Rn contamination. Inside the purge barrier, a 23 cm layer of low-activity lead protects against gammas, inside which is a layer of ancient lead. The inner neutron shield is made of 40 cm HDPE to shield against neutrons. The detector towers are housed within a vacuum-sealed cryostat container, referred to as SNOBOX, consisting of six copper cans. The operating temperature of the detectors is $\sim 15\text{ mK}$ and is obtained by a dilution refrigerator. These cans are thermally coupled to the thermal stages of the dilution refrigerator via bars through the C-stem. The entire shielding system is mounted on a seismic platform to absorb seismic and mine-related vibrations. Electronics connected to the 4 K stage pass through the E-stem from the detectors out of the SNOBOX.

Despite the shielding and the low-background environment at SNOLAB, the SuperCDMS experiment still has backgrounds that need to be addressed. Key ER backgrounds include β -decay from cosmogenic tritium (^3H) contamination in detectors, γ and β particles from radioactive isotopes in surrounding materials, and decay products from cosmic-ray activation. For silicon detectors, ^{32}Si β -decay is a significant contributor. For germanium detectors, the major background contribution comes from the activation lines from electron capture [105]. Other backgrounds include coherent neutrino-nucleus scattering, α -decay from contaminants, and neutrons produced cosmogenically or in the cavern environment.

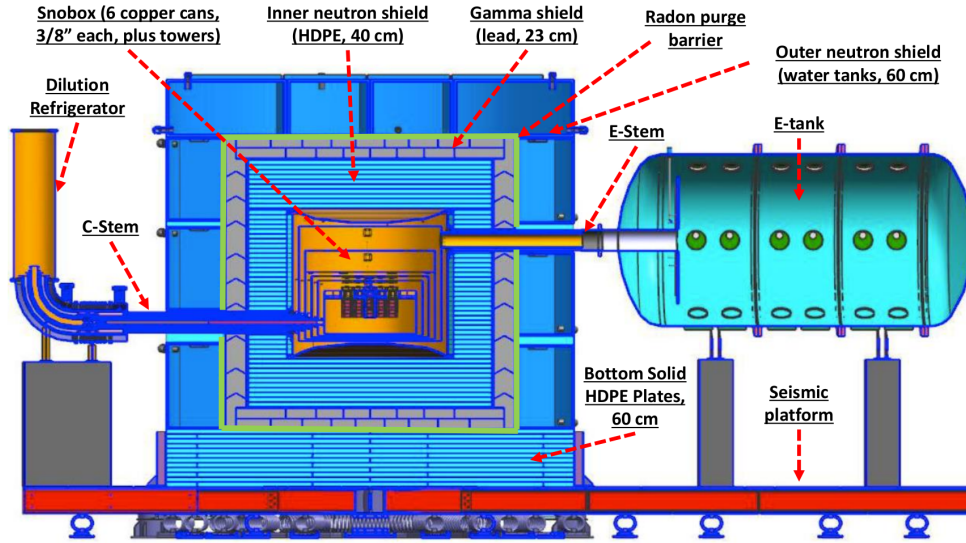


Figure 2.10.: Schematic of the SuperCDMS SNOLAB experiment setup. The detector towers are housed within the SNOBOX, which consists of six copper cans. The SNOBOX is surrounded by multiple shielding layers, including HDPE and lead, and is enclosed within a water tank. The detectors are cooled down by a dilution refrigeration system, which is connected to the SNOBOX via the C-stem. The electronics for the experiment are housed in the E-tank and are connected to the detectors through the E-stem. The entire assembly is mounted on a seismic platform to mitigate the impact of seismic activity on the setup. Figure taken from Ref. [88].

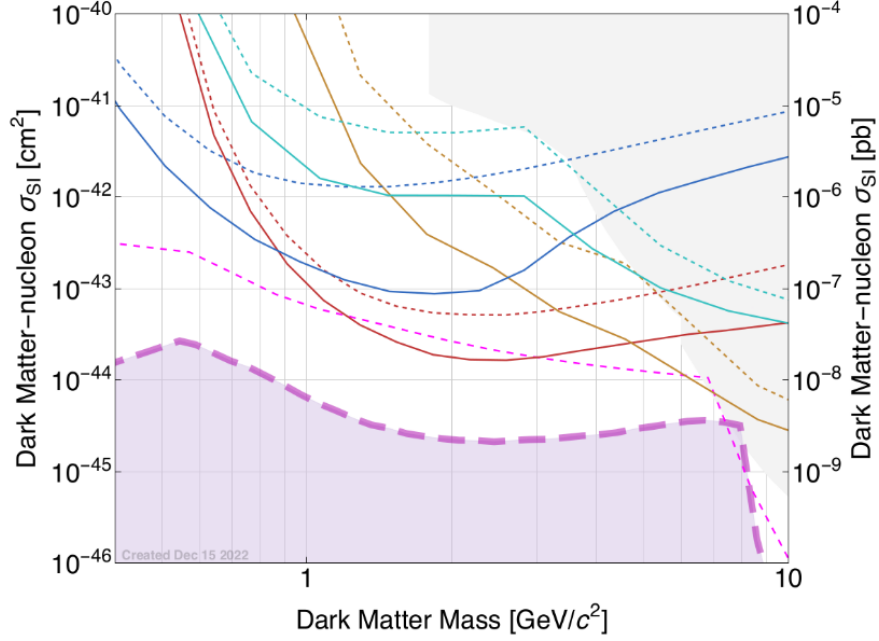


Figure 2.11.: Projected sensitivities of SuperCDMS SNOLAB experiment to spin-independent (SI) WIMP-nucleon scattering using different methods: Optimum Interval (dashed), and Profile-Likelihood Ratio (solid). The sensitivity is calculated for various detector configurations: Ge HV (red-brown), Ge iZIP (mustard), Si HV (blue), and Si iZIP (cyan). The neutrino fog [77] is shown in magenta (long dashed and shaded), and the exclusion limits as of Dec 15, 2022 are shown in the gray shaded region. The plot is taken from Ref. [108].

The SuperCDMS SNOLAB experiment is expected to begin commissioning in late 2025. The initial payload will consist of four towers, each with six detectors. The payload comprises of two HV towers and two iZIP towers, totaling 10 Ge iZIPs, 2 Si iZIPs, 8 Ge HV, and 4 Si HV detectors. Currently, all four detector towers are already in the SNOLAB and stored in the SuperCDMS clean room. The projected sensitivity of the SuperCDMS SNOLAB experiment is shown in Fig. 2.11. One of the HV towers, containing 4 Ge and 2 Si detectors, was successfully tested at the Cryogenic Underground Test (CUTE) facility at SNOLAB between 2024-2025, and analysis of this data is underway.

3. CDMSlite Run 3

During the final phase of SuperCDMS Soudan experiment, an iZIP detector was operated at high bias voltage as a part of the Cryogenic Dark Matter Search Low Ionization Threshold Experiment (CDMSlite). The high voltage application amplifies the ionization signal via NTL effect explained in Sec. 2.2. The amplification enables searches for low-mass DM particles by boosting sub-keV nuclear recoil energy deposition to surpass the threshold. With an energy threshold of 70 eV and a live-time of 36.9 kg-days, the experiment excluded the DM-nucleon cross section for DM masses up to $2 \text{ GeV}/c^2$. In this chapter, the experimental setup and an overview of the analysis on which this thesis work is based are presented. The content of this chapter is adapted from Refs. [76, 109, 110].

3.1. Description of the experiment

The experiment was conducted from February to May 2015, with a total live-time of 60.9 days. For this experiment, the top detector in the second tower (T2Z1) was selected because of its operational stability when applying high voltage and its better baseline resolution due to its reduced susceptibility to instrumental noise which enabled the low threshold analysis. During this run, the detector had a voltage bias of 75 V applied to one side while the other side was grounded for the readout. The schematic of the detector with the phonon sensor layout is shown in Fig. 3.1. Waveforms from the charge collecting electrodes were also recorded but were only used to remove events with bad noise in the charge channel.

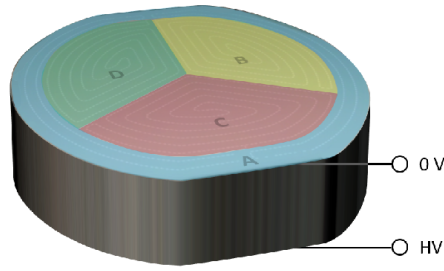


Figure 3.1.: Schematic of the CDMSlite detector showing the layout and labeling of phonon channels. In this configuration, the high voltage (HV) was applied to the bottom side, and the phonon channels from the top side were readout.

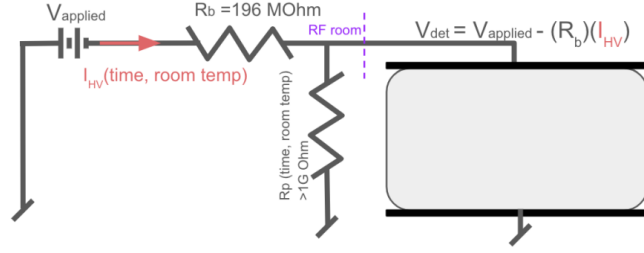


Figure 3.2.: Schematic of the HV circuit illustrating the applied voltage V_{applied} and the detector voltage V_{det} , which differs from the applied voltage due to parasitic resistance around the detector that draws some current.

Pre-biasing: A pre-biasing procedure was performed in which the detector voltage bias was temporarily increased to 85 V for 10 minutes prior to the start of collecting each data series where the operating voltage was set to 75 V. This procedure helped to mitigate charge leakage and short noise.

High voltage monitoring: The determination of detector bias voltage was crucial for determining the energy scale and energy reconstruction, and it depended on the bias current I_{HV} which was measured to monitor detector voltage. The schematic of the HV supply circuit is shown in Fig. 3.2. The voltage at the detector V_{det} was different from the applied voltage V_{applied} due to the parasitic resistance R_p which caused a voltage drop across a bias resistor R_b . The voltage at the detector is given by

$$V_{\text{det}} = V_{\text{applied}} - I_{HV} R_b. \quad (3.1)$$

where I_{HV} is the nominal current which should be zero due to infinite detector resistance, but the parasitic resistance R_p introduced current in the nanoampere (nA) range. The bias resistor R_b had a resistance of 196 M Ω and protected the cryostat from excessive current flow but reduced the voltage applied to the detector.

During the experimental run, R_p fluctuated over time and was correlated with the temperature fluctuations in the room that housed the electronics. In April 2015 modifications were made in the RF room ventilation, impacting the temperature of the room which consequently increased the parasitic resistance while reducing leakage current which stabilized the detector at 75 V. Prior to this modification, the detector voltage drifted between 50 and 70 V, which resulted in variations in the energy corrections. To address this, temperature corrections were applied, as explained in Sec. 3.3.3.

When the detector voltage was above >72 V, the trigger rate increased, as shown in Fig. 3.3. This was caused by the increased baseline noise. After the operational voltage was increased to 75 V on April 1st to stabilize the detector, the noise performance deteriorated, indicating that the optimal operating voltage was slightly less than 75 V.

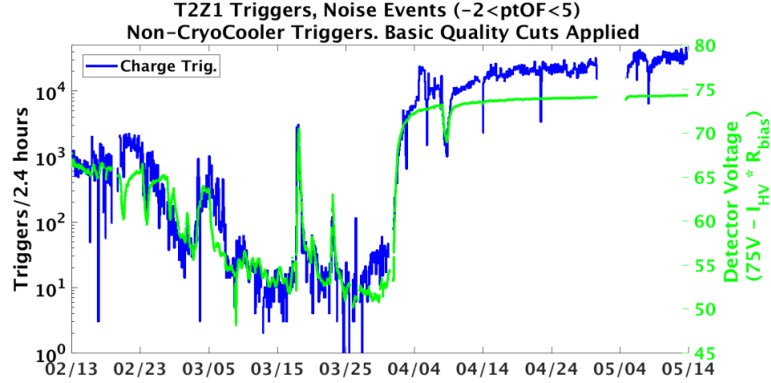


Figure 3.3.: Histogram showing the correlation between the trigger rate and the detector voltage. The charge triggers and detector voltage are shown in blue and green, respectively. The events are constrained between the energy range between -2 keV and 5 keV. The plot is taken from [109].

Consequently, the experimental run data was divided into periods before and after April 1st, named periods A and B, respectively. The hardware trigger threshold also varied between 50 and 70 eV during the course of the run.

Vibration monitoring: The broadband increase of noise was observed in low-frequency regions in the PSDs of the phonon signals. This noise surge originated from vibrations generated by the cryocooler used for cooling the electronics. These vibrations detrimentally impacted the detector baseline, which also led to more frequent triggering. Accelerometers were installed to time the vibration and found that the high event rate corresponded with vibration from the cryocooler. The vibration-induced noise, referred to as Low-Frequency Noise (LFN), caused more triggers compared to any other background source. Section 3.4.3 describes the method used to remove the LFN for this analysis. The motivation for this thesis work is to effectively eliminate this noise while preserving signal efficiency.

3.2. Data processing

The Data Acquisition (DAQ) issued a trigger when the waveform height crosses a predefined threshold. When a trigger was issued, the raw waveform was digitized with an Analog-to-Digital Converter (ADC) and stored. The digitized waveform was stored with 4096 samples per event with the trigger bin set at the 500th bin. Physics triggers occurred when the pulse height exceeded the threshold, while random triggers were issued by the DAQ at random times, saving data regardless of the energy deposited. These random triggered data were primarily noise, but occasionally also contained signal. Fig. 3.4 illustrates examples of randomly triggered waveforms and signal-containing waveforms.

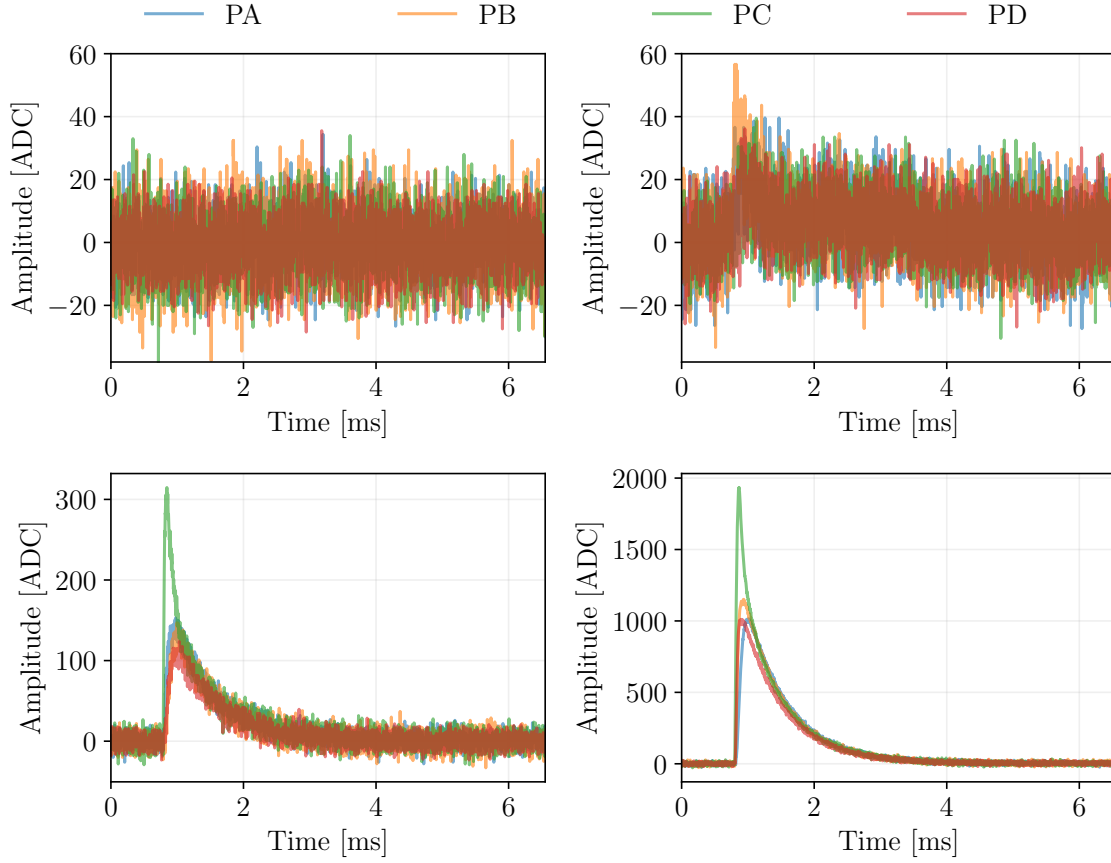


Figure 3.4.: Raw traces from the detector phonon channels PA, PB, PC, and PD for various events. Top left: a baseline noise event which was randomly triggered. Top right: one of the low-energy events that barely passed the trigger threshold. Bottom left: an event from the L-shell peak from ^{70}Ge activation. Bottom right: an event from the K-shell peak from ^{70}Ge activation.

The waveforms were then processed to determine the pulse amplitude which translates to the deposited energy. Throughout this work, the waveforms are referred to as “traces”.

Optimal Filter (OF)

Optimal Filter (OF) algorithms [111, 112, 113] determines the pulse amplitude by fitting a template to a measured trace and de-weighting the noise frequency components while maximizing the signal-to-noise ratio. This section is based on information from Ref. [79]. A measured event $S(t)$ is assumed to be the sum of an expected signal $A(t)$ (also called the signal template) scaled by an amplitude a and Gaussian noise $n(t)$:

$$S(t) = a \cdot A(t) + n(t). \quad (3.2)$$

The fit is optimized by minimizing the goodness of fit χ^2 . The optimum filtering algorithm performs the fitting in the frequency domain, the χ^2 is given by:

$$\chi^2(a) = \int_{-\infty}^{\infty} \frac{|S(f) - a A(f)|^2}{J(f)} df, \quad (3.3)$$

where $J(f)$ is the Power Spectral Density at each frequency. In the frequency domain, the OF determines the amplitude for which χ^2 in Eq. 3.3 is the lowest by de-weighting the noise frequencies with the knowledge of power spectral density.

The standard OF assumes that the measured trace and signal template are aligned in time and does not allow for time shifts of the template. This version is referred to as OF0 in this thesis and in general the SuperCDMS processing context. The OF0 estimates amplitude at a constant time bin and is typically used for estimating the noise amplitudes in a dataset. When one considers the signal template starts at $t=0$, and a measured trace with a shifted time by t_0 takes the form:

$$A(f) \rightarrow e^{-i2\pi f t_0} A(f). \quad (3.4)$$

Allowing for a time offset, the best amplitude and time shift are found by minimizing χ^2 :

$$\chi^2(a, t_0) = \int_{-\infty}^{\infty} \frac{|S(f) - a e^{-i2\pi f t_0} A(f)|^2}{J(f)} df. \quad (3.5)$$

In this thesis and in the general SuperCDMS processing context, this version of the Optimal Filter is referred to simply as OF, without the “0” used in the previous case.

The standard OF procedure assumes a constant pulse shape, however signal pulses vary in shape depending on the event location in the detector volume. For instance, if an event occurs near a phonon channel, the channel’s pulse shape will exhibit a “peaky” rise time compared to other channels. Therefore, each channel’s pulse shape varies based on the event’s position in the detector. Thus the phonon template is divided into fast and slow components - the fast component is position-sensitive, while the

slow component measures energy. The OF is performed using both the fast and slow templates and is referred to as 2TOF(2-Template Optimal Filtering). In addition to improvement in the energy reconstruction using 2TOF, the best-fit fast amplitude was used to apply a correction to the energy estimator which based on the slow template. This is detailed in Sec. 3.3.3. The fast and slow template for the CDMSlite Run 3 data analysis is showed in Fig. 3.5.

So far, the assumption has been that the noise in the traces is stationary, but the observed data exhibited time-dependent noise, so the variability of noise shape over time needed to be addressed. The Non-Stationary Optimal Filter (NSOF) generalizes the optimal filter to account for this time-dependent noise behavior. The initial part of the pulse can vary in shape depending on the event’s location in the detector, biasing the OF energy estimate. This variability in the shape of the first few bins, exhibited as “peakiness”, mainly arising from the TES layout sensitive to position. The NSOF treats positional dependence as time-dependent noise, effectively de-weighting the first 200 μs of the pulse for better template-signal agreement. This is achieved by first creating a template from averaging multiple pulses with the same start time and subtracting this template from the traces to obtain residuals. These residuals are treated as time-dependent noise. The χ^2 for NSOF is written as:

$$\chi^2 = \sum_{ff'} \left[S^*(f) e^{i2\pi f t_0} - a A^*(f) \right] \left(V(f, f')^{-1} \right) \left[S(f') e^{-i2\pi f' t_0} - a A(f') \right], \quad (3.6)$$

where $V(f, f')^{-1}$ is the non-diagonal covariance matrix. The NSOF provided better energy estimation compared to the standard OF.

Template generation:

Developing precise phonon templates for the OF/NF/2T algorithms plays a crucial role in the accuracy of energy and position estimators. The slow template for OF, NF, and 2T-slow algorithms was generated by normalizing the traces, followed by the removal of outliers from the χ^2 distribution. This process was iterative and involved evaluating the K-shell peak width of the ^{70}Ge activation at each step. The slow template (or standard OF template) was generated by averaging many high-quality pulses. The 2T-fast template is obtained by subtracting the slow template from the traces, taking the absolute value of the result, and then averaging them. This provides insights into the pulse peakiness, thereby offering positional information crucial for fiducial volume cuts. The templates for the CDMSlite Run 3 are shown in Fig. 3.5. An example of the OF fit applied to a raw trace is shown in Fig. 3.7 with the fit results displayed in the text box. The standard OF with time delay using the OF (slow) template was used for the fit. The χ^2 value in the fit results indicates a good fit performance.

Noise PSDs:

Noise PSDs were generated using randomly triggered events, with traces containing signal pulses removed before calculating the PSDs for each data series. Fig. 3.6 shows

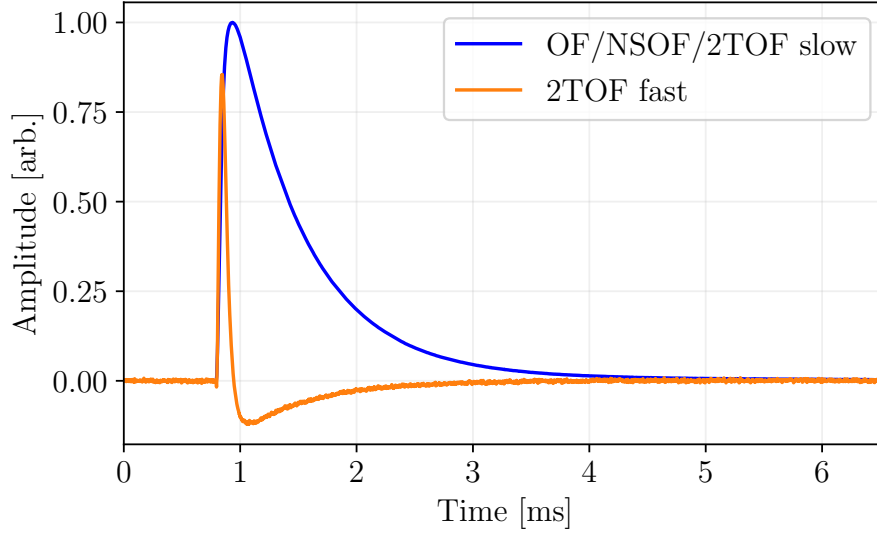


Figure 3.5.: The standard phonon templates utilized for the OF algorithms: Standard OF, NSOF, and 2TOF (slow), alongside the fast template used for 2TOF (fast). The slow template is employed initially for energy estimation, while the fast template is used to refine the determination of the event position.

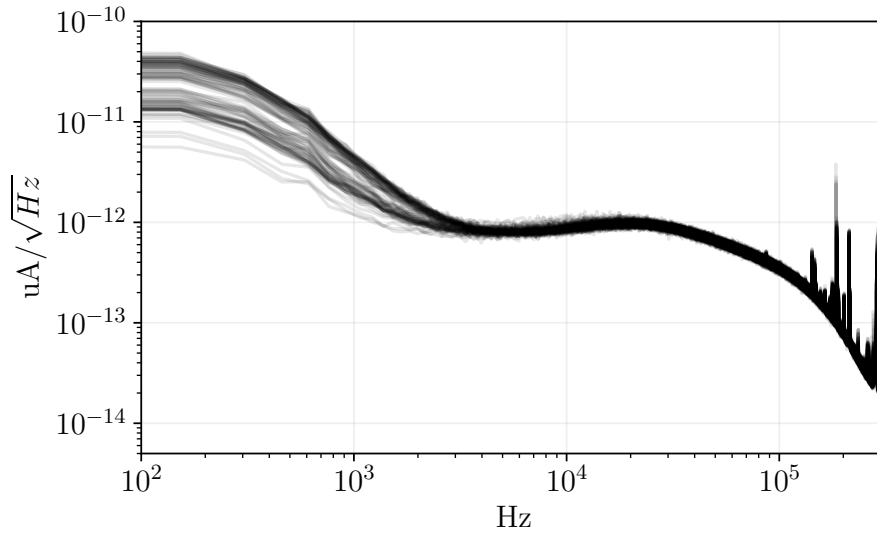


Figure 3.6.: Power Spectral Densities (PSDs) of various noise events randomly picked throughout the run, displaying the amplitude of noise with respect to the frequencies.

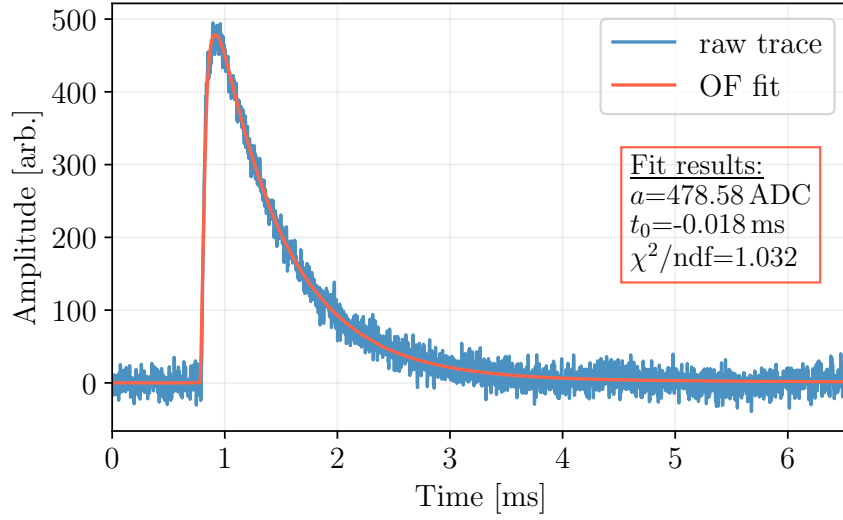


Figure 3.7.: An example of the OF fit: the raw trace is shown in blue, and the OF fitted shown in orange. The fit results are displayed in the orange text box, where a is the amplitude, t_0 is the time delay relative to the template, χ^2/ndof is the reduced χ^2 of the OF fit.

selected PSDs from the CDMSlite Run3, with noise amplitude on the y-axis and frequency on the x-axis.

The processing of raw traces from each channel included fitting the traces with OF0, OF, and 2TOF. In addition to this, the amplitude estimation using OF0, OF, 2TOF, and NSOF was performed for the PT channel. The fit results for individual channels and PT are stored as Reduced Quantities (RQs). Apart from the calculated OF quantities, some important information about the event such as trigger information, event timing, HV, and base temperature of the setup were stored as RQs.

3.3. Energy reconstruction

This section details how the Optimal Filter (OF) amplitude of the total phonon pulse, which is the sum of the four phonon channels, is converted to the measured energy. Section 3.3.1 describes the energy scale used in the data analysis. The conversion of the OF amplitude to energy using the ^{71}Ge electron-capture events is explained in Sec. 3.3.2. Corrections for environmental and detector conditions after applying the calibration factor are presented in Sec. 3.3.3.

3.3.1. Energy scale

The “electron-equivalent” energy units (eV_{ee}) is the electron recoil energy required to produce the observed phonon energy in the detector. keV_{ee} is a convenient unit for the CDMSlite analysis due to the predominance of Electron Recoil (ER) events in the observed backgrounds. To convert the recoil energy in Eq. 2.2 from electron-equivalent units to Nuclear Recoils (NR) equivalent units $E_{r,nr}$, the difference in the ionization yield for NR and ER recoils needs to be accounted for.

$$E_{r,nr} = E_{r,ee} \left(\frac{1 + e\Delta V/\epsilon_{eh}}{1 + Y(E_{r,nr})eV_{det}/\epsilon_{eh}} \right), \quad (3.7)$$

where e is the elementary charge, ϵ_{eh} is the average energy required to generate a single electron-hole pair; for Germanium $\epsilon_{eh} = 3\text{ eV}$ [22]. The yield for ER is 1 as all of the deposited energy is converted into producing electron-hole pairs. For NR events, the yield as a function of NR energy using the Lindhard model [114] is given by

$$Y(E_{r,nr}) = \frac{k \cdot g(\epsilon)}{1 + kg(\epsilon)}, \quad (3.8)$$

where $g(\epsilon) = 3\epsilon^{0.15} + 0.7\epsilon^{0.6} + \epsilon$, $\epsilon = 11.5 \cdot E_{r,nr}(\text{keV})Z^{-7/3}$, Z is the atomic number of the detector material. For this analysis, $k = 0.157$ from the Lindhard model prediction was used, which is consistent with yield measurements in Ge [76, 115, 116], and the spread in k was considered as a systematic uncertainty.

3.3.2. Energy calibration

The energy calibration of the detector is done using a ^{252}Cf neutron source that primarily induces nuclear recoils. The activation of ^{70}Ge by neutron capture produces ^{71}Ge , which decays to ^{70}Ge (plus ν_e) by electron capture with a 11.43 day half-life [105]. ^{71}Ga in turn decays to gammas and electrons. The reaction is $n + ^{70}\text{Ge} \rightarrow ^{71}\text{Ge}^*$, $^{71}\text{Ge}^* + e^- \rightarrow ^{71}\text{Ga} + \gamma + e^-$. These decays produce peaks at the K-, L-, and M -shell binding energies of ^{71}Ga at 10.37, 1.30, and 0.16 keV, respectively. The most prominent of these electron capture peaks is the K shell peak which was used for calibrating the energy scale and to monitor changes in the detector response with time. The corrections and calibration determined with the K shell peak were consistent with the less prominent L and M shell peaks, confirming detector response linearity across the energy range [76].

3.3.3. Energy Corrections

An initial calibration was performed to convert the uncalibrated energy estimator from OF amplitudes to energies in eV units. The final energy spectrum is determined after applying several corrections to account for factors such as bias voltage, base temperature, and event position within the detector. The energy corrections were applied sequentially: first, corrections for HV current dependence, followed by adjustments for the base tem-

perature, and finally, the energy is adjusted by considering the position dependence of the event. These corrections are necessary to improve the energy resolution. This section details these energy corrections.

Current correction

Observation of the 10 keV activation line showed that the amplification drifted during the run and was caused by inconsistency in the voltage at the detector. By measuring the HV current I_{HV} , the detector voltage relative to the applied voltage can be found, assuming all leakage current is downstream of the bias resistor, R_b . This assumption was verified by observing that the measured phonon energy of electron recoil calibration peaks depends on the detector voltage in Eq. 3.1. The current correction procedure involves solving for the expected total phonon energy of the 10 keV calibration line using V_{det} . The measured total energy of electron recoil events is given by:

$$E_{pt} = E_{recoil} \cdot (1 + V_{det} \cdot e/\epsilon). \quad (3.9)$$

Upon solving for the expected total phonon energy of the 10 keV calibration line using a V_{det} corrected by the measured leakage current (I_{HV}), the expected total phonon energy of the 10 keV calibration line and the observed drift in the data were in good agreement.

Base temperature correction

Following the current correction, a linear relationship between energy and base temperature was observed, although the underlying cause remains unidentified. Subsequently, this dependency was removed through a straight line fit which improved the energy resolution.

$$E_{corr}(T_{base}) = E_{uncorr} \cdot (1 + (T_{base} - \text{mean}(T_{base})) \cdot C_{base}), \quad (3.10)$$

where $\text{mean}(T_{base})$ is the mean base temperature of the run and C_{base} is the slope of the aforementioned line fit.

Position correction

The fast template is indicative of the pulse peakiness. During CDMSlite operation, the position-dependent variation in pulse shape is diminished, as the NTL phonon signal, which is more collimated in the z-direction, masks the initial recoil phonons from which the positional information can be derived. There are still variations in pulse shape characterized by the peaky vs. non-peaky pulse shape correlating with the proximity of the initial recoil to the phonon channels on the detector. Following the base temperature correction, a positive linear relationship between energy and the fast template amplitude was observed. Thus the following correction based on a linear fit was performed:

$$E_{corr}(2T_{fast}) = E_{uncorr} \cdot (1 + (2T_{fast} - \text{mean}(2T_{fast})) \cdot C_{2T_{fast}}), \quad (3.11)$$

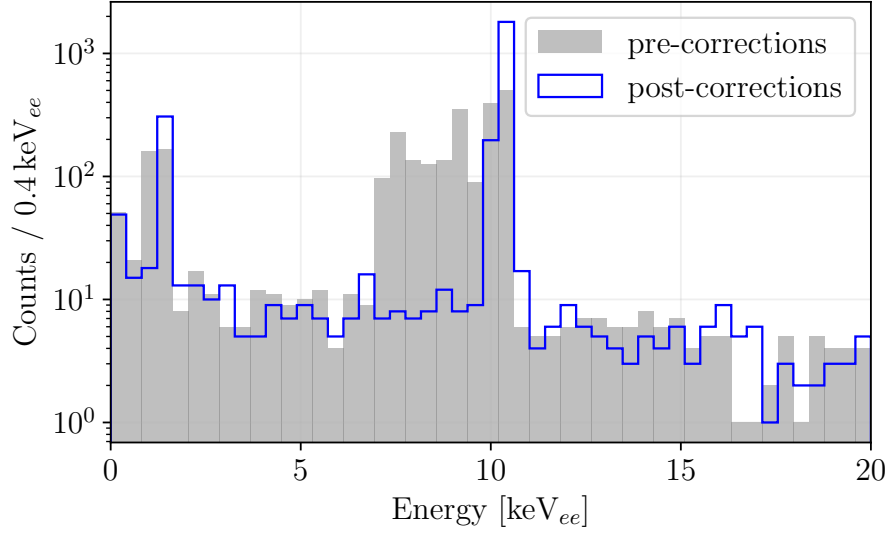


Figure 3.8.: Histogram comparing the energy distribution before and after the energy corrections. Both histograms show events that passed the quality cuts described in 3.4. The corrections were applied to address environmental factors such as detector voltage variations, base temperature fluctuations, positional effects, and the NR energy scale correction. The plot illustrates the impact of these corrections on the dataset.

where $\text{mean}(2T_{fast})$ is the fast template amplitude of 10 keV events during the run and $C_{2T_{fast}}$ is the slope of the aforementioned linear fit.

NR energy scale correction

For converting the total phonon energy (E_{pt}) to nuclear recoil energy, one needs correction terms due to the variable detector voltage V_{det} in the run. The HV current corrections were applied to electronic recoils (ER) which have an ionization yield of 1, while nuclear recoils (NR) require consideration of the yield $Y(E_{nr})$ for accurate conversion.

$$E_{pt} = E_{recoil} \cdot (1 + V_{det} \cdot Y(E_{nr}) \cdot e/\epsilon) \quad (3.12)$$

Not performing this correction could lead to errors up to 12% in the NR energy estimate. These energy corrections were applied to energies calculated from the OF, NSOF, and 2TOF estimators. After implementing these corrections and applying the data-quality cuts described in Sec. 3.4, the energy distribution of the CDMSlite Run 3, with and without corrections, is depicted in Fig. 3.8. The post-correction energy distribution reveals a more refined activation peak compared to the pre-corrected distribution.

3.3.4. Energy resolution model

The energy deposited in a detector is not always the energy measured, as various factors can affect the measured energy. Energy resolution, a measure of the detector's ability to accurately measure energy, reflects the uncertainty in these measured energies. A proper energy resolution is essential for accurately calculating the expected energy spectra of DM events.

The energy resolution of the detector is influenced by three primary components. First, the baseline resolution, which arises from the electronics and the experimental environment. Second, the discrete nature of charge production, governed by the Fano factor [117]. Lastly, the position dependence arising from different phonon channel responses, electric field distributions, etc. Therefore, the detector response model is parameterized as a function of the recoil energy (in electron-equivalent units) $E_{r,ee}$ such that total energy resolution $\sigma_T(E_{r,ee})$ is written as follows:

$$\sigma_T(E_{r,ee}) = \sqrt{\sigma_E^2 + \sigma_F^2(E_{r,ee}) + \sigma_{PD}^2(E_{r,ee})}, \quad (3.13)$$

where σ_E^2 is energy independent baseline resolution. σ_F is Fano resolution term which is proportional to $\sqrt{E_{r,ee}}$, and σ_{PD} is position dependence term, which is proportional to $E_{r,ee}$. With constants A and B, Eq. 3.13 can be rewritten using the three model parameters σ_E^2 , A and B [76]

$$\sigma_T(E_{r,ee}) = \sqrt{\sigma_E^2 + B \cdot E_{r,ee} + (A \cdot E_{r,ee})^2}. \quad (3.14)$$

Baseline resolution is found using the zero-delay-OF amplitude OF0 of baseline noise traces (called the zero energy resolution). It varied due to changes in the environmental conditions in April and the bias voltage discussed in Sec. 3.1. Thus, the values for the constants are separately calculated for the periods A and B as shown in Table 3.1.

Due to the prominence of the Ge activation peaks, the widths of these peaks were used to calculate the detector energy resolutions and the energy dependence of the resolution. After applying all the quality cuts explained in Sec. 3.4, the resolution was calculated by fitting a Gaussian with a linear background. The final parameters are shown in Table 3.2.

The best-fit parameters for the model in Eq. 3.14 is shown in Table 3.3. The coefficients A and B are consistent within uncertainties between the two periods, but the B parameter is larger than the predicted measurements of Fano factor that is 0.39 [76, 118, 119]. The uncertainties in this model were propagated as systematic uncertainties in the profile likelihood fit in Ref. [76], In this thesis work, they were incorporated when calculating the uncertainty on the final limits using the Optimum Interval (OI) method, as discussed in Sec. 5.4.

Baseline	Energy [keV _{ee}]	Resolution [eV _{ee}]
Period A	0.0	9.87±0.04
Period B	0.0	12.67±0.04

Table 3.1.: The baseline resolution of CDMSlite Run 3 for the period A and period B. Taken from Ref. [76].

Peak	Energy [keV _{ee}]	Resolution [eV _{ee}]
K shell	10.354±0.002	108±2.0
L shell	1.328±0.003	36.3±2.0
M shell	0.162±0.002	13.9±2.0

Table 3.2.: The resolution of reconstructed energies for the K, L, and M shell peaks from the activation of ⁷⁰Ge using ²⁵²Cf. Taken from Ref. [76].

	σ_E [eV _{ee}]	B [eV _{ee}]	A ($\times 10^3$)
Period A	9.87 ± 0.04	0.87 ± 0.12	4.94 ± 1.27
Period B	12.70 ± 0.04	0.80 ± 0.12	5.49 ± 1.13

Table 3.3.: The best-fit parameters of the energy resolution model in Eq. 3.14 for period A and B. Taken from Ref. [76].

3.4. Data-quality cuts

Data-quality cuts are essential to eliminate periods of undesirable detector behavior, instrumental noise, and poorly reconstructed events. These cuts fall into two categories: live-time cuts and efficiency cuts. Live-time cuts exclude events during bad time periods of the run, while efficiency cuts remove unwanted traces.

3.4.1. Live-time cuts

Live-time cuts are essential for eliminating periods of unsuitable conditions arising during an experimental run. These cuts ensure that only data from stable and reliable periods were included in the analysis, thereby improving the overall quality and accuracy of the results at the expense of total exposure. The following are the live-time cuts applied for the data analysis.

High voltage cut: During the operation of the T2Z1 detector, the detector voltage measured wasn’t always 75 V, especially during the beginning of this run and during pre-biasing. Thus this cut ensured that only data measured while the detector had a bias voltage of -75 V were analysed.

NuMI cut: This cut removed time periods coincident with neutrinos at the Main Injector (NuMI) beam at Fermilab within 200 μ s of a NuMI beam spill.

Pre-trigger baseline cut: The raw data has 4096 time bins, and the time bins before the 500th bin are defined to as “pre-pulse”. For each data series, the standard

deviation of the pre-pulse was fit to Gaussian. Events with pre-pulse variance larger than 4σ from the average variance of the data series were removed by this cut.

Trigger cuts: This cut removed time periods that encountered high trigger rates either due to anomalous detector behavior or due to poor noise performance from experimental conditions such as cold trap cleaning, following detector calibrations or when the threshold was set lower after the start of the run. Events not triggering the detector of interest T2Z1 were also removed.

Glitch cuts: Electronic noise caused glitch events characterized by a fast rise time and a fast fall time, which triggered multiple detectors. These events were identified by looking at the number of phonon and charge triggers for each event and the cut was applied on the number of phonon and charge triggers plane. Events with glitches in charge channels were removed by placing a flat cut on the charge χ^2 .

GPS cut: Events with inaccurate GPS time information, where the difference between the GPS time and the event time exceeded 17 s, were removed by this cut.

3.4.2. Pulse shape cuts

Pulse shape cuts are designed to eliminate events characterized by poorly-fit OF traces such as saturated events, pile-up events, and those with high noise levels. This was achieved by placing a quartic cut on the OF fit amplitude and the OF fit χ^2 [109].

Another set of pulse shape cuts aimed at removing events such as glitch events, square pulses, and low-frequency noise events that differ in shape compared to the expected phonon signal traces. Glitch events have fast rise and fall times and are caused by electronic noise from the TES bias line. Square pulse events also have fast rise and fall times but plateau for approximately $200\mu\text{s}$ at the maximum. Due to their unique pulse shapes that do not resemble the signal pulse shape, templates were created for glitch events and another for square pulses to effectively remove them. The glitch noise template and one of the LFN template is shown alongside the OF template in Fig. 3.9. However, low-frequency noise has varying shapes, so three templates were developed. This is explained in detail in 3.4.3.

To remove a noise type, the traces were fitted with both the standard phonon template and the noise template, resulting in two χ^2 values, one for each template fit. The difference $\Delta\chi^2$ shown in Eq. 3.15 was used to identify and remove these events based on the concept that events that are more noise-like would have a positive positive $\Delta\chi^2$ as the χ^2 using the standard phonon template will be higher, and vice versa for non-noise events i.e. signal-like events.

$$\Delta\chi^2 = \chi_{\text{standard template}}^2 - \chi_{\text{noise template}}^2 \quad (3.15)$$

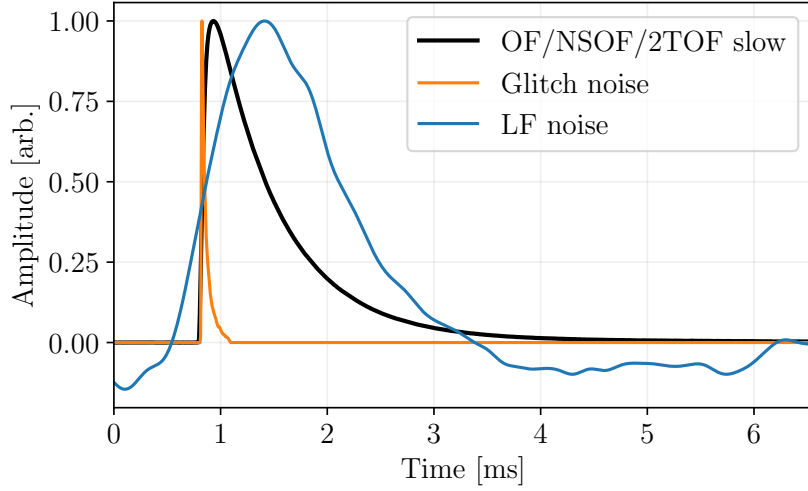


Figure 3.9.: Standard phonon template alongside noise templates such as glitch and low-frequency noise (LFN). Variations in shape among these templates were utilized for pulse shape cuts to remove these noise events.

