

**ADVANCES IN GERMANIUM DETECTOR SIGNAL MODELING WITH GPU  
COMPUTING AND DEEP LEARNING TECHNIQUES**

Kevin Hargovindbhai Bhimani

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in  
partial fulfillment of the requirements for the degree of Doctor of Philosophy in the  
Department of Physics and Astronomy in the College of Arts and Sciences.

Chapel Hill  
2025

Approved by:  
Julieta Gruszko  
Amy Nicholson  
David Radford  
Fabian Heitsch  
John Wilkerson

©2024  
Kevin Hargovindbhai Bhimani  
ALL RIGHTS RESERVED

## ABSTRACT

Kevin Hargovindbhai Bhimani: Advances in Germanium Detector Signal Modeling with GPU Computing and Deep Learning Techniques  
(Under the direction of Julieta Gruszko)

The Large Enriched Germanium Experiment for Neutrinoless Double-Beta Decay (LEGEND) collaboration is searching for neutrinoless double-beta ( $0\nu\beta\beta$ ) decay in  $^{76}\text{Ge}$  using modular strings of enriched germanium detectors. Pulse shape simulations (PSS) are essential for modeling the analysis techniques used to distinguish signal events from multi-site and surface event backgrounds, and for generating reliable simulations of energy spectra. Together, these capabilities are critical for identifying the sources of background in current experiments and projecting background levels in future searches. Two longstanding obstacles have hindered reliable PSS: incomplete models of charge collection from surface event interactions and the complexity of replicating the detector's electronic response. This dissertation addresses both issues using advanced computational techniques.

EH-Drift<sup>1</sup> is a novel simulation of charge carriers in germanium that incorporates surface drift, diffusion, and self-repulsion to model surface events. EH-Drift accurately simulates the passivated surface events while reproducing the experimental results from Germanium detector test stands. Using parallel computing on GPUs EH-Drift achieves 62.5x run-time improvement over CPU implementations, making their use for L-200 background modeling computationally feasible.

Cyclic Positional U-Net (CPU-Net)<sup>2</sup> is a neural network architecture that performs translations of simulated pulses so that they closely resemble measured detector signals. Using a CycleGAN framework, this Ad-hoc Translation Network learns a data-driven mapping between simulated and measured pulses with high fidelity and computational efficiency. Using data from an HPGe detector, we show that CPU-Net effectively captures and reproduces critical pulse shape features, allowing more realistic simulations without detector-specific tuning.

---

<sup>1</sup><https://github.com/kevinhbhimani/EH-Drift>

<sup>2</sup><https://github.com/aobol/CPU-Net>

To my loving parents, Hargovindbhai and Jyotsnaben, whose sacrifices and belief in me made this journey possible, and to my sister Niyati, who supported me every step of the way.

## ACKNOWLEDGMENTS

This work would not have been possible without my advisor, Julieta Gruszko, to whom I owe my deepest gratitude. Thank you for your patience, encouragement, and believing in my potential when I did not.

I am grateful to David Radford for trusting me with the EH-Drift project. Your foundational work provided me with a great opportunity to learn.

Many thanks to Aobo Li for being a kind mentor and giving me the CPU-Net project. Your knowledge of machine learning has shaped my learning in the field.

I extend my thanks to the entire ENAP group at UNC, particularly John Wilkerson and Reyco Henning, who cultivated a supportive and learning environment. I would like to recognize the LEGEND collaboration for being an excellent and fun group to collaborate with. Thank you to Tom Caldwell for teaching me the fundamentals of GPU computing. I also want to thank Alex Leder for the continuous collaboration and mentoring on the EH-Drift project.

Personally, I want to thank my best friend and colleague Esteban Leon, who was a vital part of my entire journey. I am also grateful to my friends in Chapel Hill, Waleed, Yuma, Quique, David, Vale M, Vale R, Vera, Morgan, and Guido, whose company and support made my six years in Chapel Hill a wonderful experience.

I also would like to thank my partner Leah. Thank you for always being there to support me and help me grow every step of the way.

I want to thank my parents Hargovindbhai and Jyotsnaben. I am here only because they took a chance on me, using their life savings to send me to the United States ten years ago. My thanks also go to my sister, Niyati, for her constant encouragement.

UNC Longleaf Cluster and the NERSC computing resources were instrumental to this work. Finally, I also want to acknowledge the funding agencies that made this research possible. This work is supported by the U.S. DOE and the NSF, the LANL, ORNL and LBNL LDRD programs; the European ERC and Horizon programs; the German DFG, BMBF, and MPG; the Italian INFN; the Polish NCN and

MNiSW; the Czech MEYS; the Slovak SRDA; the Swiss SNF; the UK STFC; the Canadian NSERC and CFI; the LNGS, SNOLAB, and SURF facilities; and the DE-SC0022339, DE-FG02-97ER41041, DE-FG02-97ER41033, and PHY-1812374 grants.

## TABLE OF CONTENTS

<b>LIST OF FIGURES .....</b>	<b>xi</b>
<b>LIST OF TABLES .....</b>	<b>xix</b>
<b>LIST OF ABBREVIATIONS AND SYMBOLS .....</b>	<b>xx</b>
<b>CHAPTER 1: Neutrinoless Double Beta Decay .....</b>	<b>1</b>
Section 1.1: Introduction.....	1
Section 1.2: Neutrino .....	1
Section 1.3: Weak interactions .....	2
Section 1.4: Neutrino Mass .....	4
1.4.1: Dirac Mass Mechanism .....	4
1.4.2: Majorana Mass Mechanism .....	4
1.4.3: See-Saw Mechanism .....	4
1.4.4: Measuring neutrino mass .....	5
Section 1.5: Double Beta Decay .....	7
Section 1.6: Neutrinoless Double-Beta Decay .....	9
Section 1.7: Detecting Neutrinoless Double Beta Decay .....	11
<b>CHAPTER 2: High Purity Germanium Detectors .....</b>	<b>15</b>
Section 2.1: Introduction.....	15
Section 2.2: Signal formation in Germanium .....	15
2.2.1: Germanium properties .....	15
2.2.2: P-type Point Contact Germanium Detectors.....	16
2.2.3: Shockley-Ramo Theorem .....	18

Section 2.3: The MAJORANA DEMONSTRATOR .....	19
Section 2.4: The GERmanium Detector Array.....	20
Section 2.5: The LEGEND experiment .....	21
2.5.1: LEGEND-200.....	21
2.5.2: LEGEND-1000 .....	23
Section 2.6: LEGEND-200 background model .....	24
<b>CHAPTER 3: Surface Event Modeling and Pulse Shape Simulations .....</b>	<b>29</b>
Section 3.1: Introduction.....	29
Section 3.2: Properties of Alpha Interactions .....	29
Section 3.3: Alpha Background Observations .....	31
Section 3.4: Challenges in Modeling Surface Events .....	31
Section 3.5: Surface Charge Effect .....	33
Section 3.6: Current Status in Pulse Shape Simulations .....	35
Section 3.7: EH-Drift .....	36
3.7.1: Charge Transport Equation .....	36
3.7.2: Initial Setup.....	37
3.7.3: Diffusion .....	39
3.7.4: Drift in Electric Field .....	42
3.7.5: Geometric Factors .....	45
3.7.6: Surface Drift .....	45
3.7.7: Courant Number and Adaptive Time Step .....	46
3.7.8: Impurity Correction .....	47
3.7.9: Signal Calculation.....	48
3.7.10: Density Snapshots .....	49
3.7.11: Model Parameters .....	49
3.7.12: Input and Output .....	52

<b>CHAPTER 4: GPU Acceleration .....</b>	<b>55</b>
Section 4.1: Introduction.....	55
Section 4.2: Iterative Relaxation Algorithms for Solving 2-D Poisson Equation .....	56
Section 4.3: GPU Programming with CUDA C++ .....	59
Section 4.4: GPU Performance Comparisons and Benchmarking .....	63
<b>CHAPTER 5: Surface Simulations Results .....</b>	<b>68</b>
Section 5.1: Introduction.....	68
Section 5.2: Post-Processing and Activeness Maps .....	68
Section 5.3: Reproduction of Results from Established Test Stands.....	74
Section 5.4: Efficiency Estimations in the Region of Interest .....	76
Section 5.5: Generating Background Spectra .....	77
<b>CHAPTER 6: Neural Net Based Electronics Response Modeling .....</b>	<b>81</b>
Section 6.1: Introduction.....	81
Section 6.2: LEGEND-200 Readout Electronics.....	81
Section 6.3: Electronics Influence on Waveforms .....	83
Section 6.4: Challenges in Modeling Electronics Response .....	84
Section 6.5: Deep Learning Approach to Electronics Modeling .....	86
6.5.1: Mathematical Formalism .....	86
6.5.2: Positional U-Net .....	87
6.5.3: RNN with Attention .....	90
<b>CHAPTER 7: Data Preparation &amp; Model Training .....</b>	<b>92</b>
Section 7.1: Introduction.....	92
Section 7.2: Data selection .....	92
Section 7.3: Simulations .....	93
Section 7.4: Post Processing .....	95
Section 7.5: Network Training .....	96
Section 7.6: Loss Functions .....	99

Section 7.7: Hyperparameter Tuning .....	101
<b>CHAPTER 8: CPU-Net Results .....</b>	<b>103</b>
Section 8.1: Introduction .....	103
Section 8.2: Training progression .....	103
Section 8.3: Waveform Translation .....	103
Section 8.4: Validation of Key Waveform Parameters .....	104
8.4.1: Drift Time Distribution .....	105
8.4.2: Current Amplitude .....	107
8.4.3: Tail Slope .....	111
<b>CHAPTER 9: Conclusion .....</b>	<b>115</b>
Section 9.1: Summary of Results .....	115
Section 9.2: Future Directions .....	116
9.2.1: Validation of Results .....	116
9.2.2: Improving EH-Drift Activeness Maps .....	117
9.2.3: Future for Pulse Shape Simulations .....	117
9.2.4: Challenges in CycleGAN based Model .....	118
9.2.5: CPU-Net Model Improvements .....	118
Section 9.3: Implications for the LEGEND Physics Program .....	119
9.3.1: Passivated Surface Background Modeling for LEGEND .....	119
9.3.2: Beyond Standard Model Searches .....	119
9.3.3: LegendGeSim and CPU-Net Integration .....	120
<b>BIBLIOGRAPHY .....</b>	<b>121</b>

## LIST OF FIGURES

1.1 Comparison of the masses of fermions in the Standard Model. Neutrino masses are six orders of magnitude lower than other fermions. Neutrino masses were estimated assuming the normal mass hierarchy and a loose upper bound [7].	2
1.2 Feynman diagrams illustrating the two types of beta decay. Left: Beta-minus decay, where a neutron decays into a proton, emitting an electron and an electron antineutrino via a virtual $W^-$ boson. Right: Beta-plus decay, where a proton converts into a neutron, emitting a positron and an electron neutrino via a virtual $W^+$ boson.	3
1.3 Two possible neutrino mass hierarchies. In the normal hierarchy (left), the lightest mass state is $m_1$ , followed by $m_2$ (separated by the solar mass-squared difference $\Delta m_{21}^2$ ), with $m_3$ being the heaviest, separated by the atmospheric mass-squared difference $\Delta m_{32}^2$ . In the inverted hierarchy (right), $m_3$ is the lightest, while $m_1$ and $m_2$ are heavier. Figure from [7].	7
1.4 Feynman Diagram for a double beta decay ( $2\nu\beta\beta$ ) process.	8
1.5 Ground state mass parabola for nuclei showing the necessary conditions for double-beta decay. The plot shows nuclear charge $Z$ versus nuclear mass $M(A, Z)$ . Double-beta decay is energetically allowed when the initial even-even nucleus has a higher mass than the final nucleus, and single beta decay is forbidden. Only even-even nuclei can undergo double-beta decay due to their greater stability. Figure from [18].	9
1.6 Feynman Diagram for light neutrino exchange neutrinoless double beta decay ( $0\nu\beta\beta$ ) process.	10
1.7 Parameter space for $m_{\beta\beta}$ and $m_l$ for the normal (right) and inverted (left) mass ordering. The allowed parameter space assuming $3\sigma$ intervals of the neutrino oscillation observables from NuFIT is shown by the solid lines [21]. The marginalized posterior distributions for $m_{\beta\beta}$ and $m_l$ is shown by the color bar. This was obtained by combining equation 1.9 with neutrino oscillation measurements. Plot assumes QRPA NMEs and the absence of mechanisms that drive $m_{\beta\beta}$ and $m_l$ to zero. Source: [22]	11
1.8 Simulated $0\nu\beta\beta$ signal in a $^{76}\text{Ge}$ detector. The $0\nu\beta\beta$ signal is shown with an assumed half life of $10^{25}$ [23] years.	12
1.9 The exposure needed for $3\sigma$ confidence level discovery of neutrinoless double beta decay for different half life limit, given the inverted hierarchy. Various lines represent scenarios with different background levels in ROI. Red line indicates the LEGEND-1000 background goal. The blue band shows the bottom of the inverted ordering.	13

2.1	A plot showing a timeline of past and future HPGe $0\nu\beta\beta$ decay experiments. The half-lives probed by them is shown along with their background index. Next generation experiments aim to continue improve half life limit while having the lowest background index. Picture credit: Luigi Pertoldi.	16
2.2	Left: Simulated weighting potential inside a LEGEND ICPC style point contact detector. Right: Collected signal from a detector showing sharp step shaped waveform.	19
2.3	The MAJORANA Demonstrator setup (left) and its Ge detector arrays (right). Source: MAJORANA Collaboration.	20
2.4	GERDA setup (left) with Ge detectors (right) arranged in vertical strings. Source: GERDA Collaboration.	21
2.5	Left: LEGEND-200 Ge detectors strings surrounded by wavelength-shifting fibers for LAr veto. Right: The L-200 design. The detector array is mounted in the center of the LAr cryostat. Muon veto is achieved by placing the cryostat in a water tank and detecting Cherenkov radiation with photo-multipliers. Source: LEGEND Collaboration.	22
2.6	The three point contact detector geometries used in LEGEND: MAJORANA DEMONSTRATOR PPC detectors (left), GERDA BEGe detectors (middle), and newly developed ICPC detectors (right).The color bar depicts weighting potential within each detector's cross section.The thick black and gray lines are the $p^+$ and $n^+$ electrode, respectively. The yellow points are locations of simulated energy depositions. The holes travel along the white lines to the $p^+$ electrode, while electrons travel along white line to the $n^+$ electrode. Figure from [35].	23
2.7	LEGEND-200 Projected background contributions near $Q_{\beta\beta}$ after all analysis cuts. Grey bands indicate uncertainties in assays and background rejection.	24
2.8	Signal (black) and current (blue) for four different types of events in LEGEND-200. The point contact geometry enables long and unique drift times that allow for pulse shape discrimination. The maximum of the current signal can be used to reject the three background events and accept the $2\nu\beta\beta$ events. Credit: The LEGEND Collaboration.	25
2.9	LEGEND-200 background contributions near $Q_{\beta\beta}$ after all analysis cuts. Credit: The LEGEND Collaboration.	26
3.1	The specific energy loss in a material for an alpha particle with several MeV initial energy. Plots are shown for a single alpha particle track and for the average behavior of a parallel beam of alpha particles. Alphas lose most of their energies in a small region. Picture from [31]	30
3.2	An alpha waveform (red) compared to a bulk waveform (blue) in the MJD PPC detector. The DCR effect results in a slow charge collection after the fast rise of the signal, as shown by the insert plot in the box. [40]	32