Energy dependent cut for each of the noise types were placed on the $\Delta\chi^2$ as it is difficult to remove these instrumental background events at low energies.

The $\Delta\chi^2$ distribution for “good” events varies with time due to changes in the noise environment throughout the run, leading to splitting the data into blocks of series with custom cuts for each block. The series are group together based on the Cryocooler cycle variable explained in Sec. 3.4.3.

3.4.3. Low-frequency noise removal

More than 75% of the events that triggered the detector during this experimental run were caused by the Low-Frequency Noise (LFN) fluctuations (in the kHz frequency range) in the baseline caused by vibrations. These fluctuations in the baseline caused the energy deposition from the signal events at low energies to be unidentifiable. The LFN events are difficult to remove as the shape of these events is inconsistent and the bandwidth of these events overlaps the signal’s dominant frequencies. This section discusses the methods used in the data analysis to remove this noise.

Cryocooler variable

The source of vibration was found to be from the Gifford-McMahon cryocooler on the E-stem, as explained in Sec. 2.3, which was used to keep the electronics cold. The vibrations occurred due to the deterioration of the cryocooler. Accelerometers were installed to time the cryocooler vibrations, displaying distinct cryocooler chirps

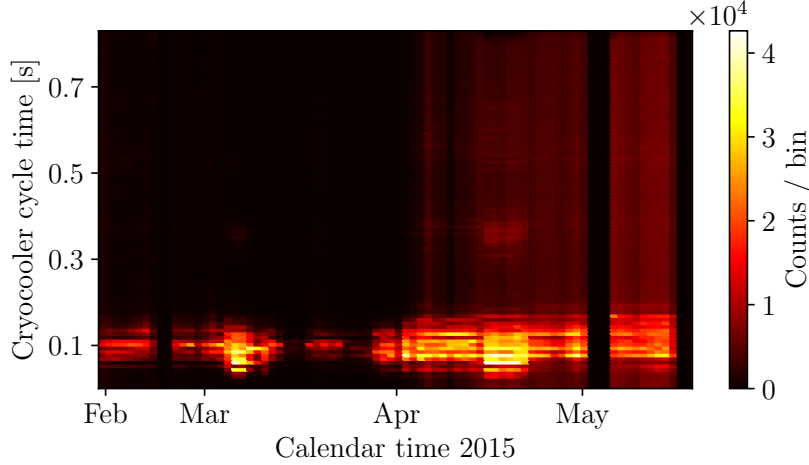


Figure 3.10.: A heatmap illustrating the cryocooler cycle variable CryoPreTime with a 0.83s modulus over the run period in calendar time, with the color intensity representing the number of trigger events. High trigger rates, particularly notable at CryoPreTime of 0.1s, correspond to vibrations occurring 0.1s after each cryocooler chirp. This high trigger rate correlates with low-frequency noise (LFN) in the data, attributed to cryocooler vibration, which deteriorated during the run.

occurring at regular 0.83-second intervals. High LFN was observed 0.1s after each chirp. A variable that was developed based on the cryocooler vibration measurement is the "CryoPreTime".

The CryoPreTime variable is defined as the time elapsed since the last cryocooler chirp. To address instances where the accelerometer failed to detect a chirp and a period lasted $n \times 0.83$ seconds, 0.83s modulus of the CryoPreTime variable was used [110]. Fig. 3.10 shows the number of triggered events on a plane of the CryoPreTime variable and the calendar time of this run. It indicates a high number of events at 0.1s of the cryocooler cycle time, caused by vibration-induced LFN. The figure also shows a progressively increasing number of events towards the end of the run compared to the beginning, due to the deterioration of the cryocooler.

Pulse shape cut

Due to its varying shape, one template was not enough to remove the LFN with $\Delta\chi^2$ method. Thus, a set of three LFN templates were developed to identify range of LFN shapes, which are shown in Fig. 3.11. The $\Delta\chi^2$ cut based on series blocks was applied for events with energy above 5 keV_t ($\sim 250 \text{ eV}_{ee}$) since the signal-to-noise ratio are high enough in these energies. For events with energies less than 5 keV_t ($\sim 250 \text{ eV}_{ee}$), removing LFN is challenging when the signal-to-baseline noise ratio is low. Under these condi-

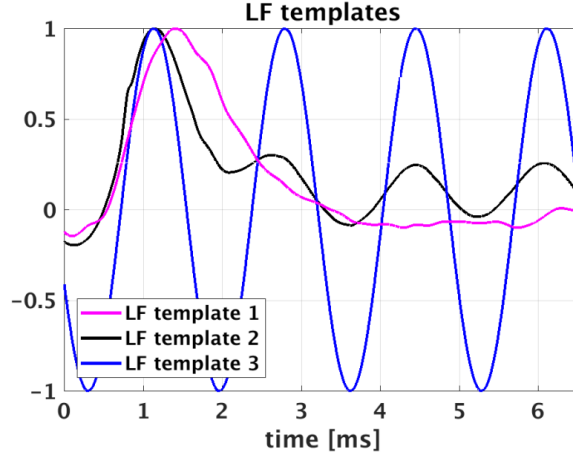


Figure 3.11.: The three low-frequency noise (LFN) templates used for the pulse shape cuts to eliminate LFN above 5 keV_t . The plot is taken from [120].

tions, distinguishing between an actual energy deposition and LFN fluctuations becomes difficult, as the signal from an energy deposition and the LFN are both buried under baseline noise. To address this, a cut based on bifurcated analysis was developed, which is explained below.

Bifurcated analysis cut

In cases where the LFN-background rate is difficult to model, a bifurcated analysis was employed. This method utilizes sideband information to estimate the background leakage after the quality cuts. The objective was to achieve less than 1 LFN-background event passing the cut out of all events. The number of background events leaking past a cut is denoted as

$$N_{\text{leak}} = N_{\text{LFN}} \cdot P(\text{cuts}), \quad (3.16)$$

where N_{LFN} is the number of LFN-background events, $P(\text{cuts})$ is the passage fraction of the cuts. Since both N_{LFN} and $P(\text{cuts})$ are unknown, an estimation of both N_{LFN} and $P(\text{cuts})$ can be achieved if there are two uncorrelated sets of cuts that are both sensitive to LFN events. These uncorrelated cuts are labeled as A and B, and the known signal efficiencies for these cuts are denoted as ϵ_A and ϵ_B , respectively. The unknown leakage fractions of LFN events past these cuts are represented as L_A and L_B . The number of good events is denoted as N_G . The events passing cut A and B can be written as

$$N_{AB} = \epsilon_A \epsilon_B N_G + L_A L_B N_{\text{LFN}}. \quad (3.17)$$

The number of events passing the cut A can be written as

$$N_{AB} + N_{A\bar{B}} = \epsilon_A N_G + L_A N_{\text{LFN}}, \quad (3.18)$$

where $N_{A\bar{B}}$ is number of events that pass cut A when cut B is not applied. The number of events passing the cut B can be written as

$$N_{AB} + N_{A\bar{B}} = \epsilon_B N_G + L_B N_{LF}, \quad (3.19)$$

where $N_{\bar{A}B}$ is number of events that pass cut B when cut A is not applied. With 100% signal efficiency, the number of LFN events leaking past the cuts can be written as

$$N_{\text{leak}} = \frac{N_{A\bar{B}} N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} \quad (3.20)$$

For non-100% signal efficiency, a revised method was employed with corrections applied to Eq. 3.20. Further information can be found in Ref. [76].

For the two different uncorrelated cuts, there were three parameters that were used to narrow down low-frequency noise:

1. $\Delta\chi^2$: using the difference in the LFN pulse shape from the standard signal trace.
2. CryoPreTime variable: Time since the last cryocooler cycle correlated with the high trigger rate caused by LFN.
3. Correlation of raw signal between T2Z1 and other neighboring detectors: The events triggered due to LFN produce triggered events in other detectors due to the vibration source affecting all of them.

The bifurcated cut A uses parameter 1, and cut B uses parameters 2 and 3 to identify LFN events. In this analysis, Boosted Decision Trees (BDTs) [121, 122] were employed to reduce multiple parameters into a single BDT score. The BDTs were trained using events from LFN-rich periods as background. For signal events, good phonon pulses were added to baseline noise extracted from randomly triggered events. Cut values were placed on BDT A and BDT B scores such that total low-frequency noise leakage remained below 1 event. The “box relaxation” technique verified the uncorrelated nature of variables; relaxing a cut resulted in new events entering the signal box, thereby changing the leakage estimate. With uncorrelated cuts the estimate grows proportionally with the number of new events.

3.4.4. Other efficiency cuts

In addition to the efficiency cuts discussed earlier, several other cuts were applied to further refine the data quality

Good randoms cut: Random triggers were implemented at the beginning, during and the end of the series to determine noise characteristics and the detector resolution. These events were removed for the final DM search analysis.

Muon veto cut: This cut removed events caused by interacting muons that triggered the muon veto panel and caused a global trigger within $50\,\mu\text{s}$ of the veto history buffer trigger.

Good phonon start time cut: This cut removed events that were outside of the OF search window which is $-200\,\mu\text{s}$ to $100\,\mu\text{s}$ around the trigger bin.

Multiple scatter cut: Assuming that the WIMP doesn't interact with multiple detectors, events triggering multiple detectors are removed. This cut selects multiple scatter events by first checking whether detectors other than T2Z1 were triggered and verified the possibility of a glitch or low-frequency noise, which were not considered as multiple scattering.

3.4.5. Fiducial volume cut

Due to non-uniformities in the electric field as shown in Fig. 3.12, events at the larger radii had a reduced NTL/Luke amplification as the voltage experienced by charge carriers is less than the nominal voltage of 75 V. For most events, the potential difference ΔV experienced by the charge carriers was the potential difference between the detector faces resulting in maximal NTL amplification. However, having low ΔV , sidewall events had reduced amplification leading to significantly lower reconstructed energy. This created Reduced NTL-Gain Events (RNTLs) in the signal region, and they had to be identified and removed. This section outlines the method used to remove these RNTLs. This cut is explained in detail in Ref. [123]. While the exact position of an event could not be determined to correct for the reduced NTL amplification, a parameter correlated with the radial position was developed. The radial parameter ξ takes advantage of the phonon channel layout, i.e. one outer channel and three wedge-shaped inner channels, by comparing the fast and slow template and the start time between the outer and the inner channels. In this written work, the terms fiducial volume cut and radial cut are used interchangeably and refer to the same event selection criterion.

Reduced NTL-gain events (RNTLs)

The characterization of RNTLs was performed by modeling their distribution as a function of reconstructed energy and ξ through a multi-step process. First, the energy response of the detector was determined, considering the NTL gain as a function of position inside the detector. Next, the rates of events contributing RNTLs into the signal region below $2\,\text{keV}_{ee}$ were evaluated. Subsequently, the distribution of RNTLs in ξ space was modeled. Then, the resolution of ξ as a function of energy and ξ was modeled. Using these models, a Monte Carlo simulation was constructed to determine the expected distribution of RNTLs in the energy- ξ plane. More information of this method can be found in [76] [124].

Cut Criteria

For energies below $2\,\text{keV}_{ee}$, a cut was placed on the distribution of ξ as a function of

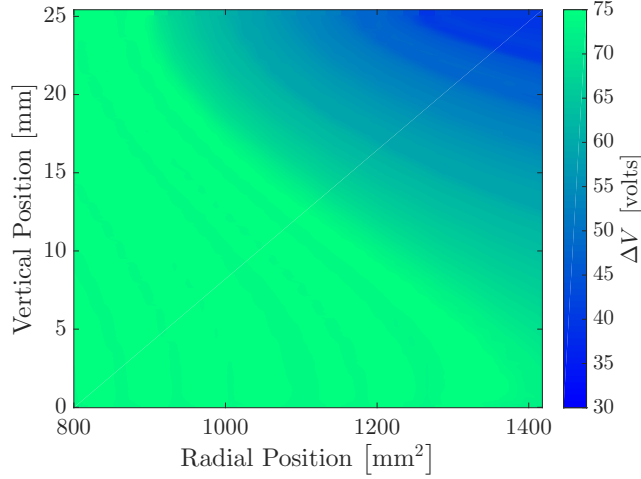


Figure 3.12.: Contour plot showing the calculated bias voltage distribution across a 2D plane of vertical position against radial position. The color gradient denotes varying voltage bias across positions, revealing electric field non-uniformities. The nominal voltage applied across the detector is 75 V and the regions with voltages below 75 V had reduced NTL amplification, leading to fiducial volume cuts. The plot is taken from [76].

energy. The cut boundary was chosen such that the expected distribution of RNTLs passing the cut is uniform in energy between 0.07 and 2 keV_{ee} to limit RNTLs to fewer than 0.125 events. This radial parameter cut imposes an analysis threshold of 70 eV_{ee} , determined by the lowest well-defined bound of the radial resolution model. For energies above 2 keV_{ee} , the threshold of the fiducial volume cut was set differently due to the inability to model ξ accurately because of saturation effects in the phonon pulse shape. Instead, a restrictive cut at -4×10^{-5} in ξ was set above 2 keV_{ee} to ensure zero RNTLs in this region. The result of this cut application on the CDMSlite Run 3 data is shown in Fig. 3.13

3.5. Live-time and efficiencies

The amount of time the detector was live and recorded data, referred to as live-time, is essential for determining the expected DM interaction rate during this period and is detailed in Sec. 3.1. Ignoring annual modulation effects, the number of events observed from a DM signal increases linearly with live-time.

Quality cuts were developed to remove background events, but they also remove signal events, leading to a signal efficiency of less than 100%. Calculating this signal efficiency is essential for scaling the expected DM interactions in the detectors after cuts, particularly at low energies where signal-background discrimination is difficult.

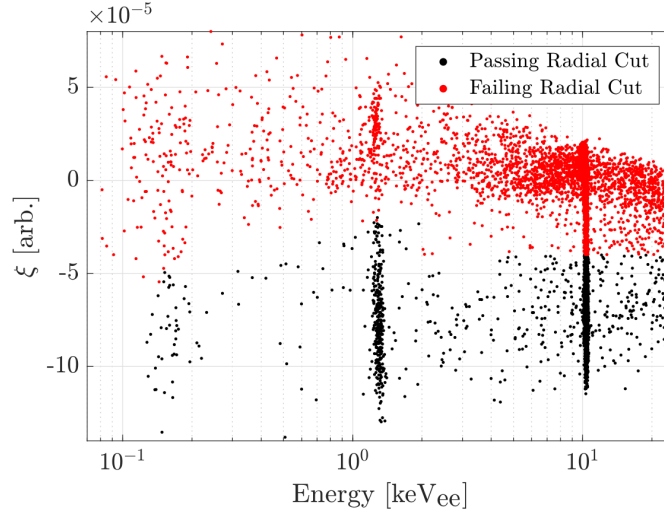


Figure 3.13.: CDMSlite Run 3 events on the energy- ξ plane with events passing the radial cut shown in black and the events failing the cut is shown in red. The plot is taken from [76].

Therefore, the signal efficiency was calculated as a function of energy.

Additionally, the hardware trigger only records events with a sufficiently high signal-to-noise ratio, i.e., sufficiently high energy, requiring the calculation of trigger efficiency as a function of energy. This is outlined in Sec. 3.5.2.

3.5.1. Live-time estimate

The cuts applied in the data analysis were categorized into live-time and efficiency cuts due to their distinct roles in eliminating events from the dataset. Live-time cuts removed periods and events of unfavorable experimental conditions. These cuts are the high voltage (HV), prepulse baseline, charge χ^2 , bad series, GPS time, bad trigger rate, and the T2Z1 trigger cuts, as explained in Sec. 3.4.1. After applying the live-time cuts, 95% of the Run 3 events were retained leading to an exposure of 60.9 days, which corresponds to 36.4 ± 0.3 kg-days with a detector mass of 0.597 ± 0.005 kg.

3.5.2. Trigger efficiency

The DAQ hardware triggers when an event exceeds a certain energy threshold, typically significantly above the baseline noise level. While high-energy events are triggered in all cases, events with lower energy slightly distinguishable above baseline noise exhibit a non-zero and non-unity trigger efficiency. Thus the trigger efficiency was studied as a function of energy. Calibration data was used for the trigger efficiency calculations due

to high statistics. The trigger efficiency was calculated by determining the fraction of events that trigger T2Z1, but also triggered other detectors. Quality cuts were applied to this dataset ensure the trigger efficiency was calculated based on signal-like events.

3.5.3. Data-quality cuts efficiencies

Efficiencies for pulse shape cuts, including glitch cut, phonon χ^2 cut, square pulse cut, and low-frequency noise cut were determined through a pulse simulation using a software called BatFaker, originally developed for salting. In BatFaker, simulated pulses were generated by combining noise from randomly triggered events with phonon templates. The phonon templates have fast and slow components by combining a slow and a fast template scaled to the user-defined amplitude. Slow amplitudes were uniformly distributed to ensure event energies were uniformly distributed between 0 and 60 keV_{ee}. The relative amplitude between the fast and slow templates was extracted from the DM search data. These simulated traces were processed to determine the cut variables, and energy. To measure cut efficiencies as a function of energy, events were generated with a uniform energy distribution. CryoPreTime was also recreated for simulated traces for the low-energy LFN cut by choosing noise traces from uniformly distributed CryoPreTime. Efficiencies were evaluated separately for periods A and B, and uncertainties were propagated through Monte Carlo simulations.

3.5.4. Fiducial volume efficiency

The efficiency of the fiducial volume cut was measured using a Monte Carlo simulation based on the resolution model of ξ , simulating the radial parameter distribution for events with full NTL amplification. The ξ distribution is modeled after events with reconstructed energies in the L-shell line, with statistical subtraction of non-⁷¹Ge-backgrounds and deconvolution of the radial-parameter resolution at 1.3 keV_{ee}. This results in the expected “true” distribution of ξ for L-shell energy events, which is then scaled according to energy to create an energy-dependent probability distribution for ξ . Applying the fiducial volume cut to these simulated distributions provided the efficiency of the fiducial volume cut for fully amplified NTL events. The final efficiency was obtained by multiplying the fiducial volume cut by the percentage of non-RNTLs (86%).

3.5.5. Combining efficiencies

Efficiencies of the trigger, efficiency cuts, and fiducial volume cut are combined by multiplying the mean of each efficiency and using Monte Carlo simulations to propagate uncertainties. The effect of the fiducial volume cut is to reduce the difference in the efficiencies between periods A and B, necessitating a live-time weighted average to merge these periods into a single efficiency curve.

These efficiency curves were parameterized using a functional form to keep track of parameter uncertainties and covariances. This parameterization was performed for the purpose of the profile likelihood fit which is beyond the scope of this thesis work. For the

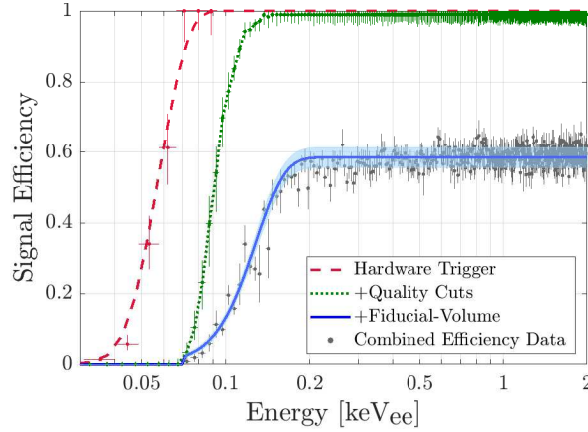


Figure 3.14.: The plot shows the efficiency curves obtained through the sequential application of trigger efficiency, quality cuts efficiency, and fiducial volume cut efficiency. The final efficiency curve in blue is derived via a parameterization using a three-parameter fit with an error function. The $\pm 1\sigma$ uncertainty is depicted as an uncertainty band around the final efficiency. The plot is taken from [76].

main analysis, the efficiency curves were parameterized using a three-parameter error function as follows:

$$f(E) = \mu_1 * (1 - \text{erf}(\frac{E - \mu_2}{\sqrt{2} \cdot \mu_3})). \quad (3.21)$$

As the fiducial volume cut imposes a threshold of 70 eV_{ee}, the efficiency is set to zero below this energy. Figure 3.14 shows the signal efficiency after applying the trigger, quality, and fiducial volume cuts. The trigger efficiency is 100% above 0.09 keV_{ee}. The fiducial volume cut presents the major constraint on signal efficiency, while the quality cuts also play a significant role, particularly at low energies. This is attributed to poor data quality arising from a lowered threshold and instrumental noise which lead to a stringent cut to remove them, such as the low-frequency noise cut which allowed only less than one noise event to pass the cut. The motivation for this thesis work is to replace this LFN cut with a Machine Learning (ML) based cut which results in a higher signal efficiency. The impact of the fiducial volume cut on the signal efficiency gained from the ML-based LFN cut, as well as its effect on the final results, is discussed in Chapter 5 and Chapter 6.

4. Low-Frequency Noise Identification using Machine Learning

As seen in the previous chapter, the vibration-induced Low-Frequency Noise (LFN) has led to stronger event selection criteria to remove the LFN events which set a higher analysis threshold and thus reduced the sensitivity of the experiment to low-mass dark matter candidates. This chapter presents a machine learning-based approach for removing the LFN events while preserving the low-energy signal events. Section 4.1 and Section 4.2 describe the sample generation methods, followed by a brief introduction to neural network algorithms in Sec. 4.3. The classification of LFN using single-channel information, the Phonon Total (PT) channel, and multi-channel information using Convolutional Neural Networks (CNNs) and CNN with Long Short-Term Memorys (LSTMs) Section 4.5 are detailed in Sec. 4.4 and Sec. 4.5, respectively.

4.1. Sample preparation

The dataset for the machine learning method consists of signal sample containing signal features, and the LFN-background sample ideally consisting of triggered LFN events. This section details the sample generation process for both signal and LFN-background sample.

4.1.1. LFN-background samples

The LFN-background events cannot be simulated as the shape of these events varies and the nature of these events is not well known. Thus, the LFN-background samples are selected from the CDMSlite Run 3 experimental data, specifically events from periods of elevated LFN events. The selection is done by choosing periods of intense vibration during the experimental runtime, as determined by the CryoPreTime quantity shown in Fig. 3.10. Specifically, the periods between March 4 and 5, and between April 2 and 30, are selected. Then the events in this period are filtered by passing them through the LFN cut developed for the analysis of this run explained in Sec. 3.4.3. Events failing this cut are then used as the LFN-background events for this analysis. It should be emphasized that these samples from elevated LFN periods may potentially contain signal events, however they are predominantly triggered by LFN events. Figure 4.1 shows the OF amplitudes of the PT channel and energy of the LFN events. It can be seen that most of the events have amplitudes between ~ 15 and ~ 40 ADC units in the PT channel with corresponding energy between 0.07 keV_{ee} and $\sim 0.15 \text{ keV}_{ee}$. Raw traces of selected LFN events are shown in Fig. 4.2, showing the varying shape of this noise.

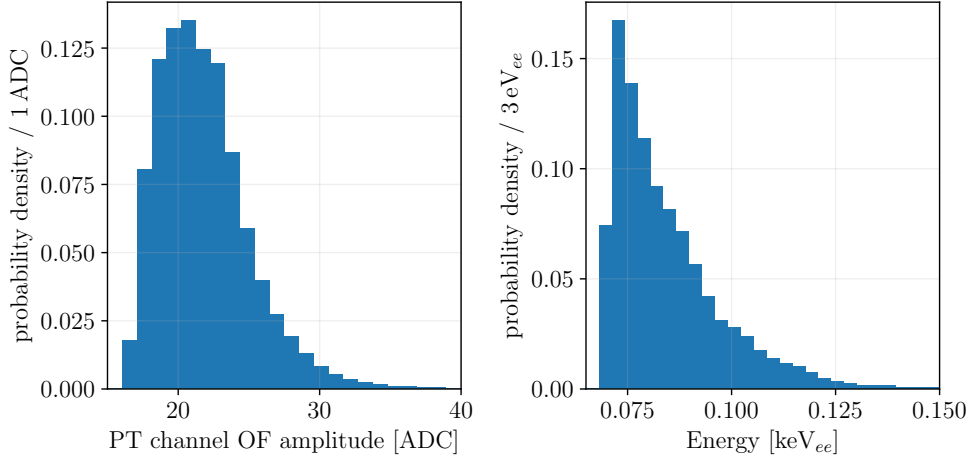


Figure 4.1.: PT channel OF amplitude (left) and energy (right) distributions of the LFN events. The events are chosen from vibration-intense periods from CDMSlite Run 3 experiment, and filtered using the LFN cut developed for this analysis described in Sec. 3.4.3. The histogram shapes result from applying energy corrections to the amplitudes to derive the energy.

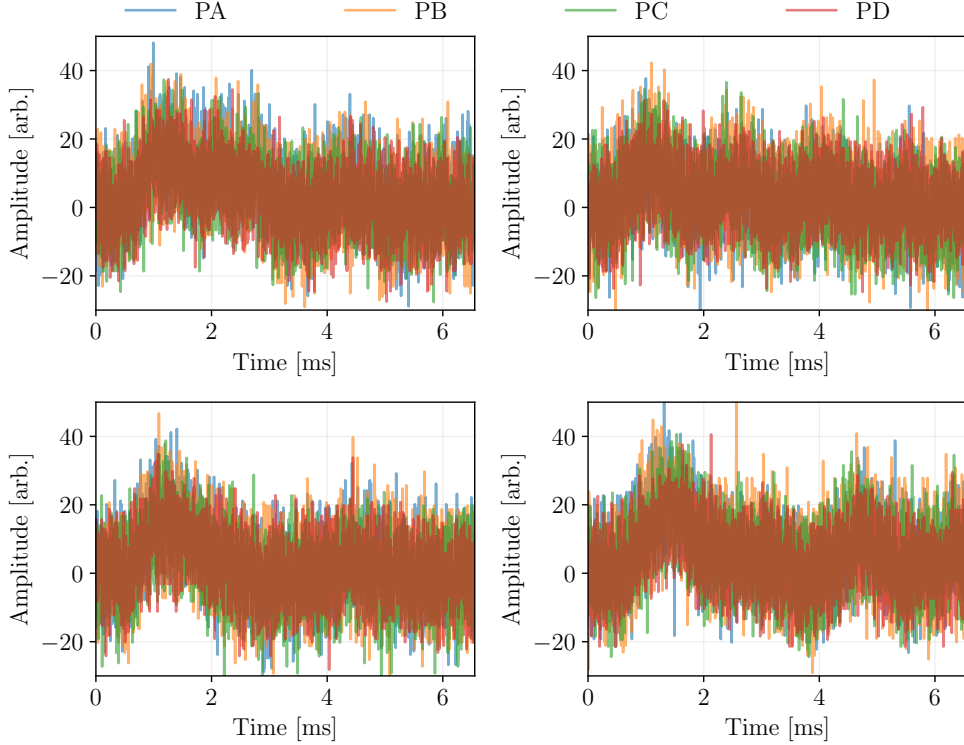


Figure 4.2.: Examples of raw traces from LFN-background dataset with four phonon channels PA, PB, PC, and PD. The traces show varying shapes of the noise events caused by the cryocooler’s vibration.

4.1.2. Signal samples

The main idea behind the signal samples generation is to inject phonon templates of various amplitudes into the baseline noise traces. The baseline noise events are extracted from randomly triggered events in the CDMSlite Run3 experimental data. The randomly triggered events may contain signal events, and to eliminate these, a cut based on the standard deviation of the raw traces was applied, as signal traces tend to a higher standard deviation compared to the noise traces. Raw traces of the randomly triggered events with a standard deviation larger than 15 ADC units were removed by this cut. This cut also effectively removed traces with very poor noise with high amplitudes, which could submerge low-amplitude signal features. Although these high-amplitude noisy events resemble LFN, they are excluded from this analysis due to their amplitudes. This cut preserves LFN in the baseline events, ensuring that the signal samples have LFN in the baseline. As a result, the ML-models are trained to identify signal features within traces that contain LFN in the baseline. Some events removed by this cut are shown in Fig. 4.3, showing events with poor noise and an event with a signal. With this cut, about 0.16 % noise traces were removed. Figure 4.4 shows examples of raw traces that passed this cut and were used in generating the signal samples. However, it is important to note that these raw traces may still contain LFN in the baseline.

The majority of LFN event amplitudes fall within the range of ~ 18 and ~ 25 ADC units in the PT channel as seen in Fig. 4.1, and the established cut in the Run3 analysis effectively eliminated all events within this amplitude range, so these amplitudes are chosen for signal samples to train and test the performance of the neural network.

The template used for “spiking” the baseline noise depends on whether the classification is based on the individual channels (PA, PB, PC, PD), the PT channel, or all the channels together (also called multi-channel). Figure 4.5 shows the pulse templates of individual channels as well as the PT pulse template. It can be seen the individual channel templates are identical, however, the PT pulse template is slightly different from the individual channel templates. It is important to note that the shape of individual channel traces varies based on the position of the interaction in the detector volume, thus the template is inadequate in capturing these features. This is explained in detail in Sec. 4.2. The PT pulse template, shown in Fig. 4.5, is derived from the sum of all the channels, thus washing out this position dependency. Most of the CDMSlite Run3 analysis of this dataset is done with the PT pulse template. The schematic of the “spiking” is shown in Fig. 4.6 where it can be seen that the OF template is added to a random noise trace to get the fake signal samples.

Unlike the individual channels and PT channel, the templates for the multi-channel are not straightforward as traces in individual channels differ based on the location of the interaction. Therefore, a single set of templates cannot represent all channels. Thus the simulation of events in the detector is essential for multi-channel classification

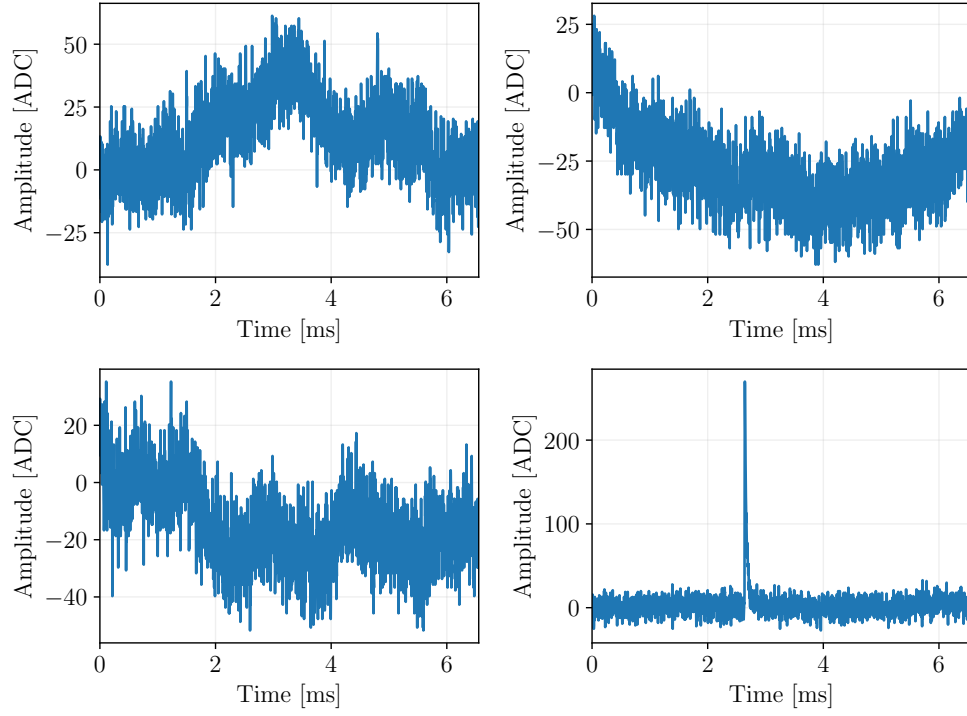


Figure 4.3.: Examples of raw traces of one phonon channel from randomly triggered events that did not pass the standard deviation cut. The top left, top right, and bottom left traces show baseline noise with poor noise quality which when spiked would submerge the signal features. The bottom right trace contains a signal, and thus the trace is not suitable to be spiked with the template.

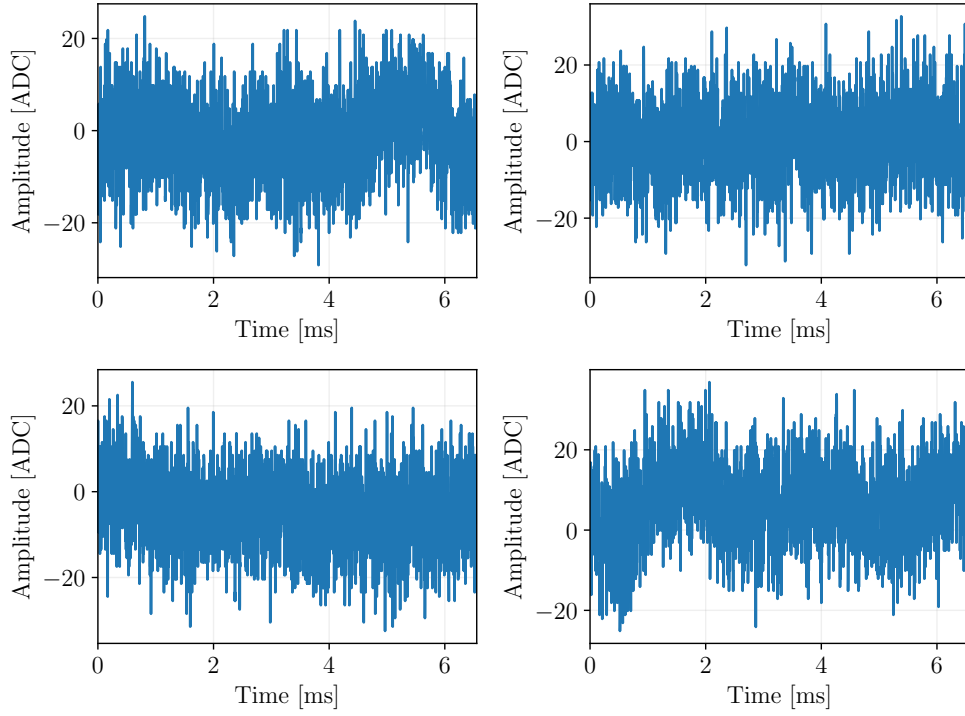


Figure 4.4.: Examples of raw traces from one phonon channel of randomly triggered events that passed the standard deviation cut. These traces are used for “spiking” in order to generate the signal samples.

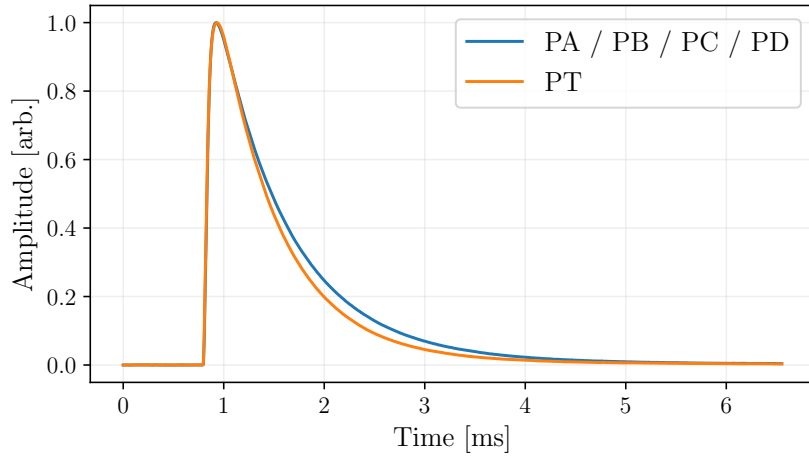


Figure 4.5.: The pulse templates of the individual phonon channels PA, PB, PC, and PD along with the pulse template for the PT channel. The individual channel templates are the same.

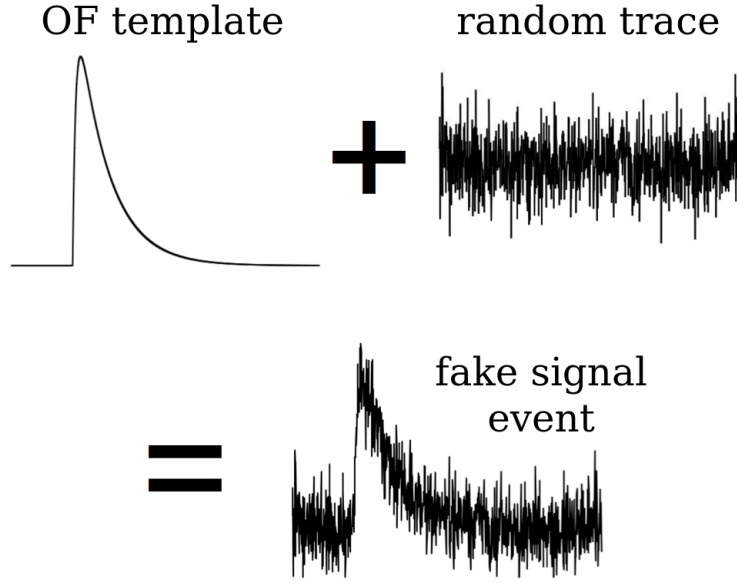


Figure 4.6.: Schematic of the signal samples generation. The OF pulse template is added to baseline noise from randomly triggered events to generate fake signal events.

with uniformly distributed interactions in the detector volume. Section 4.2 details the method used for signal samples generation for the multi-channel classification.

4.1.3. Pre-processing

The traces that are used to train and test the network can either be pre-processed with filters or can be used without any pre-processing. The following are the two methods that are considered for this LFN analysis, with results presented in Sec. 4.4.

Low-pass filter

A low-pass filter removes the high-frequency components of the trace and only lets the low-frequency components pass through. As our signal is prominent in lower frequencies, this filter effectively preserves the signal features.

Moving average

Moving average smooths the trace by performing a series of averages on the subsets of data within the full trace. This downsampling flattens the small fluctuations, i.e. the high frequency components. While a small window smooths out a small portion leaving the major fluctuations intact, a large moving average window would wash out some of the important components of the trace such as the rising edge of the pulse. This is not only important in localizing the event in the detector volume but also important in

identifying signal events as LFN events tend to have a larger rise time compared to the signal events.

Figure 4.7 shows raw traces of two LFN events and a signal event, all unprocessed (top row) and after pre-processing: low-pass filter (second row) and moving average (third row). The pre-processed traces reveal the underlying features more clearly compared to the unprocessed raw traces.

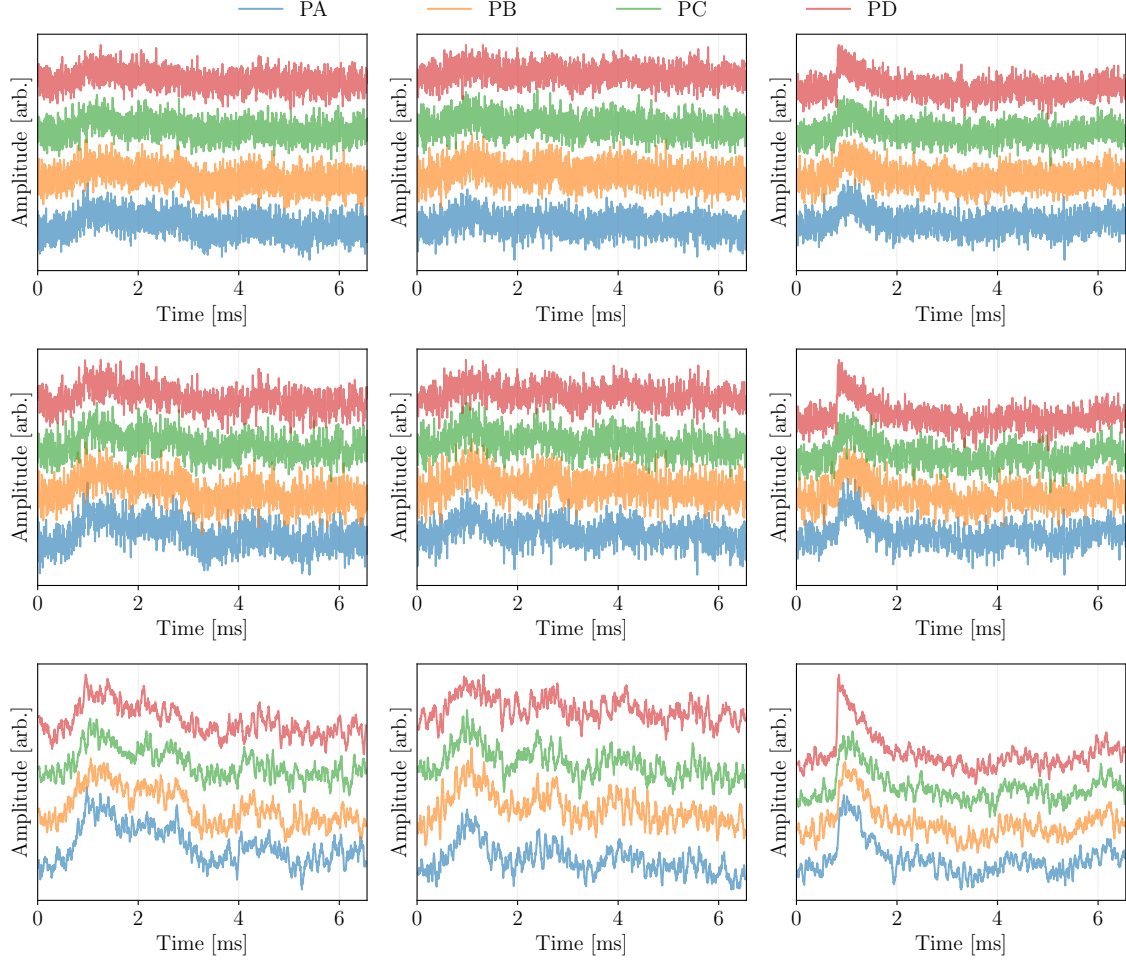


Figure 4.7.: Examples of raw traces from each phonon channel for two LFN events are displayed in the left and middle columns, with an observed signal event shown in the right column. The unprocessed traces are presented in the top row, the low-pass filtered traces in the middle row, and the traces with the moving average applied are shown in the bottom row.

4.2. Simulation Using Pulse shapes and Regression (SUPR)

As the SuperCDMS simulation based on G4CMP [125] was not tuned for older experimental setups like CDMSlite, a data-based simulation is used for signal samples generation in the multi-channel classification. The basis of this Simulation Using Pulse Shapes and Regression (SUPR) method is to capture the pulse shapes from existing data in the CDMSlite dataset.

For an event, each channel has different features based on the position of the interaction in the detector. For example, if an event occurred close to channel PB, the trace PB will have the highest amplitude of the four channels since the number of phonons absorbed in that channel is higher compared to the other channels. Additionally, the rise time will be shorter for channel PB, as the phonons arrive on this channel earlier than other channels. The other three channels will have a larger rise time based on their proximity to the event location. For accurate simulations, interactions distributed throughout the detector is required to prevent bias toward any specific direction. So, using calibration source data is not ideal as they are predominantly positionally non-uniform. On the other hand, events from the activation peaks, as explained in Sec. 3.3.2, are expected to be uniformly distributed across the detector. Therefore, they are used to determine the templates corresponding to the position of an event within the detector.

This section presents a novel method for generating traces using the pulse shapes based on event locations. The pulse shapes are fitted with an exponential function, and the resulting fit parameters are used to correlate with event localization through a regressor. This correlation is used to generate additional events.

4.2.1. K-shell events

Of the activation peaks in Ge, the K-shell peak has the highest branching ratio [126], thus producing a large number of events in the detector on the order of $\mathcal{O}(10^3)$ events. Among the activation peaks in Ge, the K-shell peak has the highest branching ratio, resulting in a significant number of events in the detector, approximately $\mathcal{O}(10^3)$ events. These events were used for the study of other analysis topics like the radial cut, energy scale corrections, and calculating signal efficiencies.

Event selection

The mean of the K-shell peak is $10.354 \pm 0.002 \text{ keV}_{ee}$ and the resolution of the peak is $108 \pm 2.0 \text{ eV}_{ee}$ [76]. The Run 3 low-background events which pass all of the quality cuts explained in Sec. 3.4, and have energies within the range of $10.354 \pm 0.108 \text{ keV}_{ee}$ are chosen.

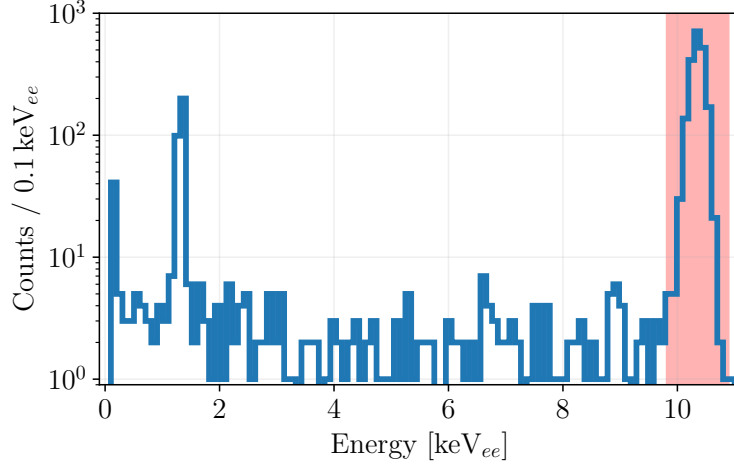


Figure 4.8.: The distribution of event energies from the CDMSlite Run3 analysis that passed all of the cuts with the selected K-shell events highlighted in red. The K-shell events are selected by choosing events within the energy range of $10.354 \pm 0.108 \text{ keV}_{ee}$ [76].

Characteristics of the K-shell events

The energy distribution of CDMSlite Run3 events that pass all the quality cuts, with the region corresponding to the K-shell event selection, is shown in Fig. 4.8. The event location within the detector volume is a quantity that can not be directly measured from the event data. However, parameters like the amplitude and rise time correlations between channels provide insights into the event’s location. One key parameter for localizing events is the energy partition, which is calculated from the amplitudes of inner channels that are wedge-shaped. The X energy partition X_{OF} and Y energy partition Y_{OF} are computed using

$$X_{OF} = \frac{\cos(30^\circ)D_{OF} + \cos(150^\circ)B_{OF} + \cos(270^\circ)C_{OF}}{D_{OF} + B_{OF} + C_{OF}}, \quad (4.1)$$

$$Y_{OF} = \frac{\sin(30^\circ)D_{OF} + \sin(150^\circ)B_{OF} + \sin(270^\circ)C_{OF}}{D_{OF} + B_{OF} + C_{OF}}, \quad (4.2)$$

where B_{OF} , C_{OF} , and D_{OF} are the OF amplitudes of channels PB, PC, and PD, respectively. Plotting the X_{OF} and Y_{OF} energy partitions, called a partition plot, can be used to verify the homogeneity of the event distribution in the XY plane. Figure 4.9 shows the energy partitions of the selected K-shell events. It can be seen that they are uniformly distributed in the XY plane, as demonstrated by the even distribution throughout the partition triangle.

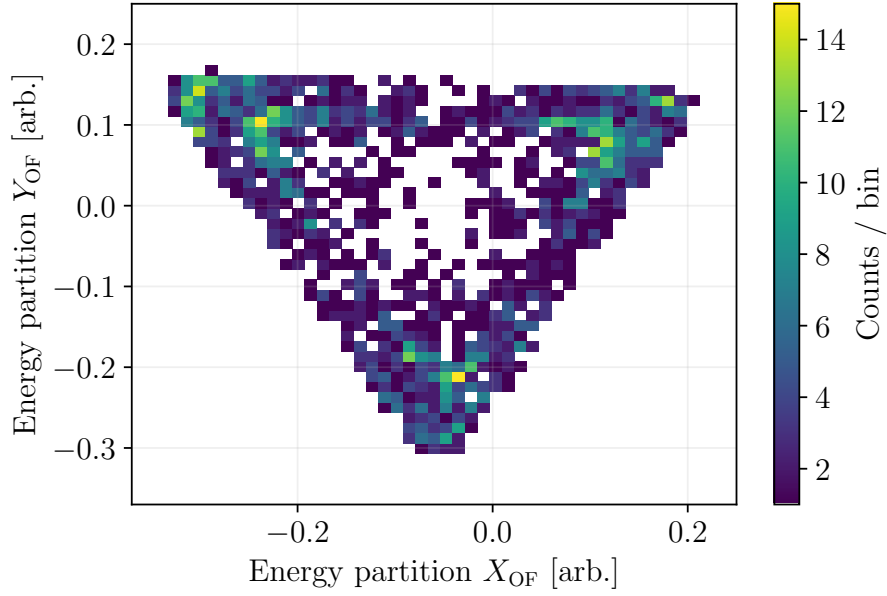


Figure 4.9.: The partition plot of K-shell events, which is a representation of the localization of the events in the XY plane, using the energy partition information calculated using Eqs. 4.1 and 4.2.

4.2.2. Fitting traces

The purpose of fitting the raw traces is to quantify their shape by measuring the amplitude, rise time, and fall time of each channel's raw trace in an event. Initially, a simple double exponential fit with one rise constant and one fall time constant was used. However, it will be shown later in this section, this approach was insufficient to capture the complexity of the pulse shape. Consequently, multiple rise and fall times were introduced to improve the fit. The following sections describe these methods and their results.

Double exponential fit with one rise time constant and one fall time constant

For the double exponential fit, the fit function y as a function of time t can be written as

$$y(t) = a(e^{-(t+t_0)/\tau_f} - e^{-(t+t_0)/\tau_r}), \quad (4.3)$$

where a , τ_f , τ_r , and t_0 are the fit parameters, representing the amplitude, fall time constant, rise time constant, and the offset in the time, respectively. Although t_0 is calculated during fitting, it is not specifically used for new trace generation. For new trace generation, x-offset values were drawn from a uniform distribution between 470 and 540 time bins, resembling the K-shell events offset.

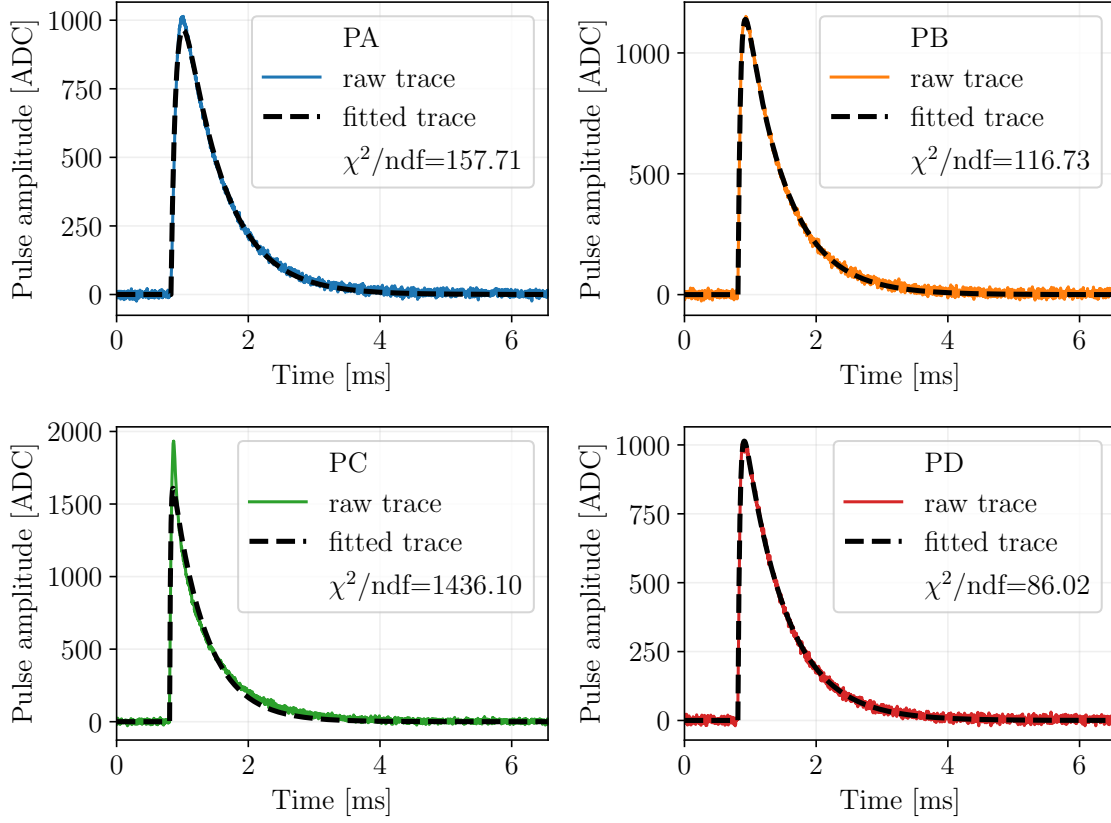


Figure 4.10.: The raw traces from the four channels of a K-shell event were fitted using an exponential fit based on Eq. 4.3 with the corresponding reduced χ^2 shown in the legend. The leading channel PC shows an inadequate fitting using this version of the exponential fit.

Figure 4.10 shows the four raw traces from the four channels of a K-shell event, along with the fitted traces using Eq. 4.3. The channel closest to the interaction, PC, exhibits a “peaky” shape and a high amplitude, and will be referred to as the “leading channel”. The fit for this channel does not accurately represent its shape, as reflected by the higher reduced χ^2 value compared to the other traces. To better capture the pulse shape, a fit function with two rise time constants and two fall time constants is used, as explained in the next section.

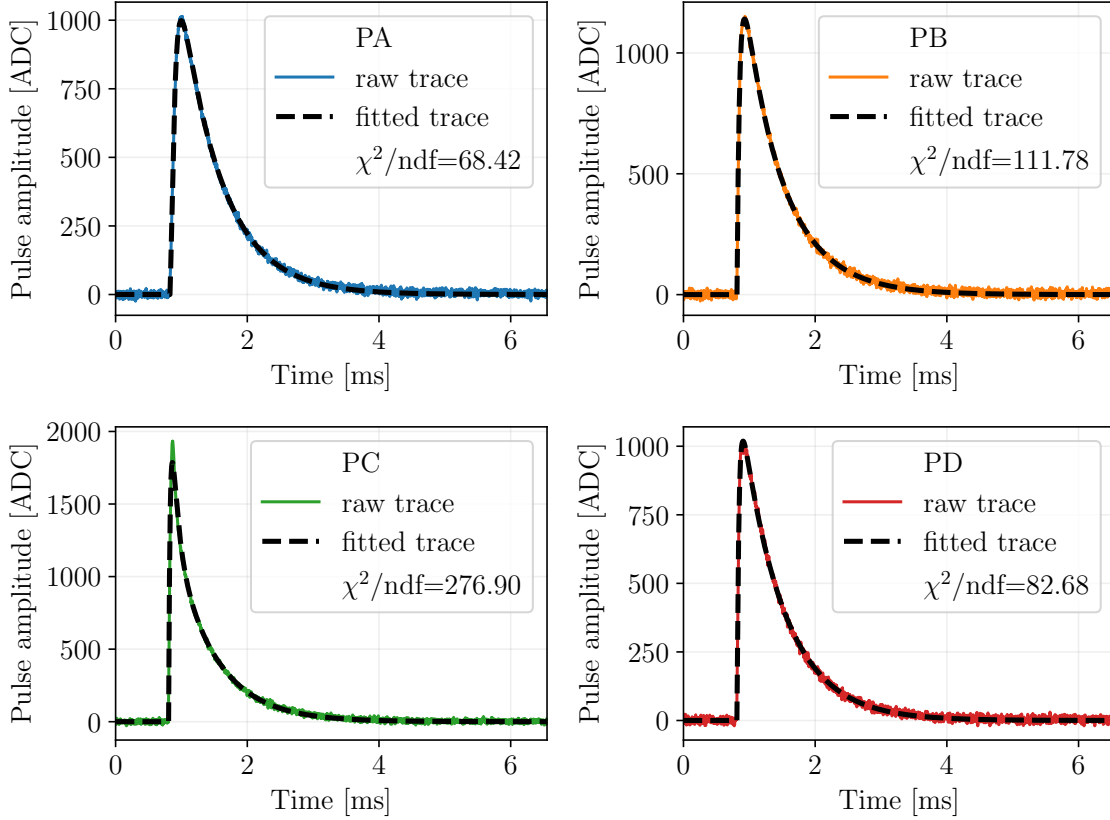


Figure 4.11.: The raw traces from the four channels of a K-shell event were fitted using an exponential fit based on Eq. 4.4 with the corresponding reduced χ^2 is shown in the legend. The leading channel shows an improved fit using this version of the fit equation compared to the fit based on Eq. 4.3 seen in Fig. 4.10.

Double exponential fit with two rise time constants and two fall time constants

For the modified double exponential fit with two rise time constants and two fall time constants, the fit function y as a function of time t is written as

$$y(t) = a(e^{-(t+t_0)/\tau_{f1}} + e^{-(t+t_0)/\tau_{f2}} - e^{-(t+t_0)/\tau_{r1}} - e^{-(t+t_0)/\tau_{r2}}), \quad (4.4)$$

where a is the amplitude, τ_{f1} and τ_{f2} are the fall time constants, τ_{r1} and τ_{r2} are the rise time constants, and t_0 is the offset in the time. Similar to the previous section, the offset t_0 is calculated but not specifically used for new trace generation. Figure 4.11 shows the four raw traces from the four channels of the same event in Fig. 4.10, along with the fitted traces using Eq. 4.4. The channel closest to the interaction point PC exhibits a “peaky” shape, and the fit represents the shape more accurately than in the previous

case, as evidenced by the lower reduced χ^2 value. The χ^2 is still high because noise fluctuations are not accounted for unlike OF fit, and the version of fit does not fully capture the peak. While additional fit parameters could improve the χ^2 , this version is considered sufficient for the current analysis.

4.2.3. Using the fit parameters for the traces generation

The goal of this method is to map the pulse shape of the events, obtained from the double exponential fit, to the location of the interaction in the detector volume, estimated from the partition. This section details the approach used for this correlation utilizing a decision tree regressor using the scikit-learn package [127].

Decision tree regressor

A decision tree regressor [128] is a supervised learning method that uses a series of conditions to predict outcomes. It consist of nodes that represent a condition based on feature values, which then leads to further conditions on further nodes or predicted values on leaves. This hierarchical structure enables us to understand how inputs influence the predicted output. Further information can be found in Ref. [129].

The regressor is trained with the partition information shown in Fig. 4.9 with the fit parameters as the target. In this analysis, only the fitting method with two rise times and two fall times shown in Eq. 4.4 is used to find the correlation between the position of the interaction event to the shape of the waveforms. The training was performed with 1400 samples. The performance of the model is evaluated using a Mean Square Error (MSE) given by

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2, \quad (4.5)$$

where N is the number of samples, y_i is the true value, and p_i is the predicted value. Due to its squaring of the penalties instead of applying linearly, MSE results in higher error for outliers. [130]

4.2.4. Generation of new traces

For generating traces, a uniform distribution of X_{OF} and Y_{OF} values within the partition triangle should be chosen. The partition of K-shell events in Fig. 4.9 shows that events form a band of triangles with events concentrating around the corners and very few number of events in the middle. To approximate this distribution, rough lines were drawn to imitate these bands. Then, a uniform set of partition values was generated within the band, with events between the corners randomly removed to replicate the observed distribution.

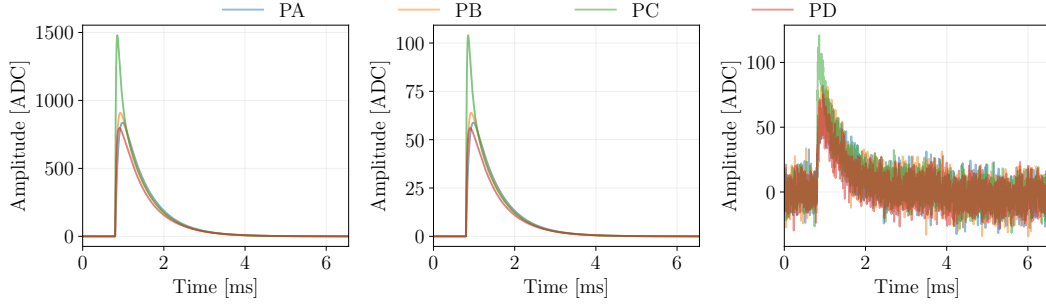


Figure 4.12.: The traces generated from the predicted fit parameters using the SUPR method are shown on the left. In the middle, the traces are rescaled to the desired PT amplitude, and on the right baseline noise is added to the rescaled trace.

The partition values are then passed through the trained regression model to obtain the fit parameters, which were subsequently used to generate new traces. These generated traces were then rescaled to the desired PT amplitudes and combined with baseline noise to produce the final traces for the analysis. Figure 4.12 shows the process of generating a signal event using a SUPR method trace. An event generated from the SUPR method is shown on the left, in the middle it is rescaled such that the PT amplitude is 200 ADC, and the right plot shows the scaled trace combined with the baseline noise, resulting in a signal event trace. This method using pulse shapes and regressor serves as a proof of principle, and it can be fine-tuned in the future for improved performance. The schematic of the signal sample generation process is illustrated in Fig. 4.13. To generate events, X and Y partition values are provided as input, and the corresponding pulse shape fit values are predicted by the trained regressor. These fit values are then used to generate the templates to which baseline noise is added. Using this method, approximately 28,000 events were generated, limited by the number of available baseline noise traces.

4.3. ML-based classification approach

For the LFN removal task, two types of network architectures were studied: Convolutional Neural Networks (CNNs) [131] and Long Short-Term Memory (LSTM) [132] models. The networks were developed using the TensorFlow package [133]. This analysis primarily focuses on CNNs, with results of CNN combined with LSTM i.e. CNN-LSTMs used for comparison. This section provides a brief introduction to CNNs and LSTMs, followed by an overview of the loss functions and metrics used in this study. Additionally, activation maps, which help identify features important for class decisions, are discussed. Finally, the classification results using single-channel information, the PT channel, and multi-channels are presented.

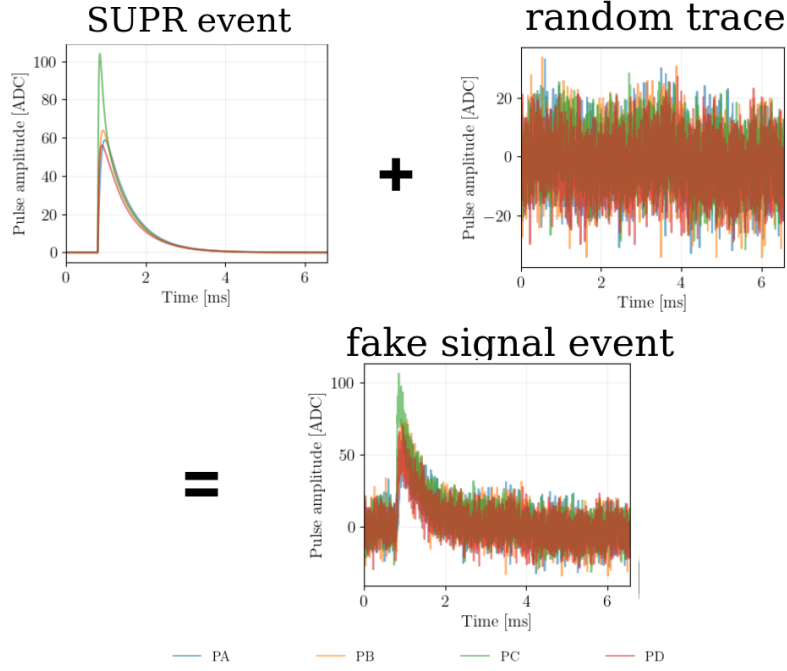


Figure 4.13.: The schematic for generating signal samples for the multi-channel network involves adding events generated by the SUPR method to baseline noise extracted from randomly triggered events to obtain the signal event.

4.3.1. Convolutional neural network (CNN)

Convolutional Neural Networks (CNNs) [131] are a type of deep neural network known for their performance in computer vision problems [134]; however, they can also be effectively utilized in any application with a grid-like topology. The key component of CNNs is the convolution operation performed by a convolution layer which transforms data into feature maps. In a 1D convolution, the convolution operation is applied along the trace using a specified kernel size. When a pattern is detected within a segment of the trace, the same pattern will be recognized later in the sequence regardless of its position. This property is known as translation invariance which ensures that the convolutional layer identifies features consistently across different regions of the trace. This is especially important as the signal features, like the sharp rising edge, vary slightly in position along the trace due to the trigger procedure, and this is why a neural network consisting of fully connected neurons such as a Multi-Layer Perceptron (MLP) [135] is not favored for this classification. A CNN architecture usually contains one or more convolution layers, followed by regularization and pooling, and fully connected layers.

Convolution: A convolution layer contains filters (or kernels) that slide along the input data and perform convolutions, which is a dot product of the input and the filter. The output of a convolution layer is a feature map, which is a representation of the

input. By applying convolutions, the convolution layer captures local dependencies in a sequence. The option “strides” determines the steps taken along the data samples as the filter moves across the input array. The option “padding” is used to control the output length of the feature map, ensuring that the shape of the output matches the input array by adding zeros around the input. This is done to prevent the loss of border information. Given an input sequence $\mathbf{x} = [x_1, x_2, \dots, x_n]$ and a filter $\mathbf{w} = [w_1, w_2, \dots, w_k]$, the convolution operation outputs a sequence $\mathbf{y} = [y_1, y_2, \dots, y_{n-k+1}]$, where each element y_i is computed as:

$$y_i = \sum_{j=1}^k w_j \cdot x_{i+j-1} \quad (4.6)$$

for $i = 1, 2, \dots, n - k + 1$. When the output sequence are fed to the successive convolution layers, the network learns increasingly abstract features to get a comprehensive understanding of the data. The size of the convolution kernel is important, as a large kernel can pick up features in the larger area in a trace, but they are computationally expensive. For the LFN analysis, having a larger kernel is favored as the LFN is usually spread over a larger area in the trace. After the first convolution, the kernel sizes are reduced to pick out the high-level features.

Pooling: Pooling layers are incorporated to reduce the dimensionality of the feature maps without losing information and provide translation invariance. They usually follow the convolution layer, to retain or pick out high-level features while discarding less important information, thereby making the subsequent layers more efficient in learning higher-level representations. The following are the different types of pooling used in the deep neural networks.

Max Pooling: Max pooling [136] selects the maximum value within a sliding window of the feature map and captures the most prominent feature in that region. It is commonly used to downsample and reduce the dimensionality of feature maps.

Average Pooling: Average pooling [136] computes the average of all values within the sliding window, providing a smoother downsampling method by considering the overall feature intensity rather than focusing on the maximum value.

Global Average Pooling: Global Average Pooling (GAP) [137] reduces each feature map to a single value by averaging all the elements in that feature map. It is often used in a network after the convolution layer to convert the spatial dimensions into a vector before passing it to fully connected layers.

Global Max Pooling: Similar to GAP, global max pooling reduces each feature map to a single value, but it selects the maximum value across the entire feature map, emphasizing the most dominant feature present.

Regularization techniques: Regularization techniques, such as L1/L2 Regularization [138, 139], are used to prevent overtraining by adding constraints to the model’s complexity. L2 regularization penalizes large weights by adding a cost term to the loss function. Dropouts are a type of generalization where weights are randomly dropped, in other words, it randomly discards the output of the neurons, to prevent overfitting. This is especially useful in cases with a large number of parameters where overfitting is expected.

Fully connected network: Finally, the extracted features are passed through fully connected layers with appropriate activation in the final layer to make predictions. This part of the architecture takes care of mapping the information from the features to output predictions using non-linear activations.

4.3.2. Long Short-Term Memory (LSTM)

Recurrent Neural Networks (RNNs) are a type of neural network typically used for sequential data. The distinguishing feature of an RNN, compared to a feedforward network like a CNN, is its ability to take information from prior inputs and incorporate it into the current step. This is often referred to as “memory”. The memory is updated at each step based on the current output and the previous step’s information. Long Short-Term Memorys (LSTMs) are a type of RNNs developed by Hochreiter and Schmidhuber in 1997 [132]. LSTMs are designed to capture long dependencies in the time series data using “memory cells”. A memory cell consist of input, forget, and output gates that determine which information to add to the memory, remove from the memory, and determine the output, respectively. These memory cells manage the flow of information. At each time step, LSTM takes in the current input and information from the previous step in memory to produce output output.

CNN-LSTM In this type of architecture, LSTMs are combined with CNNs. The convolutional layers are used to extract features that represent the data. These feature maps are then passed through LSTM layers, which capture long-term temporal dependencies by maintaining the memory of past information. This combination of pattern recognition using convolutions and sequence learning with LSTMs is tested for the LFN classification task.

4.3.3. Model training

Model training involves multiple iterations called epochs. The dataset used for network training is typically split into two parts: the training samples, which are used during the training process, and the validation samples, which are used only for validation. In each epoch, the training samples pass through the entire network and the loss is calculated based on the difference between predicted and true labels. This step is called the forward pass. Then the loss function is propagated back and the gradients of the loss functions with each weight are calculated using the chain rule. This step, called

backpropagation, determines how much each weight has contributed to the loss. The given optimization algorithm updates the weights based on the gradients such that the loss is reduced according to the learning rate. The learning rate determines the step size of the weights updated. A large learning rate leads to a large step size but the optimal solution may be skipped, while a small learning rate makes the training process longer but likely achieves optimal solution.

This process is repeated either until the given number of epochs is reached or stopped early when the model shows convergence. The model reaches convergence when the loss begins to plateau, which implies no further improvement. Not every loss reduction means that the model is learning, sometimes it means that the model is learning too much from the data by picking up on small fluctuations, commonly called overtraining, and the generalization capacity of the model is lost. This can be observed when the loss on the validation dataset increases with the epoch. It is ideal to stop the training at this point, and it is usually done with early stopping options. Thus, by looking at the training and validation loss, it helps us understand whether the model has effectively learned the data and make accurate predictions.

4.3.4. Loss function

The loss function used for this classification is the binary cross-entropy loss function [140]. This loss function measures the performance of the model by comparing the predicted probabilities with the actual class labels. It is defined as:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)], \quad (4.7)$$

where N is the number of samples, p_i represents the prediction of i^{th} event, y_i denotes the true label of the i^{th} event.

4.3.5. Metrics

For evaluating the performance of the neural network, several metrics can be used. In this analysis, the Receiver Operating Characteristic (ROC) curve, the confusion matrix, and efficiency and purity as a function of amplitudes are used. The nomenclature used in this section is shown below:

TP	True Positive	True signal predicted as signal
TN	True Negative	True LFN-background predicted as LFN-background
FP	False Positive	True LFN-background predicted as signal
FN	False Negative	True signal predicted as background

The ROC curve

The Receiver Operating Characteristic (ROC) curve is a metric that uses a probabilistic approach tool for evaluating the classification performance by illustrating their ability to differentiate between two classes. The True Positive Rate (TPR) is plotted against the False Positive Rate (FPR) across various thresholds, where the TPR and FPR are defined below:

$$TPR = \frac{TP}{TP + FN} \quad (4.8)$$

$$FPR = \frac{FP}{FP + TN} \quad (4.9)$$

The Area Under the Curve (AUC) of the ROC curve quantifies overall separability, with a higher AUC denoting a better separation between true positives and true negatives. The ideal point in a ROC curve is at the top-left corner where TPR equals 1 and FPR equals 0, thus the AUC is 1. AUC values range from 0 to 1, with 1 indicating perfect classification, 0.5 indicating no discriminatory power, and 0 representing perfect inverse classification.

Confusion matrix

A confusion matrix summarizes the correct and incorrect predictions by comparing the actual and predicted classes for a given threshold, as shown in Fig. 4.14. The top-left quadrant shows the TN, the top-right quadrant is the FP, the bottom-left quadrant is the FN, and the bottom-right quadrant is the TP.

True Class	Background (0)		
	Signal (1)		
		TN	FP
		FN	TP
		Background (0)	Signal (1)
		Predicted Class	

Figure 4.14.: Illustration of the confusion matrix.

Signal efficiency and purity

As the amplitude decreases, it becomes increasingly difficult to distinguish between LFN and signal events, as their shapes become almost indistinguishable. Therefore, evaluating the classification performance as a function of amplitude is essential. However, the ROC curve and the confusion matrix don't tell us the performance as a function of amplitudes. Thus, efficiency and purity as a function of amplitudes are used.

Recall or signal efficiency: Recall or signal efficiency is defined as the ratio of the true positives to all positives. So if the TP is higher, then the ratio would be higher, increasing the efficiency. For each amplitude, this quantity gives an estimate of the signal efficiency at that amplitude. For a given amplitude a , it is written as

$$\text{Efficiency}(a) = \frac{\text{TP}(a)}{\text{TP}(a) + \text{FN}(a)}. \quad (4.10)$$

Precision or signal purity: Precision or signal purity is the ratio of the true positives to all predicted positives. If there is no FP, the purity would be 1; if the FP is higher, then the purity would be reduced. Thus, signal purity shows the impact of LFN-background leakage on signal identification. For a given amplitude a , the signal purity as a function of amplitude, is written as

$$\text{Purity}(a) = \frac{\text{TP}(a)}{\text{TP}(a) + \text{FP}(a)}. \quad (4.11)$$

Signal efficiency and purity are crucial in the dark matter (DM) search analysis. Better signal efficiency enables sensitivity to search of low-mass DM, while higher purity improves LFN-background rejection, enabling the setting of limits on smaller cross-sections. As will be explained in Sec. 5.4, the DM exclusion limits are set using Optimum Interval (OI) method assuming a signal-only hypothesis. Therefore, in this analysis, purity is given higher importance, as the objective is to minimize the presence of LFN events in the final energy spectrum.

4.3.6. Class activation maps

The Class Activation Maps (CAMs) [141] are used to understand which features from the convolution layers were given high importance during classification. To enable this, a GAP layer is applied between the last convolutional layer and the fully connected part of the network. For an output of the convolution layer A_i of the i^{th} filter, the Class Activation Map is calculated as

$$\text{CAM}(t) = \sum_i w_i \cdot A_i(t), \quad (4.12)$$

where w_i is the weight of the i^{th} filter. When these activation maps are overlaid on the trace, they highlight the temporal regions in the trace that were crucial for the decision.

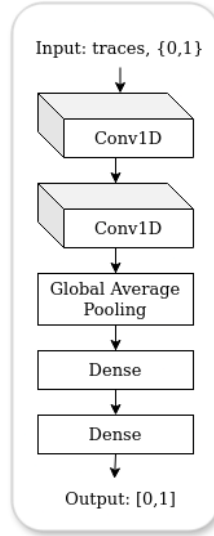


Figure 4.15.: A schematic of the CNN architecture for LFN classification. The input is the raw trace and the targets are 0s and 1s. It consists of two convolution layers followed by the global average pooling layer and a fully connected network consisting of dense layers, with the final dense layer for determining the output between 0 and 1.

4.4. LFN classification using CNN

Prior to the hyperparameter optimization (detailed in Appendix A), various architectures were tested. This section presents the best-performing model, which was selected through iterative testing of different architectures and a manual search for hyperparameters. The results from this model serve as a benchmark and are compared with the results from the hyperparameter optimization.

Fig. 4.15 illustrates the CNN architecture used in this section for LFN classification using single-, PT-, and multi-channels. The model consists of two convolutional layers: The first convolution layer contain 32 filters with a kernel size of 500, followed by the second convolution layer with 64 filters with a kernel size of 100. For both convolution layers, padding is set to “same” to maintain the spatial dimensions, Rectified Linear Unit (ReLU) [142] is used as the activation function. Following these convolutional layers, a GAP layer reduces each feature map to a single value, before feeding into a dense layer of 64 units with ReLU activation. The usage of GAP is beneficial in this case because the outputs can be used to get the activation maps [141] discussed in Sec. 4.3.6. A dropout layer with a 50 % rate is applied for regularization before the final layer, which is a dense layer with a single unit and sigmoid activation [143] to produce binary probabilistic predictions. The schematic of the architecture is shown in Fig. 4.15. The model is compiled with the Adam optimizer [144] with a

learning rate of 0.001, binary cross-entropy (explained in Sec. 4.3.4) as the loss function, and accuracy as the evaluation metric. Table 4.1 shows the model summary of this architecture.

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 4096, 32)	16,032
conv1d_1 (Conv1D)	(None, 4096, 64)	204,864
global_average_pooling1d (GlobalAveragePooling1D)	(None, 64)	0
dense (Dense)	(None, 64)	4,160
dropout (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 1)	33

Table 4.1.: The model summary of the CNN classifier.

A threshold is placed on the output value, also called network score or signal probability, such that an event with an output greater than this threshold value will be classified as a signal, and less than the score will be classified as an LFN-background.

4.4.1. Single-channel

This section presents the classification of LFN using single-channel approach, in other words, using information from raw traces of one phonon channel at a time. Prediction based on a single-channel trace is made independently of the other channels i.e. depending solely on the information from a single trace without considering data from additional channels, making the classification channel-independent. However, this presents challenges due to the inadequately defined pulse template, which is the same for all four channels as is shown in Fig. 4.5, and doesn't capture the variations in the pulse shape based on the interaction position within the detector. Consequently, the signal samples may not adequately describe the real data. The training dataset comprises 80000 samples with equal proportions of signal and LFN-background, with 20% of these samples randomly selected as the validation dataset. This selection process is repeated for each training epoch to ensure that the validation is conducted on different subsets at each step. The test dataset contains 131983 (37983 signal samples, and 94000 LFN-background) samples which are not used during training, and thus are used to evaluate the performance of the network. There are four phonon channels in the detector, so the single-channel classifier is applied on individual channel traces. If all four channels are predicted as signal, the final classification for the event is considered a signal. If any channel predicts background, the final classification will be LFN-background, ensuring a more cautious approach to event classification.

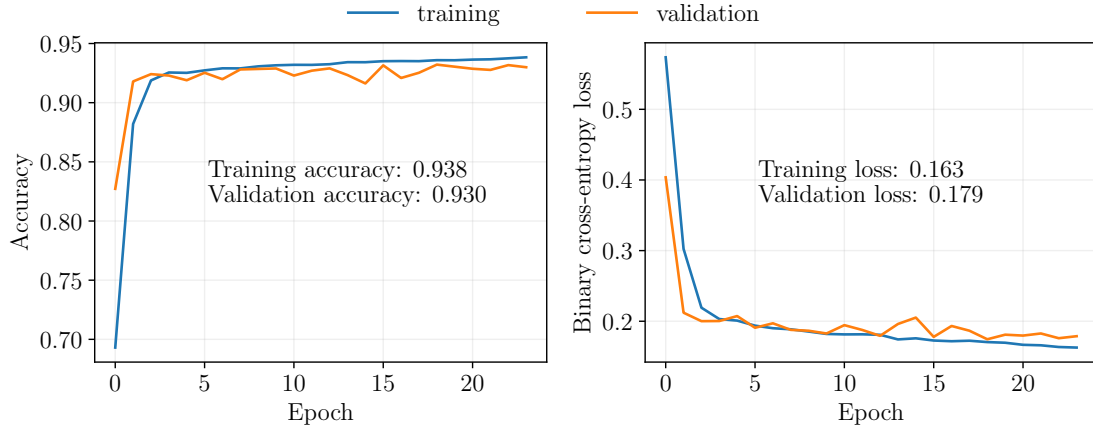


Figure 4.16.: Training and validation accuracy (left) and loss curves (right) of the single-channel classifier training plotted against epochs. The training set is shown in blue and the validation is shown in orange, with the accuracy and loss of each set in the text.

Classifier training

Figure 4.16 presents the training of the single-channel CNN. The accuracy as a function of epochs is shown on the left, where both training and validation accuracy rapidly increase during the initial epochs, stabilizing from epoch 5 onwards. The binary cross-entropy loss as a function of epochs is shown on the right, where both training and validation losses decrease sharply during the first few epochs, and then losses continue to decline gradually, with some oscillations observed in the validation loss. Early stopping was employed to avoid over-training after the stabilization. The final training and validation accuracy of this model is 93.8 % and 93 %, respectively.

Classifier performance

Figure 4.17 displays the network performance using the test samples which the network had not seen during training. On the left is the ROC curve and the signal probability distribution is on the right. The ROC curve achieves an AUC of 0.993, indicating good performance in classification. The signal probability distributions for LFN-background (red) and signal (blue) classes show good separation between the classes, indicated by the LFN-background distribution peaking at lower network scores and the signal distribution peaking at higher scores with minimal events in the overlapping regions.

Figure 4.18 shows the confusion matrix of the classification task using the test samples, placing a cut on signal probability at 0.5 and 0.9. The matrix shows that 94.7 % and 99.1 % of LFN-background events are properly identified by placing a threshold at 0.5 and 0.9, respectively. Similarly, 96.7 % and 90.8 % of signal events are correctly identified

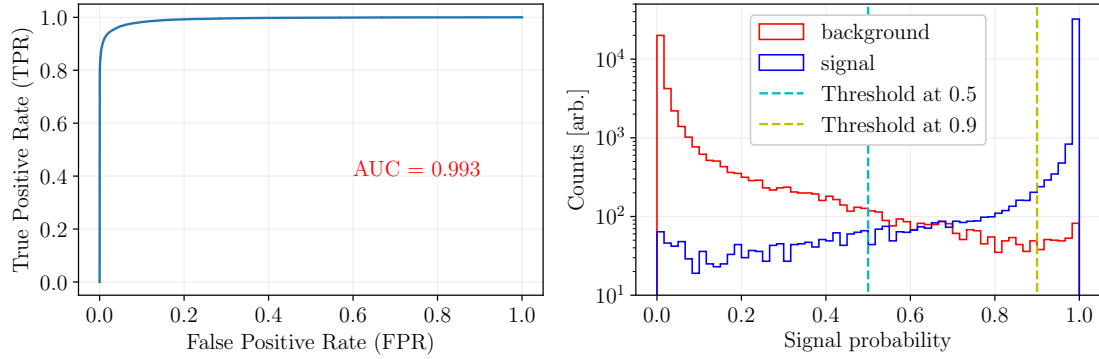


Figure 4.17.: The ROC curve (left) and the signal probability distribution (right) for the single-channel classifier tested on unseen test samples. The threshold on the signal probability at 0.5 (cyan) and 0.9 (yellow) is shown in dotted lines. The AUC of the ROC is shown in the text.

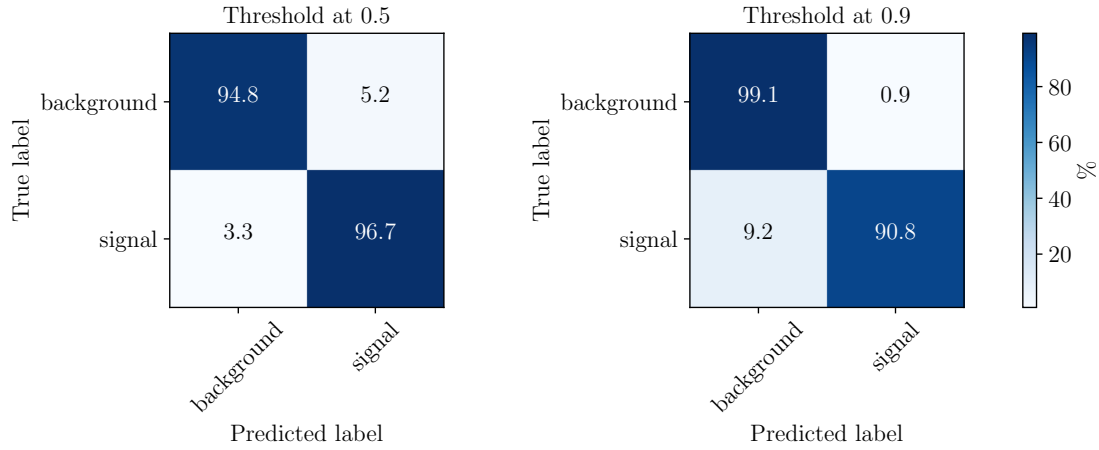


Figure 4.18.: The confusion matrix of the single-channel classifier based on an unseen test samples by placing a threshold on signal probability at 0.5 (left) and at 0.9 (right). The values are in percentage.

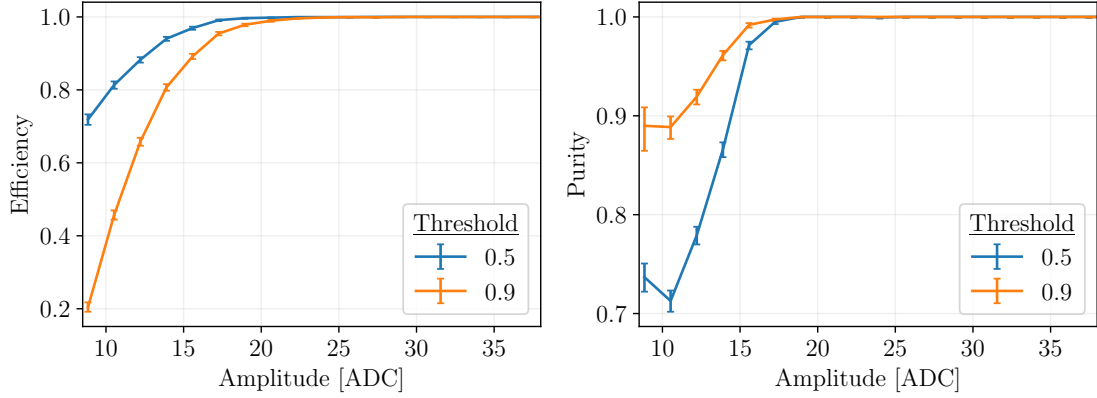


Figure 4.19.: The signal efficiency (left) and purity (right) of the single-channel classifier when the threshold on signal probability is placed at 0.5 (blue) and 0.9 (orange) as a function of the channel amplitude. The error bars denote the statistical uncertainty.

as signal when placing a threshold at 0.5 and 0.9, respectively. These results indicate a high level of accuracy for the model in distinguishing between LFN-background and signal events, with a particularly strong performance in identifying LFN-background events correctly when placing the threshold at 0.9.

For the classification using the test samples, Fig. 4.19 shows the signal efficiency on the left and the signal purity on the right, calculated using Eq. 4.10 and Eq. 4.11, respectively. Both the signal efficiency and the signal purity decrease with decreasing amplitude. The signal efficiency shows 100 % for both cuts at 0.5 and 0.9 for amplitudes above ~ 25 ADC, below which it decreases. The signal purity shows 100 % purity for amplitudes until ~ 20 ADC, below which it decreases. The classification with a cut at 0.5 has a higher signal efficiency than when the cut is placed at 0.9, while the signal purity is higher when the cut is placed at 0.9, which is explained by the lesser number of LFN events with high network scores. A notable dip in the purity is observed around 12 ADC, where the purity decreases temporarily before rising again. This dip is caused by more LFN-background events in this region, leading to a slight reduction in the purity before stabilizing.