3.3	LEGEND-200 physics spectrum. The events above 3000 keV after the muon and multiplicity cut are primarily alphas. These events can be effectively removed using PSD cuts as shown in green. Credit: LEGEND collaboration .....	32
3.4	Electric field magnitude and field lines in a region near the passivated surface of the PONaMa-1 PPC detector with different surface charges. The white arrows shows the direction of electric field lines. Geometry information for the detector can be found in table 5.1. Charges on the passivated surface can pull electrons and holes to the surface and cause energy loss and slow charge collection. Top figure has no surface charge and the field lines are parallel to surface. Middle figure has a surface charge of $-0.5 \times 10^{-10}$ e/cm <sup>2</sup> which pulls the filed lines to the surface. Bottom figure has surface charge of $0.5 \times 10^{-10}$ e/cm <sup>2</sup> that repels the field lines form the surface. ....	34
3.5	Effect of surface charge on depletion voltage for LEGEND PPC detector P00661C. Depletion voltages calculated using <code>siggen</code> software. The nominal depletion voltage is 1694 V at zero surface charge. Surface charge can alter the depletion voltage and potentially cause a detector to become undepleted after being deployed. ....	35
3.6	Diagram showing a simple setup and coordinates in EH-Drift for a PPC detector. The n <sup>+</sup> electrode is kept at high voltage while p <sup>+</sup> is kept at zero. Passivated region is the region in between the two electrodes.....	38
3.7	A flow chat illustration the EH-Drift steps. The initial steps to calculate intrinsic parameters of the detector (e.g. weighting potential, capacitance, etc.) are performed once. Then subsequent loops step through each time step and each grid point to track the movement of charge clouds. ....	40
3.8	Germanium crystal unit cell, showing the crystal axis and the crystallographic basis axes. The atoms on the corners, faces and inside the cube are shown in black, orange and blue circles respectively. ⟨001⟩, ⟨010⟩, and ⟨100⟩ are the axes of Ge crystal unit cell. And ⟨100⟩, ⟨110⟩, and ⟨111⟩ are the crystallographic basis axes. The crystallographic basis axes are the direction used to describe physical properties such as mobility and energy band structure. Source: [46,47] .....	43
3.9	Relation of drift velocity versus electric field in EH-Drift. The points shows the experimental values adapted from [49]. To show how EH-Drift estimates the values we create 500 samples between 0 and 5000 V/cm and performed the piecewise linear interpolation between the points. Only < 100 > direction velocities are used in the current model. ....	44
3.10	Maximum values of terms in the charge transport equation 3.3 at a given time step for an event in EH-Drift. The drift terms are about two orders of magnitude large than the diffusion terms throughout the time evolution of the charge cloud. This justifies the use of drift velocity to define the timestep used for Courant number calculation.....	47

3.11	Drift of electron and hole charge clouds in EH-Drift with zero surface charge. The projected densities show how the charges are distributed along the radius. The densities have two peaks, one due to the fast moving component in bulk and another due to the slow moving component on the passivated surface. ....	50
3.12	Drift of electron and hole charge clouds in EH-Drift with negative surface charge. The negative surface charge pulls the holes onto the surface, which then move at a slower speed. ....	51
3.13	Drift of electron and hole charge clouds in EH-Drift with positive surface charge. The positive surface charge pulls the electrons onto the surface which move at a slower speed. ....	51
3.14	A comparison of waveforms generated by EH-Drift for a LEGEND-200 PPC detector as surface charge ( $\sigma$ ) and relative surface drift velocity (SD) are varied. The events were at $r=15$ mm $z=0.02$ mm. ....	53
4.1	Run time for a single EH-Drift event on CPU. Run time based on a single compute server in the Longleaf cluster at the University of North Carolina at Chapel Hill. ....	56
4.2	Updating grid points in the SOR method. Red grid points represent all the points that have already been updated. A given grid point is updated using a mixture of old and new values. ....	58
4.3	Updating grid points in Red-Black SOR. All the red grid points are updated in parallel using old values, and the black grid points are updated using the new red grid points. ....	59
4.4	A solution to the Electric Potential using Red-Black SOR in EH-Drift. The white lines show electric field lines. ....	60
4.5	A depiction of the GPU architecture. Source: NVIDIA CUDA C++ Programming Guide .....	61
4.6	Comparison of waveform generated by CPU program versus GPU program in EH-Drift at the default converge threshold. The event was at $r=15$ mm and $z=0.10$ mm. ....	64
4.7	Difference in waveforms produced by the CPU and GPU programs. Left plots shows the difference at default converge threshold for RB-SOR algorithm. Right plot shows the difference at numerical convergence which is set to $8 \times 10^{-8}$ ....	64
4.8	Comparison of run time for an EH-Drift event on GPU and CPU. Run times are based on a single CPU node on the Longleaf cluster at UNC-Chapel Hill and an NVIDIA A100 node of the Perlmutter supercomputer at the NERSC. ....	65
4.9	Program run time on various GPU models for a $20\mu$ grid. The faster GPU results in lower run time. ....	66

5.1	A simulated waveform and the resulting trapezoidal filtered waveform for a highly energy degraded surface waveform from EH-Drift simulations. The filter had a rise time of $10\mu s$ , flat time of $3\mu s$ , and the energy estimate is based on the value of the filtered waveform $13\mu s$ after the start of the pulse rise.....	69
5.2	Activeness versus initial energy for event at $r=14\text{mm}$ and $z=1.00\text{mm}$ in PONaMa-1 detector. The lower activeness loss for higher energy could be due to self-repulsion pushing charges away from surface thus leading to smaller fraction of total charges on surface.....	71
5.3	Activeness values to map a PPC detector EH-Drift with an initial energy 5000 keV and negative surface charge ( $-0.30 \times 10^{-10} \text{ e/cm}^2$ ). The radius points were uniformly spaced while the height points were densely sampled close to the surface. ....	71
5.4	Interpolated activeness map for a PPC detector using EH-Drift with negative surface charge ( $-0.30 \times 10^{-10} \text{ e/cm}^2$ ). For a larger radius events there is a higher chance of having the hole component pulled onto the surface, causing higher loss of activeness. ....	72
5.5	Interpolated activeness map for a PPC detector using EH-Drift with zero surface charge. Without surface charge there is nearly full collection. Right plot shows the same map with using narrow color scale. ....	72
5.6	Interpolated activeness map for a PPC detector using EH-Drift with positive surface charge ( $0.30 \times 10^{-10} \text{ e/cm}^2$ ). The right plot is shown using narrow color scale. At a smaller radius, there is a higher chance of the electron component being pulled onto the surface causing higher loss of activeness. The activeness loss is smaller than negative surface charge since electrons are minority signal component. ....	73
5.7	Interpolated activeness map for a Mirion ICPC detector using EH-Drift with negative charge ( $-0.30 \times 10^{-10} \text{ e/cm}^2$ ). The right plot is shown in a narrow color scale. The ICPC detector has significant less reduction in activeness than PPC detectors. The activeness loss also does not depend on the surface charge and the same activeness is found for positive and negative surface charge. ....	73
5.8	Fitting the activeness in the TUBE scanner with EH-Drift. A positive surface charge can help explain the increase in collected energy with radius. A 350 keV energy offset and 2.0 mm radial offset is needed to match the data, which could occur if there was a thin fully dead region either in the source or at the detector surface. Experimental values from [40] .....	74
5.9	Fitting the activeness in GALATEA scanner with EH-Drift. A negative surface charge can help explain the decrease in collected energy with radius. Experimental values from [41] .....	75