As seen in the confusion matrix in Fig. 4.18, the amount of FN events is 3.3 % and 9.2 % when placing the threshold at 0.5 and 0.9 respectively, where the true signal events were misclassified as LFN-background. Three randomly selected raw traces corresponding to false negatives in both threshold cases are shown in Fig. 4.20, with a moving average applied to show the prominent features. As it can be observed in the left and the middle plot, the signal feature is small, i.e. the amplitude is very low. The trace on the right has noise in the signal's rising edge, leading to a

triangular-looking trace with slow rise-time which is typical of LFN. Therefore, it can be concluded that the low amplitudes coupled with noise in the signal rising feature lead to misidentification and are challenging for the model to accurately identify signal events.

Figure 4.21 shows the class activation maps (CAMs) corresponding to the false negative traces shown in Fig. 4.20. These CAMs, calculated using Eq. 4.12, show the temporal regions of the raw traces that were most important to the model’s decision. It can be seen that the trace on the left was predicted as LFN-background mainly based on the second bump at around 5 ms. The trace on the right shows that the rising edge of the trace was given high importance as we predicted before. The middle plot shows a more uniform activation spread, with lower activation throughout the trace, possibly explaining why the model failed to identify key signal characteristics.

The confusion matrix in Fig. 4.18 shows that 5.3% and 0.9% of events are classified as false positive when placing the threshold at 0.5 and 0.9, respectively. Three randomly selected raw traces corresponding to false positives in both threshold cases are shown in Fig. 4.22, with the moving average applied to show the prominent features. The traces exhibit distinct signal-like features, such as sharp rising edges, which resemble signal events. The class activation maps for these traces are shown in Fig. 4.23. The plot on the left and the middle show moderately high activations near the rising edge of the signal feature, but overall show low activations throughout. The right plot displays activations spread across the entire trace with no clear structure, highlighting the model’s uncertainty.

It is crucial to note that LFN-background samples may contain signal events, as they are LFN enriched samples. So the possibility that these false positives contain true signal events cannot be ruled out.

Comparison of model performance with pre-processed input data

Thus far, the results presented were based on the single-channel classifier using unprocessed data for both training and testing. This section introduces the classifier trained on traces processed using techniques such as low-pass filtering and moving average, as outlined in Sec. 4.1.3. The performance of these classifiers is compared with that of the classifier trained with unprocessed data, as shown in Fig. 4.24. The efficiency and purity as a function of amplitudes for the classifier using processed data compared with the classifier using the unprocessed raw data are shown on the left and right, respectively. The efficiency curves show that the low-pass filtered input performs slightly better than the unprocessed and moving-averaged inputs in low amplitudes. In contrast, the purity plot demonstrates that the unprocessed data provide better purity than the processed inputs. As referenced earlier in Sec. 4.1.3, while pre-processed traces display more pronounced features, this has little effect on the network’s performance. Given that purity is more important than efficiency in this analysis, as mentioned previously, the unprocessed data offers optimal performance with minimal loss in efficiency without the need for additional processing.

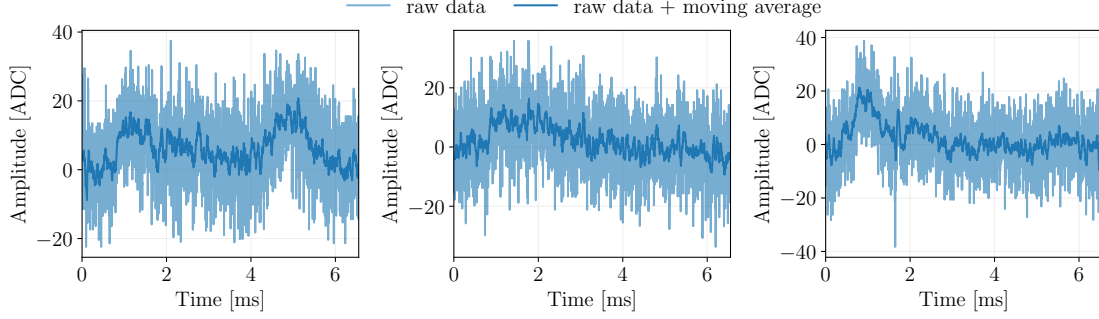


Figure 4.20.: Raw traces of randomly selected false negatives of the single-channel classifier shown in light blue and with the moving average shown in dark blue.

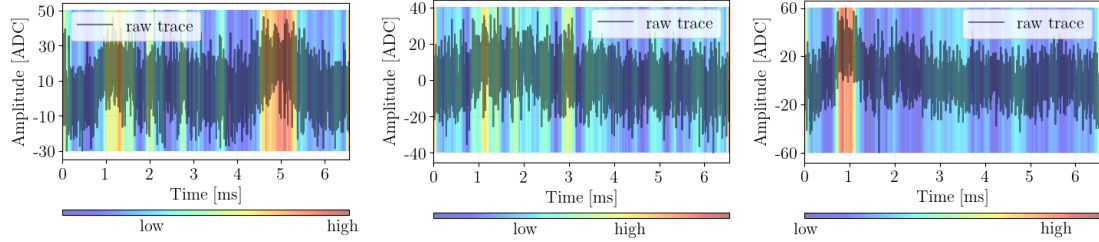


Figure 4.21.: The activation maps of the randomly selected false negatives of the single-channel classifier shown in Fig. 4.20. The raw traces are shown in transparent black over which the color maps representing the feature importance for the prediction is overlaid. The activation maps are calculated using Eq. 4.12.

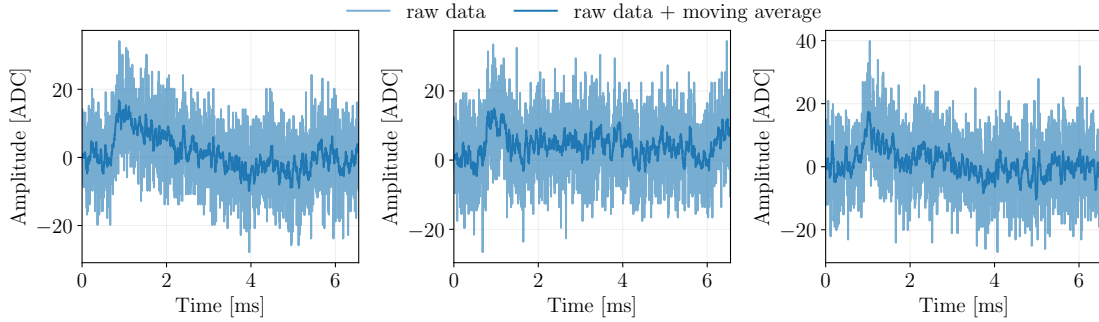


Figure 4.22.: Raw traces of randomly selected false positive of the single-channel classifier shown in light blue and with the moving average shown in dark blue.

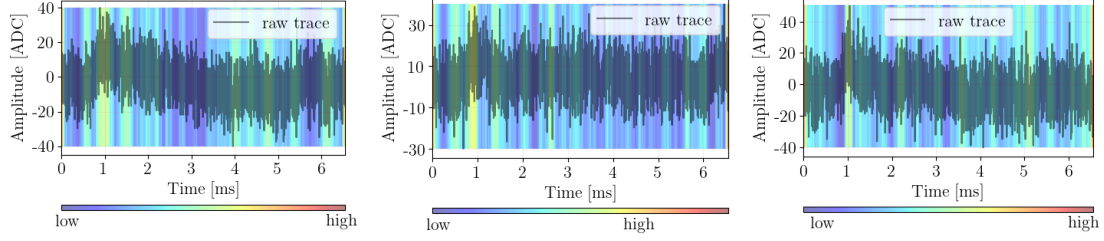


Figure 4.23.: The activation maps of the randomly selected false positives of the single-channel classifier shown in Fig. 4.22. The raw traces are shown in transparent black over which the color maps representing the feature importance for the prediction are overlaid. The activation maps are calculated using Eq. 4.12.

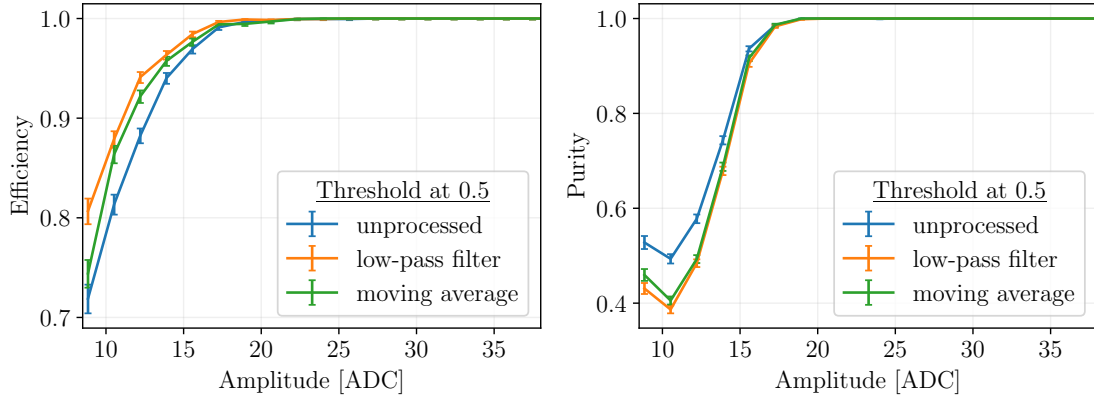


Figure 4.24.: Comparison of single-channel classifier performance trained on processed data, such as low-pass filtered (orange) and moving average (green), versus unprocessed input data (blue). Signal efficiency (left) and purity (right) are evaluated on test samples processed in the same way as their respective training data, with a threshold of 0.5 on signal probability. The error bars denote the statistical uncertainty.

4.4.2. PT-channel

The LFN removal using the Phonon Total (PT) channel is explained in this section. This method is more reliable than the single-channel approach, as the pulse template generated for PT is representative of signal events. Most of the CDMSlite analysis was done using the PT channel, including the pulse shape cut for removing the LFN events discussed in Sec. 3.4.2. To obtain the PT channel for events, individual channels were weighted using relative calibration determined based on their respective properties. The weighted traces were then summed to form the PT channel. For the signal, baseline noise was similarly weighted and summed, and was injected with the PT pulse template scaled by a desired amplitude. The LFN event traces were weighted using the relative calibration and summed to obtain the LFN-background events. The training dataset comprised 40000 samples with equal proportions of signal and LFN-background, with 20% of these samples randomly selected as the validation dataset. The test dataset contains 27178 (3678 signal samples, and 23500 LFN-background) samples which are not used during training, and thus are used to evaluate the performance of the network. This approach was data-limited as it reduced the number of samples by a factor of four compared to using single-channel data, where individual channels were treated independently.

Classifier training

The PT-channel CNN training was completed by early stopping at 29 epochs as shown in Fig. 4.25. It can be seen that the training and validation loss stayed close throughout the training, with some fluctuations.

Classifier performance

Figure 4.26 shows the ROC curve on the left using the test sample, and has an AUC of 0.992, which indicates very high separability. The network score distribution of the test samples on the right of Fig. 4.26 shows two distinct peaks at 0 and 1 for LFN-background and signal samples. There is a peak at 0 consisting of signal events exhibiting misidentification leading to some signal loss. These events correspond to the lowest amplitudes, which is reflected in the signal efficiencies.

Figure 4.27 shows the confusion matrix in percentage when placing a cut on the signal probability distribution at 0.5 and 0.9. The 98.9 % of the LFN-background and 89.1 % of the signal events are correctly identified for the threshold at 0.5. 99.7 % of the background and 84.2 % of the signal events are correctly identified for the threshold at 0.9. One has to note that the background is not true background but from a LFN-background enriched samples, and this is important when looking at FP which could contain potential signal events. In this classification model, the LFN-background rejection is better than the signal identification.

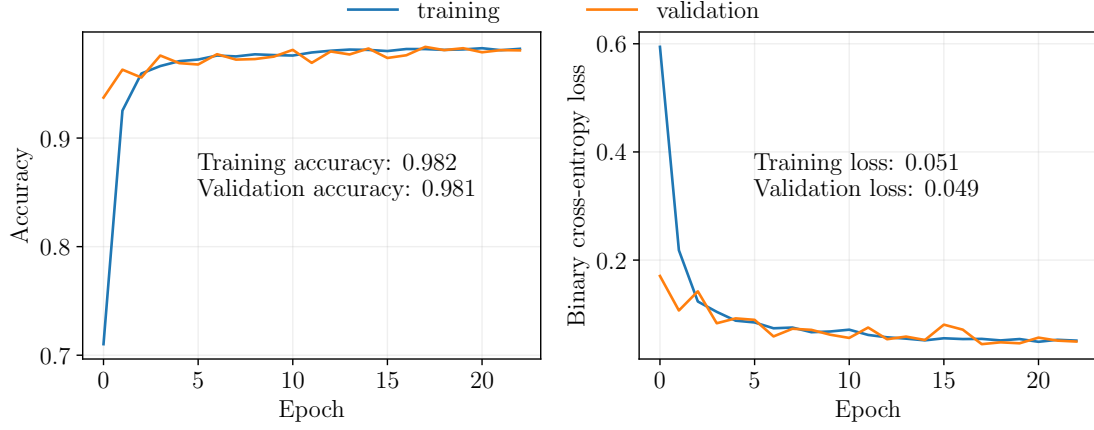


Figure 4.25.: Training process of the PT-channel classifier shows the accuracy (left) and loss (right) per epoch of the training (blue) and validation (orange) set. The training is stopped using early stopping when the validation loss plateaus.

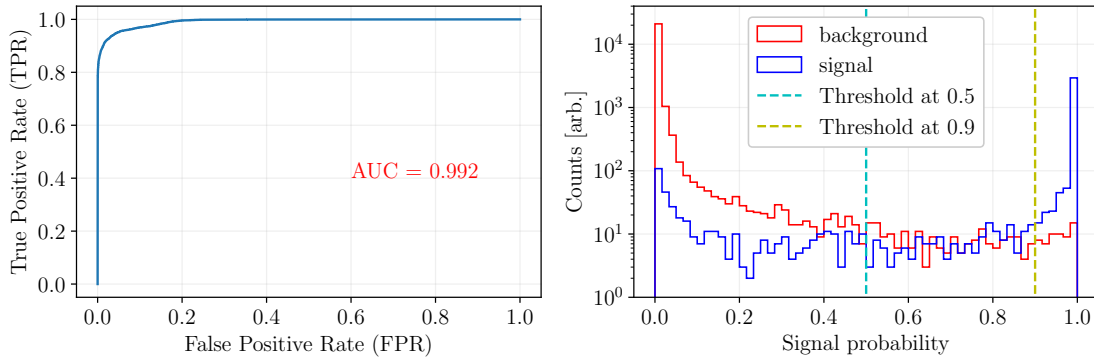


Figure 4.26.: The ROC curve (left) and signal probability distribution (right) of the PT-channel classifier based on an unseen test samples. The threshold on the signal probability at 0.5 (cyan) and 0.9 (yellow) is shown in dotted lines. The model achieves an AUC of 0.992 as seen in the text.

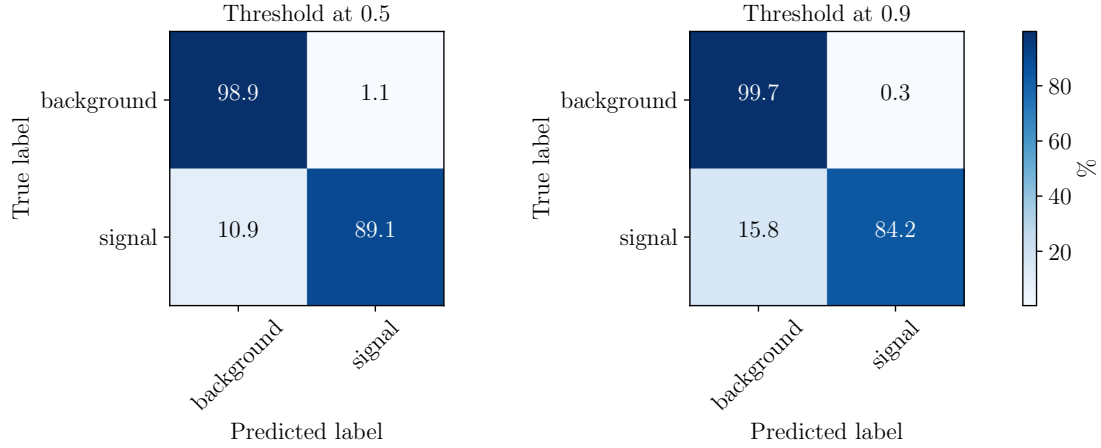


Figure 4.27.: Confusion matrix of the PT-channel classifier of the test samples for the threshold at 0.5 (left) and 0.9 (right). The values are in percentage.

Figure 4.28 shows the signal efficiency on the left and signal purity on the right derived from Eq. 4.10 and Eq. 4.11, respectively. The signal efficiency and the purity decrease as the amplitude decreases. For a threshold at 0.5 and 0.9, the signal efficiency is 100 % for amplitudes from ~ 45 ADC, and below which the efficiency drops as the amplitude decreases. The signal purity is 100 % for amplitudes from ~ 35 ADC for a cut placed at 0.5 and at 0.9, below which the purity drops drastically. A notable dip is observed around 22 ADC, where the purity decreases temporarily before rising again. This dip is caused by more LFN-background events in this region, leading to a slight reduction in purity before stabilizing.

The confusion matrix in Fig. 4.27 shows that 10.9 % and 15.8 % of events are classified as false negatives when placing a threshold of 0.5 and 0.9 on the signal probability, respectively. Fig. 4.29 shows randomly chosen false negative traces in both cut scenarios applied to show features underneath. The left plot shows a trace with the rising edge of the signal that is in combination with a noise feature. The middle and the right plots show events with a very small signal that cannot be seen predominantly. The CAMs for these traces is shown in Fig. 4.30. In all the three plots, the activations highlight the rising edge of the pulse, corresponding to the primary signal feature. However, the left and middle plots also display activations in neighboring time bins. Notably, the middle plot shows a higher activation for a bump near 2 ms, which suggests an additional feature in the trace that the model is considering in its classification.

The confusion matrices in Fig. 4.27 show that 1.1 % and 0.3 % of events are classified as false positives when placing a cut at 0.5 and at 0.9, respectively. Raw traces of randomly chosen events from these false positives in both threshold cases are shown

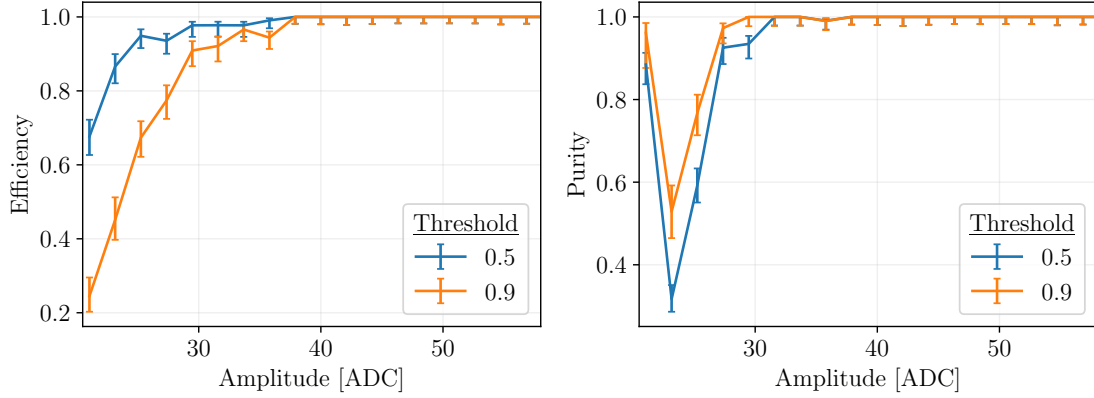


Figure 4.28.: The signal efficiency (left) and purity (right) of the PT-channel classifier of the test samples when the threshold is placed at 0.5 (blue) and 0.9 (orange). Both efficiency and purity were made from the test samples with amplitudes between 20 and 70 ADC units. The error bars denote the statistical uncertainties.

in Fig. 4.31 with a moving average to show the underlying features. It can be seen that all three traces have rising features like signal events, with the left plot showing noise near this feature.

The CAMs for these false positive traces are shown in Fig. 4.32 overlaid on the raw traces. All three CAMs for the traces show high activations near the pulse’s rising edge, indicating that the model predicts based on this signal-like feature. The activation map in the left plot shows broader activations across the trace, suggesting some level of uncertainty or influence from additional features. It is crucial to note that LFN-background samples may contain signal events, as they are LFN-enriched samples. So the possibility that these false positives containing a true signal events cannot be ruled out.

Comparison of model performance with pre-processed input data

So far, the results have been obtained from the PT-channel classifier using unprocessed data for training and testing. In this section, the classifier trained on traces processed using techniques like the low-pass filter and moving average discussed in Sec 4.4 is presented. Fig. 4.33 compares the performance of the classifier by plotting the results obtained using the unprocessed input data and pre-processed data (low-pass filter and moving average) in terms of efficiency (left) and purity (right) as a function of amplitude when placing a threshold at 0.5. Both the unprocessed and low-pass filtered data achieve higher purity compared to the moving average, though the latter provides slightly better efficiency. In this case, both the unprocessed data and low-pass filtered are favored.

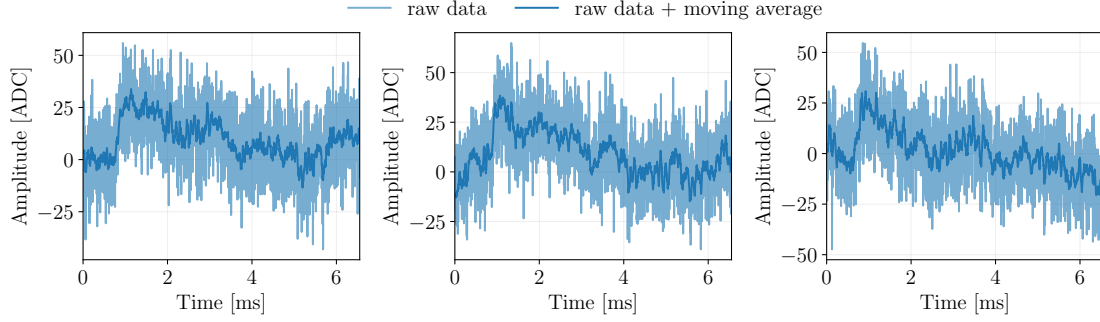


Figure 4.29.: Raw traces of randomly selected False Negatives (FN) of the PT-channel classifier shown in light blue and with the moving average shown in dark blue.

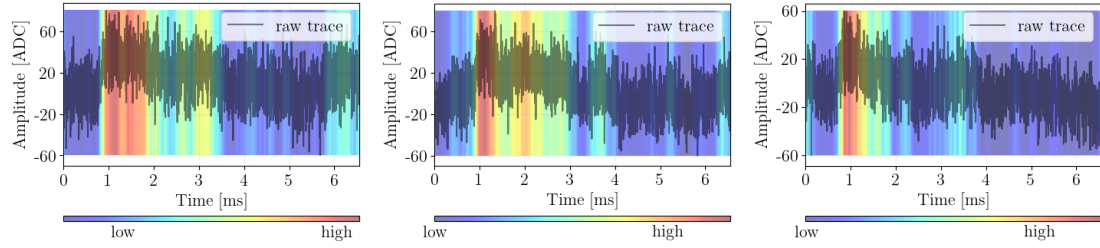


Figure 4.30.: The activation maps of the randomly selected false negatives in both threshold cases of the PT-channel classifier shown in Fig. 4.29. The raw traces are shown in transparent black over which the color maps representing the feature importance for the prediction are overlaid. The activation maps are calculated using Eq. 4.12.

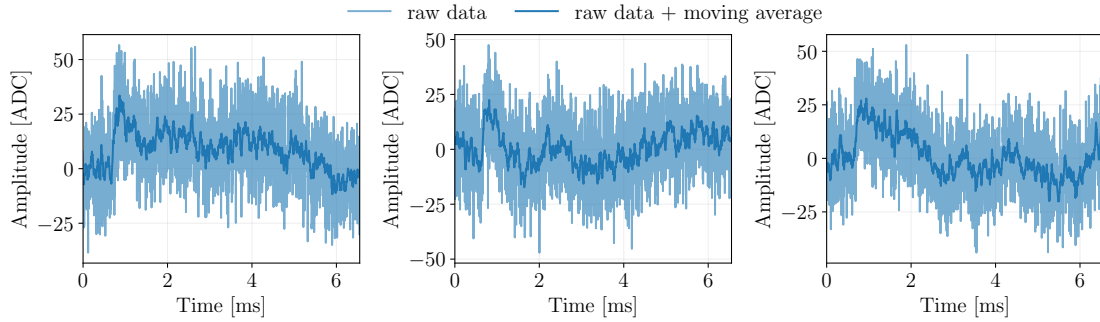


Figure 4.31.: Raw traces of randomly selected False Positives (FP) of the PT-channel classifier shown in light blue and with the moving average shown in dark blue.

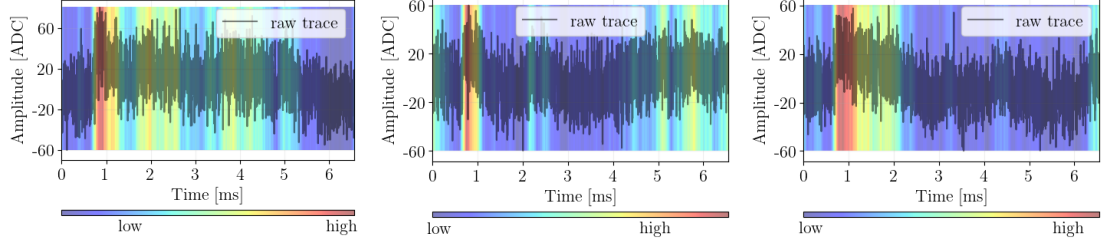


Figure 4.32.: The activation maps of the randomly selected false positives of the PT-channel classifier shown in Fig.4.31. The raw traces are shown in transparent black over which the color maps representing the feature importance for the prediction are overlaid. The activation maps are calculated using Eq. 4.12.

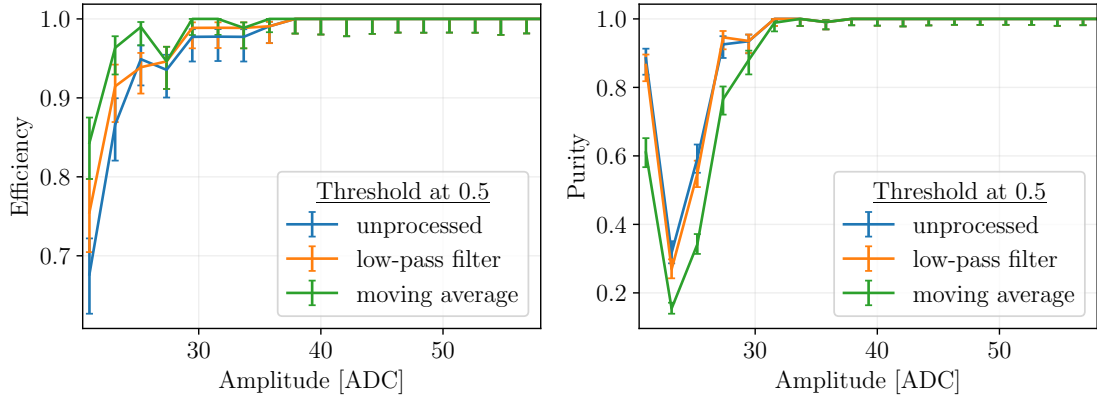


Figure 4.33.: Comparison of the PT-channel classifier performance trained on processed data, such as a low-pass filter (orange) and a moving average (green), versus the unprocessed input data (blue). The signal efficiency (left) and purity (right) are evaluated on test samples processed in the same way as their respective training data, with a threshold of 0.5 on the signal probability. The error bars denote the statistical uncertainty.

4.4.3. Multi-channel

This section discusses the approach of multi-channel classification for removing LFN events by utilizing all four channels simultaneously. Unlike the single-channel approach, where each channel is analyzed independently, this method combines information from all four channels. By processing the combined input, the model can better capture the spatial and temporal correlations across the channels, leading to a more robust determination of signal and LFN events. The training dataset comprised 40000 samples with equal proportions of signal and LFN-background, with 20% of these samples randomly selected as the validation dataset. The test dataset contains 26086 (6487 signal samples, and 19599 LFN-background) samples which are not used during training, and thus are used to evaluate the performance of the network. The approach also reduces the number of samples by a factor of four compared to a single-channel model, similar to the PT classification.

Classifier training

Figure 4.34 presents the training performance of the multi-channel CNN classifier showing the accuracy on the left and loss on the right at each epoch of the training. The training and validation loss stays close throughout the training and decreases as the training progresses with some fluctuations. The training stopped at 18 epochs by early stopping.

Classifier performance

The performance of the trained network is evaluated using test samples generated with the SUPR method, the samples that the network hasn't seen before, and is shown in Fig. 4.35. The ROC curve on the left shows the separability performance for various thresholds and has an AUC of 0.981. The signal probability distribution on the right shows distinct peaks of LFN-background and signal events at 0 and 1, respectively, with minimal events in between. This is also reflected in the confusion matrix in Fig. 4.36. With the test samples and a cut on the signal probability at 0.5, 99.5% of LFN-background and 81.7% of signal events are correctly identified, and for a cut on signal probability at 0.9, 99.9% of LFN-background and 72.1% of signal events are correctly identified. One has to note that the LFN-background samples are LFN-enriched samples and it is uncertain whether the samples represent a true signal-free background. The amplitudes of the test samples generated for the multi-channel model differ from those of the PT-channel, making a direct comparison of their confusion matrices incompatible. However, efficiency and purity, calculated as functions of amplitude, are suitable for comparison.

The signal efficiency and purity as a function of PT amplitude for a cut on signal probability at 0.5 and at 0.9 is shown in Fig. 4.37. The PT amplitude is used in this case as the classification is done event by event, and the channel-wise

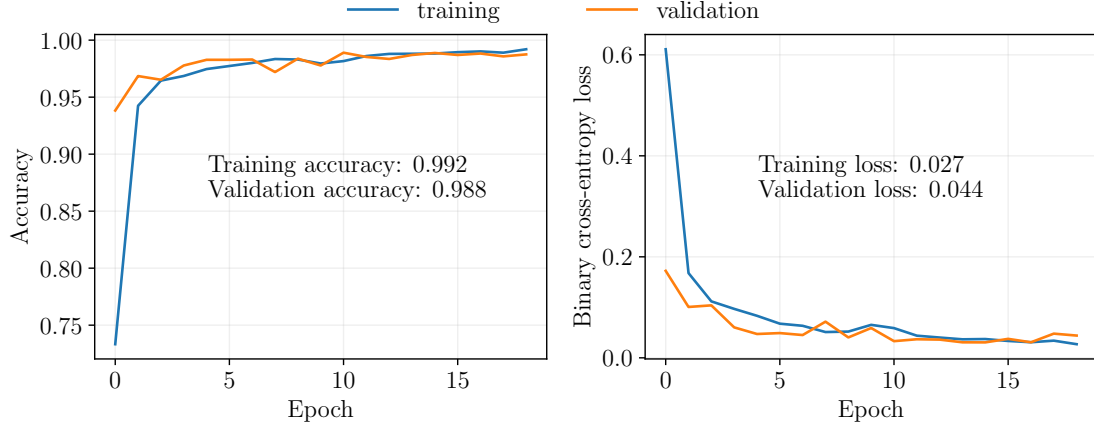


Figure 4.34.: Training performance of the multi-channel classifier as the function of epoch. Accuracy (left) and binary cross entropy loss (right) for the training (blue) and validation (orange) datasets per epoch are shown.

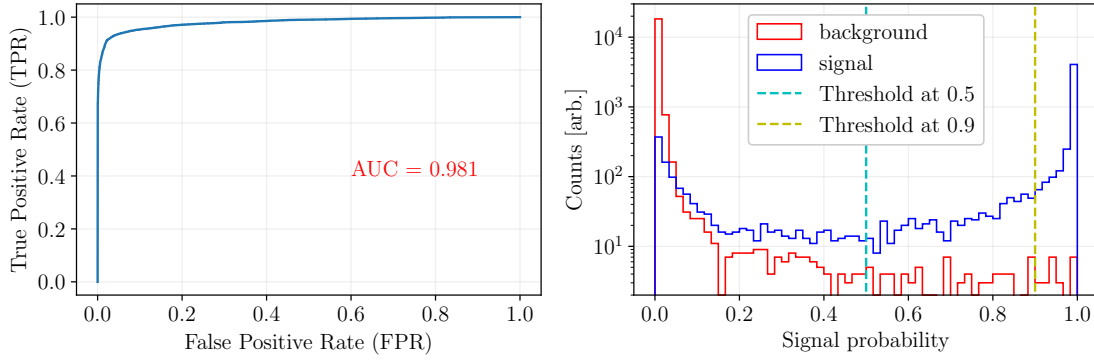


Figure 4.35.: The ROC curve (left) and signal probability (right) distribution of the multi-channel classifier with the test samples. The AUC of 0.981 shows the good separability of the classes. The signal probability distribution shows the LFN-background samples in red and the signal samples in blue. The threshold on the signal probability at 0.5 (cyan) and 0.9 (yellow) is shown in dotted lines.

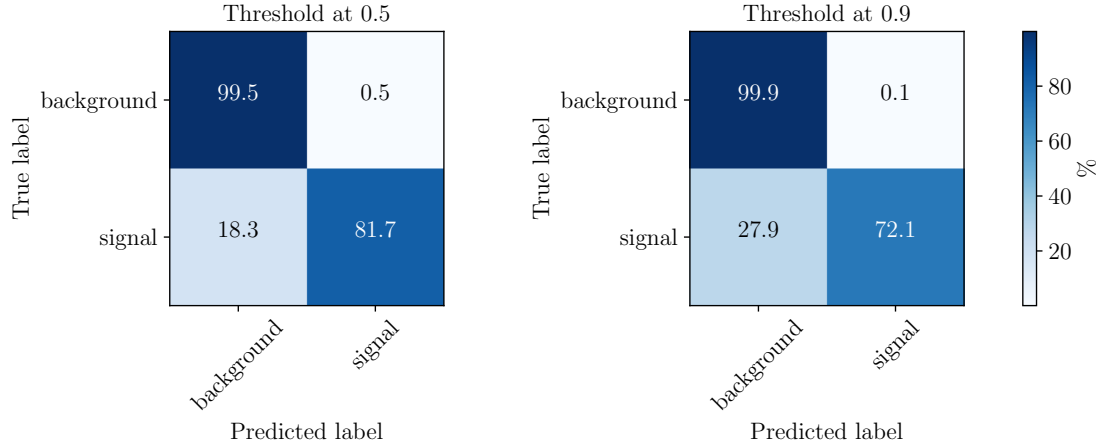


Figure 4.36.: Confusion matrix of the multi-channel classifier using the test samples for a threshold placed on the signal probability at 0.5 (left) and at 0.9 (right). The values are in percentage.

amplitude is not required in this case, as the model takes all four channels into account simultaneously.

The signal efficiency and the signal purity reduce as the amplitude reduces. Above ~ 45 ADC, the signal efficiency is 100% for both cut values at 0.5 and 0.9 on signal probability. The signal purity is 100% for amplitudes above ~ 32 ADC, below which the purity decreases for both the cut values, with 0.9 cut performing better than 0.5 cut. The confusion matrix in Fig. 4.36 shows 18.3% and 27.9% of events are false negatives when placing a threshold at 0.5 and 0.9. Raw traces of three randomly chosen events from these false negatives in both threshold cases are shown in Fig. 4.38 with a moving average applied to highlight the underlying features. The trace on the left shows a signal peak on the PB channel, and the other channel do not have signal features observable by eye. The middle and the right plots show that noise fluctuations are too high that they submerged the underlying injected signal event, possibly leading to misidentification.

The confusion matrix also shows that 0.5% and 0.1% of events are false positives when placing a cut at 0.5 and 0.9, respectively. Raw traces of three randomly chosen events from these false positives in both threshold cases are shown in Fig. 4.39, with a moving average applied to highlight the underlying features. As can be seen, all the traces show a sharp rising edge like a signal event, with the left and the middle traces showing high noise fluctuations. It is important to note that the LFN-background samples is an LFN-enriched samples and could potentially contain signal events, which is relevant when considering false positives.

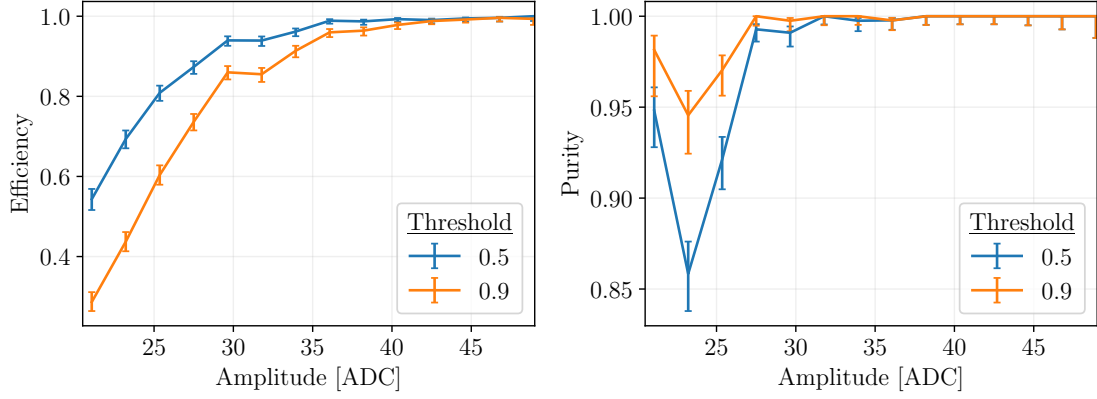


Figure 4.37.: The signal efficiency (left) and purity (right) of the multi-channel classifier calculated with the test samples. The thresholds placed at 0.5 and 0.9 are shown by the blue and orange curves, respectively. The error bars denote the statistical uncertainty.

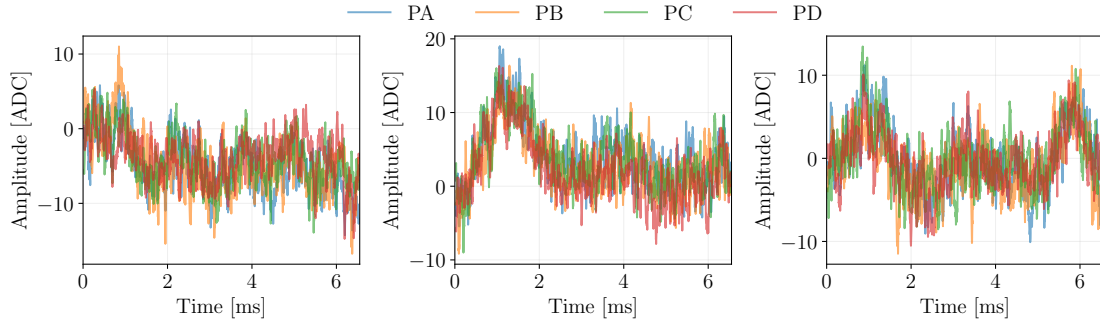


Figure 4.38.: Raw traces of randomly selected False Negatives (FN) of the multi-channel classifier, showing four phonon channels: PA (blue), PB (orange), PC (green), and PD (red), with moving average applied.

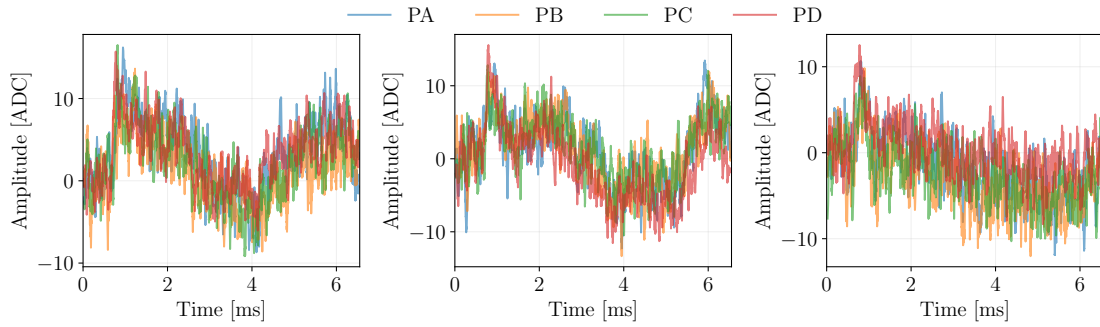


Figure 4.39.: Raw traces of randomly selected False Positives (FP) of the multi-channel classifier, showing four phonon channels: PA (blue), PB (orange), PC (green), and PD (red), with moving average applied.

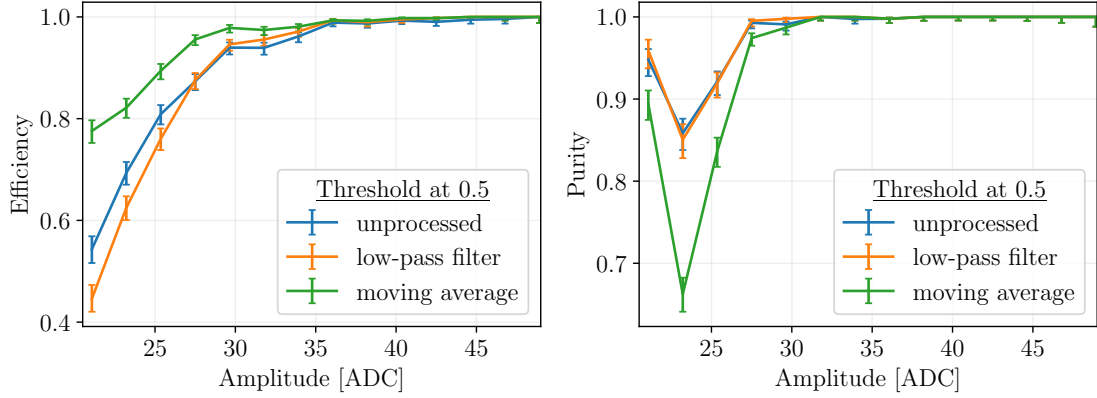


Figure 4.40.: Comparison of multi-channel classifier performance trained on processed data, such as low-pass filtered (orange) and moving average (green), versus unprocessed input data (blue). The signal efficiency (left) and purity (right) are evaluated on test samples processed in the same way as their respective training data, with a threshold of 0.5 on signal probability. The error bars denote the statistical uncertainty.

Comparison of model performance with pre-processed input data

Until now, the results provided have been based on the multi-channel classifier using unprocessed data for both training and testing. This section describes classifiers trained on pre-processed traces using a low-pass filter and moving average, as outlined in Sec. 4.1.3. Their performance is compared with the classifier based on the unprocessed data, and the results are shown in Fig. 4.40. The efficiency (left) and purity (right) as a function of amplitude when placing a threshold at 0.5 shows the performance of the multi-channel classifiers trained with unprocessed and pre-processed data. The unprocessed and low-pass filtered cases perform similarly in terms of purity, however moving average case has better efficiency than the other two cases. As having high purity is favored, the unprocessed case is favored as it also has better efficiency than the case with a low-pass filter.

4.4.4. Hyperparameter optimized CNNs

Hyperparameter optimization is an important step in building machine learning models. To achieve a state-of-the-art model for a specific task, it is essential to select the optimal set of parameters. In this analysis, several hyperparameters are considered, including the number of convolutional layers, number of filters, kernel size, strides, number of dense layers, use of normalization and regularization techniques, and dropout rates. A comprehensive hyperparameter search was performed using random search [145] and Bayesian optimization [146] for single-, PT-, and multi-channel models. The parameter combination yielding high accuracy with minimal deviation when retrained was selected. The

optimized hyperparameters for single-, PT-, and multi-channel are shown in Table. 4.2.