5.10	Binary activeness for PONaMa-1 detector and Mirion ICPC detector with negative surface charge. A 3 keV energy resolution threshold is defined. If energy is degraded above it, it is labeled zero, else 1. The purple dots show where $0\nu\beta\beta$ events would be not recorded for the detectors. ....	77
5.11	Efficiency in the ROI versus surface charge calculated using the binary activeness. ....	78
5.12	Electric magnitude and field lines near the ditch of a Mirion ICPC detector with negative surface charge. In Mirion ICPC, the flat p <sup>+</sup> point contact, where the field lines end, extend to the start of ditch at r=12. The n <sup>+</sup> contact starts at end of ditch at r=15, shown by the origin of field lines. The field lines attract electron component to the surface which drift slowly. The reduction in activeness is fraction of other detector geometries. ....	78
5.13	Degradation of the alpha-particle energy spectrum under various surface charges. The location of hits of alpha particles were generated using a Geant 4 simulation of planar alpha source of 5 MeV on PONaMa-1 detector. Energy degradation for an alpha event was calculated from interpolated map for each surface charge. The unit of surface charge is in $10^{-10}$ e/cm <sup>2</sup> .....	79
5.14	Degradation of the $^{42}\text{K}$ energy spectrum in PONaMa-1 detector. The detector surface had a surface charge of $-0.5 \times 10^{-10}$ e/cm <sup>2</sup> . The activeness was calculated assuming 2039 keV energy map. This result does not include a model for the dead layer created by the lithiated n <sup>+</sup> contact. ....	80
6.1	A schematic of the LEGEND-200 signal readout chain. [55].....	82
6.2	Pictures of the readout electronics of LEGEND-200. Left shows a close up of the detector with the LMFE, middle shows the detector strings and front end electronics, with the CC4s located at the top of the array. Below them are the connectors for the high voltage cable. Right shows the Flashcam crates in the DAQ rack. Photo: Michael Willers .....	82
6.3	Comparison of a simulated waveform with a real data waveform. Both pulses were randomly sampled from datasets and are aligned for clarity. Three components of the waveforms are labeled. The data pulses contain electronics effects such as RC decay tail, electronics noise and preamplifier integration. ....	84
6.4	Layer-wise breakdown of the Positional U-Net. The blue lines represent the skip connections between contracting and expanding paths of the U-Net. Positional encoding layers are also shown for all levels. Credit: Aobo Li .....	88
6.5	Components of the Conv1d Module in the PU-Net from figure 6.4. Credit: Aobo Li .....	89
6.6	The attention coupled RNN discriminator. Waveforms are embedded into 128 dimension space and then passed to a GRU unit. The outputs of the GRU are used to calculate the attention score. Credit: Aobo Li .....	90

7.1	The simulation geometry of the ORNL characterization setup, showing the detector in green, the source in red, the aluminum holder in dark grey, and the lead shielding in light grey. ....	94
7.2	Calibrated energy spectrum from a $^{228}\text{Th}$ source at ORNL compared to the Monte Carlo simulation results. Key peaks from $^{228}\text{Th}$ source are labeled. ....	94
7.3	Production of energy weighted waveform in simulations for four events. Left figures show the location of hit locations for each event. The magnitude of energy deposited for hits in each event is shown in middle figure. Right plot show the energy-weighted waveforms for the corresponding events. Hits produced using Geant4 simulations and waveforms are generated using siggen software. ....	95
7.4	Input data and simulated waveforms. The waveforms are aligned by 99.9% rise time. Data pulses are normalized by aligning the tail and baselines. ....	97
7.5	Training in CPU–Net. ATN transfers simulated waveforms to data-like waveforms, while the IATN transfers data waveforms to simulation-like waveforms. Discriminator A differentiates data-like waveforms outputs from the ATN from true data. Discriminator B differentiates simulation-like waveforms outputs from IATN from true simulations. ....	98
8.1	Training losses for CPU–Net. Curves are smoothed using a moving average of 10 samples for clarity. The identity and cycle losses rapidly converged to zero while the generator and discriminator losses fluctuate. ....	104
8.2	An example ATN output (blue) generated by the input simulated waveform (red). A detector waveform (dotted green), manually selected from data to resemble the ATN output, is included for reference. The orange region depicts the area where the preamplifier integration effect is most visible. The blue region shows the impact of the RC circuit discharge effect. Grey region shows the baseline region with electronics noise. A residual panel comparing the ATN output to the reference waveform is included below. A 0 residue is not expected, as the reference waveform is not a one-to-one match with the simulated waveform. ....	105
8.3	Cycle consistency in waveform translation in the forward direction. Starting with 10 simulated waveforms, the ATN first translates them into detector-like waveforms, and then IATN translates them back to the simulation like waveforms. ....	106
8.4	Cycle consistency in waveform translation in the backward direction. Beginning with 10 detector waveforms, the IATN translates them into simulated-like waveforms, and then ATN translates back to detector-like waveforms. ....	106
8.5	Calculating the time points of the waveform. $T_{Drift}$ of the waveform is the time between the two points. ....	107
8.6	The distribution of the $T_{Drift}$ on SEP dataset. ....	108

8.7	The distribution of the $T_{Drift}$ on DEP dataset. ....	108
8.8	Steps involved in calculating the maximum current amplitude of the waveform. First the discrete derivative of the waveform is calculated which is then up-sampled by a factor of 16 to extract finer detail. Then the current is smoothed to reduce the noise. The maximum value is $I_{max}$ . The ratio of the $I_{max}$ , to energy is used as a PSD parameter. ....	109
8.9	Distribution of maximum current amplitude ( $I_{max}$ ) on SEP validation datasets.....	110
8.10	Distribution of maximum current amplitude ( $I_{max}$ ) on DEP validation datasets. ....	110
8.11	Scatter plot between $I_{max}$ of simulated waveforms and ATN translated waveforms. ATN shifts the amplitude distribution to align with data while maintaining the relative ordering of individual events. ....	111
8.12	Calculating the tail slope ( $c_{tail}$ ) of the waveform. We take the logarithm of last 300 samples of the waveform and linear fit it to find the slope. ....	112
8.13	Distribution of $c_{tail}$ for simulations, data, and ATN translated waveforms.....	113
8.14	Distribution of tail slope ( $c_{tail}$ ) between data and ATN translated waveforms. ....	113

## LIST OF TABLES

3.1	Input parameters to EH-Drift .....	54
4.2	Profiling results of custom GPU kernels during simulation.....	67
5.1	Key geometry dimensions for the PONaMa-1 detector .....	70
5.2	Estimated $0\nu\beta\beta$ efficiency for three different germanium detectors at different surface charges. ....	76
7.1	Overview of Loss Calculations, Loss Types, and Optimizers in CPU-Net training. ....	100
7.2	Hyperparameters used for CPU-Net training. ....	101

## LIST OF ABBREVIATIONS AND SYMBOLS

$0\nu\beta\beta$	Neutrinoless double beta decay
$2\nu\beta\beta$	Two neutrino double beta decay
AC	Anti-Coincidence
ATN	Ad-hoc Translation Network
BEGe	Broad energy Germanium detector
BI	Background index
COAX	Semi-coaxial detector
CUDA	Compute Unified Device Architecture
CPU	Central Processing Unit
CSA	Charge-Sensitive Amplifier
DAQ	Data acquisition
DEP	Double escape peak
DSP	Digital signal processing
FEP	Full energy peak
FWHM	Full width at half maximum
GERDA	GERmanium Detector Array
GPU	Graphics Processing Unit
HPC	High performance computing
HPGe	High purity Germanium
HV	High voltage
IATN	Inverse Ad-hoc Translation Network
ICPC	Inverted coaxial point contact detector
IO	Inverted ordering of neutrino masses
KL	Kullback-Leibler
L-200	Initial phase of LEGEND with 200 kg of HPGe detectors
L-1000	Subsequent phase of LEGEND with 1,000 kg of HPGe detectors
LEGEND	Large Enriched Germanium Experiment for Neutrinoless $\beta\beta$ Decay
LNGS	Laboratori Nazionali del Gran Sasso

MJD	MAJORANA DEMONSTRATOR
NERSC	National Energy Research Scientific Computing Center
NME	Nuclear Matrix Element
NO	Normal ordering of neutrino masses
ORNL	Oak Ridge National Laboratory
PPC	P-type point contact detector
PSD	Pulse shape discrimination
RB-SOR	Red-Black Successive Over-Relaxation
SOR	Successive Over-Relaxation
SURF	Sanford Underground Research Facility

## CHAPTER 1: NEUTRINOLESS DOUBLE BETA DECAY

### 1.1 Introduction

According to the well-established Standard Model of physics, the Big Bang should have created equal amounts of matter and antimatter; however, the observable universe has a significant predominance of matter over antimatter. This imbalance is one of the most profound mysteries in modern physics. To explain this asymmetry, Andrei Sakharov in 1967 proposed three conditions that a particle must meet to solve the imbalance [1]. These conditions, known as the ‘Sakharov conditions’, are a violation of baryon number conservation, a violation of the C and CP symmetries, and a departure from thermal equilibrium.