Layer	Layer parameters	single	PT	multi
Conv1D	number of filters (x 32)	1	3	3
	kernel size	1000	376	200
	stride	1	15	18
	kernel regularizer (L2)	0.01	0.01	0.01
	bias regularizer (L2)	-	0.01	0.01
	activation	relu	relu	relu
Conv1D	number of filters (x 32)	2	6	6
	kernel size	500	188	100
	stride	1	1	1
	kernel regularizer (L2)	0.01	0.01	0.01
	bias regularizer (L2)	-	0.01	0.01
	activation	relu	relu	relu
GAP	-			
Dense	units (x32)	2	6	6
	kernel regularizer (L2)	0.01	0.01	0.01
	bias regularizer (L2)	-	0.01	0.01
	activation	relu	relu	relu
Dropout	rate	0.1	0.276	0
Dense	units (x32)	1	3	3
	kernel regularizer (L2)	0.01	0.01	0.01
	bias regularizer (L2)	-	0.01	0.01
	activation	relu	relu	relu
Dropout	rate	0.1	0.276	0
Dense	units	1	1	1
	activation	sigmoid	sigmoid	sigmoid

Table 4.2.: The best-performing parameter combination from the hyperparameter-optimization of the CNN architecture for single-, PT-, and multi-channels. The optimization details are provided in Appendix A.1. The model was complied with the Adam optimizer using a learning rate of 0.001.

Thus far, the results from Secs. 4.4.1, 4.4.2, and 4.4.3 serve only as benchmarks. This section presents the results of the final models that use the best-performing parameters for each of the single-channel, PT-channel, and multi-channel configurations.

Single channel

Figure 4.41 compares the efficiency and purity of the hyperparameter-optimized CNN with the previously used CNN for signal-channel classification as a function of amplitude. The top row shows the results for a threshold of 0.5 on the signal probability, while the

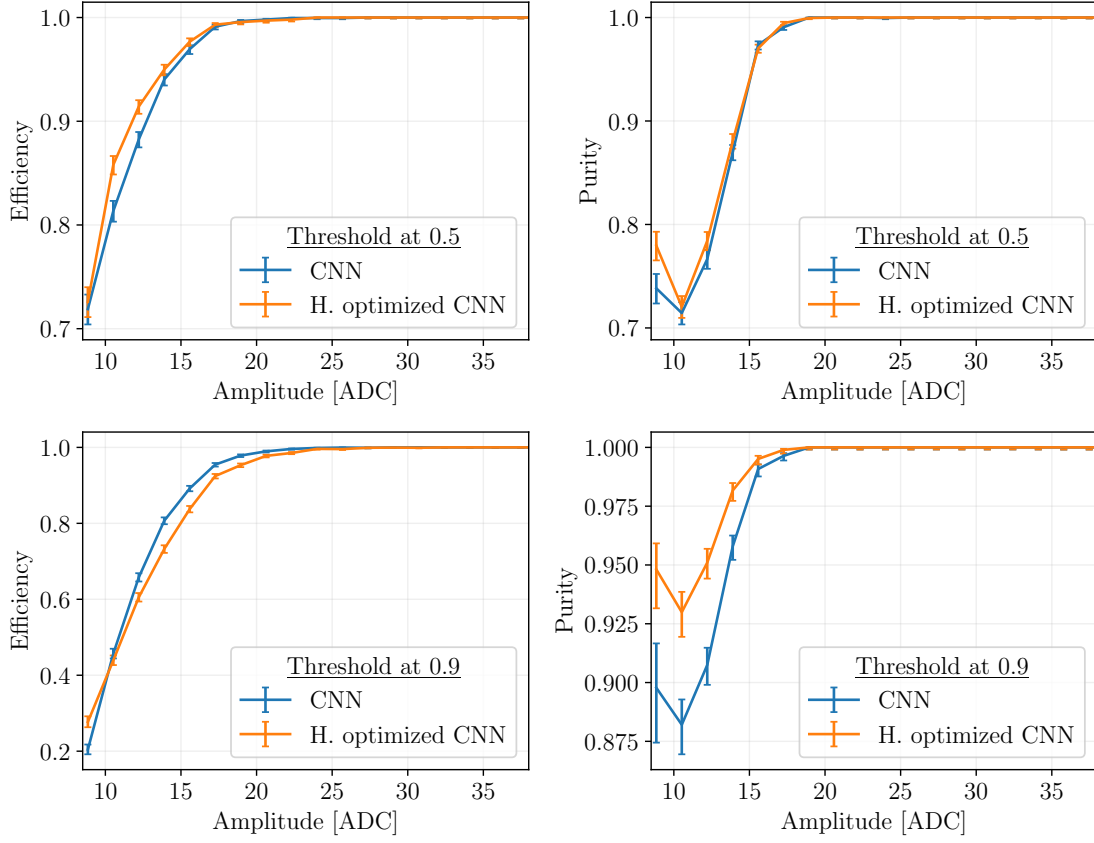


Figure 4.41.: Performance of the hyperparameter-optimized CNN (orange) with the previously discussed CNN (blue) for the single-channel classification. The efficiency (left) and purity (right) as functions of the amplitude were calculated for thresholds at 0.5 (top row) and 0.9 (bottom row), with error bars showing the statistical uncertainty.

bottom row displays results for a threshold of 0.9. The signal efficiency is comparable between both models at both thresholds. However, the hyperparameter-optimized CNN demonstrates a significant improvement in purity, particularly at lower signal amplitudes. The hyperparameter-optimized model maintains over 90 % purity for the 0.9 threshold across the amplitude range and is chosen out of the two models.

PT channel

The comparison between the hyperparameter-optimized CNN and the previously used CNN for the PT-channel classification is presented in Fig. 4.42, where the efficiency and purity are shown as a function of amplitude for both thresholds at 0.5 (top) and 0.9 (bottom). The previously used CNN demonstrates a notable improvement in efficiency at low amplitudes for both thresholds. The purity plot shows a similar performance

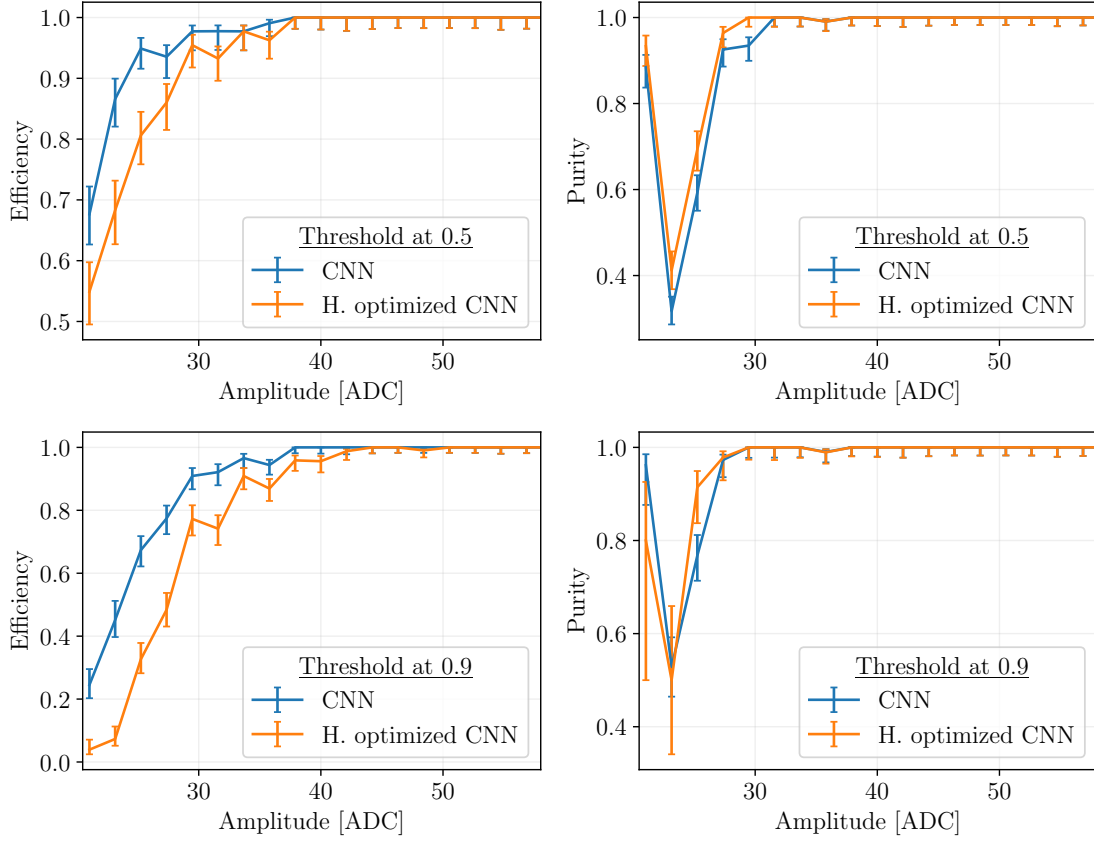


Figure 4.42.: Performance of the hyperparameter-optimized CNN (orange) with the previously discussed CNN (blue) for the PT-channel classification. The efficiency (left) and purity (right) as functions of the amplitude were calculated for thresholds at 0.5 (top row) and 0.9 (bottom row), with error bars showing the statistical uncertainty.

between these two models with the hyperparameter-optimized model exhibiting slight improvement in purity at low energies, particularly between 25 ADC and 30 ADC. The hyperparameter-optimized CNN includes bias and kernel regularization, which stabilizes the training process. Despite the loss in efficiency, the hyperparameter-optimized CNN was selected due to its purity between 25 ADC and 30 ADC and improved training stability, which was achieved through the use of kernel and bias regularization in the model.

Multi channel

For multi-channel classification, Fig. 4.43 compares the hyperparameter-optimized CNN with the previously used CNN. The efficiency curves for both the 0.5 and 0.9 thresholds show the hyperparameter-optimized CNN consistently performing better than the previ-

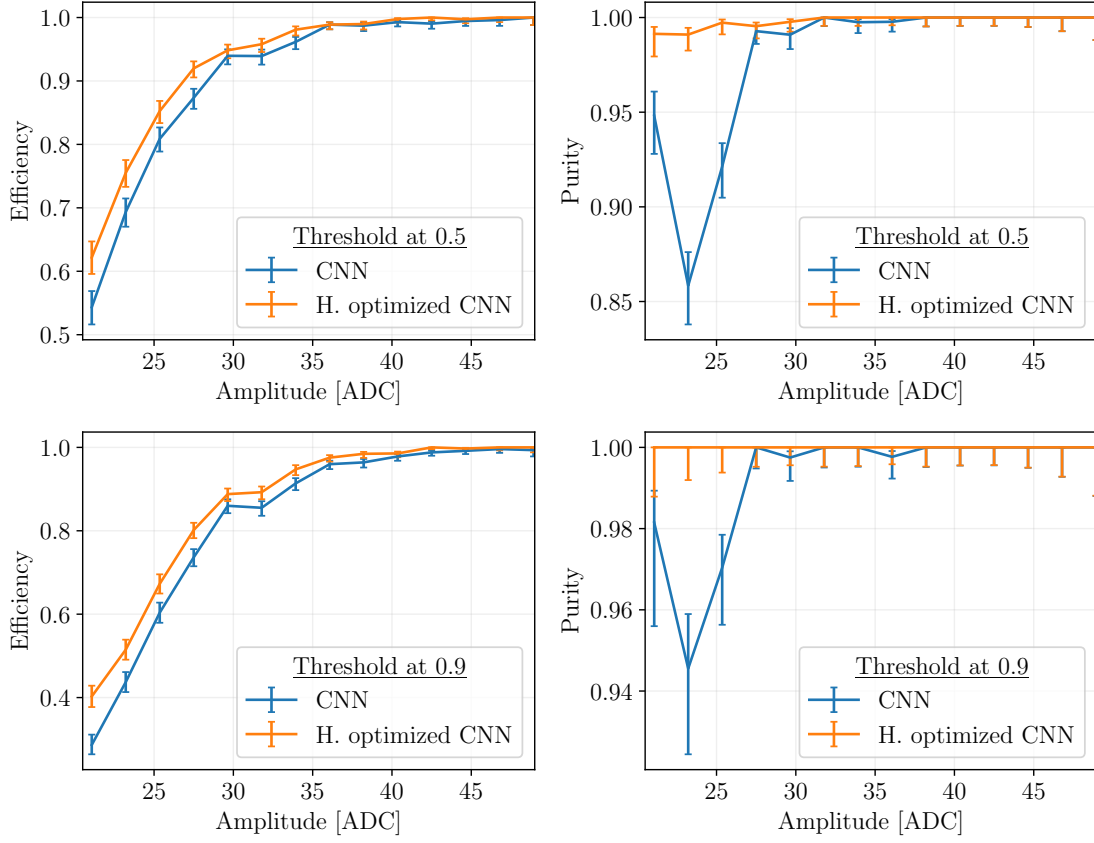


Figure 4.43.: Performance of the hyperparameter-optimized CNN (orange) with the previously discussed CNN (blue) for the multi-channel classification. The efficiency (left) and purity (right) as functions of the amplitude were calculated for thresholds at 0.5 (top row) and 0.9 (bottom row), with error bars showing the statistical uncertainty.

ous model across all amplitudes. In terms of purity, the hyperparameter-optimized CNN demonstrates substantial improvements, especially at low amplitudes, for both thresholds. Notably, at the 0.9 threshold, the hyperparameter-optimized CNN achieves 100 % purity, indicating the zero LFN-background events are expected to be misidentified as signal events. Moreover, it should be noted that the false positives with signal-like features, such as sharp rising edges, that were previously seen are removed with this model at a 0.9 cut. Out of these two models, the hyperparameter-optimized CNN is selected for its performance in LFN-background rejection, as reflected in purity.

4.5. LFN classification using CNN-LSTM

The CNN-LSTM model combines a 1D CNN with LSTM layers for binary classification. The architecture includes two convolution layers, followed by LSTM layers and a fully connected network. Batch normalization [147] is applied after each layer to improve training stability. Dropout [148] is incorporated after each layer to reduce overfitting. The final output layer has a single unit with a sigmoid activation function to produce output as signal probability. The CNN-LSTM architecture is illustrated in Fig. 4.44. The model is compiled using the Adam optimizer and binary cross-entropy loss, consistent with the CNN architecture. The optimized hyperparameters for single-, PT-, and multi-channel are shown in Table. 4.3. Details of the hyperparameter optimization process for the CNN-LSTM model are provided in Appendix A.2. This section compares the performance of the CNN and CNN-LSTM architectures for single-, PT-, and multi-channel configurations.

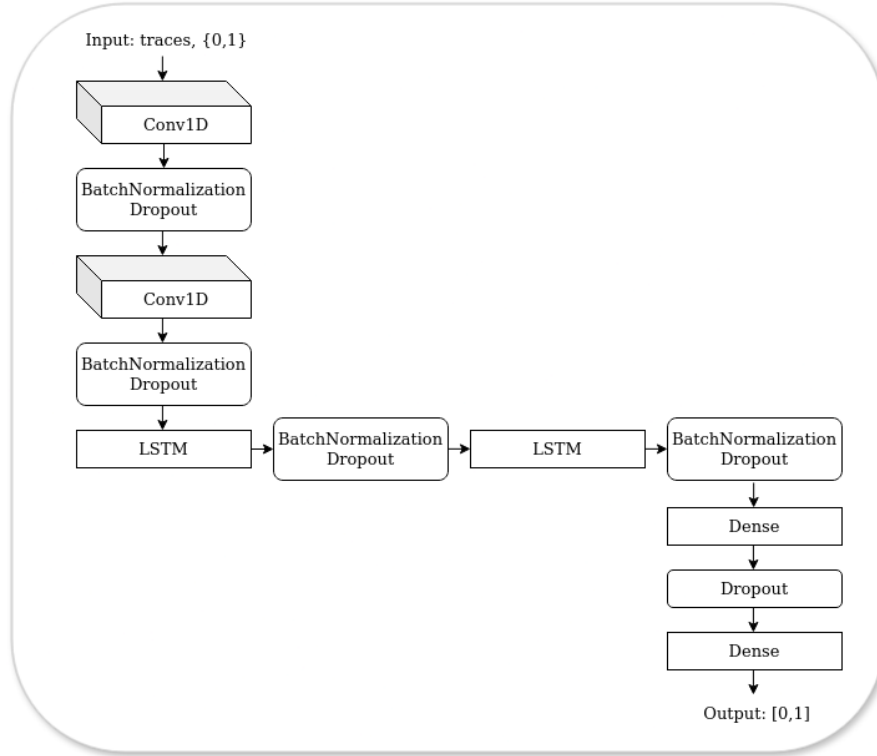


Figure 4.44.: A schematic of the CNN-LSTM architecture for LFN classification. The input is the raw trace and the targets are 0s and 1s. It consists of convolution layers followed by LSTM layers, and a fully connected network consisting of dense layers with the final dense layer for the output between 0 and 1.

Layers	Layer parameters	single	PT	multi
Conv1D	number of filters (x 32)	1	1	2
	kernel size	600	600	200
	stride	20	20	20
	kernel regularizer (L2)	-	0.01	0.01
	bias regularizer (L2)	-	0.01	0.01
	activation	relu	relu	relu
BatchNormalization		No	Yes	Yes
Dropout	rate	0.1	0.1	0.2
Conv1D	number of filters (x 32)	2	2	4
	kernel size	300	300	100
	stride	1	1	1
	kernel regularizer (L2)	-	0.01	0.01
	bias regularizer (L2)	-	0.01	0.01
	activation	relu	relu	relu
BatchNormalization		No	Yes	Yes
Dropout	rate	0.1	0.1	0.2
LSTM	units (x32)	2	2	4
	return sequence	True	True	True
BatchNormalization		No	Yes	Yes
Dropout	rate	0.2	0.2	0.1
LSTM	units (x32)	2	2	4
	return sequence	True	True	True
BatchNormalization		No	Yes	Yes
Dropout	rate	0.2	0.2	0.3
Dense	units (x32)	2	2	4
	kernel regularizer (L2)	-	0.01	0.01
	bias regularizer (L2)	-	0.01	0.01
	activation	relu	relu	relu
Dropout	rate	0.3	0.3	0
Dense	units	1	1	1
	activation	sigmoid	sigmoid	sigmoid

Table 4.3.: The best-performing parameter combination from the hyperparameter-optimization of the CNN-LSTM architecture for single-, PT-, and multi-channels. The optimization details are provided in Appendix A.2. The model was compiled with the Adam optimizer using a learning rate of 0.001 for both single- and PT-channels, and 0.01 for the multi-channel configuration.

4.5.1. Single-channel

Figure 4.45 shows the efficiency on the left and purity on the right as a function of amplitude for the hyperparameter-optimized CNN and CNN-LSTM models, with a threshold

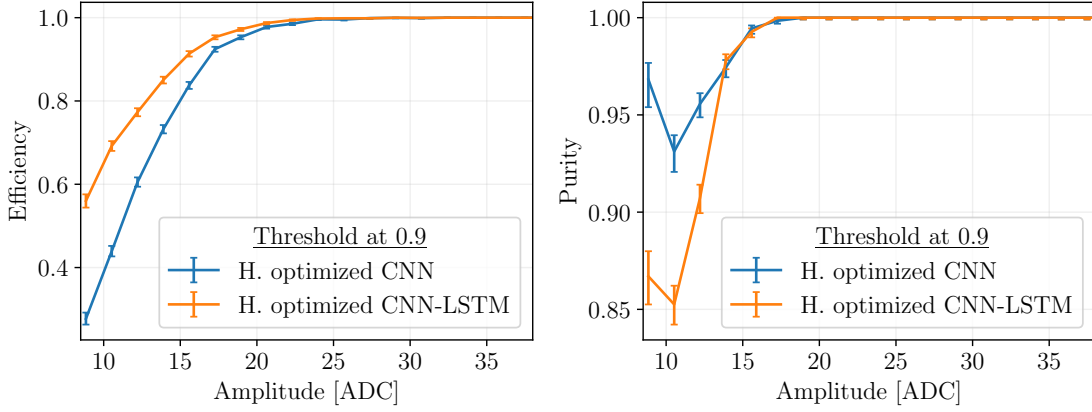


Figure 4.45.: Performance of the hyperparameter-optimized CNN (blue) with the hyperparameter-optimized CNN-LSTM (orange) for the single-channel classification. The efficiency (left) and purity (right) as functions of the amplitude are calculated for thresholds at 0.9 of the signal probability, with error bars showing the statistical uncertainty.

set at 0.9 on signal probability. Both models achieve 100 % efficiency at amplitudes above 25 ADC, with CNN-LSTM exhibiting a higher efficiency at lower amplitudes. The signal purity for both models is 100% for amplitudes above ~ 20 ADC, however they decrease below this amplitude, with the CNN model performing better than the CNN-LSTM model. Since high purity is more critical for this analysis than efficiency, the CNN model is chosen, as the addition of LSTM does not provide a significant improvement in purity.

4.5.2. PT-channel

The performance of the different model architectures, CNN and CNN-LSTM, using the PT channel, are displayed in the Fig. 4.46. The efficiency as a function of the amplitude on the left illustrates that both models have 100 % efficiency for amplitudes above ~ 50 ADC, with CNN-LSTM achieving a better efficiency at lower signal amplitudes compared to CNN. The purity as a function of amplitude on the right shows that both models perform similarly throughout, with 100 % purity above ~ 40 ADC. However, the purity of both models converges to nearly identical levels with exhibit a dip at ~ 24 ADC, with the CNN model performing better than the CNN-LSTM at low-amplitudes, and thus CNN model is favored for this channel.

4.5.3. Multi-channel

Figure 4.47 compares the performance of the CNN and CNN-LSTM models for the multi-channel classification with efficiency as a function of the amplitude on the left and purity as a function of the amplitude on the right. The threshold for

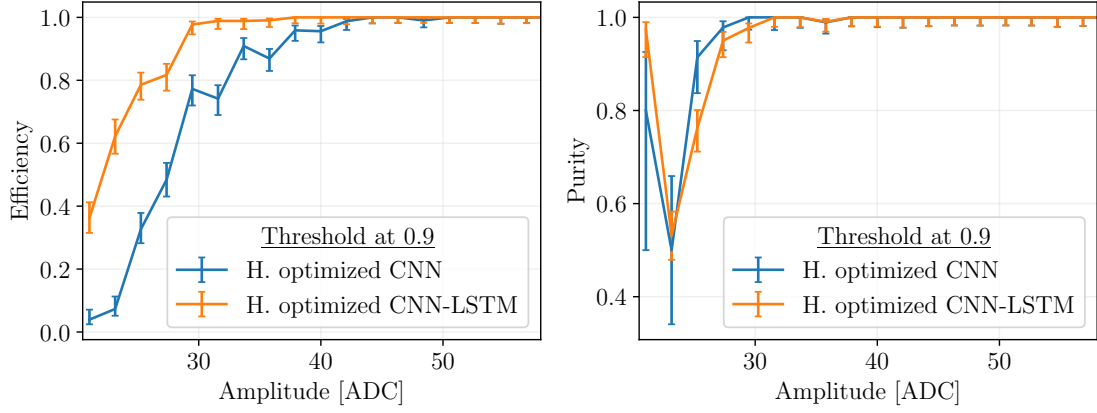


Figure 4.46.: Performance of the hyperparameter-optimized CNN (blue) with the hyperparameter-optimized CNN-LSTM (orange) for the PT-channel classification. The efficiency (left) and purity (right) as functions of the amplitude are calculated for thresholds at 0.9 of the signal probability, with error bars showing the statistical uncertainty.

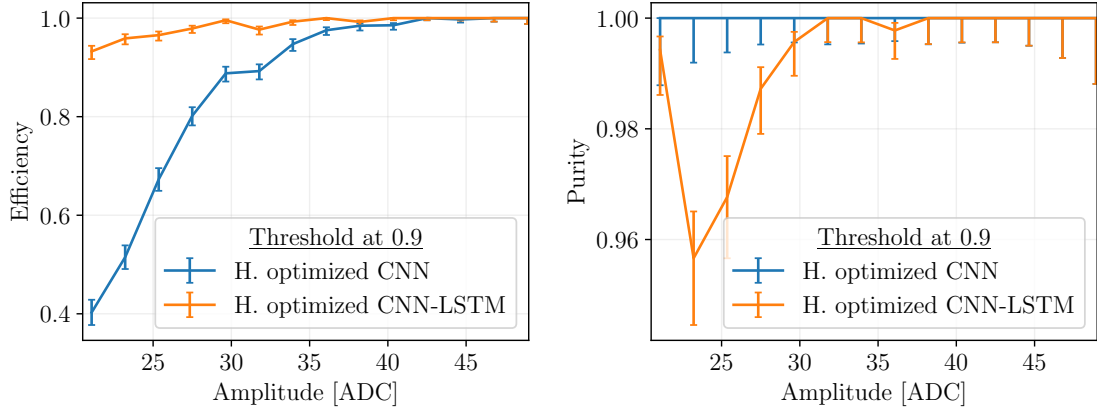


Figure 4.47.: Performance of the hyperparameter-optimized CNN (blue) with the hyperparameter-optimized CNN-LSTM (orange) for the multi-channel classification. The efficiency (left) and purity (right) as functions of the amplitude are calculated for thresholds at 0.9 of the signal probability, with error bars showing the statistical uncertainty.

comparison is placed at 0.9 on the signal probability. The efficiency shows that the CNN-LSTM model outperforms CNN model over all amplitudes. Both models show 100 % purity above ~ 38 ADC, with the CNN model maintaining 100% purity across all amplitudes. As a result, the CNN model is chosen for the multi-channel classification.

In conclusion, hyperparameter-optimized CNN models perform better than CNN-LSTM models in terms of purity across all channel-configurations: single-channel, PT-channel, and multi-channel. While CNN-LSTM models exhibit higher efficiency, particularly at low amplitudes, making them suitable for low-mass DM searches, the final DM-signal efficiency is limited by the radial cut, as is explained in Sec. 5.2.1. Therefore, a better purity is preferred, as more LFN-background would lead to setting limits on larger cross-sections. In total, there are three hyperparameter-optimized CNN models based on three different channel configurations. To select the best channel-configuration, it is essential to explore the impact of these models in the DM parameter space. In the next chapter, a sensitivity study will be conducted to finalize the channel-configuration to remove LFN events from CDMSlite Run3 data for the final DM analysis.

5. CDMSlite Run 3 Dark Matter Search

This chapter presents the final results of the CDMSlite Run 3 analysis incorporating the new ML-based cut to remove the Low-Frequency Noise (LFN) events. The previous chapter presented the three configurations for removing LFN events using ML: single-channel, Phonon Total (PT) channel, and multi-channel. In this chapter, the selected LFN cut method is applied to the experimental data to set limits on the WIMP parameter space. Sec. 5.1 outlines the limit-setting method for this analysis, while Sec. 5.2 finalizes the ML-model selection among the three channel configurations to remove the LFN noise, based on sensitivity studies. The application of the ML-based LFN cut to the CDMSlite Run 3 and corresponding observed energy spectra are discussed in Sec. 5.3. Lastly, Sec. 5.4 presents the dark matter exclusion limits from the analysis, and Sec. 5.5 investigates the new observed events identified by the ML-based LFN cut.

5.1. Limit setting method

There are several methods used in the DM community to exclude regions of the DM parameter space. Direct DM searches commonly use the Poisson counting method, maximum gap method, Optimum Interval (OI) method, and Profile Likelihood Ratio (PLR) method. Each of these methods allows for parameter exclusion by defining a confidence interval that reflects the confidence in the cross-section excluded. The maximum gap and OI methods set limits based on a signal-only hypothesis, while PLR includes background modeling and is used when background can be estimated. The CDMSlite Run 3 published analysis applied the PLR method to set limits [76]. However, in this thesis, the focus is on investigating the impact of the ML-based LFN cut. For this purpose, the OI method is chosen for its simplicity and the influence of event counts is directly observable in this method due to the limits being set on event counts. In this section, the maximum gap method, which forms the basis of the OI method, and the OI method are explained.

5.1.1. Concept of parameter space exclusion

The OI method and the maximum gap method [149] are statistical techniques for setting upper limits in the case where the background distribution in the observed parameter is not fully understood or cannot be subtracted. Thus the underlying concept is a signal-only hypothesis where all the observed events are interpreted as DM signals. These methods assume a model where signal events are distributed across an observable parameter, with the event rate directly proportional to the DM cross-section. In this case,

the observable parameter is the energy. The goal is to identify the interval that satisfies the specified confidence level.

5.1.2. Maximum gap method

For a dataset with events sorted by some observable variable, the maximum gap method identifies intervals that contain no events, in other words “gaps”. The main idea is to find the largest of these gaps as they correspond to the smallest cross-section σ that can be excluded, leading to a strong limit [149].

For an energy spectrum with event rate (dN/dE) , two adjacent events E_i and E_{i+1} form an interval or a gap. The expected number of events within the gap represented by the “size” of the gap and is given by:

$$x_i = \int_{E_i}^{E_{i+1}} \frac{dN}{dE} dE. \quad (5.1)$$

To determine the largest gap x , the observed energy spectrum is transformed to a new parameter space such that dN/dE is uniform, with the total expected number of events μ corresponding to the given cross-section. For a cross-section σ , the probability $C_0(x, \mu)$ of a random experiment finding fewer events than the observed maximum gap is calculated using

$$C_0(x, \mu) = \sum_{k=0}^m \frac{(kx - \mu)^k e^{-kx}}{k!} \left(1 + \frac{k}{\mu - kx} \right), \quad (5.2)$$

where $m \leq \mu/x$ is the highest integer, and x is the largest gap. In other words, $C_0(x, \mu)$ represents the probability of a random experiment observing a gap larger than the observed maximum gap x for a given number of expected events μ . For a 90 % confidence level (C.L.) upper limit, the cross-section σ is increased until the condition $C_0=0.9$ is met, leading to the smallest, and the strongest, upper limit on σ .

5.1.3. Optimum Interval (OI)

The Optimum Interval (OI) method is an extension of the maximum gap method to account for intervals with multiple events, which enables setting the upper limit in the case of high event counts. Similar to the maximum gap method, for each interval containing n observed events, the probability $C_n(x, \mu)$ of a random experiment observing fewer events than n events is calculated [149].

For a specific cross-section σ , $C_n(x, \mu)$ is calculated for all intervals with n events, and the interval with the maximum $C_n(x, \mu)$ i.e. the interval where $C_n(x, \mu) = C_{\max}$, provides the strongest rejection of σ and is the “optimum interval”. For a 90 % C.L. upper limit, the cross-section σ is increased until the probability of a random experiment yielding $C_{\max}$ less than the observed $C_{\max}$ value reaches 90 %. In other words, 90 %

of the random experiment's $C_{\max}$ is smaller than the observed $C_{\max}$. The OI method is particularly effective in situations where background levels are low, as it allows for strong constraints on the signal hypothesis in cases where the background is unknown.

5.2. Model Selection for final analysis

As shown in the previous chapter, three types of the LFN cut were developed based on different channel configurations: single-, PT-, and multi-channel. In this section, “signal” refers to non-LFN events, as in the previous chapter, and not the DM signal unless explicitly stated as “DM signal”. Each channel configuration uses hyperparameter-optimized CNNs, as discussed in Sec. 4.4.4, with the signal probability threshold set at 0.9, yielding high purity. To select the model for the final analysis, sensitivity studies were performed. Each ML-model has different signal efficiencies and purities, both of which are crucial: signal efficiency at low energies enables sensitivity to search of low-mass DM, while signal purity improves exclusion of parameters space. Therefore, sensitivity studies were performed to assess the impact of the ML-based LFN cut on the exclusion of DM-nucleon cross-section parameter space for WIMPs and to select the optimal model for the final analysis. The sensitivity projections were calculated using the OI method.

5.2.1. Signal efficiency

The signal efficiencies presented in the previous chapter were calculated using raw signal data generated by adding pulse templates (shown in Fig. 4.5) to baseline noise for the single- and PT-channel models and by using SUPR events for the multi-channel models. Due to the lack of direct methods to calculate the energies of these raw traces from the OF amplitudes, these samples were not used for the final DM signal efficiency calculation. Instead, a SuperCDMS software was used for generating raw signal data, which was then processed through the official reconstruction pipeline to obtain event energies for the DM signal efficiency calculations.

The samples from BatFaker were first passed through the quality cuts such as the glitch cut, phonon start time cut, phonon χ^2 cut, as described in Sec. 3.4, and finally the ML-based LFN cut. The efficiency, calculated as the ratio of the number of events passing these cuts to the total number of events at a given energy, was then combined with the trigger efficiency to get the quality cut efficiency. Figure 5.1 presents the quality cut efficiencies based on the ML-based LFN cuts, compared with those from the previous analysis. The PT- and multi-channel efficiencies perform better than the previous analysis, however, the single-channel efficiency performs worse than the previous analysis, except in a narrow energy range around $\sim 0.07 \text{ keV}_{ee}$. The quality cut efficiency was then multiplied by the radial cut efficiency and parameterized using Eq. 3.21 to determine the final DM signal efficiency. Detailed information on this method is provided in Sec. 5.3.1. Figure 5.2 presents the final efficiencies of the different ML-models in

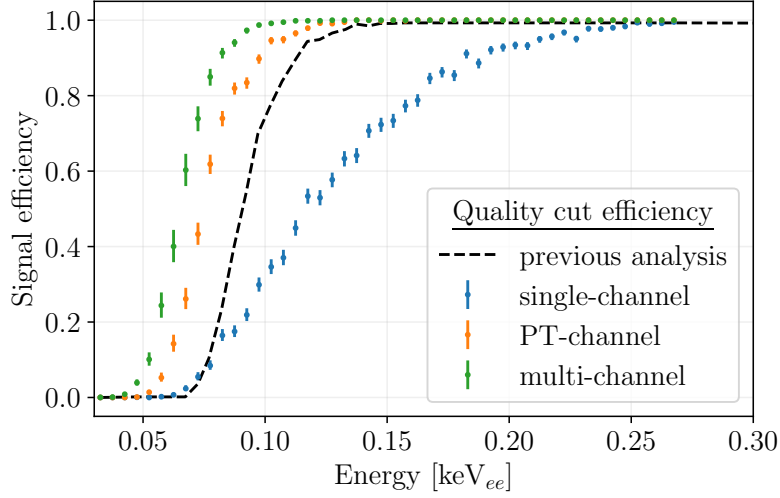


Figure 5.1.: Quality cut efficiency as a function of energy for the three ML-models to remove LFN events. The blue, orange, and green data points represent the single-channel, PT-channel, and multi-channel ML-models, respectively. The error bars denote the statistical uncertainty. The quality cut efficiency of the previous analysis is shown in black (dashed), taken from Ref. [76].

comparison with the efficiency from the previous analysis. The systematic uncertainties depicted in the figure are discussed in detail in Sec. 5.4.1. The results indicate that the PT- and multi-channel final efficiencies mostly overlap, except for slight deviations at low energies around 0.08 keV_{ee} . They perform better than the previous analysis efficiency up to $\sim 0.12 \text{ keV}_{ee}$, after which until $\sim 0.24 \text{ keV}_{ee}$, the previous analysis has high efficiency. In contrast, the single-channel efficiency is significantly lower than that of the PT- and multi-channel configurations, as well as the previous analysis, up to $\sim 0.24 \text{ keV}_{ee}$; beyond this energy, it plateaus similarly to the other two efficiencies. All efficiencies reach a plateau at a DM signal efficiency of ~ 0.58 , attributed to the radial cut's effect on the DM signal efficiency, as will be shown in Sec. 5.3.1. Although the quality cut efficiencies with the PT- and multi-channel models are higher than those in the previous analysis (seen in Fig. 5.2), applying the radial cut diminishes the overall efficiencies.

5.2.2. Expected energy spectra

The expected energy spectra for this sensitivity study were derived from background models estimated for the CDMSlite Run 3 PLR analysis [76]. The backgrounds modeled were from tritium, Compton scattering, activation peaks of ^{71}Ge and ^{68}Ga , and the three components of the surface events (Top Lid (TL), Housing (H), and Germanium (Ge)) [104]. Figure 5.3 shows the best-fit background models for each source, overlaid on the CDMSlite Run 3 event spectrum from the previous analysis. The probability density function (pdf) of the total background was used to generate toy experiments for

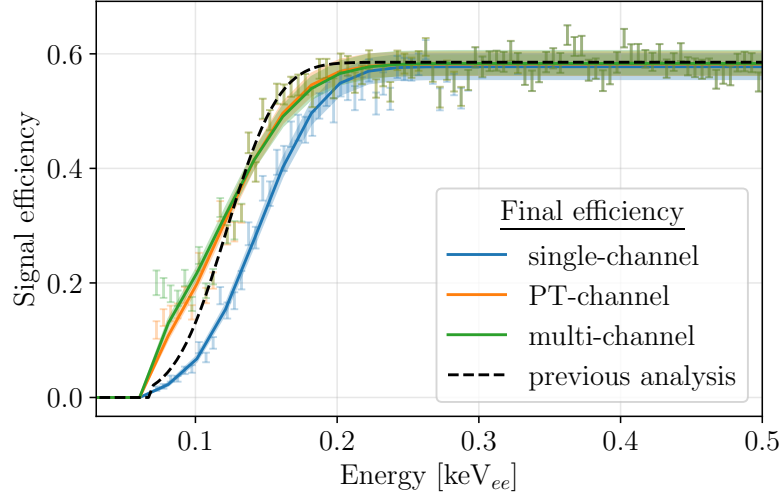


Figure 5.2.: The DM signal efficiency obtained by consecutively applying the trigger efficiency, the quality cut efficiency calculated using the three ML-models, and the fiducial volume (radial cut) efficiency. The parameterization of the combined efficiency data using Eq. 3.21 is shown by the solid lines. The uncertainties, which include statistical and systematic uncertainties, are shown as a band around the parameterized curve. The blue, orange, and green colors represent the single-, PT-, and multi- channels, respectively. The final DM signal efficiency of the previous analysis is shown in black (dashed), taken from Ref. [76].

this sensitivity study. The public data for this run has 401 events (from the previous analysis), and therefore this value was used as the mean when drawing a number from a Poisson distribution for event generation. Monte Carlo toy experiments were generated with 100 trials with the number of events drawn from a Poisson distribution.

The background models shown in Fig. 5.3 didn't account for the contributions from LFN events as the previous analysis assumed that less than one LFN event passed through the bifurcated LFN cut. However, with the ML-based LFN cut, LFN events may leak into the energy spectra if the signal purity is below 100 %, thereby contributing to the expected spectra. So, based on the purity of the ML-model being considered and convolving it with the pdf of the LFN-background events shown in Fig. 4.1, the LFN events contribution was added to the generated spectra. The uncertainties in the DM signal efficiency were not considered for this sensitivity study as the objective was to select the ML-model for the final analysis. Thus, the sensitivity band was solely based on Poisson fluctuations. The expected spectra for the three models are shown in Fig. 5.4, where the lowest energy peak in the single- and PT-channel arises from the LFN contribution. It is absent in the multi-channel, which has 100 % signal purity.

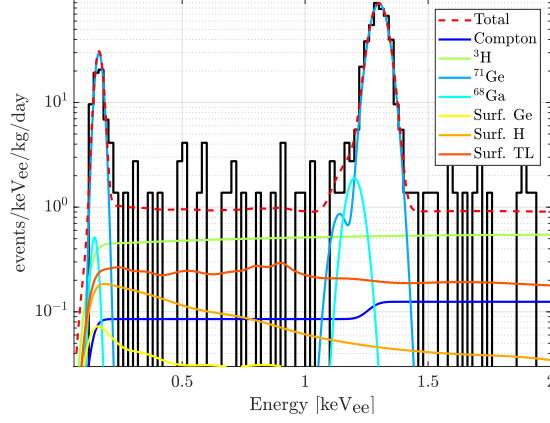


Figure 5.3.: The CDMSlite Run 3 best-fit background models overlaid on the previous analysis energy spectrum. The background contributions are from Compton scattering, tritium, ^{71}Ge and ^{68}Ga , and surface events from the Top Lid (TL), Housing (H), and Germanium (Ge). The total background estimate in red is used to generate toy experiments. This plot is taken from Ref. [76].

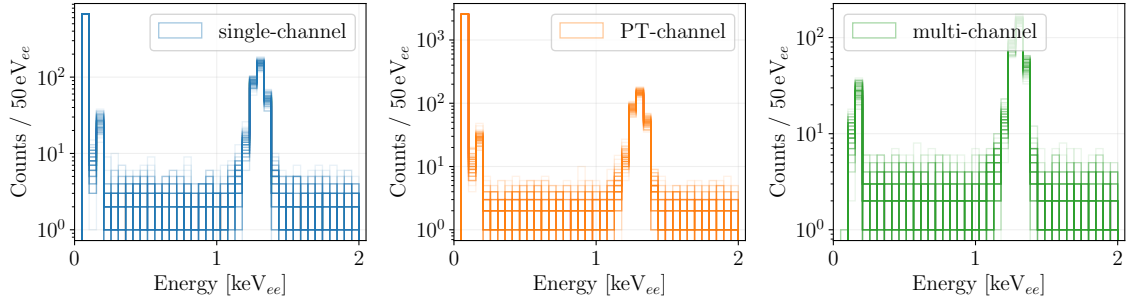


Figure 5.4.: Toy experiments generated using the total background estimation shown in Fig. 5.3 and the LFN contribution added using the purity of each of the channel configurations. The expected energy spectra for single-channel is shown on the left, PT-channel in the middle, and multi-channel on the right.

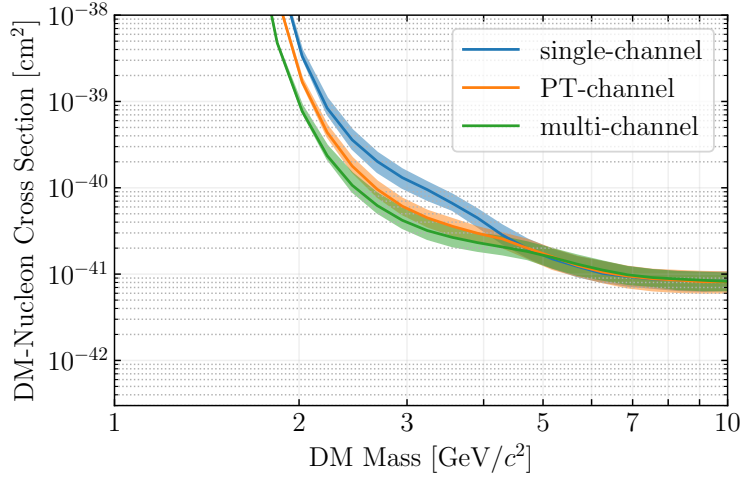


Figure 5.5.: Sensitivity projections using the single-channel (blue), PT-channel (orange), and multi-channel (green) ML-models for removing the LFN noise events. The 1σ uncertainty band is shown around the median of each case.

5.2.3. Sensitivity projections

The sensitivity projections were calculated using the OI method. The expected DM signal was scaled according to the DM signal efficiency of each model and smeared using the energy resolution discussed in Sec. 3.3.4. Based on the expected DM signal and the expected energy spectra, sensitivity projections for the toy experiments were calculated using the OI method. The median of the sensitivities along with 1σ uncertainty band for the single-, PT- and multi-channel are shown in Fig. 5.5. It is evident that the multi-channel model outperforms the other two models by excluding smaller cross-sections for DM masses below $\sim 5 \text{ GeV}/c^2$, which can be understood by the better purity of this model. Its ability to probe lower DM masses correlates with the high DM signal efficiency of the multi-channel model at the low-energies. As a result, the multi-channel network was selected for the final analysis.

5.3. Analysis results

In the previous section, the multi-channel ML-model for the LFN event removal was selected for the final analysis. This section presents the resulting DM signal efficiency, the outcomes of applying this model to the CDMSlite Run 3 data, and the corresponding final energy spectrum.

5.3.1. DM signal efficiency

The motivation of any event selection is to remove unwanted events like noise and backgrounds, but in the process, these cuts can also remove DM signal events. Quantifying

this loss of the signal events is important in calculating the expected DM spectrum since not all of the signal events will make it through the cuts. This section describes the DM signal efficiency formulation for the DM search analysis.

Sample generation

For calculating the DM signal efficiency, one needs a data sample containing signal-like events. The multi-channel LFN cut was developed and evaluated using samples generated with the SUPR method, as explained in Sec. 4.2, and the output raw data from this method was stored as arrays. However, the SuperCDMS reconstruction pipeline requires the raw data to be stored in a specific binary format. As it was not possible to convert the SUPR raw data arrays into the binary format due to technical difficulties, the SuperCDMS software called “BatFaker” was used to generate the signal-like sample.

BatFaker uses the pulse templates and baseline noise events to generate fake signal events in the binary format. The BatFaker was used in several analyses mainly using the PT template to inject the baseline noise. However for the ML-based LFN removal, one needs to correctly reconstruct events in all 4 phonon channels, thus using PT template was not the correct option. Therefore, in this analysis, the signal-like samples were generated using the fast and slow templates, which capture the positional dependencies, for calculating the final DM signal efficiencies. Furthermore, This was the method used to generate samples for the signal efficiency calculations for the previous analysis.

For the baseline noise, randomly triggered events from the beginning of the data series were chosen. The fast and slow template amplitudes were extracted for each phonon channel from the K-shell events and were rescaled to the desired amplitude. These amplitudes were then used to scale the fast and slow template as shown in Fig. 3.5, and were then added to the baseline noise. The samples were then processed through the SuperCDMS reconstruction pipeline required to get the event energies.

DM signal efficiency parametrization

The data quality cuts were first applied to the signal samples, which include the glitch cut, phonon χ^2 cut, phonon start time cut, square pulse cut, as described in Sec. 3.4, and the ML-based LFN cut. All of these cuts were applied to the signal samples except the square pulse cut since the implementation of this cut was no longer possible. As this cut has 100% signal efficiency, it was ignored as all the events will pass the cut anyway. The efficiency obtained from applying the quality cuts was multiplied by the trigger efficiency to obtain the quality cut efficiency.

The quality cut efficiency was combined with the radial cut efficiency to obtain the final DM signal efficiency. The radial cut was developed separately for periods A and B, as the noise conditions changed between these periods, as explained in Sec. 3.1. The

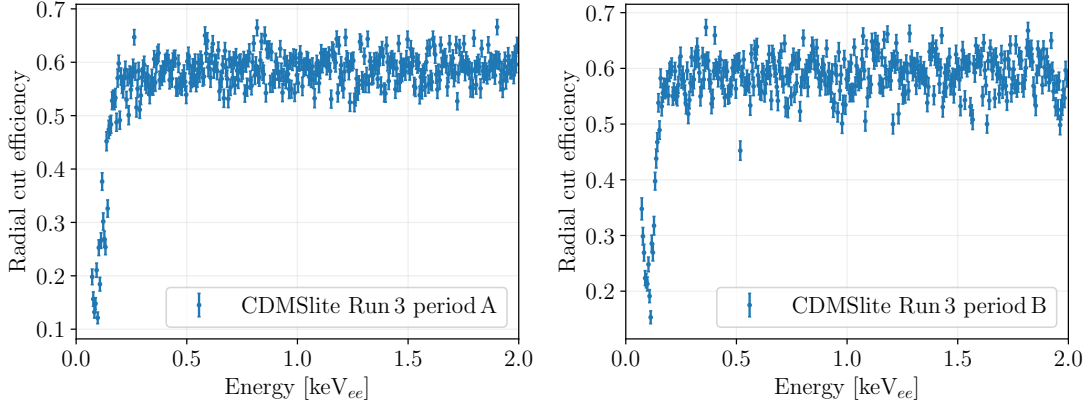


Figure 5.6.: The radial cut efficiency for Run 3 period A (left) and period B (right) as a function of energy. The plots taken from [124].

efficiencies of the radial cut are shown in Fig. 5.6, where it is seen that the efficiency is $\sim 58\%$ for event energies above $\sim 0.25 \text{ keV}_{ee}$ and below this energy, the efficiencies drop drastically. As explained in Sec. 3.4.5, the radial cut sets a hard threshold at an energy of 0.07 keV_{ee} . The ML-based LFN cut was developed irrespective of the time periods, thus there is only one quality cuts efficiency for the whole of Run 3. The quality cut efficiency was multiplied with the radial cut for periods A and B separately, producing two final efficiencies which are then combined using a weighted live-time average. The DM signal efficiency was parameterized using Eq. 3.21 where μ_1 , μ_2 , and μ_3 are the parameters to optimize. The Minuit package [150] was used for the parameterization of the efficiency data points with error propagation. The optimized values for μ_1 , μ_2 , and μ_3 are 0.292 ± 0.002 , 0.114 ± 0.002 , and 0.047 ± 0.003 , respectively.

In the CDMSlite Run3 analysis, the correlated systematic uncertainty arising from the radial efficiency was not included initially. An uncorrelated fit was performed for the efficiency. Since the μ_1 parameter controls the vertical scale of the efficiency curve, the correlated systematic uncertainty was incorporated into the covariance matrices derived from the uncorrelated fit [123, 124]. However, as the correlated systematic uncertainty could not be found, the quantity was empirically determined by comparing the CDMSlite DM signal efficiency covariance with the calculated covariance of this analysis. The systematic uncertainty was found to be approximately $1/0.125$, which was then used to scale the uncertainty on μ_1 . Figure 5.7 shows the hardware trigger efficiency, quality cut efficiency, and the combined efficiency data with the parameterized final DM signal efficiency. The jump around 0.08 keV_{ee} arises from the radial cut, which sets a hard threshold at 0.07 keV_{ee} . The combined efficiency data points in the figure show an increased efficiencies below 0.1 keV_{ee} , which results from the radial cut efficiency shown in Fig. 5.6.

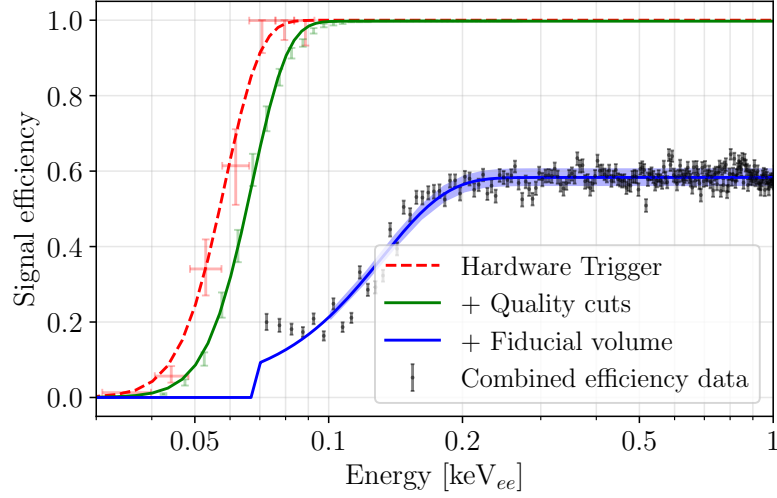


Figure 5.7.: The final DM signal efficiency of CDMSlite Run 3 calculated by applying the trigger efficiency (red), quality cut efficiency (green), and fiducial volume i.e. radial cut efficiency (black) successively. The quality cut is calculated using the ML-based LFN cut. The parameterized final efficiency is shown in blue with 1σ uncertainty band consisting of systematic and statistical uncertainties. The hard threshold at 0.07 keV_{ee} is enforced by the radial cut efficiency.

5.3.2. Application of the new LFN cut to CDMSlite Run 3 data

There were a total of $\sim 28\text{M}$ events in CDMSlite Run 3, of which $\sim 24.7\text{M}$ events are physics-triggered events. These $\sim 24.7\text{M}$ events were passed through the ML-model and the network predictions were extracted. Table 5.1 summarizes the number of events passing each selection criterion, including the new ML-based cut used in this analysis. The major event loss occurs due to the LFN cut (old and new) and the radial cut. The new LFN cut retains more events compared to the previous bifurcated LFN cut due to its higher DM signal efficiency at low energies (as seen in Fig. 5.2) compared to the previous LFN cut, thus allowing more events to be retained. After applying these cuts, the number of events remaining in the region of interest (ROI) between 0.07 keV_{ee} and 2 keV_{ee} is 401 for the previous analysis using the bifurcated LFN cut and 406 for the new analysis using the ML-based LFN cut.

Name of the event selection	Number of events passed	Fraction of events passed [%]
High voltage	24216313	97.87
NuMI beam	24739506	99.99
Pre-trigger baseline	24721938	99.92
Charge χ^2	24664257	99.68
Bad series	24047906	97.19
Trigger glitch	24733487	99.96
GPS	24741113	99.99
Trigger rate	24741036	99.99
T2Z1 triggered	24298565	98.21
Good randoms	24742506	100.00
Muon veto	24492864	98.99
Multiple scatter	24622315	99.51
Good phonon start time	22832570	92.28
Phonon χ^2	24647762	99.62
Glitch	24119133	97.48
Square Pulses	24693265	99.80
Low-Frequency Noise (old)	64045	0.26
Low-Frequency Noise (new)	83335	0.34
Radial	361319	1.46

Table 5.1.: The number of events passing each selection criterion individually, including results for both the old and new LFN cuts in the whole analysis energy range.

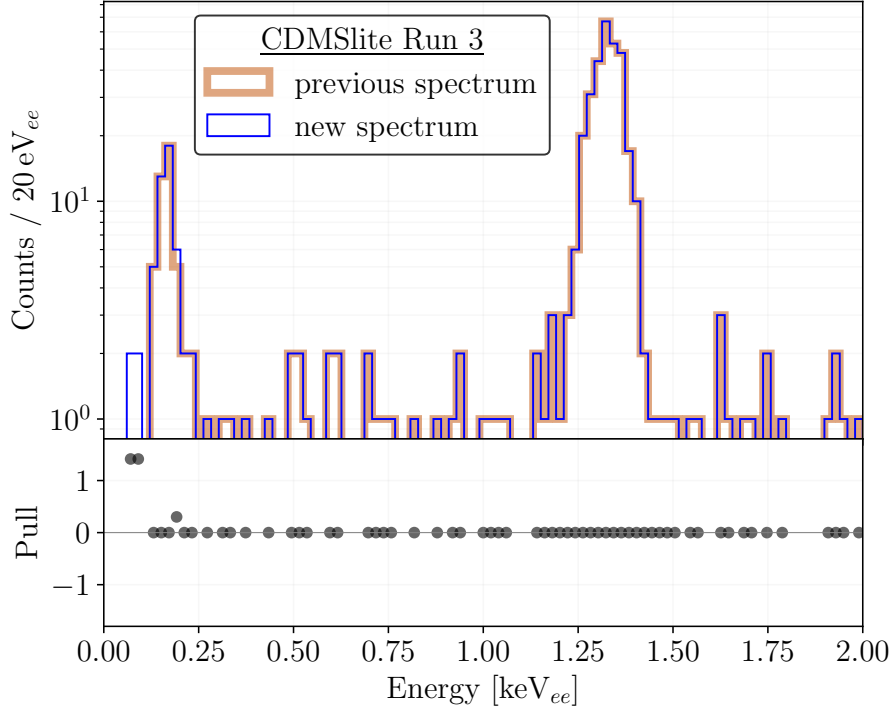


Figure 5.8.: The final energy spectra of CDMSlite Run3 using the previous analysis with the bifurcated LFN cut (orange), and the new analysis with ML-based LFN cut (blue). Bottom panel shows pull, which is defined as difference between the previous and the new analysis spectrum in each bin divided by the uncertainty.

5.3.3. Observed energy spectra

The final energy spectra obtained using the bifurcated LFN cut and the new ML-based LFN cut are shown in Fig. 5.8. There are notably more events below 0.2 keV_{ee} in the new analysis. The residual plot below further quantifies the differences between the two spectra for events below 0.2 keV_{ee} . These additional observed events are discussed in Sec. 5.5. The M- and L-shell activation peaks at $\sim 0.16 \text{ keV}_{ee}$ and $\sim 1.3 \text{ keV}_{ee}$ are visible in the spectra. There are five additional events in the new analysis energy spectrum when compared to the old analysis energy spectrum. Upon investigating these events, an issue with energy reconstruction for two of these five events was found. Addressing the problem with the energy reconstruction scripts is beyond the scope of this work, however, the findings are discussed in Appendix. B.

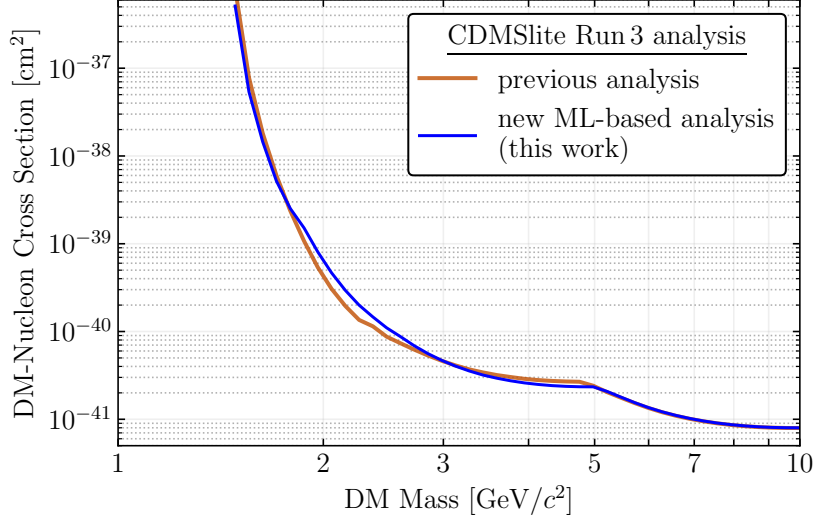


Figure 5.9.: The CDMSlite Run 3 90 % C.L. OI limits on the spin-independent WIMP-nucleon cross-section using the previous analysis with the bifurcated LFN cut (orange) and the new analysis with the ML-based LFN cut (blue).

5.4. DM exclusion limit with OI

The 90 % C.L. OI exclusion limits on the spin-independent WIMP-nucleon cross-section were calculated for this analysis with (1) the bifurcated LFN cut and (2) the ML-based LFN cut. The results are shown in Fig. 5.9. While the overall profiles of both the limits are similar, the new limits performs slightly worse around DM masses of $\sim 2 \text{ GeV}/c^2$. This is due to the OI selecting a larger interval in the previous analysis compared to the interval selected for the new analysis, which is constrained by the presence of additional events. The bump that occurs at a DM mass of $\sim 5 \text{ GeV}/c^2$ arises from the presence of the activation peak at 0.16 keV, which prevents OI from having a smooth transition of intervals to higher masses. Beyond a DM mass of $\sim 5 \text{ GeV}/c^2$, both limits overlap, which is explained by having no difference in the observed spectra at high energies in the old and new analysis.

5.4.1. Systematic uncertainties

Similar to the previous CDMSlite Run 3 analysis, the systematic uncertainties considered in this analysis were from the DM signal efficiency, energy resolution, and Lindhard k factor. The DM signal efficiency was parameterized using Eq. 3.21 with the three parameters μ_1 , μ_2 , and μ_3 . The uncertainty of the signal efficiency is propagated using the uncertainties of these three parameters discussed in Sec. 5.3.1. The signal resolution consists of three quantities: the baseline resolution, the position dependence term, and the Fano resolution term reflected in parameters σ_E , A , and B in Eq. 3.14. The uncertainties in both the DM signal efficiency and resolution follow a Gaussian distribution,

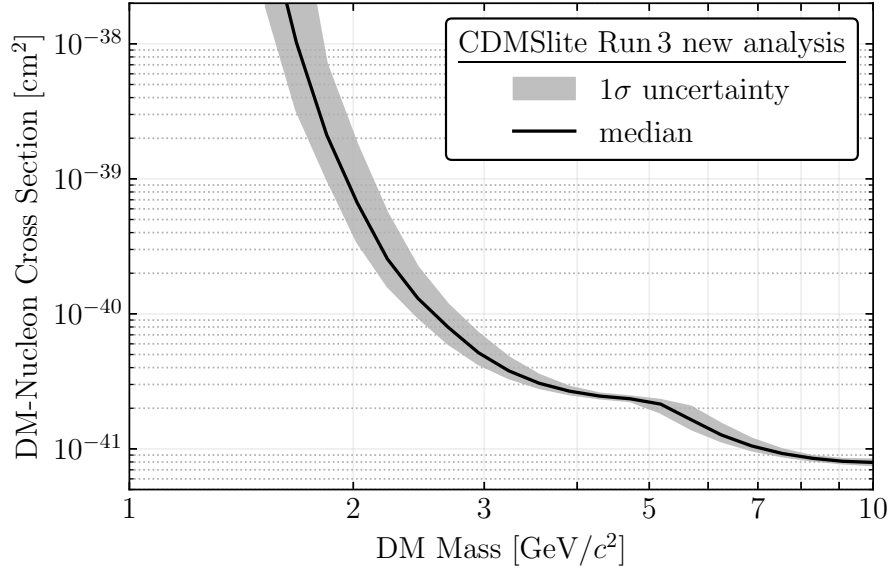


Figure 5.10.: The systematic uncertainty estimation for the 90 % C.L. OI limits on the spin-independent WIMP-nucleon cross-section using the new analysis with the ML-based LFN cut. The 1σ band is shown in gray with the median shown in black.

while the systematic uncertainty for the Lindhard k parameter is represented by a uniform distribution between 0.1 and 0.2.

2000 trials were performed, leading to a median and 1-sigma uncertainty band shown in Fig. 5.10. It should be noted that for DM masses below $\sim 1.5 \text{ GeV}/c^2$, limits with k values below ~ 0.125 were inaccessible, i.e. small k values make the analysis insensitive to DM masses below $\sim 1.5 \text{ GeV}/c^2$, and thus they are not included in the uncertainty band calculation.

5.5. Investigation of the additional observed events

As shown in Fig. 5.8, there are additional events in the final spectra with the new analysis compared to the previous analysis spectra. These events passed the new ML-based LFN cut but failed the previous bifurcated LFN cut. As mentioned in Sec. 5.3.3, five new observed events were identified. The raw traces of these events are shown in Fig. 5.11, and their energy and ML predictions are shown in Table 5.2.

In Fig. 5.11, event 1 shows an event peak around 5.9 ms instead of the trigger window, which was set around 0.8 ms (500th time bin). However, for the CNN, the location of the signal-feature is irrelevant due to its translational invariance. Events 2 and 3 are events

Event	Energy [keV _{ee}]	ML prediction
1	0.0881	0.9985
2	0.1905	1.
3	0.0907	1.
4	0.0731	0.9621
5	0.0714	0.9781

Table 5.2.: The energy and network predictions for the newly identified events.

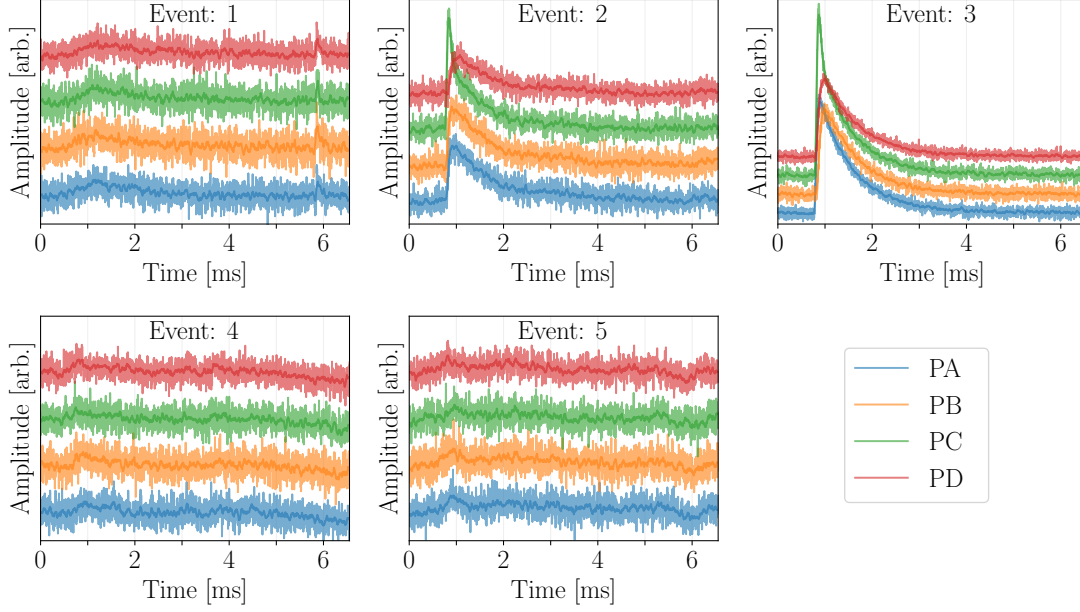


Figure 5.11.: Raw traces of the new observed events with the four phonon channels PA, PB, PC, and PD. The new signal events passed the new ML-based LFN cut but failed the previous bifurcated LFN cut.

with one channel exhibiting a distinct “peakier” pulse compared to the other channels, indicating the location of the event. These events have ML-score of 1 and have a high amplitude, which can be deduced from the noise levels. Events 4 and 5 exhibit small pulse-like features around 1 ms, with lower network scores of 0.96 and 0.98, respectively. The Class Activation Maps (CAMs) for these new observed events is shown in Fig. 5.12.

For event 1, the peak around 5.9 ms is given the highest importance in the classification decision. For events 2 and 3, the rising part, which is a very differentiating signal-like feature, is given the high importance. Event 4 visually exhibits the peaky nature of the pulse to some extent. However, for event 5, this characteristic is not visually evident due to the very small energy deposition. In any case, both events require further investigation to quantify the signal-like characteristics.

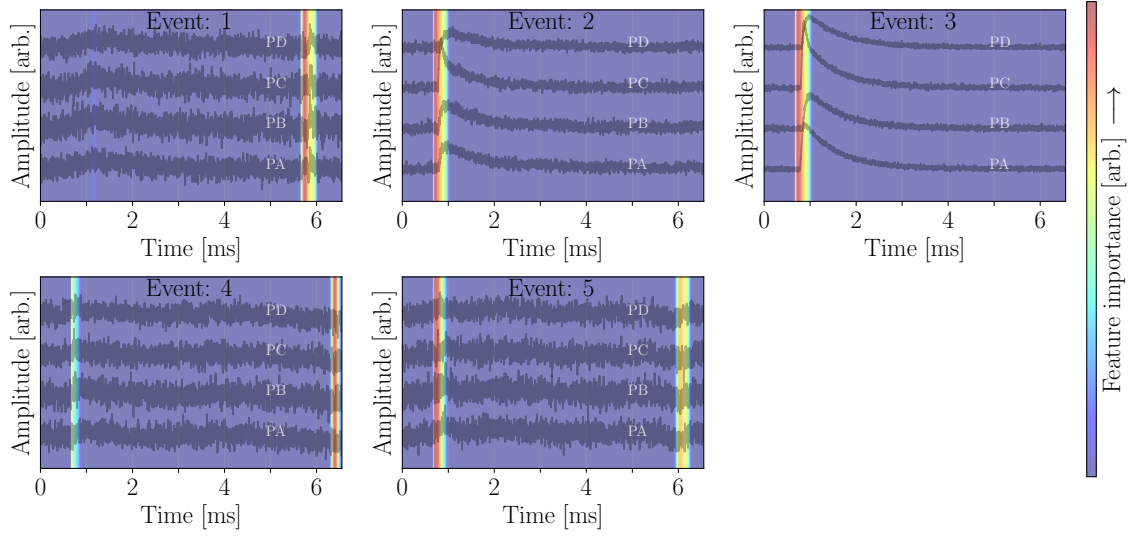


Figure 5.12.: The activation maps of the new observed events with the raw traces shown in the transparent black over which the color maps representing the features that are given high importance for the prediction are overlaid. The activation maps are calculated using Eq. 4.12.

5.6. DM exclusion limit with OI without mis-reconstructed events

Events 2 and 3 in Fig. 5.11 are events with a high amplitude that are mis-reconstructed to low energies. In this section, the 90 % C.L. OI exclusion limits were calculated after removing these two events from the observed energy spectrum. This was done to accurately calculate the limits, as these events should not be present at the energies of 0.19 keV_{ee} and 0.09 keV_{ee} , respectively. The result is shown in Fig. 5.13. The performance deficit observed around DM masses of $\sim 2 \text{ GeV}/c^2$ in Fig. 5.9 is absent after removing the mis-reconstructed events. This is because the optimum interval used to set the limits for these DM masses is between 0.073 keV_{ee} and 0.131 keV_{ee} , where the two mis-reconstructed events were present. The presence of these events at those energies was not “expected” based on DM signal efficiency, which determines expected number of events. The new analysis shows a slight improvement in limits between DM masses of $2.36 \text{ GeV}/c^2$ and $5 \text{ GeV}/c^2$. The optimum intervals for these DM masses are also between 0.07 keV_{ee} and 0.131 keV_{ee} , where the DM signal efficiency with the new analysis performs better than the old analysis, as seen in Fig. 5.2 (multi-channel), leading to a slight improvement in the limits. The overall profiles of both the old and the new limits remain similar below DM masses of $2.36 \text{ GeV}/c^2$, and above $5 \text{ GeV}/c^2$.

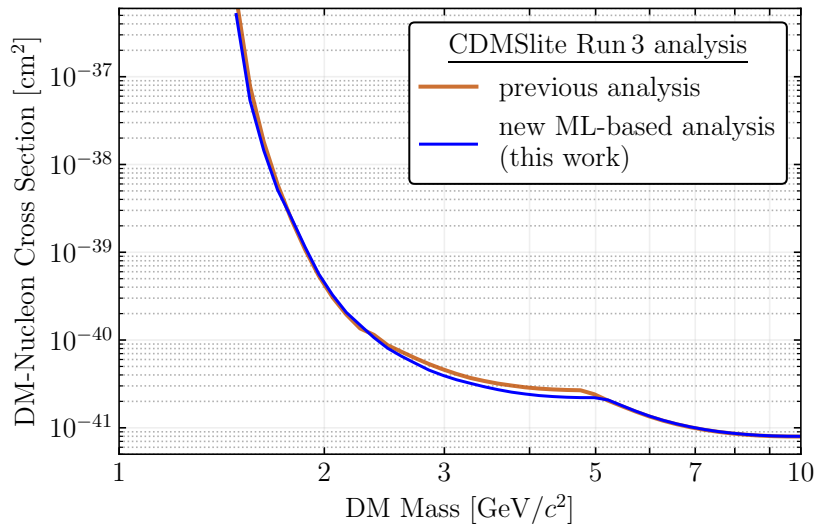


Figure 5.13.: The CDMSlite Run 3 90 % C.L. OI limits on the spin-independent WIMP-nucleon cross-section using the previous analysis with the bifurcated LFN cut (orange) and the new analysis with the ML-based LFN cut (blue). The misreconstructed events (Event 2 and 3 in Fig. 5.11) were removed.

6. Case Studies

In the previous chapter, it was seen that the radial cut had a significant impact on the final DM signal efficiency. While the new ML-based LFN cut improved signal efficiencies, its effectiveness was constrained by the radial cut. To explore this further, a case study without the radial cut was conducted and discussed in Sec. 6.1. Another case explored the ML-based LFN cut's potential under the assumption of no observed events, focusing solely on signal efficiency. Sec. 6.3 details this scenario with and without the radial cut. Additionally, the impact of applying a stringent cut with and without the radial cut is presented in Sec. 6.2.

6.1. Impact of radial cut

The radial cut was implemented to remove RNTL events caused by HV field non-uniformities, as explained in Sec. 3.4.5. This cut had the efficiency of $\sim 58\%$ for events with energies above 0.25 keV_{ee} , but showed a sharp decline in efficiency for lower energies, significantly reducing the final DM signal efficiency. To evaluate the impact of not applying the radial cut, this case study was conducted by applying all the live-time and quality cuts to the CDMSlite Run 3 data, but the radial cut was omitted. Both the previous bifurcated LFN cut and the new ML-based LFN cut were applied independently for comparison. Therefore, in this case, the final DM signal efficiency was the quality cuts efficiency showed in Fig. 3.14 for the bifurcated LFN cut and in Fig. 5.7 for the ML-based LFN cut. The analysis thresholds were set at $\sim 0.069 \text{ keV}_{ee}$ for the previous analysis, and $\sim 0.034 \text{ keV}_{ee}$ for the new analysis. Further details of this case are documented in Appendix C.

The observed spectra for this case study are shown in Fig. 6.1. The spectrum with the ML-based LFN cut shows an excess of events around 0.05 keV_{ee} . Upon inspection of these events, they consist of: signal-like features in the non-trigger region (trigger is set around 0.8 ms i.e. 500^{th} time bin) like event 1 in Fig. 5.11, noise-like events (e.g., glitch, square pulse, or LFN) in the non-trigger region, signal-like events mis-reconstructed in energy, and signal-like events with properly reconstructed energies.

It is important to note that the ML models were trained and tested on LFN-background samples that had passed the radial cut, which ensured event energies above 0.07 keV_{ee} . The peak at 0.05 keV_{ee} was not anticipated during the development of this cut, as the purity was calculated only for amplitudes with corresponding energies above 0.07 keV_{ee} . However, the network was trained with signal events that have energies

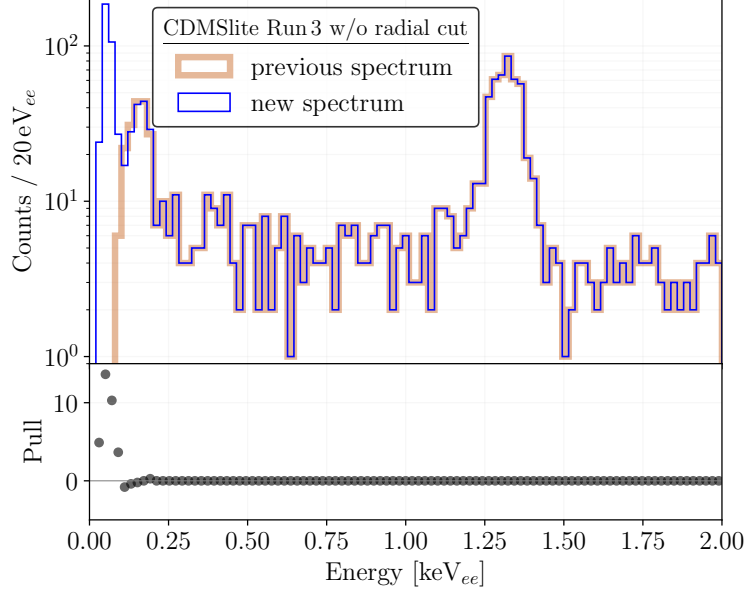


Figure 6.1.: The energy spectrum of CDMSlite Run 3 using the previous analysis with the bifurcated LFN cut (orange), and the new analysis with ML-based LFN cut (blue) without the radial cut. Bottom panel shows pull, which is defined as difference between the previous and the new analysis spectrum in each bin divided by the uncertainty.

down to $\sim 0.04 \text{ keV}_{ee}$. In the region between $\sim 0.04 \text{ keV}_{ee}$ and 0.07 keV_{ee} , it is possible for the network predictions to be biased toward higher signal probabilities due to the absence of LFN-background events in this energy range during training. Validating this would require a pure LFN-background sample within the energy range of $\sim 0.04 \text{ keV}_{ee}$ to 0.07 keV_{ee} . However, obtaining such a sample poses a challenge, as no established methods currently exist to isolate LFN events in this range. This introduces uncertainty in the signal efficiency for energies below 0.07 keV_{ee} . Therefore, this case study is not part of the main analysis but rather a case study that requires further investigation in the future. The energy spectrum resulting from the bifurcated LFN cut does not have these low-energy events due to its analysis threshold at $\sim 0.069 \text{ keV}_{ee}$, effectively eliminating those low-energy events entirely.

The corresponding DM sensitivities are shown in Fig. 6.2, comparing the results with and without the radial cut. In both cases, mis-reconstructed events were not removed. The results for the case with the radial cut are provided only as a reference, and thus the findings discussed below correspond to the case without the radial cut.

- For DM masses $> 4 \text{ GeV}/c^2$: Both the previous analyses and new analysis perform similarly, excluding large cross-sections in comparison to the case with radial cut. This arises from the overall increase in the number of events in the spectrum when the radial cut is omitted.

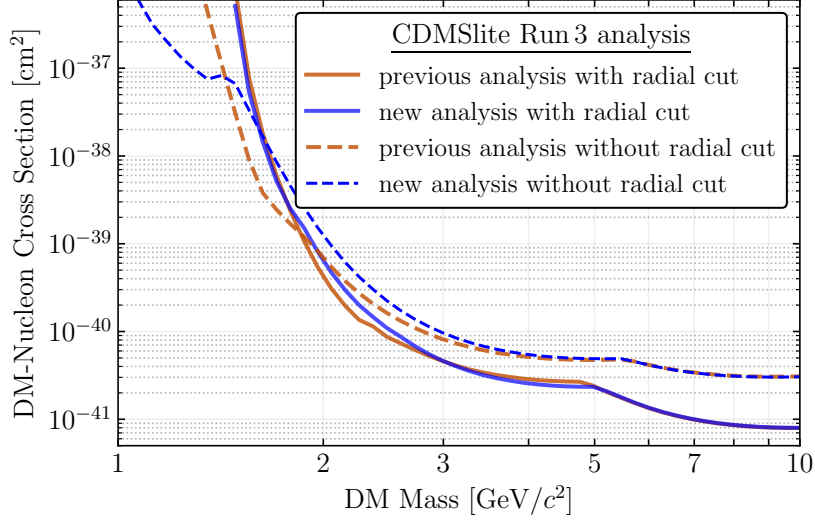


Figure 6.2.: Effect of the radial cut on the 90 % C.L. OI-based sensitivities on the spin-independent WIMP-nucleon cross-section for CDMSlite Run 3 using the previous analysis with the bifurcated LFN cut (orange) and the new analysis with the ML-based LFN cut (blue). The solid lines represent the cases where the radial cut is applied, and the dashed lines represent the cases without the radial cut. The mis-reconstructed events were not removed in this case study.

- For DM masses between $1.3 \text{ GeV}/c^2$ and $4 \text{ GeV}/c^2$: The previous analysis excludes smaller cross-sections than the new analysis due to the peak at 0.05 keV_{ee} in the latter, which led to excluding higher cross-sections.
- For DM masses $< 1.3 \text{ GeV}/c^2$: The new analysis outperforms the previous one due to its lower analysis threshold, which allows it to explore DM masses as low as $0.8 \text{ GeV}/c^2$ —an important goal of the ML-based LFN cut. The improvement in the sensitivity for DM masses below $1.2 \text{ GeV}/c^2$ is up to three orders of magnitude.

The study emphasizes that obtaining better results depends on resolving issues caused by training ML models with the samples that have already passed through the radial cut.

6.2. Placing a stringent cut

In Sec. 6.1, events at energies below $\sim 0.01 \text{ keV}_{ee}$ were observed when the radial cut was not applied. The proper approach to handling the case without radial cut is to re-train the neural network using events that have not been filtered through the radial cut. Alternatively, with the current neural network for LFN cut, the signal probability threshold value can be increased to allow only highly signal-like events to

pass. This modification minimizes the presence of events below $\sim 0.01 \text{ keV}_{ee}$, which exhibit square-pulse-like and glitch-like features outside the trigger region. However, this approach constitutes a post-unblinding study.

The impact of increasing the signal probability threshold to 0.99 is discussed in this section. It is important to note that the signal samples used to calculate the signal efficiencies in the previous analysis based on bifurcated LFN cut are not the same as those used in the new analysis based on ML-based LFN cut. The signal samples used for the previous analysis are unavailable, and it is currently not possible to recalculate the previous analysis efficiencies with the current sample due to the unavailability of the necessary computational facilities. Therefore, the results presented in this chapter should be interpreted with caution. Despite these limitations, this chapter serves as a proof of principle, highlighting the impact of the ML-based LFN cut with a higher threshold and motivating future work to incorporate this approach.

6.2.1. With radial cut

This case study involved applying a signal probability threshold of 0.99 to remove the LFN events with the radial cut. Based on this case study, this section presents the final DM signal efficiencies, the resulting spectra, and sensitivities on the SI WIMP-nucleon cross-section.

Signal efficiency

The final DM signal efficiency for this high-threshold case was determined using the same method described in Sec. 5.3.1. The final efficiency was obtained by sequentially applying the hardware trigger efficiency, quality cut efficiency, and radial cut efficiency. The efficiency was parameterized using Eq. 3.21. The final DM signal efficiency using the ML-based LFN cut with a signal probability threshold of 0.99 is shown in Fig. 6.3.

The comparison of the final DM signal efficiency for the new analysis using the ML-based LFN cut with thresholds at 0.99 and 0.9 is shown in Fig. 6.4, alongside the previous analysis using the bifurcated LFN method. The signal probability thresholds at 0.9 and 0.99 show similar efficiencies compared to the previous analysis above energies of 0.24 keV_{ee} . Between 0.13 keV_{ee} and 0.24 keV_{ee} , the previous analysis exhibits higher efficiency compared to the ML-based signal efficiencies. Below 0.13 keV_{ee} , the new analyses provide higher signal efficiency than the previous analysis, with the signal probability threshold of 0.9 showing higher efficiency than the threshold at 0.99.

WIMP-nucleon sensitivities

The ML-based LFN cut with the signal probability threshold of 0.99 was applied to the CDMSLite Run3 dataset, and the resulting final energy spectrum is shown in Fig. 6.5. In this case, only three additional events are observed compared to the

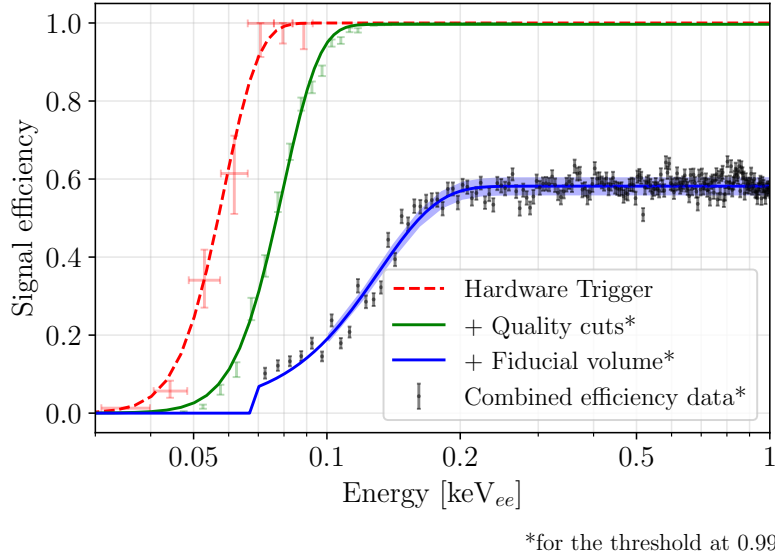


Figure 6.3.: The DM signal efficiency of CDMSlite Run 3 calculated by applying the trigger efficiency (red), quality cut efficiency (green), and fiducial volume or radial cut efficiency (black) successively. The quality cut is calculated using the ML-based LFN cut with a signal probability threshold of 0.99. The parameterized final efficiency is shown in blue with 1σ uncertainty band consisting of systematic and statistical uncertainties. The hard threshold at 0.07 keV_{ee} is imposed by the radial cut efficiency.

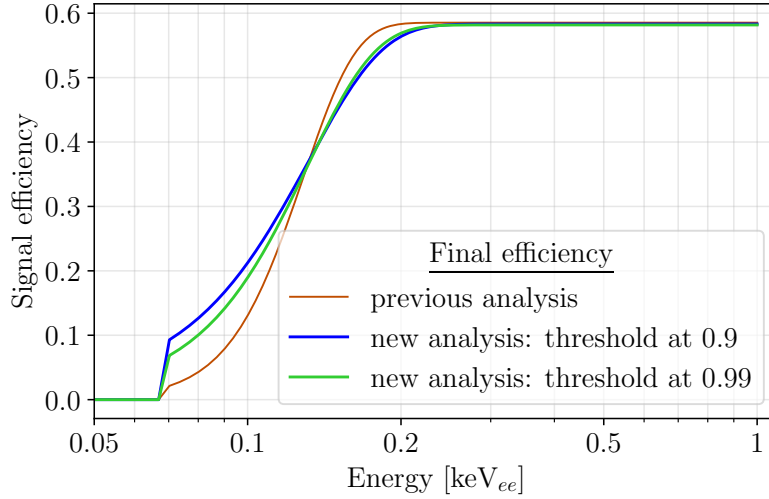


Figure 6.4.: Comparison of the DM signal efficiencies of CDMSlite Run 3 for the previous analysis using bifurcated LFN cut (orange), and new analysis using ML-based LFN cut by placing a threshold on signal probability at 0.9 (blue) and at 0.99 (green). The uncertainty band is omitted for visual clarity.

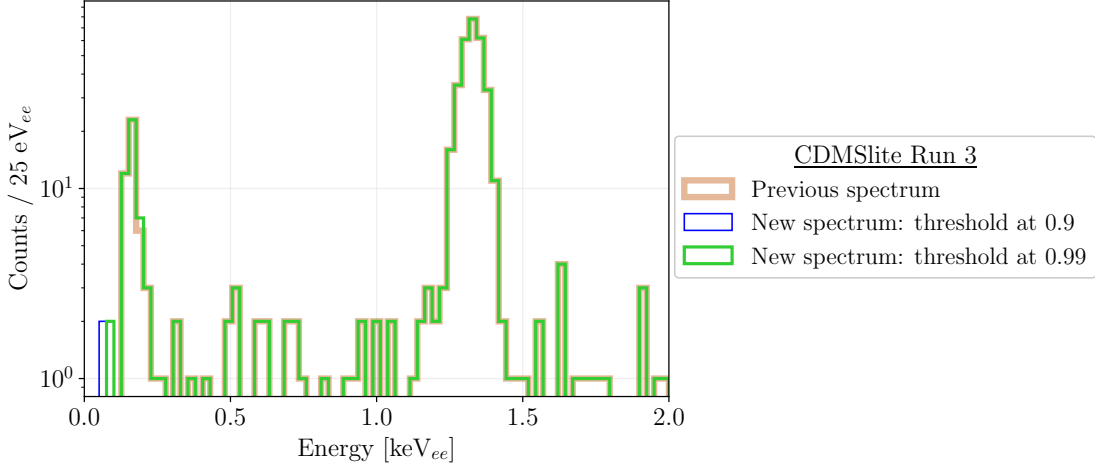


Figure 6.5.: The energy for of CDMSlite Run 3 using the previous analysis with bifurcated LFN cut (orange), and the new analysis with ML-based LFN cut using signal probability thresholds of 0.9 (blue) and 0.99 (green), shown after applying the radial cut.

previous spectrum using the bifurcated LFN cut. These additional events are Events 1, 2, and 3 in Fig. 5.11. The events with mis-reconstructed energies, Event 2 and Event 3, were removed.

The 90% C.L. OI-based sensitivity on the spin-independent WIMP-nucleon cross-section for this case study is shown in Fig. 6.6. The threshold at 0.99 results in overall comparable sensitivities to the threshold at 0.9 and the previous analysis, with a slight improvement for DM masses below $\sim 1.9 \text{ GeV}/c^2$. This improvement is due to the reduction in the number of events (Event 4 and 5 in Fig. 5.11) in the final spectrum compared to the analysis with a signal probability threshold of 0.9, as well as the enhanced signal efficiency at low energies compared to the previous analysis.