The neutrino is a very promising candidate that could meet all three Sakharov conditions and explain matter-antimatter asymmetry using a mechanism called leptogenesis. In this process, a heavy right-handed neutrino decays in a CP-violating way to create more leptons than antileptons. This lepton asymmetry is converted into a baryon asymmetry in sphaleron processes, which violate the baryon number. The expansion of the universe prevents the inverse reactions from balancing the asymmetry, so these decays occur out of thermal equilibrium, meeting the three Sakharov conditions. Neutrinos are thus fundamental to understanding the origin of matter in the universe. In this chapter, we introduce the neutrino and then the weak interaction that governs neutrino behavior. We then describe some mechanisms through which neutrinos could obtain their mass. Finally, we describe the neutrino processes that can help to understand the matter-antimatter asymmetry.

### 1.2 Neutrino

In the early parts of the twentieth century, physicists were surprised to find that the electrons in beta decay seemed to violate momentum conservation. Instead of carrying the full Q-value of the decay, the electron carried a range of energies. To explain this anomaly, Wolfgang Pauli in 1930 postulated a new, light, and electrically neutral particle. Fermi later called it the neutrino because it was electrically neutral and because its rest mass was thought to be very small or zero. It took 25 years to directly

observe the neutrino in the 1956 Cowan–Reines experiment [2]. Since then, neutrinos have continued to surprise physicists. The Homestake experiment in 1960 only detected about a third of the predicted solar neutrinos [3]. This puzzling neutrino deficit suggested that there could be more than one type of neutrino. This led to a review of Pontecorvo’s neutrino oscillation theory, which could explain how neutrinos oscillate between different flavors [4]. Neutrino oscillations were first observed by the Sudbury Neutrino Observatory [5] and Super-Kamiokande [6] in the 1990s and early 2000s.

Through the development of theory and experimental observation, neutrinos have been found in three flavors: electron ( $\nu_e$ ), muon ( $\nu_\mu$ ), and tau ( $\nu_\tau$ ), corresponding to the three generations of charged leptons they are associated with. Each neutrino flavor state is a linear combination of three discrete mass eigenstates, and neutrinos oscillate between the flavors as they travel through space. The experimental limits on the neutrinos’ mass eigenstates are orders of magnitude smaller than those of all other Standard Model fermions, as shown in figure 1.1. To date, it is not known what the absolute masses of neutrinos are and why their masses are so small. Neutrinos are hard to study because they are only theorized to interact via gravity and weak interaction. Most neutrino experiments are based on studying weak neutrino interactions, as the small mass makes gravitational neutrino studies practically impossible. In the next section, we provide a brief review of the weak interaction.

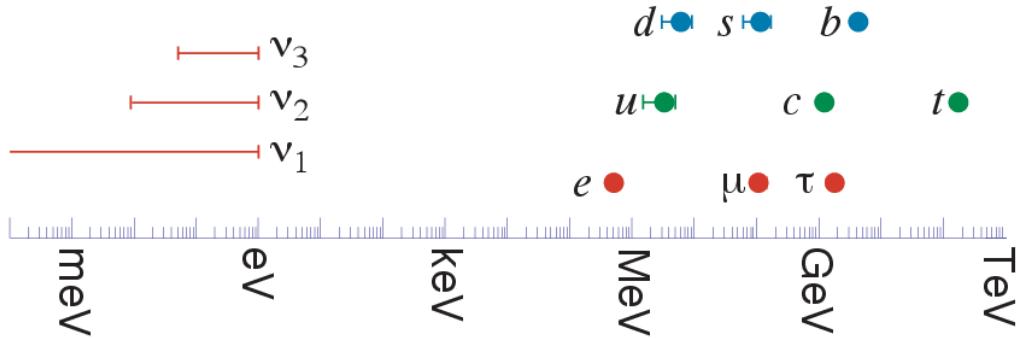


Figure 1.1: Comparison of the masses of fermions in the Standard Model. Neutrino masses are six orders of magnitude lower than other fermions. Neutrino masses were estimated assuming the normal mass hierarchy and a loose upper bound [7].

### 1.3 Weak interactions

The weak interaction is one of the four fundamental forces described in the Standard Model. It has a short effective range and is responsible for the radioactive decay of nuclei. In the weak interaction,

fermions can exchange three types of force carriers:  $W^+$ ,  $W^-$ , and  $Z$  bosons. The interactions associated with the charged  $W^+$  and  $W^-$  bosons are known as charged current interactions, whereas those associated with the neutral  $Z$  boson are known as neutral current interactions. Together, these vertices form the basis for weak interactions between leptons in the Standard Model. The weak interaction does not preserve parity symmetry [8].

During a weak decay, different final-state particles are produced depending on the charge and mechanism of the exchanged  $W$  boson. This can be understood by considering two types of beta decay illustrated by the Feynman diagram 1.2. A beta-minus decay with no initial charge converts a neutron into a proton at the same vertex. Thus, the interacting boson has to carry a negative charge and produces an electron at that vertex. Similarly, in a beta-plus decay, the boson is  $W^+$  and thus a positron is formed. Neutrinos are created on the basis of an intermediate boson that carries the interaction. Thus, there are three neutrino flavors based on the three charged leptons (electrons, muons, and tau) and their corresponding antineutrinos.

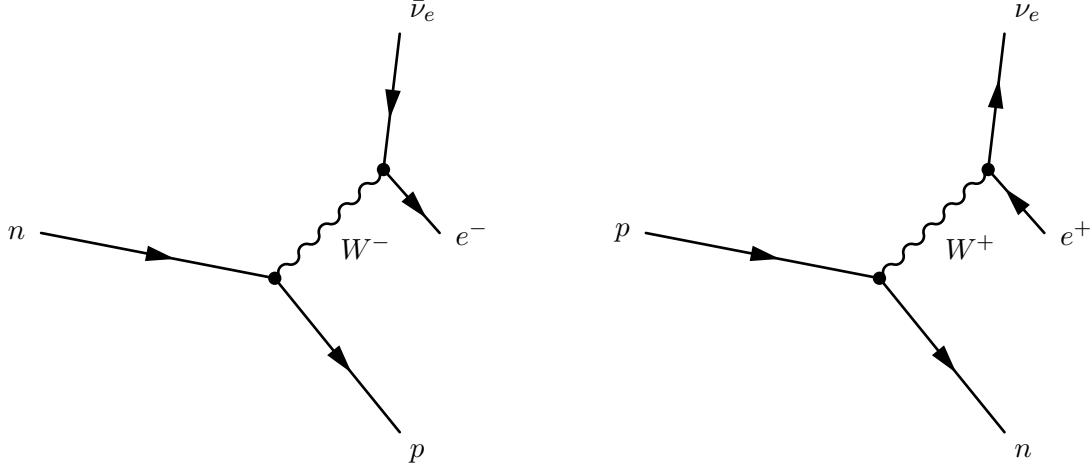


Figure 1.2: Feynman diagrams illustrating the two types of beta decay. Left: Beta-minus decay, where a neutron decays into a proton, emitting an electron and an electron antineutrino via a virtual  $W^-$  boson. Right: Beta-plus decay, where a proton converts into a neutron, emitting a positron and an electron neutrino via a virtual  $W^+$  boson.