6.2.2. Without radial cut

This case study examined the effect of applying a signal probability threshold of 0.99 for removing the LFN events without the radial cut, and the results are shown in this section. This study is particularly essential as it examines how the events with energies below $\sim 0.01 \text{ keV}_{ee}$ discussed in Sec. 6.1 are impacted with the higher threshold.

Signal efficiency

The final DM signal efficiency without the radial cut was obtained from the quality cut efficiencies. Fig. 6.7 shows the DM signal efficiency without the radial cut for the previous analysis using the bifurcated LFN cut, as well as the new analyses using the ML-based

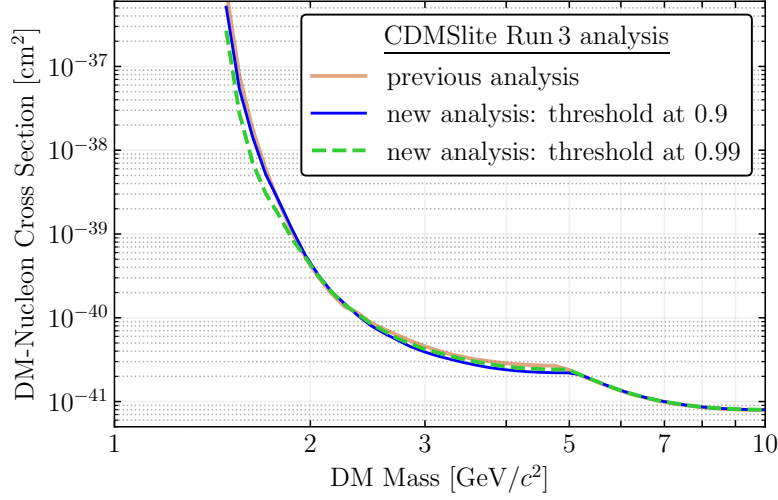


Figure 6.6.: Effect of applying a stringent signal probability threshold with the radial cut for the ML-based LFN cut on the 90 % C.L. OI-based sensitivities on the spin-independent WIMP-nucleon cross-section for CDMSlite Run 3. Results from the previous analysis with the bifurcated LFN cut (orange) and the new analysis with the ML-based LFN cut by placing a threshold on signal probability at 0.9 (blue) and at 0.99 (green, dashed) are shown.

LFN cut for signal probability thresholds at 0.9 and 0.99. The efficiency data points for the new analyses are shown with the parameterized curve, while for the previous analysis, only the parameterized curve is shown, as the data points are unavailable. As expected, the signal efficiency is lower for the signal probability threshold of 0.99 compared to 0.9. However, it should be noted the signal efficiency for the threshold at 0.99 is still higher than that of the previous analysis, indicating a gain even with a more stringent threshold.

WIMP-nucleon sensitivities

The final spectrum of the CDMSlite Run3 using the ML-based LFN cut with a signal probability threshold of 0.99 and without the application of radial cut is shown in Fig. 6.8. It is evident that the low-energy events below $\sim 0.1 \text{ keV}_{ee}$ observed in the spectrum with a threshold of 0.9 are significantly reduced in the spectrum with the threshold of 0.99. Further information on the remaining events is provided in Appendix C. In addition to less number of events below $\sim 0.1 \text{ keV}_{ee}$, there are fewer events in the M-shell peak at 0.16 keV_{ee} compared to the spectrum with a threshold of 0.9 and the previous analysis using the bifurcated LFN cut. In total, this threshold removed 50 events that had previously passed the bifurcated cut, which requires further investigation.

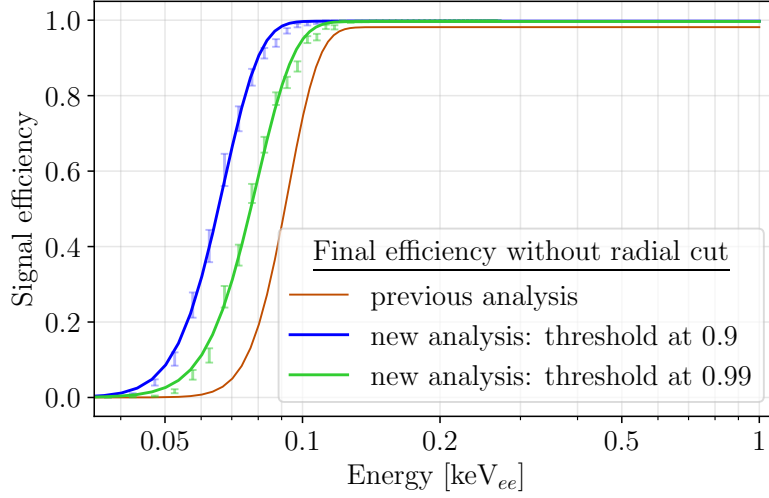


Figure 6.7.: Comparison of the DM signal efficiencies of CDMSlite Run3 without the application of the radial cut for the previous analysis using bifurcated LFN cut (orange), and new analysis using ML-based LFN cut with signal probability thresholds of 0.9 (blue) and 0.99 (green). The shown efficiencies are parameterized from the efficiency data points shown for the new analysis (blue for 0.9 threshold and green for 0.99 threshold); efficiency data points for the previous analysis are not shown as the data points are unavailable.

The 90% C.L. OI-based sensitivity on the spin-independent WIMP-nucleon cross-section for the new analysis using the ML-based LFN cut with a signal probability threshold of 0.99, without applying the radial cut, is shown in Fig. 6.9. For DM masses below $\sim 1.35 \text{ GeV}/c^2$, the threshold at 0.9 performs better than the 0.99 threshold due to better signal efficiency at the lowest energies for the 0.9 threshold.

A substantial improvement in sensitivity is observed for DM masses between $\sim 1.35 \text{ GeV}/c^2$ and $10 \text{ GeV}/c^2$. This improvement is attributed to a higher signal efficiency compared to the previous analysis, a reduced number of events with energies below $\sim 0.18 \text{ keV}_{ee}$, and the optimal intervals for the entire mass range being below 0.37 keV_{ee} . The events removed by the high-threshold LFN cut, which previously passed the bifurcated LFN cut and the 0.9 threshold with the ML-based cut, have energies between 0.08 keV_{ee} and 0.17 keV_{ee} . The optimal intervals for all masses are within this energy range, leading to a better sensitivity due to low number of events.

However, it is important to note that the ML model was not trained for energies below 0.07 keV_{ee} , so results in this range should be interpreted with caution. This study highlights the impact of the radial cut on the ML-based LFN cut’s performance and demonstrates the ML-based cut’s unhindered potential. While the stringent threshold of 0.99 results in some loss of signal efficiency, it effectively removes LFN and other

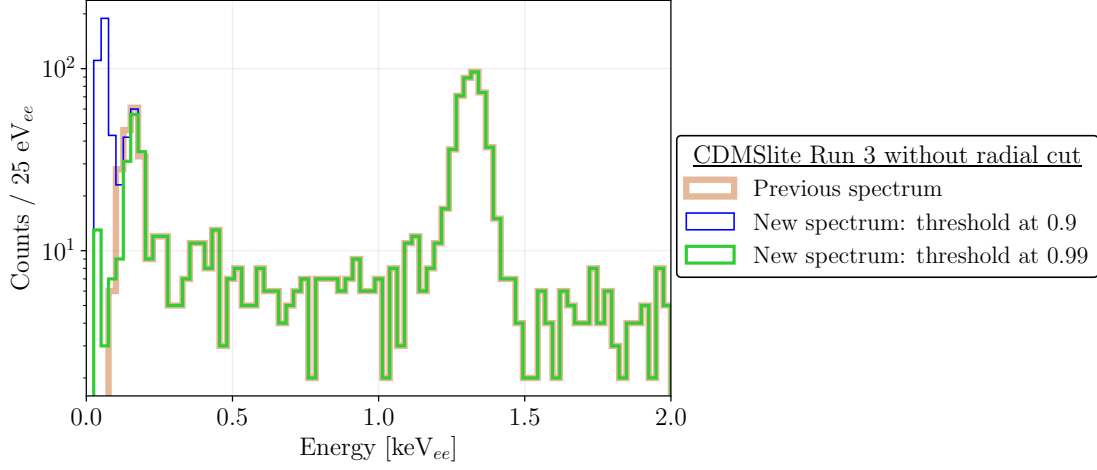


Figure 6.8.: The energy spectrum of CDMSlite Run 3 from the previous analysis with bifurcated LFN cut (orange), and the new analysis with ML-based LFN cut (blue) without the application of the radial cut.

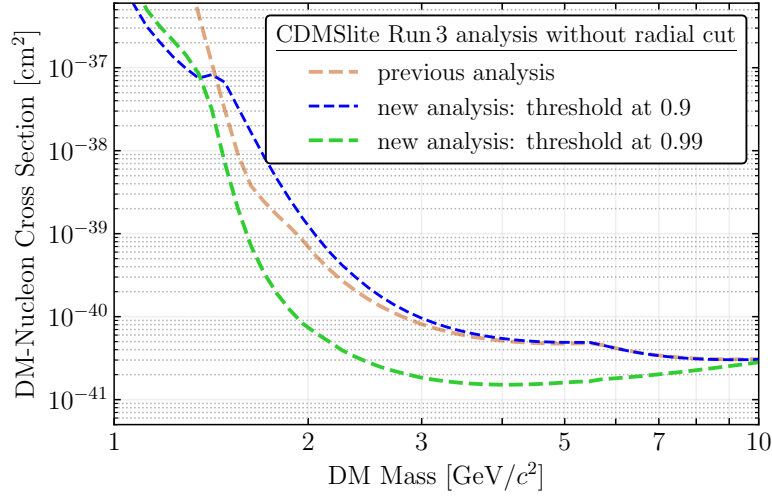


Figure 6.9.: Effect of applying a signal probability threshold at 0.99 and excluding the radial cut on the 90 % C.L. OI-based sensitivities on the spin-independent WIMP-nucleon cross-section for CDMSlite Run 3. Results from the previous analysis with the bifurcated LFN cut (orange) and the new analysis with the ML-based LFN cut by placing a threshold on signal probability at 0.9 (blue) and at 0.99 (green) are shown.

low-energy events below $\sim 0.01 \text{ keV}_{ee}$, representing a trade-off between efficiency and background reduction.

6.3. In the absence of observed events

To fully understand the extent of the ML-based approach, a case study was conducted under the assumption that no events were observed. In this scenario, the results directly reflect the signal efficiencies. This case was similar to the main results shown in Sec. 5.4 in that the DM signal efficiencies and expected DM spectra remain unchanged, but the number of observed events was set to zero. The study was performed both with and without the radial cut to examine its impact on the results. The sensitivity for this case study is shown in Fig. 6.10.

With the radial cut applied, there is no improvement in the reach of low-mass DM search because the signal efficiencies are constrained by the energy threshold of the radial cut at 0.07 keV_{ee} . Both the previous and the new analysis with the radial cut yield similar sensitivities, although the new analysis performs slightly better for DM masses below $3.5 \text{ GeV}/c^2$.

Without the radial cut, both the previous and the new analyses show improvement in the low-mass DM search, with the new analysis demonstrating substantial improvement in the sensitivity down to a DM mass of $0.8 \text{ GeV}/c^2$, reflecting its ability to explore events in low-energy regime. For mid-range DM masses between $2 \text{ GeV}/c^2$ and $10 \text{ GeV}/c^2$, both analyses have sensitivity to smaller cross-sections, with the new analysis performing better than the previous analysis.

This case study highlights the potential of the ML-based LFN cut approach compared to the previous bifurcated LFN cut. The results underscore the impact of the radial cut on the signal efficiency, which in turn affects the sensitivity to DM at lower masses. Without the radial cut, the new analysis demonstrates its capability without the influence of the event excess in the low energies observed in Sec. 6.1, providing a clear view of the method's capabilities.

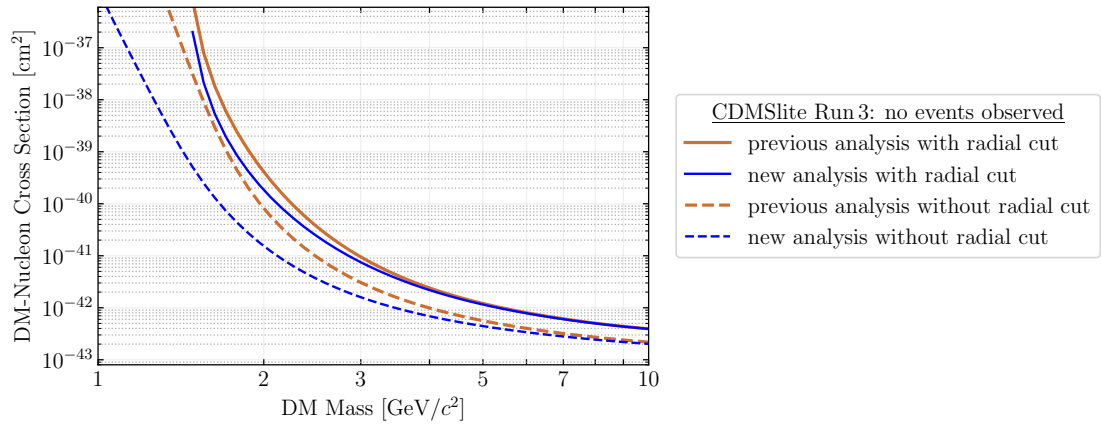


Figure 6.10.: 90% C.L. OI-based sensitivities on the spin-independent WIMP-nucleon cross-section for CDMSlite Run 3 with the assumption of no observed events. The previous analysis using the bifurcated LFN cut and the new analysis with the ML-based LFN cut are shown in orange and blue, respectively. The solid lines represent the cases where the radial cut is applied, and the dashed lines represent the cases without the radial cut.

7. Conclusion and Outlook

The primary focus of this thesis work was the development of a machine learning (ML)-based cut for the Low-Frequency Noise (LFN) induced by vibrations during the CDM-Slite Run 3 experiment. This written work began with an introduction to Dark Matter (DM) in Chapter 1, followed by an overview of the SuperCDMS experiment in Chapter 2. Chapter 3 outlined the CDM-Slite Run 3 experiment and its analysis. The development of the ML-based LFN cut is presented in Chapter 4, and its implementation and results are discussed in Chapter 5. Chapter 6 explored extensions of the analysis through different case studies. This final chapter summarizes the thesis in Sec. 7.1 and concludes with an outlook in Sec. 7.2.

7.1. Summary

The cryocooler that provided supplemental cooling deteriorated over the course of CDM-Slite Runs, as discussed in Sec. 3.1. Particularly in Run 3, it caused vibration-induced LFN, which triggered the detector often. Due to the similarity in pulse shape between LFN traces and low-energy signal traces, removing these LFN events from experimental data was challenging. The previous analysis applied a bifurcated cut to remove the LFN events, as detailed in Sec. 3.4.3. However, this method resulted in poor signal efficiency at low-energies. Therefore, the motivation for this work was to develop a new method based on ML to remove LFN events with an improved signal efficiency while maintaining a high signal purity.

Three channel configurations were explored: single-channel, Phonon Total (PT)-channel, and multi-channel (4 channels). LFN-background samples were data-driven for all configurations, as described in Sec. 4.1.1. For the signal samples, phonon templates scaled by desired amplitudes were added to the baseline noise traces for the single- and PT-channel configurations, as explained in Sec. 4.1.2. Due to the lack of proper SuperCDMS simulation that's needed for the multi-channel configuration, a novel method called Simulation Using Pulse shapes and Regression (SUPR) to generate templates based on event positions was developed, as detailed in Sec. 4.2. The ML models considered for this analysis were Convolutional Neural Networks (CNNs) and CNN with Long Short-Term Memorys (LSTMs). Section 4.5 demonstrated that the hyperparameter-optimized CNNs performed better than the CNN-LSTMs in terms of signal purity across all channel configurations. High signal purity was crucial for this analysis, as the Optimum Interval (OI) method was used to set limits, which is based on a signal-only hypothesis. Sensitivity studies, explained in Sec. 5.2, were conducted

to determine the best channel configuration for the LFN cut. The multi-channel configuration was found to perform better than the single- and PT-channel configurations, leading to its selection for the final implementation of the LFN cut.

Application of the new ML-based cut, along with other quality cuts, to the CDMSlite Run3 dataset revealed five additional events in the final energy spectrum within the ROI of 0.07 keV_{ee} to 2 keV_{ee} . Upon investigation, two of these five events have mis-reconstructed energies, as detailed in Secs. 5.3 and 5.5.

Exclusion limits were set on the spin-independent WIMP-nucleon cross-section based on the OI method. Two cases were analyzed: (A) including the mis-reconstructed events in the final spectrum, discussed in Sec. 5.4, and (B) excluding the mis-reconstructed events in the final spectrum, discussed in Sec. 5.6. The results show that, overall, the analysis based on the new ML-based LFN cut performed similarly to the previous analysis based on the bifurcated LFN cut. However, in case A, the limits (shown in Fig. 5.9) displayed slightly worse performance for DM masses around $\sim 2\text{ GeV}/c^2$, as the additional events resulted in limits being set at higher cross-sections. In case B, where the mis-reconstructed events were excluded, the performance deficit was not observed, confirming that the mis-reconstructed events caused poor performance at those mass ranges. Additionally, both cases demonstrated an improvement of up to $\sim 17.5\%$ over the previous results for DM masses around $5\text{ GeV}/c^2$ due to the improvement in the signal efficiency.

The radial cut imposed a threshold at 0.07 keV_{ee} , and limited the signal efficiency gains with the use of the ML-based LFN cut developed. Sec. 6.1 presented a case study where the radial cut was removed. The study revealed a significant number of events below $\sim 0.1\text{ keV}_{ee}$ in the energy spectrum of the new analysis using the ML-based LFN cut. As the developed neural network was not trained for energies below 0.07 keV_{ee} and used samples already filtered by the radial cut, all the case studies without the radial cut should be interpreted with caution. Sec. 6.2 examined a higher signal probability threshold. This study demonstrated improved sensitivities in the low-mass DM, highlighting the potential of the ML-based LFN cut. Additionally, a case study assuming no events were observed was discussed in Sec. 6.3, which demonstrated the potential of the ML-based LFN cut based on its signal efficiency.

7.2. Outlook

First and foremost, the solution to mitigate vibration-induced LFN events is to implement a vibration isolation system. The Cryogenic Underground TEST (CUTE) facility at SNOLAB, where SuperCDMS detector towers were tested, implemented a suspension system that decouples the detectors from the pulse tube of the dilution refrigerator, effectively dampening vibrations. This is expected to reduce the LFN events significantly for future experiments.

The ML-based approach developed in this analysis has broader potential applications, such as removing other types of noise, including glitch noise and square pulse noise. To adapt this method, the LFN-background sample should be replaced with corresponding noise samples. Ongoing efforts in positional reconstruction are particularly significant, as accurate position reconstruction will improve the ability to discriminate surface events and properly fiducializing events.

It was evident in the CDMSlite Run3 analysis that the radial cut significantly limited the final DM signal efficiency, and constrained the ability to search for low-mass DM. Improvements in the HV field uniformity are essential to reduce the occurrence of RNTL events, which the radial cut primarily addressed. Furthermore, developing a more accurate positional reconstruction method are also expected to enhance event fiducialization without severely impacting DM signal efficiency.

However, removing the radial cut in this analysis led to an excess of events in the low-energy region in the case of the new analysis based on the ML-based LFN cut. This outcome was attributed to the ML model being trained and tested on LFN-background samples that had already passed the radial cut. The LFN-background samples therefore had energies starting from 0.07 keV_{ee} , and the purity was also calculated for energies above 0.07 keV_{ee} . Consequently, the network was not exposed to lower-energy events. The proper solution for this case involves training the ML model with events that are not passed through the radial cut, which allows for a more robust response to low-energy events.

To use the current ML model for the case without the radial cut application, placing a stringent cut on the signal probability to remove low-energy events can be a workaround, as most of these events have a signal probability below 1. Although such a stringent cut reduces signal efficiency, it is still expected to perform better than the previous bifurcated LFN cut. A short study, discussed in Sec. 6.2, examined the effects of applying a signal probability cut at 0.99. The events passing through the signal probability cut at 0.9 need to be studied to identify potential issues with the signal purity. One needs to determine whether these events were removed with a more stringent cut or if the stringent cut merely resulted in poor efficiency without achieving the purity necessary for the DM search analysis.

A robust simulation of samples is essential for ML-based analyses. The lack of such samples initially posed a big challenge during the development of this project, leading to the development of the SUPR method. This method is planned to serve as a tool for comparing simulated traces from other data generation methods for future analyses. While the samples from the official SuperCDMS simulation for this analysis were not adequately developed due to the outdated experiment and limited use of its data, ongoing simulation efforts for the SuperCDMS SNOLAB experiment aim to

provide robust simulation datasets.

Another critical issue identified in this work was the mis-reconstruction of energies for certain events, which negatively impacted limit-setting by introducing events in energy regions where they should not be present. For the case with the radial cut, two mis-reconstructed events in low-energies were observed, which passed the ML-based LFN cut but failed the previous bifurcated LFN cut. At first, the cause of these events failing the previous cut has to be identified. This was not done in this analysis as the necessary computing facilities to reapply the cuts are no longer operational, preventing a more detailed investigation within this thesis timeline. Additionally, the final energy spectrum contained fewer than five mis-reconstructed events, which also require further study. Similar issues were noted in the case without the radial cut, so a thorough investigation is needed to find the extent and causes of this energy mis-reconstruction and its broader impact on this and other analyses using this dataset.

Closing remarks:

In conclusion, the ML-based LFN cut demonstrated high signal efficiency and purity, enabling the identification of additional events and setting competitive limits compared to the previous analysis using the bifurcated LFN cut. This analysis also revealed issues with energy reconstruction. Furthermore, the SUPR method was developed for trace generation based on event position during this analysis and is planned to serve as a validation tool for other simulation efforts. The gain in the signal efficiency was limited by the radial cut application. Thus case study when radial cut was not applied was investigated, it showed improved sensitivity upto $0.8 \text{ GeV}/c^2$. The gain in signal efficiency was particularly evident in the case study assuming no observed events, highlighting the ability of this approach to improve sensitivity for low-mass DM searches.

Appendices

A. Hyperparameter Optimization

The hyperparameter optimization was performed for both the CNN and CNN-LSTM models, as explained in Sec. 4.4 and Sec. 4.5, respectively. The optimization procedure involved using random search [145], Bayesian optimization [146], or a combination of both, depending on computing availability at the time. Random search picks parameter combinations randomly to find the best parameter combination; however, Bayesian optimization builds a probabilistic model to obtain the optimal parameters. The key metric for evaluation is accuracy, represented by the score. The optimization was carried out using a 3-fold cross-validation [151, 152], with the final score being the average of the cross-validation results. Cross-validation is a resampling technique used to train the network on different subsets of the training data in order to evaluate the robustness of the results. The score error mentioned in this section corresponds to the standard deviation of the cross-validation results. The models were finalized based on achieving the highest score with the lowest possible error.

A.1. CNN architecture

Among the architectures tested initially, the simplest one that yielded the highest accuracy is shown in Fig. 4.15. This architecture consists of two 1D convolution layers, followed by a Global Average Pooling (GAP) layer, two dense layers, and a final dense layer that outputs the signal probability. Table. 4.2 shows the values of the finalized model for all three channel configurations.

The details of the parameter tuning for this architecture are presented in this section. For all three channel configurations, 10000 samples consisting of equal portions of signal and LFN-background samples, were used. Both random search and Bayesian optimization were applied for the single-channel and PT-channel configurations, while only Bayesian optimization was used for the multi-channel configuration due to time and computing constraints.

A.1.1. Single-channel

The combinations tested during the hyperparameter optimization are showed in Fig. A.1. The final combination chosen is shown with a thick line, while the other combinations are represented with thin lines. In addition to achieving high accuracy, L2 kernel regularization with a regularizer value of 0.01 was applied to the convolution layers and dense layers to improve stability during training.

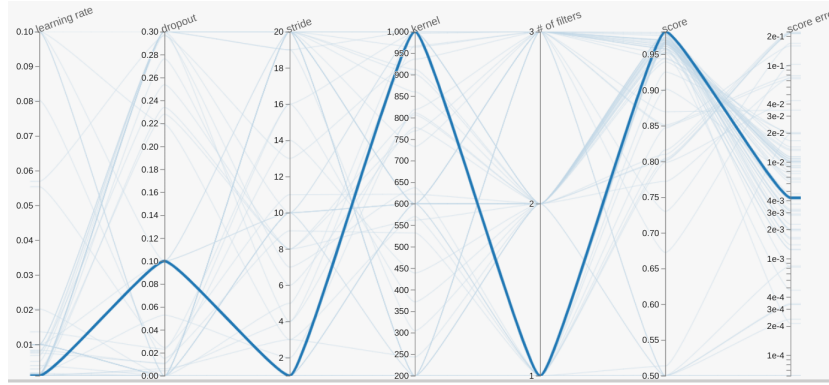


Figure A.1.: The hyperparameter optimization graph for the CNN model with the single-channel configuration. The finalized combination is highlighted with a thick line, while other combinations are shown as thin lines. The number of filters (shown as # of filters) was scaled by a factor of 32. The stride, kernel size (shown as kernel), and number of filters correspond to the first convolutional layer, whereas the second convolutional layer uses half the kernel size, twice the number of filters, and a stride of 1. Dropout is applied between the fully connected layers. The optimal combination was selected based on achieving the highest score with the lowest error.

A.1.2. PT-channel

For the PT-channel configuration, the hyperparameter optimization was performed using 10000 samples, consisting of equal portions of signal and LFN-background samples. Fig. A.2 shows the combinations tested during the hyperparameter optimization. The final combination chosen is shown with a thick line, while the other combinations are represented with thin lines. In addition to achieving high accuracy, L2 kernel and bias regularization with a regularizer value of 0.01 was applied to the convolution layers and dense layers to improve stability during training. Additionally, the samples were pre-processed using a low-pass filter, as it showed the highest accuracy.

A.1.3. Multi-channel

For the PT-channel configuration, the hyperparameter optimization was performed using 10000 samples, consisting of equal portions of signal and LFN-background samples. Fig. A.3 shows the combinations tested during the hyperparameter optimization. The final combination chosen is shown with a thick line, while the other combinations are represented with thin lines. In addition to achieving high accuracy, L2 kernel and bias regularization with a regularizer value of 0.01 was applied to the convolution and dense layers to improve stability during training.

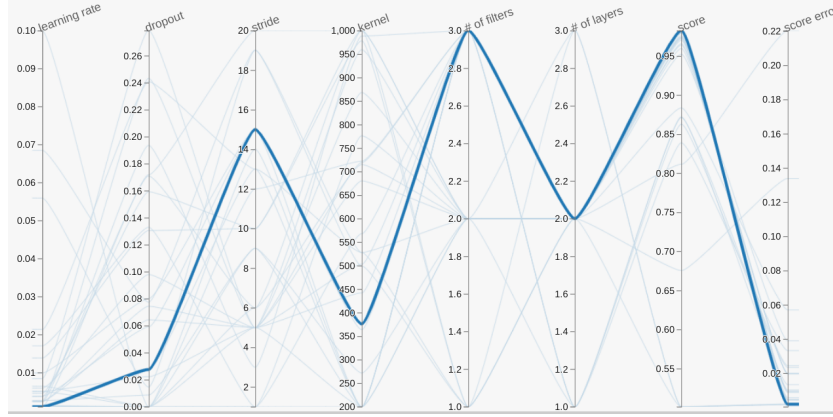


Figure A.2.: The hyperparameter optimization graph for the CNN model with the PT-channel configuration. The finalized combination is highlighted with a thick line, while other combinations are shown as thin lines. The number of filters (shown as # of filters) was scaled by a factor of 32. The stride, kernel size (shown as kernel), and number of filters correspond to the first convolutional layer, whereas the second convolutional layer uses half the kernel size, twice the number of filters, and a stride of 1. Dropout is applied between the fully connected layers. The optimal combination was selected based on achieving the highest score with the lowest error.

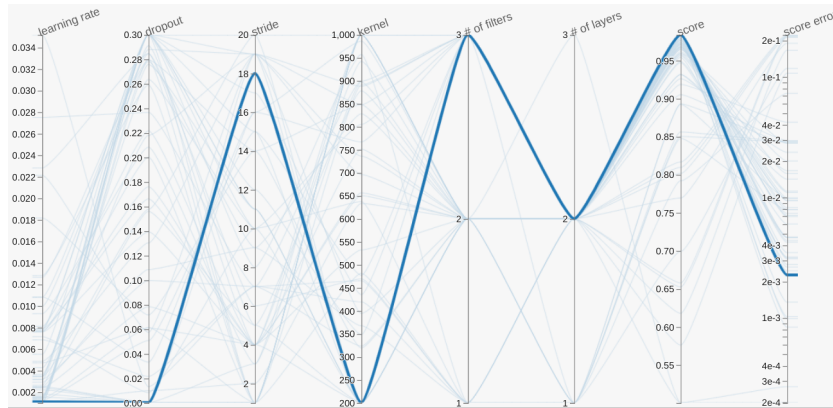


Figure A.3.: The hyperparameter optimization graph for the CNN model with the multi-channel configuration. The finalized combination is highlighted with a thick line, while other combinations are shown as thin lines. The number of filters (shown as # of filters) was scaled by a factor of 32. The stride, kernel size (shown as kernel), and number of filters correspond to the first convolutional layer, whereas the second convolutional layer uses half the kernel size, twice the number of filters, and a stride of 1. Dropout is applied between the fully connected layers. The optimal combination was selected based on achieving the highest score with the lowest error.

A.2. CNN-LSTM architecture

The CNN-LSTM architecture consists of two 1D convolution layers, followed by two LSTM layers, two dense layers, with the final layer that outputs the signal probability. Dropouts are applied between convolution, LSTM layers. Batch normalization are applied between convolution, LSTM layers for only PT- and multi-channels. Fig. 4.44 displays a schematic of the architecture. The values of the finalized model for all three channel configurations is shown in Table. 4.3.

The details of the parameter tuning for this architecture are presented in this section. Both random search and Bayesian optimization were applied for the multi-channel configuration, while only random search was used for the single-channel and PT-channel configuration due to time and computing constraints.

A.2.1. Single-channel

For the single-channel configuration, 5000 samples consisting of equal portions of signal and LFN-background samples, were used. The combinations tested during the hyperparameter optimization are showed in Fig. A.4. The final combination chosen is shown with a thick line, while the other combinations are represented with thin lines.

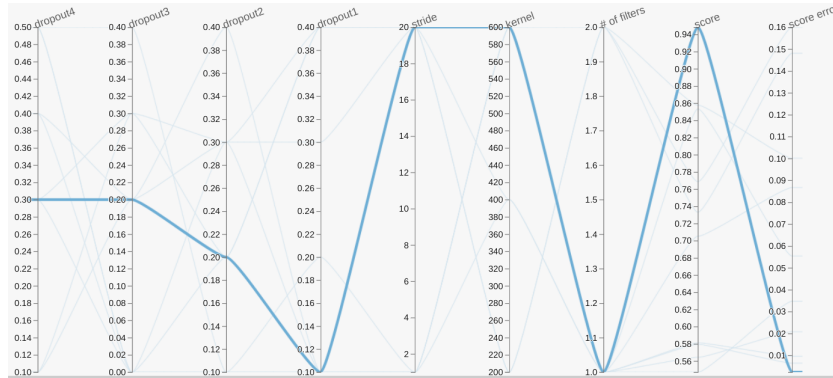


Figure A.4.: The hyperparameter optimization graph for the CNN-LSTM model without batch normalization in the single-channel configuration. The finalized combination is highlighted with a thick line, while other combinations are shown as thin lines. The number of filters (denoted as # of filters) was scaled by a factor of 32. The stride, kernel size (denoted as kernel), and number of filters correspond to the first convolutional layer, while the second convolutional layer uses half the kernel size, twice the number of filters, and a stride of 1. Dropout 1 is applied after the first and the second convolutional layers, Dropout 2 is applied after the first LSTM layer, Dropout 3 is applied after the second LSTM layer, and Dropout 4 is applied between the fully connected layers. The optimal combination was selected based on achieving the highest score with the lowest error.

A.2.2. PT-channel

For the PT-channel configuration, 8000 samples consisting of equal portions of signal and LFN-background samples, were used. The combinations tested during the hyperparameter optimization are showed in Fig. A.5. The final combination chosen is shown with a thick line, while the other combinations are represented with thin lines. In addition to achieving high accuracy, L2 kernel and bias regularization with a regularizer value of 0.01 was applied to the convolution and dense layers to improve stability during training.

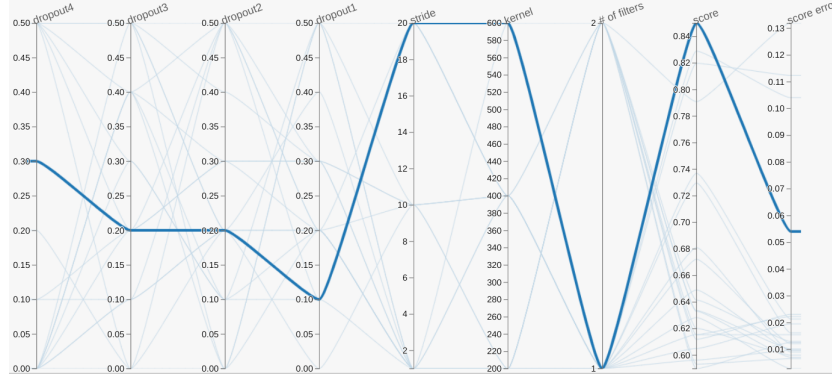


Figure A.5.: The hyperparameter optimization graph for the CNN-LSTM model with the PT-channel configuration. The finalized combination is highlighted with a thick line, while other combinations are shown as thin lines. The number of filters (denoted as # of filters) was scaled by a factor of 32. The stride, kernel size (denoted as kernel), and number of filters correspond to the first convolutional layer, while the second convolutional layer uses half the kernel size, twice the number of filters, and a stride of 1. Dropout 1 is applied after the first and the second convolutional layers, Dropout 2 is applied after the first LSTM layer, Dropout 3 is applied after the second LSTM layer, and Dropout 4 is applied between the fully connected layers. BatchNormalization is applied after each convolutional and LSTM layer and before the corresponding dropouts. The optimal combination was selected based on achieving the highest score with the lowest error.

A.2.3. Multi-channel

For the PT-channel configuration, 10000 samples consisting of equal portions of signal and LFN-background samples, were used. The combinations tested during the hyperparameter optimization are showed in Fig. A.6. The final combination chosen is shown with a thick line, while the other combinations are represented with thin lines. In addition to achieving high accuracy, L2 kernel and bias regularization with a regularizer value of 0.01 was applied to the convolution and dense layers to improve stability during training.

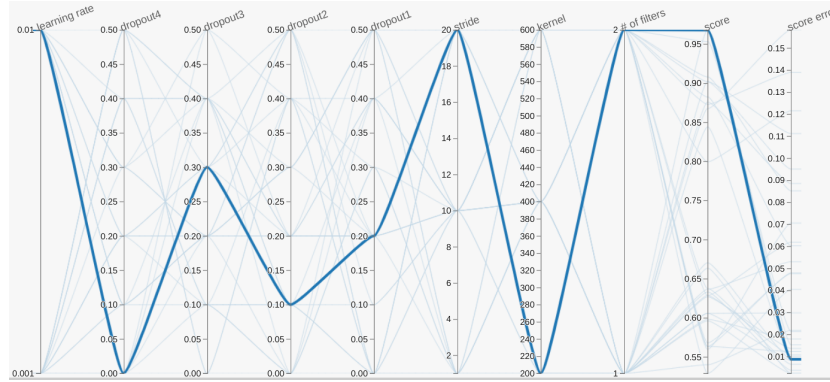


Figure A.6.: The hyperparameter optimization graph for the CNN-LSTM model with the multi-channel configuration. The finalized combination is highlighted with a thick line, while other combinations are shown as thin lines. The number of filters (denoted as # of filters) was scaled by a factor of 32. The stride, kernel size (denoted as kernel), and number of filters correspond to the first convolutional layer, while the second convolutional layer uses half the kernel size, twice the number of filters, and a stride of 1. Dropout 1 is applied after the first and the second convolutional layers, Dropout 2 is applied after the first LSTM layer, Dropout 3 is applied after the second LSTM layer, and Dropout 4 is applied between the fully connected layers. BatchNormalization is applied after each convolutional and LSTM layer and before the corresponding dropouts. The optimal combination was selected based on achieving the highest score with the lowest error.

B. Energy Mis-reconstruction

The final spectrum resulting from the ML-based LFN cut, as discussed in Sec. 5.5, revealed five additional events compared to the previous analysis using the bifurcated LFN cut. Among these, two events were improperly reconstructed, resulting in lower energies than expected. This section provides an estimation of energy for these events, which are shown in Fig. B.1. The energies of these two events are estimated using the Optimal Filter (OF), which provides amplitudes in ADC units (or amperes if the input traces are in amperes).

Figure B.2 presents the OF amplitude of the PT channel against energy in the keV_{ee} scale for events from the final spectrum of the previous analysis using the bifurcated LFN cut. Both quantities are obtained from the official reconstruction pipeline. It is important to note two points: 1) this energy estimation uses the Non-Stationary Optimal Filter (NSOF) to obtain amplitudes, however the estimated energies shown in this section are based on the standard OF amplitudes. 2) non-linearities in the conversion from ADC amplitudes to keV_{ee} arise from the energy corrections discussed in Sec. 3.3.3. Consequently, the OF amplitude of the PT channel on the x-axis is uncorrected, while the energy on the y-axis includes these corrections, as shown in Fig. 3.8 where the spectra before and after energy correction was shown. Despite these non-linearities, a rough linear relationship provides a reasonable estimate of the events' energies from amplitudes.

Event	Mis-reconstructed energy	Current estimated energy
2	0.191	0.581 - 0.945
3	0.091	2.239 - 2.522

Table B.1.: Comparison of the assigned energy and the currently estimated energy for mis-reconstructed events.

The estimated energies are presented as ranges due to the aforementioned reasons. Table B.1 provides a comparison between the assigned and estimated energies. The estimated energies are significantly higher than the assigned values, exceeding them by over three times for Event 2 and more than 25 times for Event 3.

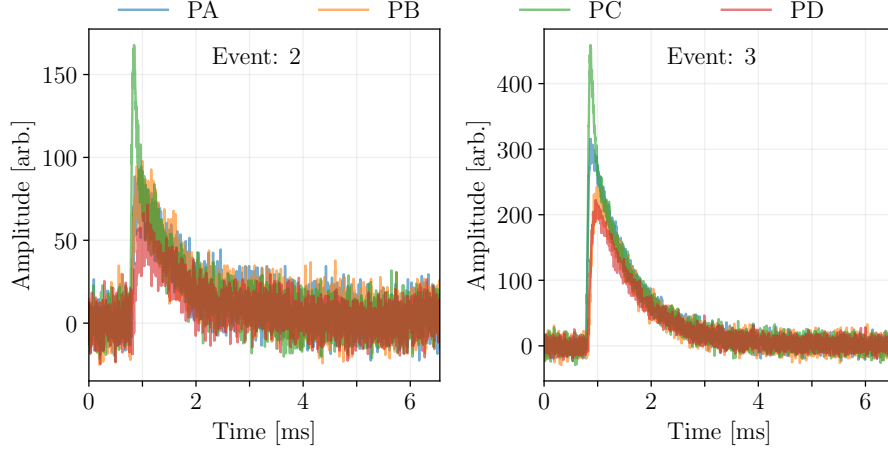


Figure B.1.: Raw traces of the additional events observed that have mis-reconstructed energies. The four phonon channels PA, PB, PC, and PD are shown in blue, orange, green, and red, respectively. The new signal events passed the new ML-based LFN cut and failed the previous bifurcated LFN cut.

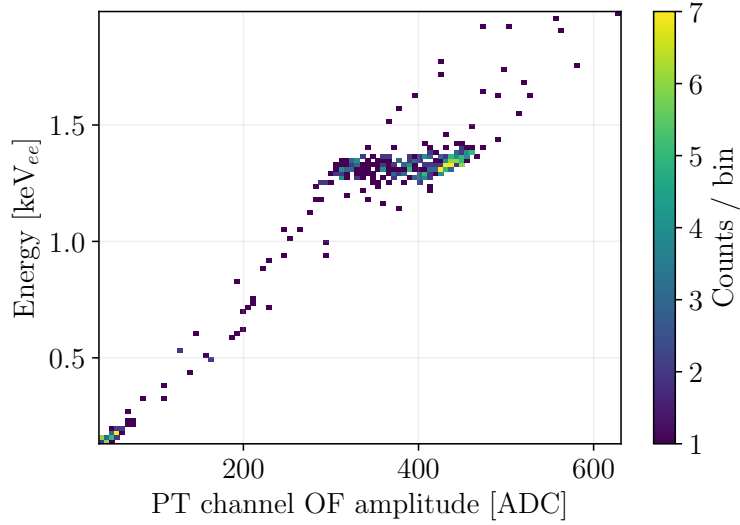


Figure B.2.: Energy in keV_{ee} scale is plotted against the OF amplitude of the PT channel in ADC units for events in the final spectrum from the previous analysis using the bifurcated LFN cut. Both quantities are obtained from the official reconstruction pipeline.

C. Supplementary Information on the Impact of Radial Cut

Section 6.1 discussed the impact of the radial cut on the final spectrum in both the previous analysis using the bifurcated LFN cut and the new analysis using the ML-based LFN cut. The new analysis revealed an excess of events at low energies below 0.1 keV_{ee} , as seen in Fig. C.1. This section provides an overview of the characteristics of these events and their raw traces for both the nominal signal probability threshold of 0.9 (used in the final analysis) and the more stringent threshold of 0.99, which was explored in a post-unblinding study described in Sec. 6.2. For the case with the radial cut applied,

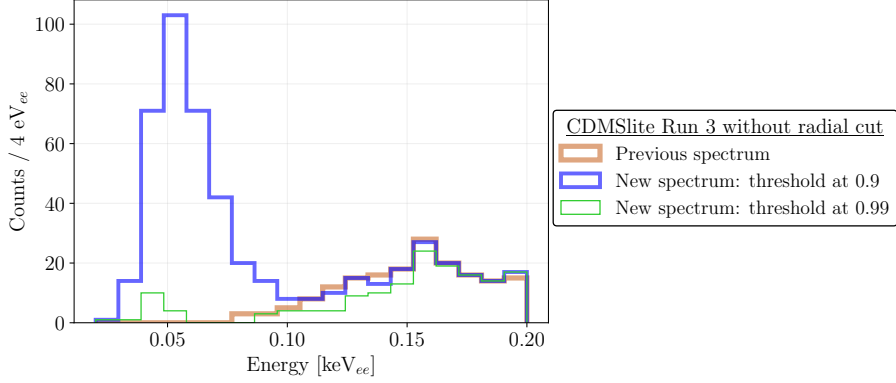


Figure C.1.: Energy spectrum of the CDMSlite Run 3 events below 0.08 keV_{ee} for previous analysis using the bifurcated LFN cut and the new analysis using the ML-based LFN cut for signal probability thresholds 0.9 and 0.99.

there are 351 additional events in the energy spectrum that passed the ML-based LFN cut with a signal probability threshold of 0.9 and failed the bifurcated LFN cut. Of these, $\sim 54\%$ have noise-like or signal-like features outside the trigger region (trigger is set around 0.8 ms i.e. 500^{th} time bin). Additionally, 10 events are high-energy events that were mis-reconstructed and have low-energies. Raw traces of randomly selected events from the energy spectrum of the new analysis with a signal probability threshold of 0.9, and that did not pass the bifurcated LFN cut, are shown in Fig. C.2, along with their reconstructed energies. The Event 4 and 9 in the figure are high-energy mis-reconstructed events, while Events 2, 3, and 7 exhibit key features outside the trigger region. Events 1, 5, 7, and 8 show signal-like features with small amplitudes, but these are predominantly affected by high baseline noise.