In particle physics, the Lepton number is the net number of leptons in a reaction. Particles are given a unit positive lepton number and antiparticles a unit negative number. Two emergent properties from the Standard Model of Physics are that the lepton flavor and total lepton number  $L$  are conserved quantities. Conservation of the lepton number has no theoretical motivation and is not associated with any gauge symmetry. The discovery of neutrino oscillations provided the first evidence that lepton flavor is not

strictly conserved, suggesting there is physics beyond the Standard Model (BSM). Understanding the mechanism by which the neutrino gets its mass can provide additional evidence for physics beyond the Standard Model (BSM) physics.

## 1.4 Neutrino Mass

### 1.4.1 Dirac Mass Mechanism

There is no mechanism in the Standard Model to give neutrino mass. One way to add mass would be to use four neutrino fields that couple via the Higgs mechanism. The Higgs mechanism requires the existence of a right-handed neutrino field because, in the Standard Model, interaction with the Higgs boson couples opposite-chirality fields. The weak interaction is parity violating, and thus only the left-handed neutrino and right-handed antineutrino are produced or observed via weak interactions. They alone cannot couple to the Higgs field. It is possible that the right-handed neutrino and left-handed antineutrino are sterile to weak interaction, such that they do not couple to W or Z bosons. However, this would not explain the relatively low observed neutrino mass limits.

### 1.4.2 Majorana Mass Mechanism

A different approach is to consider neutrinos as Majorana fermions. Then the right-handed antineutrino,  $\nu_R^-$ , could be coupled to the left-handed neutrino,  $\nu_L$ . Thus, the left-chiral field would correspond to the neutrino and the right-chiral field would correspond to the antineutrino, and the neutrino and antineutrino could only be distinguished by chirality. This mechanism was first postulated by the Italian physicist Ettore Majorana [9]. Using the See-Saw mechanism, it could provide a basis for neutrino mass, and also explain why neutrinos are much lighter than any other Standard Model particles.

### 1.4.3 See-Saw Mechanism

The See-Saw mechanism attempts to explain the relatively low observed neutrino mass limits by including both Dirac and Majorana masses in the Lagrangian. This suggests that there would be two light neutrino and two heavy neutrino terms such that the Lagrangian is given by:

$$\mathcal{L}_{mass} = -\frac{1}{2}(\bar{\nu}_L \bar{\nu}_R^c) \begin{bmatrix} m_M^L & m_D \\ m_D & m_M^R \end{bmatrix} \begin{bmatrix} \nu_L^c \\ \nu_R \end{bmatrix} + hc \quad (1.1)$$

$m_D$  is the Dirac mass term,  $m_M^L$  and  $m_M^R$  are the left- and right-handed Majorana mass terms, and  $hc$  is the Hermitian conjugate of the prior term. The left-handed Majorana mass term is forbidden by  $SU(2)$  symmetry, so we let  $m_M^L \rightarrow 0$ . Diagonalizing the resulting matrix gives two effective fields:

$$\mathcal{L}_{mass} = -\frac{1}{2}(\bar{\nu}_L \bar{N}) \begin{bmatrix} m_\nu & 0 \\ 0 & M_N \end{bmatrix} \begin{bmatrix} \nu \\ N \end{bmatrix} + hc \quad (1.2)$$

The resulting eigenvalues are  $M_\nu \approx \frac{m_D^2}{m_R}$  and  $M_N \approx m_R$ . We have thus identified the mass eigenstates for neutrinos. If one of the mass terms increases, the other has to decrease, explaining the origin of the ‘See-Saw’ nomenclature. This would explain the very small mass of observed neutrinos. The left-handed light neutrinos and right-handed light antineutrinos would be the ones observed in current experiments. The heavy right-handed neutrinos and left-handed antineutrinos would have a huge phase space, thus a shorter half-life, and would have decayed quickly in the early universe. Experiments measuring neutrino mass would give insight into this mechanism.

#### 1.4.4 Measuring neutrino mass

There are four ways to study the neutrino mass. The first is to use results from cosmology. The Lambda-CDM model-based spectral fit to the Planck data and other observations of large-scale structures in the universe can be used to measure the sum of the neutrino masses. The sum of masses is one of the model’s free parameters, and the current limit implies a  $\sum m_\nu < 0.23$  eV [10]. However, this sum is model-dependent and relies on an extensive understanding of how the universe evolved.

A model-independent approach is to measure the mass of a neutrino kinematically. This is done by studying the endpoint of a beta decay spectrum, which would have a slight variation in shape depending on the neutrino mass. The KATRIN collaboration is performing such a direct mass measurement by looking at the tritium  $\beta$ -decay spectrum and has achieved a sensitivity of  $m_\beta < 0.45$  eVc $^{-2}$  at a 90% confidence level [11].

The third approach is to use neutrino oscillation experiments, though these can only set a lower limit on  $\Sigma m_\nu$  or  $m_\beta$ . The mass eigenstates of neutrinos form a complete, orthonormal basis. Similarly, the three flavors  $\nu_e$ ,  $\nu_\mu$ , and  $\nu_\tau$  form yet another orthonormal basis. Experiments indicate that these two eigenbases are rotated relative to each other. The Pontecorvo–Maki–Nakagawa–Sakata matrix represents the unitary transformation between the two bases as shown [12]:

$$\begin{bmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{bmatrix} = \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} \quad (1.3)$$

Equation 1.3 is often expressed using three mixing angles  $\theta_{12}$ ,  $\theta_{23}$ ,  $\theta_{13}$  and three phase angles ( $\delta_{CP}$ ,  $\alpha_1$  and  $\alpha_2$ ).

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta_{CP}} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & -c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{i\frac{\alpha_1}{2}} & 0 \\ 0 & 0 & e^{i\frac{\alpha_2}{2}} \end{bmatrix} \quad (1.4)$$

where  $c_{ij} = \cos(\theta_{i,j})$  and  $s_{ij} = \sin(\theta_{i,j})$ . This gives the probability that a neutrino of flavor  $\alpha$  oscillates to flavor  $\beta$  as:

$$P_{\alpha \rightarrow \beta} = \left| \sum_i U_{\alpha i}^* U_{\beta i} e^{-i \frac{m_i^2 L}{2E}} \right|^2 \quad (1.5)$$

It can be shown that equation 1.5 depends on  $\Delta m_{i,j}^2 = m_i^2 - m_j^2$  and not on individual masses. Thus, neutrino oscillation experiments can determine the squared difference of the masses and their sign, but not the individual masses of different flavors of neutrinos. We know from experiments such as SNO that  $m_2 > m_1$ , but no information is available about the sign of  $\Delta m_{23}^2$ . Thus, it is not known which neutrino is the heaviest, nor the values of each of the mass eigenstates. This problem is called the neutrino mass hierarchy problem. If the mass  $m_3$  is heavier than the mass  $m_2$ , the hierarchy is called normal, but if it is lighter, it is said to be inverted, as shown in figure 1.3.

The fourth approach is to study double beta decay, which would probe the effective neutrino mass term  $\langle m_{\beta\beta} \rangle$ .

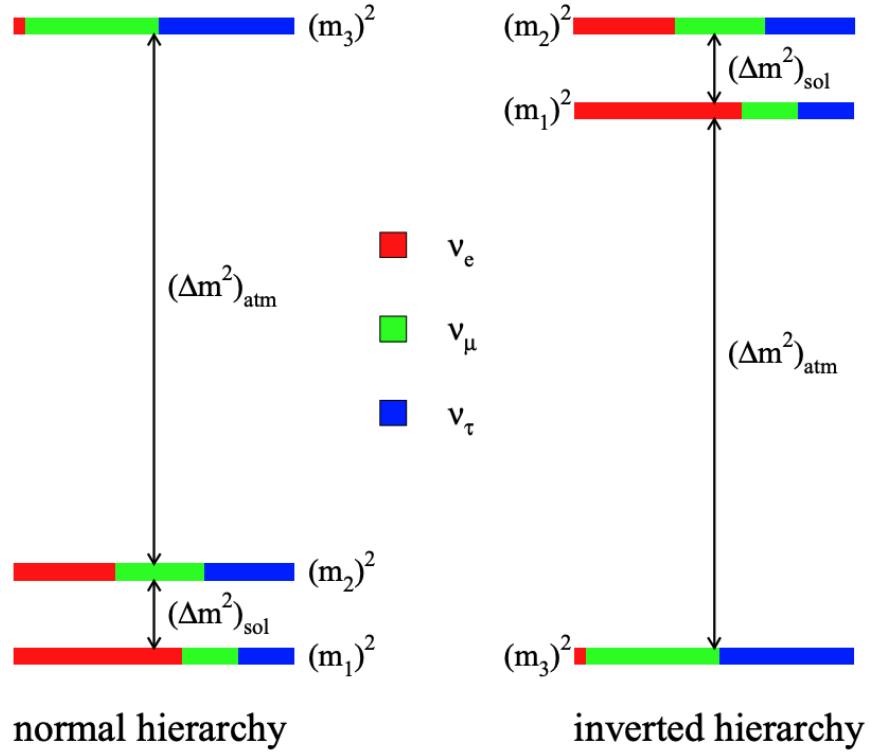


Figure 1.3: Two possible neutrino mass hierarchies. In the normal hierarchy (left), the lightest mass state is  $m_1$ , followed by  $m_2$  (separated by the solar mass-squared difference  $\Delta m_{21}^2$ ), with  $m_3$  being the heaviest, separated by the atmospheric mass-squared difference  $\Delta m_{32}^2$ . In the inverted hierarchy (right),  $m_3$  is the lightest, while  $m_1$  and  $m_2$  are heavier. Figure from [7].

## 1.5 Double Beta Decay

Some nuclei need to undergo a beta decay to move close to the ideal ratio of protons to neutrons; however, they are energetically forbidden to undergo such a decay. Instead, these nuclei may undergo double beta decay to achieve the optimal ratio of nucleons. In double beta decay, two neutrons in the nucleus are converted to protons simultaneously, producing two electrons and two electron antineutrinos. It is represented by equation 1.6 and Feynman diagram 1.4. It was first discussed by M. Goeppert-Mayer, and there are 35 potential nuclei in nature that can undergo this process [13].

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- + 2\bar{\nu}_e \quad (1.6)$$

The necessary conditions for a double beta decay are shown in figure 1.5. Candidate isotopes are neutron-rich, even-even nuclei that are forbidden to decay by beta decay but could decay via two

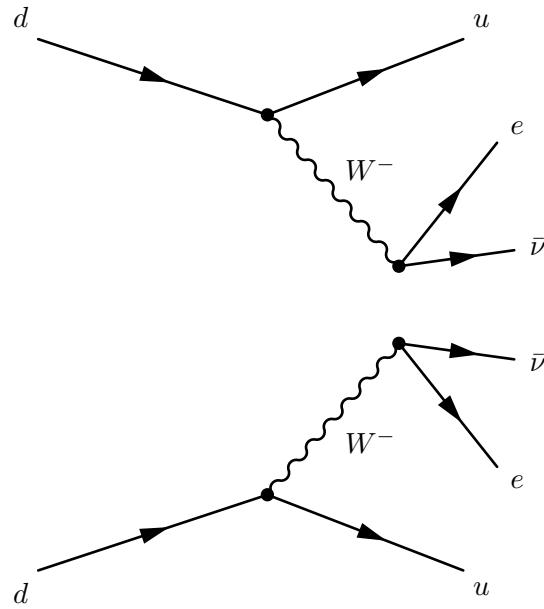


Figure 1.4: Feynman Diagram for a double beta decay ( $2\nu\beta\beta$ ) process.

simultaneous steps. Since such simultaneous decay of two nucleons in the same nucleus is very unlikely, double beta decay is a second-order weak process. It was first experimentally observed in  $^{82}\text{Se}$  with a half-life of  $(1.1^{+0.8}_{-0.3}) \times 10^{20}$  years by Elliott, Hahn, and Moe. [14]. It has been observed in many more nuclei since then, such as  $^{76}\text{Ge}$  [15],  $^{136}\text{Xe}$  [16],  $^{130}\text{Te}$  [17], etc.

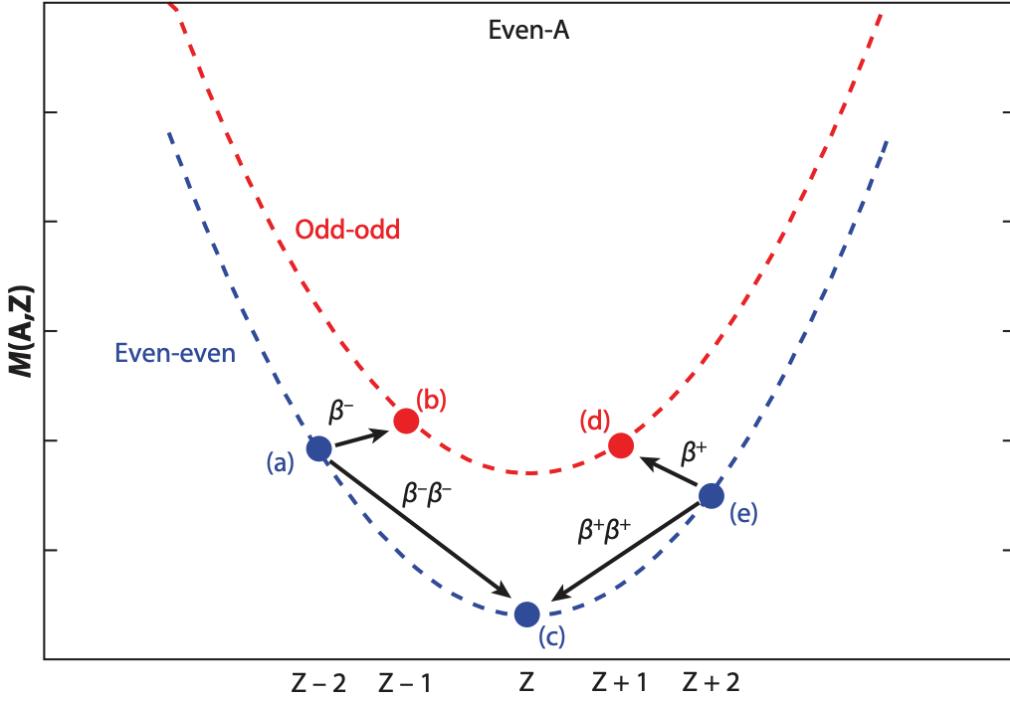


Figure 1.5: Ground state mass parabola for nuclei showing the necessary conditions for double-beta decay. The plot shows nuclear charge  $Z$  versus nuclear mass  $M(A, Z)$ . Double-beta decay is energetically allowed when the initial even-even nucleus has a higher mass than the final nucleus, and single beta decay is forbidden. Only even-even nuclei can undergo double-beta decay due to their greater stability. Figure from [18].