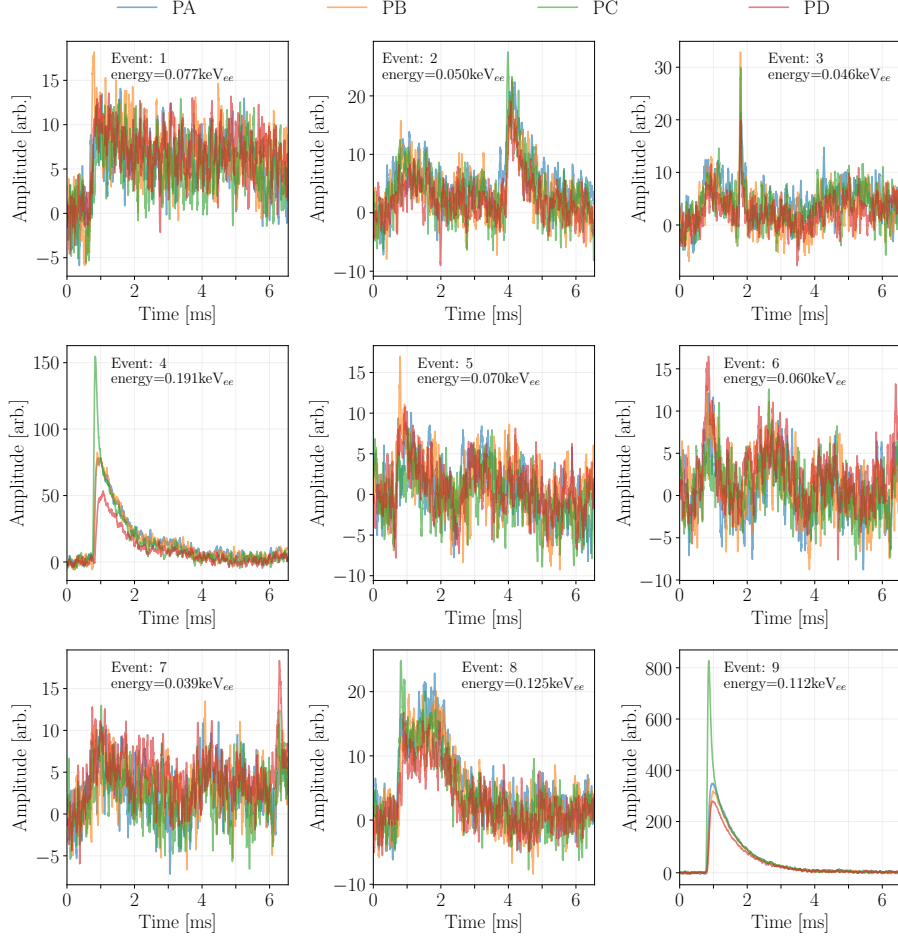


Figure C.2.: Raw traces of additional events observed with the four phonon channels (PA, PB, PC, and PD). The radial cut was not applied. These events passed the ML-based LFN cut with a signal probability threshold of 0.9 and failed the previous bifurcated LFN cut. Reconstructed energies are displayed in the text box. Moving average is applied on the raw traces for visual aid.

For the case without the radial cut applied, there are 30 additional events in the energy spectrum that passed the ML-based LFN cut with a signal probability threshold of 0.99 and failed the bifurcated LFN cut. Of these, 14 events exhibit noise-like or signal-like features outside the trigger region. The 10 high-energy events identified in the threshold 0.9 case are also present in this case due to their signal probability of 1, as expected. Raw traces of randomly selected events from the energy spectrum of the new analysis with a signal probability threshold of 0.99, which did not pass the bifurcated LFN cut, are shown in Fig. C.3, along with their reconstructed energies. Events 2, 4, 5, and 6 in the figure are high-energy mis-reconstructed events, while Events 1, 3, and 8 show key features outside the trigger region. Event 7 is difficult to interpret due to the baseline noise and any underlying feature with very low amplitude. Event 9 shows signal-like features with small amplitudes but is predominantly affected by high baseline noise. Although the events below 0.1 keV_{ee} were significantly reduced with a signal probability threshold of 0.99, this stringent threshold also excluded 50 “good events” below an energy of $\sim 0.16 \text{ keV}_{ee}$ that had passed the bifurcated LFN cut but did not pass the ML-based cut with a higher threshold.

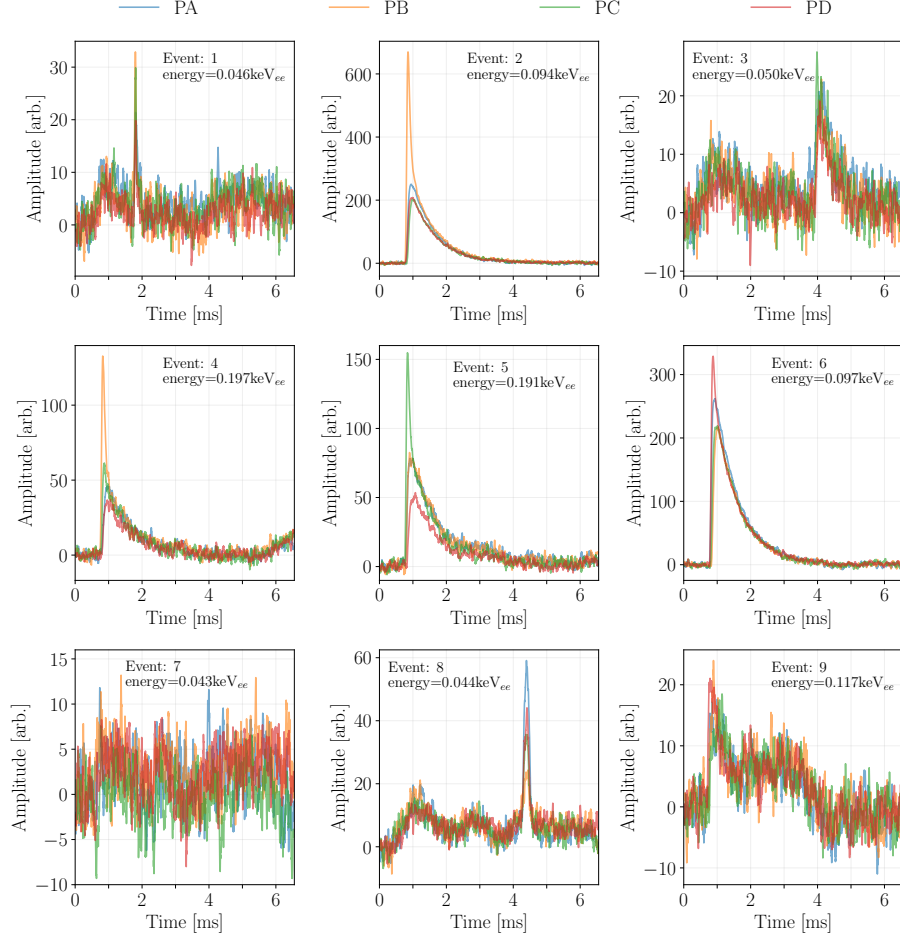


Figure C.3.: Raw traces of additional events observed with the four phoron channels (PA, PB, PC, and PD). The radial cut was not applied. These events passed the ML-based LFN cut with a signal probability threshold of 0.99 and failed the previous bifurcated LFN cut. Reconstructed energies are displayed in the text box. Moving average is applied on the raw traces for visual aid.

Bibliography

- [1] F. Zwicky, “Die Rotverschiebung von extragalaktischen Nebeln”, *Helv. Phys. Acta*, vol. 6, pp. 110–127, 1933. doi: 10.1007/s10714-008-0707-4.
- [2] Gianfranco Bertone and Dan Hooper, “History of dark matter”, *Reviews of Modern Physics*, vol. 90, Oct 2018. doi: 10.1103/revmodphys.90.045002.
- [3] Vera C. Rubin, W. Kent Ford, and Norbert Thonnard, “Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 ($R = 4\text{kpc}$) to UGC 2885 ($R = 122\text{ kpc}$)”, *The Astrophysical Journal*, vol. 238, pp. 471–487, 1980. doi: 10.1086/158003.
- [4] Vera C. Rubin and W. Kent Ford, “Rotation of the andromeda nebula from a spectroscopic survey of emission regions”, *The Astrophysical Journal*, vol. 159, pp. 379–403, 1970. doi: 10.1086/150317.
- [5] Vera C. Rubin, W. Kent Ford, Jr., and Norbert Thonnard, “Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa through Sc”, *Astrophys. J. Lett.*, vol. 225, pp. L107–L111, 1978. doi: 10.1086/182804.
- [6] 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*, vol. 6, no. 1, pp. 87–112, 1996. doi: 10.1016/S0927-6505(96)00047-3.
- [7] D. J. Fixsen, “The temperature of the cosmic microwave background”, *The Astrophysical Journal*, vol. 707, p. 916–920, Nov 2009. doi: 10.1088/0004-637x/707/2/916.
- [8] Arno A. Penzias and Robert G. Wilson, “A measurement of excess antenna temperature at 4080-mc/s”, *The Astrophysical Journal*, vol. 142, pp. 419–421, 1965. doi: 10.1086/148307.
- [9] S. Navas *et al.*, “Review of particle physics”, *Phys. Rev. D*, vol. 110, no. 3, p. 030001, 2024. doi: 10.1103/PhysRevD.110.030001.
- [10] R. Adam, *et al.*, “Planck2015 results: X. diffuse component separation: Foreground maps”, *Astronomy and Astrophysics*, vol. 594, p. A10, Sep 2016. doi: 10.1051/0004-6361/201525967.
- [11] P J.E. Peebles and J T Yu, “Primeval adiabatic perturbation in an expanding universe.”, *Astrophys. J.* 162: 815-36(Dec 1970).

- [12] Daniel J. Eisenstein and Wayne Hu, “Baryonic features in the matter transfer function”, *The Astrophysical Journal*, vol. 496, p. 605–614, Apr. 1998. doi: 10.1086/305424.
- [13] P. A. R. Ade, *et al.*, “Planck2015 results: XIII. cosmological parameters”, *Astronomy and Astrophysics*, vol. 594, p. A13, Sep 2016. doi: 10.1051/0004-6361/201525830.
- [14] Wayne Hu, Masataka Fukugita, Matias Zaldarriaga, and Max Tegmark, “Cosmic microwave background observables and their cosmological implications”, *The Astrophysical Journal*, vol. 549, p. 669–680, Mar. 2001. doi: 10.1086/319449.
- [15] N. Aghanim, *et al.*, “Planck2018 results: VI. cosmological parameters”, *Astronomy and Astrophysics*, vol. 641, p. A6, Sept. 2020. doi: 10.1051/0004-6361/201833910.
- [16] P J.E. Peebles, “Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations”, *Astrophys. J., Lett. Ed.; (United States)*, vol. 263:1, Dec 1982. doi: 10.1086/183911.
- [17] Albert Einstein, “Die Grundlage der allgemeinen Relativitätstheorie”, *Annalen Phys.*, vol. 49, no. 7, pp. 769–822, 1916. doi: 10.1002/andp.19163540702.
- [18] Y. Mellier, F. Bernardeau, and L. van Waerbeke, “Dark matter and gravitational lensing”, 1998. arxiv:9802005.
- [19] Richard Massey, Thomas Kitching, and Johan Richard, “The dark matter of gravitational lensing”, *Reports on Progress in Physics*, vol. 73, p. 086901, Jul 2010. doi: 10.1088/0034-4885/73/8/086901.
- [20] Douglas Clowe, Maruša Bradač, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky, “A direct empirical proof of the existence of dark matter”, *The Astrophysical Journal*, vol. 648, p. L109–L113, Aug 2006. doi: 10.1086/508162.
- [21] Scott W. Randall, Maxim Markevitch, Douglas Clowe, Anthony H. Gonzalez, and Marusa Bradač, “Constraints on the self-interaction cross section of dark matter from numerical simulations of the merging galaxy cluster 1E 0657-56”, *The Astrophysical Journal*, vol. 679, p. 1173–1180, Jun 2008. doi: 10.1086/587859.
- [22] Maruša Bradač, Steven W. Allen, Tommaso Treu, Harald Ebeling, Richard Massey, R. Glenn Morris, Anja von der Linden, and Douglas Applegate, “Revealing the properties of dark matter in the merging cluster MACS J0025.4-1222”, *The Astrophysical Journal*, vol. 687, p. 959, nov 2008. doi: 10.1086/591246.
- [23] Mordehai Milgrom, “A modification of the newtonian dynamics as a possible alternative to the hidden mass hypothesis”, *The Astrophysical Journal*, vol. 270, pp. 365–370, 1983. doi: 10.1086/161130.

- [24] J Bekenstein and M Milgrom, “Does the missing mass problem signal the breakdown of newtonian gravity”, *Astrophys. J.; (United States)*, vol. 286:1, Nov 1984. doi: 10.1086/162570.
- [25] Benoît Famaey and Stacy S. McGaugh, “Modified newtonian dynamics (MOND): Observational phenomenology and relativistic extensions”, *Living Reviews in Relativity*, vol. 15, Sep 2012. doi: 10.12942/lrr-2012-10.
- [26] Jacob D. Bekenstein, “Relativistic gravitation theory for the modified newtonian dynamics paradigm”, *Phys. Rev. D*, vol. 70, p. 083509, Oct 2004. doi: 10.1103/PhysRevD.70.083509.
- [27] Alexander Kusenko and Leslie J. Rosenberg, “Snowmass-2013 cosmic frontier 3 (CF3) working group summary: Non-wimp dark matter”, 2013. arXiv:1310.8642.
- [28] Lars Bergström, “Dark matter candidates”, *New Journal of Physics*, vol. 11, p. 105006, Oct 2009. doi: 10.1088/1367-2630/11/10/105006.
- [29] Jonathan L. Feng, “Dark matter candidates from particle physics and methods of detection”, *Annual Review of Astronomy and Astrophysics*, vol. 48, p. 495–545, Aug 2010. doi: 10.1146/annurev-astro-082708-101659.
- [30] Torsten Bringmann and Stefan Hofmann, “Thermal decoupling of WIMPs from first principles”, *Journal of Cosmology and Astroparticle Physics*, vol. 2007, p. 016, apr 2007. doi: 10.1088/1475-7516/2007/04/016.
- [31] Edward W. Kolb, *The Early Universe*, vol. 69. Taylor and Francis, 5 2019. doi: 10.1201/9780429492860.
- [32] Gerard Jungman, Marc Kamionkowski, and Kim Griest, “Supersymmetric dark matter”, *Physics Reports*, vol. 267, no. 5, pp. 195–373, 1996. doi: 10.1016/0370-1573(95)00058-5.
- [33] F. Donato, N. Fornengo, and S. Scopel, “Effects of galactic dark halo rotation on WIMP direct detection”, *Astroparticle Physics*, vol. 9, no. 3, pp. 247–260, 1998. doi: 10.1016/S0927-6505(98)00025-5.
- [34] R. Agnese, *et al.*, “Low-mass dark matter search with CDMSlite”, *Phys. Rev. D*, vol. 97, p. 022002, Jan 2018. doi: 10.1103/PhysRevD.97.022002.
- [35] T. Piffl, *et al.*, “The RAVE survey: the galactic escape speed and the mass of the milky way”, *Astronomy and Astrophysics*, vol. 562, p. A91, Feb 2014. doi: 10.1051/0004-6361/201322531.
- [36] Matthew James Wilson, *A New Search for Low-mass Dark Matter and an Examination and Reduction of the Uncertainty Due to The Photoelectric Absorption Cross Section Using a Cryogenic Silicon Detector With Single-charge Sensitivity*. Phd thesis, University of Toronto, 2021. Available at <https://inspirehep.net/literature/2048844>.

- [37] Matthew R. Buckley and W. Hugh Lippincott, “Spin-dependent interpretation for possible signals of light dark matter”, *Physical Review D*, vol. 88, Sep 2013. doi: 10.1103/physrevd.88.056003.
- [38] Benjamin W. Lee and Steven Weinberg, “Cosmological lower bound on Heavy-Neutrino masses”, *Phys. Rev. Lett.*, vol. 39, pp. 165–168, Jul 1977. doi: 10.1103/PhysRevLett.39.165.
- [39] Itay M. Bloch, Rouven Essig, Kohsaku Tobioka, Tomer Volansky, and Tien-Tien Yu, “Searching for dark absorption with direct detection experiments”, *Journal of High Energy Physics*, vol. 2017, Jun 2017. doi: 10.1007/jhep06(2017)087.
- [40] Steven Weinberg, “A new light boson?”, *Phys. Rev. Lett.*, vol. 40, pp. 223–226, Jan 1978. doi: 10.1103/PhysRevLett.40.223.
- [41] F. Wilczek, “Problem of strong p and t invariance in the presence of instantons”, *Phys. Rev. Lett.*, vol. 40, pp. 279–282, Jan 1978. doi: 10.1103/PhysRevLett.40.279.
- [42] Nikita Blinov, Matthew J. Dolan, Patrick Draper, and Jonathan Kozaczuk, “Dark matter targets for axionlike particle searches”, *Physical Review D*, vol. 100, Jul 2019. doi: 10.1103/physrevd.100.015049.
- [43] C. B. Adams, *et al.*, “Axion dark matter”, 2023. arXiv:2203.14923.
- [44] Scott Dodelson and Lawrence M. Widrow, “Sterile neutrinos as dark matter”, *Physical Review Letters*, vol. 72, p. 17–20, Jan 1994. doi: 10.1103/PhysRevLett.72.17.
- [45] Palash B. Pal and Lincoln Wolfenstein, “Radiative decays of massive neutrinos”, *Phys. Rev. D*, vol. 25, pp. 766–773, Feb 1982. doi: 10.1103/PhysRevD.25.766.
- [46] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens, and O. Ruchayskiy, “Sterile neutrino dark matter”, *Progress in Particle and Nuclear Physics*, vol. 104, p. 1–45, Jan 2019. doi: 10.1016/j.pnpnp.2018.07.004.
- [47] Miguel D. Campos and Werner Rodejohann, “Testing keV sterile neutrino dark matter in future direct detection experiments”, *Physical Review D*, vol. 94, Nov 2016. doi: 10.1103/physrevd.94.095010.
- [48] Edward W. Kolb, Daniel J. H. Chung, and Antonio Riotto, “WIMPZILLAS!”, 1998. doi: 10.1063/1.59655.
- [49] Gianfranco Bertone, Dan Hooper, and Joseph Silk, “Particle dark matter: evidence, candidates and constraints”, *Physics Reports*, vol. 405, p. 279–390, Jan 2005. doi: 10.1016/j.physrep.2004.08.031.
- [50] George F. Chapline, “Cosmological effects of primordial black holes”, *Nature*, vol. 253, no. 5489, pp. 251–252, 1975. doi: 10.1038/253251a0.

- [51] Karsten Jedamzik, “Primordial black hole dark matter and the ligo/virgo observations”, *Journal of Cosmology and Astroparticle Physics*, vol. 2020, p. 022, sep 2020. doi: 10.1088/1475-7516/2020/09/022.
- [52] Bernard Carr, Florian Kühnel, and Marit Sandstad, “Primordial black holes as dark matter”, *Phys. Rev. D*, vol. 94, p. 083504, Oct 2016. doi: 10.1103/PhysRevD.94.083504.
- [53] Simeon Bird, Ilias Cholis, Julian B. Muñoz, Yacine Ali-Haïmoud, Marc Kamionkowski, Ely D. Kovetz, Alvise Raccanelli, and Adam G. Riess, “Did LIGO detect dark matter?”, *Phys. Rev. Lett.*, vol. 116, p. 201301, May 2016. doi: 10.1103/PhysRevLett.116.201301.
- [54] P. Tisserand, *et al.*, “Limits on the macho content of the galactic halo from the EROS-2 survey of the magellanic clouds”, *Astronomy and Astrophysics*, vol. 469, p. 387–404, Apr 2007. doi: 10.1051/0004-6361:20066017.
- [55] A. Albert, *et al.*, “Searching for dark matter annihilation in recently discovered milky way satellites with fermi-lat”, *The Astrophysical Journal*, vol. 834, p. 110, Jan 2017. doi: 10.3847/1538-4357/834/2/110.
- [56] F. Donato, D. Maurin, P. Brun, T. Delahaye, and P. Salati, “Constraints on WIMP dark matter from the high energy PAMELA pp data”, *Phys. Rev. Lett.*, vol. 102, p. 071301, Feb 2009. doi: 10.1103/PhysRevLett.102.071301.
- [57] Ming-Yang Cui, Qiang Yuan, Yue-Lin Sming Tsai, and Yi-Zhong Fan, “Possible dark matter annihilation signal in the AMS-02 antiproton data”, *Phys. Rev. Lett.*, vol. 118, p. 191101, May 2017. doi: 10.1103/PhysRevLett.118.191101.
- [58] Nicolò Trevisani, “Collider Searches for Dark Matter (ATLAS + CMS)”, *Universe*, vol. 4, no. 11, p. 131, 2018. doi: 10.3390/universe4110131.
- [59] Marcello Campajola, “Dark sector first results at Belle II”, *Physica Scripta*, vol. 96, p. 084005, May 2021. doi: 10.1088/1402-4896/abfef2.
- [60] A. Benoit, *et al.*, “First results of the EDELWEISS WIMP search using a 320 g heat-and-ionization ge detector”, *Physics Letters B*, vol. 513, no. 1, pp. 15–22, 2001. doi: 10.1016/S0370-2693(01)00754-7.
- [61] E. Armengaud, *et al.*, “Searching for low-mass dark matter particles with a massive ge bolometer operated above-ground”, Jan 2019. doi: 10.48550/arXiv.1901.03588.
- [62] H. Kluck, *et al.*, “Search for low-mass dark matter with the CRESST experiment”, 2017. arXiv:1711.01285.
- [63] A. H. Abdelhameed, *et al.*, “First results from the CRESST-III low-mass dark matter program”, *Phys. Rev. D*, vol. 100, p. 102002, Nov 2019. doi: 10.1103/PhysRevD.100.102002.

- [64] E. Aprile, *et al.*, “Light dark matter search with ionization signals in XENON1T”, *Physical Review Letters*, vol. 123, Dec 2019. doi: 10.1103/physrevlett.123.251801.
- [65] E. Aprile, *et al.*, “First dark matter search with nuclear recoils from the XENONnT experiment”, *Physical Review Letters*, vol. 131, Jul 2023. doi: 10.1103/physrevlett.131.041003.
- [66] MengJiao Xiao, *et al.*, “First dark matter search results from the PandaX-I experiment”, *Science China Physics, Mechanics and Astronomy*, vol. 57, p. 2024–2030, Sep 2014. doi: 10.1007/s11433-014-5598-7.
- [67] Qihong Wang, *et al.*, “Results of dark matter search using the full PandaX-II exposure”, *Chinese Physics C*, vol. 44, p. 125001, Dec 2020. doi: 10.1088/1674-1137/abb658.
- [68] J. Aalbers, *et al.*, “First dark matter search results from the LUX-ZEPLIN (LZ) experiment”, *Phys. Rev. Lett.*, vol. 131, p. 041002, Jul 2023. doi: 10.1103/PhysRevLett.131.041002.
- [69] Bei Cai, “The DEAP-3600 dark matter experiment”, 2015. arXiv:1511.00949.
- [70] P. Agnes, *et al.*, “DarkSide-50 532-day dark matter search with low-radioactivity argon”, *Phys. Rev. D*, vol. 98, p. 102006, Nov 2018. doi: 10.1103/PhysRevD.98.102006.
- [71] C. Amole *et al.*, “Dark Matter Search Results from the PICO-2L C₃F₈ Bubble Chamber”, *Phys. Rev. Lett.*, vol. 114, no. 23, p. 231302, 2015. doi: 10.1103/PhysRevLett.114.231302.
- [72] C. Amole, *et al.*, “Dark matter search results from the complete exposure of the PICO-60 C₃F₈ bubble chamber”, *Physical Review D*, vol. 100, Jul 2019. doi: 10.1103/physrevd.100.022001.
- [73] Michael Crisler, Rouven Essig, Juan Estrada, Guillermo Fernandez, Javier Tiffenberg, Miguel Sofo Haro, Tomer Volansky, and Tien-Tien Yu, “SENSEI: First direct-detection constraints on sub-gev dark matter from a surface run”, *Physical Review Letters*, vol. 121, Aug 2018. doi: 10.1103/physrevlett.121.061803.
- [74] N. Castelló-Mor, “DAMIC-M experiment: Thick, silicon ccds to search for light dark matter”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 958, p. 162933, Apr 2020. doi: 10.1016/j.nima.2019.162933.
- [75] I. Alkhatib, *et al.*, “Light dark matter search with a high-resolution athermal phonon detector operated above ground”, *Phys. Rev. Lett.*, vol. 127, p. 061801, Aug 2021. doi: 10.1103/PhysRevLett.127.061801.

- [76] R. Agnese, *et al.*, “Search for low-mass dark matter with CDMSlite using a profile likelihood fit”, *Physical Review D*, vol. 99, Mar 2019. doi: 10.1103/PhysRevD.99.062001.
- [77] F. Ruppin, J. Billard, E. Figueroa-Feliciano, and L. Strigari, “Complementarity of dark matter detectors in light of the neutrino background”, *Phys. Rev. D*, vol. 90, p. 083510, Oct 2014. doi: 10.1103/PhysRevD.90.083510.
- [78] “SuperCDMS dark matter limit plotter (v5.18).” <https://supercdms.slac.stanford.edu/.../dark-matter-limit-plotter>.
- [79] Eleanor Fascione, *Searching for Electron-Interacting Dark Matter with the Super Cryogenic Dark Matter Search*. Phd thesis, Queen’s University, 2023. Available at <https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/fascione.pdf>.
- [80] D. Abrams, *et al.*, “Exclusion limits on the WIMP-nucleon cross section from the cryogenic dark matter search”, *Phys. Rev. D*, vol. 66, p. 122003, Dec. 2002. doi: 10.1103/PhysRevD.66.122003.
- [81] D. S. Akerib, *et al.*, “New results from the Cryogenic Dark Matter Search experiment”, *Phys. Rev. D*, vol. 68, p. 082002, Oct 2003. doi: 10.1103/PhysRevD.68.082002.
- [82] D. S. Akerib, *et al.*, “Exclusion limits on the WIMP-nucleon cross section from the first run of the Cryogenic Dark Matter Search in the Soudan underground laboratory”, *Physical Review D*, vol. 72, Sep 2005. doi: 10.1103/PhysRevD.72.052009.
- [83] R. Agnese, *et al.*, “Silicon detector dark matter results from the final exposure of CDMS II”, Dec. 2013. doi: 10.1103/PhysRevLett.111.251301.
- [84] Z. Ahmed *et al.*, “Dark Matter Search Results from the CDMS II Experiment”, *Science*, vol. 327, pp. 1619–1621, 2010. doi: 10.1126/science.1186112.
- [85] R. Agnese, *et al.*, “Results from the Super Cryogenic Dark Matter Search experiment at Soudan”, *Physical Review Letters*, vol. 120, Feb 2018. doi: 10.1103/PhysRevLett.120.061802.
- [86] R. Agnese, *et al.*, “Search for low-mass weakly interacting massive particles with SuperCDMS”, *Phys. Rev. Lett.*, vol. 112, p. 241302, Jun 2014. doi: 10.1103/PhysRevLett.112.241302.
- [87] R. Agnese, *et al.*, “Search for low-mass weakly interacting massive particles using voltage-assisted calorimetric ionization detection in the SuperCDMS experiment”, *Phys. Rev. Lett.*, vol. 112, p. 041302, Jan 2014. doi: 10.1103/PhysRevLett.112.041302.

- [88] R. Agnese, *et al.*, “Projected sensitivity of the SuperCDMS SNOLAB experiment”, *Phys. Rev. D*, vol. 95, p. 082002, Apr 2017. doi: 10.1103/PhysRevD.95.082002.
- [89] B. S. Neganov and V. N. Trofimov, “Calorimetric method measuring ionizing radiation”, *Otkryt. Izobret.*, vol. 146, p. 215, 1985.
- [90] P. N. Luke, “Voltage-assisted calorimetric ionization detector”, *J. Appl. Phys.*, vol. 64, no. 12, p. 6858, 1988. doi: 10.1063/1.341976.
- [91] Catalin Vitelaru, Daniel Lundin, Nils Brenning, and Tiberiu Minea, “Plasma reactivity in high-power impulse magnetron sputtering through oxygen kinetics”, *Applied Physics Letters*, vol. 103, Sep 2013. doi: 10.1063/1.4819835.
- [92] T. Saab, R.M. Clarke, B. Cabrera, R.A. Abusaidi, and R. Gaitskell, “Design of QET phonon sensors for the CDMS ZIP detectors”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 444, no. 1, pp. 300–303, 2000. doi: 10.1016/S0168-9002(99)01401-1.
- [93] K. D. Irwin, “An application of electrothermal feedback for high resolution cryogenic particle detection”, *Applied Physics Letters*, vol. 66, pp. 1998–2000, Apr 1995. doi: 10.1063/1.113674.
- [94] K.D. Irwin and G.C. Hilton, *Transition-Edge Sensors*, pp. 63–150. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. doi: 10.1007/10933596_3.
- [95] Mark David Pepin, *Low-Mass Dark Matter Search Results and Radiogenic Backgrounds for the Cryogenic Dark Matter Search*. Phd thesis, University of Minnesota, 2016. Available at <https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/pepin.pdf>.
- [96] B. D. Josephson, “The discovery of tunnelling supercurrents”, *Rev. Mod. Phys.*, vol. 46, pp. 251–254, Apr 1974. doi: 10.1103/RevModPhys.46.251.
- [97] R. C. Jaklevic, John Lambe, A. H. Silver, and J. E. Mercereau, “Quantum interference effects in josephson tunneling”, *Phys. Rev. Lett.*, vol. 12, pp. 159–160, Feb 1964. doi: 10.1103/PhysRevLett.12.159.
- [98] M. J. Wilson, *et al.*, “Improved modeling of detector response effects in phonon-based crystal detectors used for dark matter searches”, *Phys. Rev. D*, vol. 109, p. 112018, Jun 2024. doi: 10.1103/PhysRevD.109.112018.
- [99] U. Fano, “On the theory of ionization yield of radiations in different substances”, *Phys. Rev.*, vol. 70, pp. 44–52, Jul 1946. doi: 10.1103/PhysRev.70.44.
- [100] Jens Lindhard, Morten Scharff, and H. E. Schiøtt, “Range concepts and heavy ion ranges (notes on atomic collisions, II)”, 1963. url: <https://www.osti.gov/biblio/4153115>.

- [101] J Lindhard, V Nielsen, M Scharff, and P V Thomsen, “Integral equations governing radiation effects. (notes on atomic collisions, III)”, *Kgl. Danske Videnskab., Selskab. Mat. Fys. Medd.*, vol. Vol: 33: No. 10, 1 1963. url: <https://www.osti.gov/biblio/4701226>.
- [102] J Lindhard, V Nielsen, and M Scharff, “Approximation method in classical scattering by screened coulomb fields.”, *Kgl. Dan. Vidensk. Selsk., Mat.-Fys. Medd.*, 36: 31p(1968)., 1 1968. url: <https://www.osti.gov/biblio/4503382>.
- [103] Joel Sander, *Results from the Crogenic Dark Matter Search Using a Chi Squared Analysis*. Phd thesis, University of California at Santa Barbara, 2007. Available at <https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/sander.pdf>.
- [104] D’Ann Rebekah Barker, *SuperCDMS Background Models for Low-Mass Dark Matter Searches*. Phd thesis, University of Minnesota, 2018. Available at <https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/barker.pdf>.
- [105] F.G. Kondev, M. Wang, W.J. Huang, S. Naimi, and G. Audi, “The NUBASE2020 evaluation of nuclear physics properties *”, *Chinese Physics C*, vol. 45, p. 030001, Mar 2021. doi: 10.1088/1674-1137/abddae.
- [106] U. Schötzg, K. Debertin, and K.F. Walz, “Standardization and decay data of ^{133}Ba ”, *The International Journal of Applied Radiation and Isotopes*, vol. 28, no. 5, pp. 503–507, 1977. doi: 10.1016/0020-708X(77)90185-5.
- [107] M. Divadeenam and J.R. Stehn, “A least-squares fit of thermal data for fissile nuclei”, *Annals of Nuclear Energy*, vol. 11, no. 8, pp. 375–404, 1984. doi: 10.1016/0306-4549(84)90002-1.
- [108] SuperCDMS Collaboration, *et al.*, “A strategy for low-mass dark matter searches with cryogenic detectors in the SuperCDMS SNOLAB facility”, 2023. arXiv:2203.08463.
- [109] SuperCDMS collaboration, “R3 analysis summary internal documentation.” Available at https://scdms.slac.stanford.edu/cdms_restricted/dokuwiki/doku.php?id=analysis:r133:cdmslite:run3_summary.
- [110] William Alexander Page, *Searching for Low-Mass Dark Matter with SuperCDMS Soudan Detectors*. Phd thesis, University of British Columbia, 2019. Available at <https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/bpage.pdf>.
- [111] L. A. Zadeh and J. R. Ragazzini, “Optimum filters for the detection of signals in noise”, *Proceedings of the IRE*, vol. 40, no. 10, pp. 1223–1231, 1952. doi: 10.1109/JRPROC.1952.274117.

- [112] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*. Information and system sciences series, Prentice-Hall, 1979. url: <https://ieeexplore.ieee.org/document/4308806>.
- [113] E. Gatti and P. F. Manfredi, “Processing the Signals From Solid State Detectors in Elementary Particle Physics”, *Riv. Nuovo Cim.*, vol. 9N1, pp. 1–146, 1986. doi: 10.1007/BF02822156.
- [114] J Lindhard, M Scharff, and H E Schioett, “Range concepts and heavy ion ranges (notes on atomic collisions, ii)”, *Kgl. Danske Videnskab. Selskab. Mat. Fys. Medd.*, vol. Vol:33:No14, 1 1963. url: <https://www.osti.gov/biblio/4153115>.
- [115] D. Barker and D.-M. Mei, “Germanium detector response to nuclear recoils in searching for dark matter”, *Astroparticle Physics*, vol. 38, pp. 1–6, 2012. doi: 10.1016/j.astropartphys.2012.08.006.
- [116] A.K. Soma, *et al.*, “Characterization and performance of germanium detectors with sub-keV sensitivities for neutrino and dark matter experiments”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 836, pp. 67–82, 2016. doi: 10.1016/j.nima.2016.08.044.
- [117] U. Fano, “Ionization yield of radiations. II. the fluctuations of the number of ions”, *Phys. Rev.*, vol. 72, pp. 26–29, Jul 1947. doi: 10.1103/PhysRev.72.26.
- [118] S.O.W. Antman, D.A. Landis, and R.H. Pehl, “Measurements of the fano factor and the energy per hole-electron pair in germanium”, *Nuclear Instruments and Methods*, vol. 40, no. 2, pp. 272–276, 1966. doi: 10.1016/0029-554X(66)90386-7.
- [119] Hans R. Bilger, “Fano factor in germanium at 77°k”, *Phys. Rev.*, vol. 163, pp. 238–253, Nov 1967. doi: 10.1103/PhysRev.163.238.
- [120] SuperCDMS collaboration, “LFN cut internal documentation.” Available at http://scdms.slac.stanford.edu/cdms_restricted/Soudan/R133/ebook/171030_BP/R3LFDiscrimMultiVar.html.
- [121] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen, *Classification and Regression Trees*. Taylor & Francis, 1984. doi: 10.1201/9781315139470.
- [122] Y. Freund, “Boosting a weak learning algorithm by majority”, *Information and Computation*, vol. 121, no. 2, pp. 256–285, 1995. doi: 10.1006/inco.1995.1136.
- [123] Ryan Everett Underwood, *Dark Matter Searches with Germanium and Silicon Detectors in SuperCDMS and CUTE*. Phd thesis, Queen’s University, 2022. Available at <https://www.slac.stanford.edu/exp/cdms/ScienceResults/Theses/underwood.pdf>.

- [124] SuperCDMS collaboration, “Radial cut internal documentation.” Available at https://scdms.slac.stanford.edu/cdms_restricted/Soudan/R133/ebook/180521_BP/R3RadialCutEfficiencyProp.html.
- [125] M.H. Kelsey, *et al.*, “G4CMP: Condensed matter physics simulation using the geant4 toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1055, p. 168473, Oct 2023. doi: 10.1016/j.nima.2023.168473.
- [126] E. Schönfeld, “Calculation of fractional electron capture probabilities”, *Applied Radiation and Isotopes*, vol. 49, no. 9, pp. 1353–1357, 1998. doi: 10.1016/S0969-8043(97)10073-2.
- [127] F. Pedregosa, *et al.*, “Scikit-learn: Machine learning in Python”, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [128] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 ed., 2009. doi: 10.1007/978-0-387-84858-7.
- [129] S. B. Kotsiantis, “Decision trees: a recent overview”, *Artif. Intell. Rev.*, vol. 39, p. 261–283, Apr. 2013. doi: 10.1007/s10462-011-9272-4.
- [130] Aryan Jadon, Avinash Patil, and Shruti Jadon, “A comprehensive survey of regression based loss functions for time series forecasting”, 2022. doi: 10.1007/978-981-97-3245-6_9.
- [131] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. doi: 10.1109/5.726791.
- [132] Sepp Hochreiter and Jürgen Schmidhuber, “Long Short-Term Memory”, *Neural computation*, vol. 9, pp. 1735–80, Dec 1997. doi: 10.1162/neco.1997.9.8.1735.
- [133] Martín Abadi, *et al.*, “TensorFlow: Large-scale machine learning on heterogeneous systems”, 2015. arXiv:1603.04467.
- [134] Salman Khan, Hossein Rahmani, Syed Shah, and Mohammed Bennamoun, “A guide to convolutional neural networks for computer vision”, *Synthesis Lectures on Computer Vision*, vol. 8, pp. 1–207, 02 2018. doi: 10.2200/S00822ED1V01Y201712COV015.
- [135] David Rumelhart, Geoffrey Hinton, and Ronald Williams, *Learning Representations by Back-Propagating Errors*, pp. 213–222. Sep 2002. doi: 10.7551/mitpress/1888.003.0013.

- [136] Florentin Bieder, Robin Sandkuehler, and Philippe Cattin, “Comparison of methods generalizing max- and average-pooling”, Mar 2021. [arXiv:2103.01746](#).
- [137] Min Lin, Qiang Chen, and Shuicheng Yan, “Network in network”, Dec 2013. [arXiv:1312.4400](#).
- [138] Robert Tibshirani, “Regression shrinkage selection via the LASSO”, *Journal of the Royal Statistical Society Series B*, vol. 73, pp. 273–282, Jun 2011. doi: 10.2307/41262671.
- [139] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh, “L2 regularization for learning kernels.”, pp. 109–116, Jan 2009. url: <https://dl.acm.org/doi/10.5555/1795114.1795128>.
- [140] Christopher Bishop, *Pattern Recognition and Machine Learning*, vol. 16, pp. 140–155. Jan 2006. doi: 10.1117/1.2819119.
- [141] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba, “Learning deep features for discriminative localization”, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929, 2016. doi: 10.1109/CVPR.2016.319.
- [142] Vinod Nair and Geoffrey Hinton, “Rectified linear units improve restricted boltzmann machines vinod nair”, vol. 27, pp. 807–814, Jun 2010. url: <https://dl.acm.org/doi/10.5555/3104322.3104425>.
- [143] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Adaptive computation and machine learning, MIT Press, 2016.
- [144] Diederik P. Kingma and Jimmy Ba, “Adam: A method for stochastic optimization”, 2017. [arXiv:1412.6980](#).
- [145] James Bergstra and Y. Bengio, “Random search for hyper-parameter optimization”, *The Journal of Machine Learning Research*, vol. 13, pp. 281–305, Mar 2012. url: <https://jmlr.org/papers/v13/bergstra12a.html>.
- [146] Jasper Snoek, Hugo Larochelle, and Ryan Adams, “Practical bayesian optimization of machine learning algorithms”, *Advances in Neural Information Processing Systems*, vol. 4, Jun 2012. [arXiv:1206.2944](#).
- [147] Sergey Ioffe and Christian Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, 2015. url: <https://arxiv.org/abs/1502.03167>.
- [148] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting”, *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. url: <http://jmlr.org/papers/v15/srivastava14a.html>.

- [149] S. Yellin, “Finding an upper limit in the presence of an unknown background”, *Physical Review D*, vol. 66, Aug 2002. doi: 10.1103/physrevd.66.032005.
- [150] F. James, “MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1”, 1994. url: <https://cds.cern.ch/record/2296388>.
- [151] Bradley Efron and Gail Gong, “A leisurely look at the bootstrap, the jackknife, and cross-validation”, *The American Statistician*, vol. 37, no. 1, pp. 36–48, 1983. doi: 10.1080/00031305.1983.10483087.
- [152] M. Stone, “Cross-validated choice and assessment of statistical predictions”, *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, pp. 111–133, Dec. 2018. doi: 10.1111/j.2517-6161.1974.tb00994.x.
- [153] R. Agnese, *et al.*, “Improved wimp-search reach of the CDMS II germanium data”, *Phys. Rev. D*, vol. 92, p. 072003, Oct 2015. doi: 10.1103/PhysRevD.92.072003.
- [154] François Chollet *et al.*, “Keras.” <https://keras.io>, 2015.
- [155] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. doi: 10.1109/5.726791.
- [156] Evalyn I. Gates, Geza Gyuk, and Michael S. Turner, “The local halo density”, *The Astrophysical Journal*, vol. 449, Aug 1995. doi: 10.1086/309652.
- [157] Andrew R. Liddle, *An introduction to modern cosmology*. 1998. ISBN: 978-1-118-50209-9.

Acronyms and Abbreviations

ADC	Analog-to-Digital Converter	32
ALP	Axion-Like Particle	7, 10
AUC	Area Under the Curve	71
CAM	Class Activation Map	72
CCD	Charge-Coupled Device	12
CDMS	Cryogenic Dark Matter Search	14
CDMSlite	Cryogenic Dark Matter Search Low Ionization Threshold Experiment ..	30
CMB	Cosmic Microwave Background	3
CNN	Convolutional Neural Network	53, 67, 129
CUTE	Cryogenic Underground Test	29
DAQ	Data Acquisition	25, 32
DM	Dark Matter	1
ER	Electron Recoil	23, 38
FN	False Negatives	85
FPR	False Positive Rate	71
GAP	Global Average Pooling	68
HDPE	High-Density Polyethylene	27
HV	High Voltage	19
iZIP	Inter-Leaved Z-Sensitive Ionization and Phonon	17
LDM	Light Dark Matter	9
LFN	Low-Frequency Noise	12, 26, 32, 44, 53, 101, 129

LSTM Long Short-Term Memory	53, 69, 129
MOND Modified Newtonian Dynamics	7
MSE Mean Square Error	65
NR Nuclear Recoils	23, 38
NSOF Non-Stationary Optimal Filter.....	35
NTL Neganov-Trofimov-Luke	16
OF Optimal Filter	34
OI Optimum Interval.....	102
PBHs Primordial Black Holes	10
PLR Profile Likelihood Ratio.....	101
PSD Power Spectral Density.....	34
PT Phonon Total.....	81
QET Quasiparticle-Trap-Assisted Electrothermal-Feedback Transition Edge Sensor	19
RNTLs Reduced NTL-Gain Events.....	48
ROC Receiver Operating Characteristic.....	71
RQ Reduced Quantity	37
SD Spin-Dependant	8
SI Spin-Independent	8
SQUID Superconducting Quantum Interference Device	21
SUPR Simulation Using Pulse Shapes and Regression	60
TES Transition Edge Sensor	19
TPR True Positive Rate	71
WIMP Weakly Interacting Massive Particle	7