## 1.6 Neutrinoless Double-Beta Decay

If the neutrino were a Majorana particle, it would be possible to have a double beta decay without the emission of the two neutrinos. Instead, the neutrino would be exchanged as a virtual particle such that one nucleon absorbs the neutrino emitted by another. This process is called neutrinoless double-beta decay. This scenario, which is considered the simplest possible model, is termed “light neutrino mediated” decay. The general reaction would be as follows:

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- \quad (1.7)$$

Figure 1.6 shows the Feynman diagram for the process. An observable quantity of the process is the decay rate. Using Fermi’s golden rule, it can be expressed as:

$$(T_{1/2}^{0\nu})^{-1} = G^{0\nu} |M_{0\nu}|^2 \langle m_{\beta\beta} \rangle^2 \quad (1.8)$$

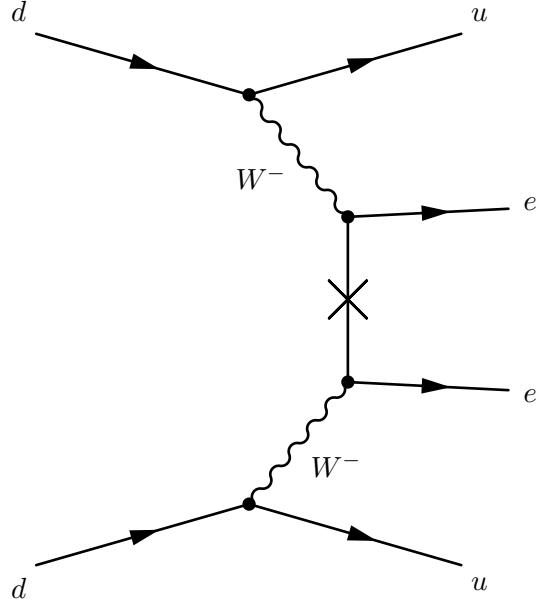


Figure 1.6: Feynman Diagram for light neutrino exchange neutrinoless double beta decay ( $0\nu\beta\beta$ ) process.

The phase space term  $G^{0\nu}$  corresponds to a phase space into which the electron can decay. The phase space can be calculated exactly for any given isotope.  $M_{0\nu}$  is the nuclear matrix element (NME), also called the ‘transition amplitude’. It encapsulates all of the nuclear physics processes occurring inside the nucleus and can be understood as the probability of the transition between the initial and daughter nuclei. It is difficult to know the initial and final wave functions of the nuclei, so calculating nuclear matrix elements is non-trivial and is an active field of research [19]. Finally,  $\langle m_{\beta\beta} \rangle$  is the effective neutrino mass term and can be represented as:

$$\langle m_{\beta\beta} \rangle = \left| \sum_{i=1}^3 U_{ei}^2 m_i \right| \quad (1.9)$$

$U_{ei}$  are the components of the PMNS neutrino mixing. This process described above assumes an exchange of a light Majorana neutrino, but there can be other exotic physics that can contribute to the decay rate (see [20] for an example). However, all of these processes still imply that the neutrino is a Majorana particle.

Equation 1.8 depends on the mass hierarchy and can help to probe the absolute neutrino mass scale. The observed decay rate would be different depending on the neutrino mass hierarchy. Figure 1.7 shows the range of  $m_{\beta\beta}$  as a function of the lightest neutrino mass  $m_l$  for the normal ordering and inverted ordering scenarios. If the neutrino masses are large compared to the size of the mass splittings, we do not

expect a significant effect on  $m_{\beta\beta}$ , as both hierarchies merge at higher masses; however, if the neutrino masses are similar in magnitude to the mass splittings, there would be a difference in  $m_{\beta\beta}$  based on the hierarchy.  $0\nu\beta\beta$  studies provide a fourth and unique way to understand neutrino mass, as they probe the effective Majorana mass  $\langle m_{\beta\beta} \rangle$  and provide a direct way to measure the ordering of neutrino mass. However, detecting this extremely rare process requires meticulous background reduction and large-mass experiments.

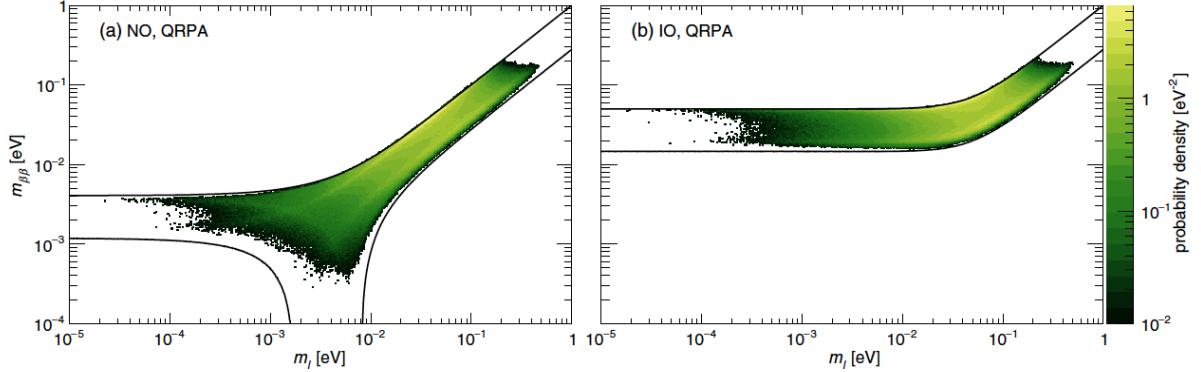


Figure 1.7: Parameter space for  $m_{\beta\beta}$  and  $m_l$  for the normal (right) and inverted (left) mass ordering. The allowed parameter space assuming  $3\sigma$  intervals of the neutrino oscillation observables from NuFIT is shown by the solid lines [21]. The marginalized posterior distributions for  $m_{\beta\beta}$  and  $m_l$  is shown by the color bar. This was obtained by combining equation 1.9 with neutrino oscillation measurements. Plot assumes QRPA NMEs and the absence of mechanisms that drive  $m_{\beta\beta}$  and  $m_l$  to zero. Source: [22]

## 1.7 Detecting Neutrinoless Double Beta Decay

Numerous experiments have searched for neutrinoless double-beta decay ( $0\nu\beta\beta$ ), and many are currently searching for it. The observable of interest is the sum of the kinetic energies of the two emitted electrons. If no antineutrino is emitted, the emitted electrons should carry the full energy difference between the final and initial nuclear states (the Q-value of the reaction). The experimental signature of such a decay would be a monoenergetic peak in energy at the end point of the two-neutrino spectrum, indicating that no antineutrinos are emitted, as shown in figure 1.8. Most  $0\nu\beta\beta$  experiments involve having a source of double-beta decay that acts as both a source and a detector. The region of interest (ROI) is defined as the narrow energy window around  $Q_{\beta\beta}$  and is selected according to the energy resolution of the detector. Using radioactive decay theory, the number  $0\nu\beta\beta$  of candidate events observed in the ROI is given by:

$$N = \ln(2) \frac{N_A}{W} \left( \frac{a\epsilon MT}{T_{1/2}^{0\nu}} \right) \quad (1.10)$$

where  $N_A$  is Avogadro's number,  $W$  is the molar mass of the source,  $a$  is the isotopic abundance of the parent isotope,  $\epsilon$  is the detector efficiency of the signal in the ROI, and  $t$  is the observed time. Using equation 1.10, we can understand how the half-life sensitivity depends on the detector:

$$T_{1/2}^{0\nu} \propto \begin{cases} aM\epsilon T & \text{(background free)} \\ a\epsilon \sqrt{\frac{MT}{B\Delta E}} & \text{(with background)} \end{cases} \quad (1.11)$$

such that  $\Delta E$  is the energy resolution of the detector, and  $B$  is the background index of the experiment normalized to the width of the ROI, source mass, and measurement time. The importance of low background is apparent as  $T_{1/2}^{0\nu}$  scales linearly with run time  $t$  for a no background experiment compared to  $\sqrt{t}$  in the presence of backgrounds.

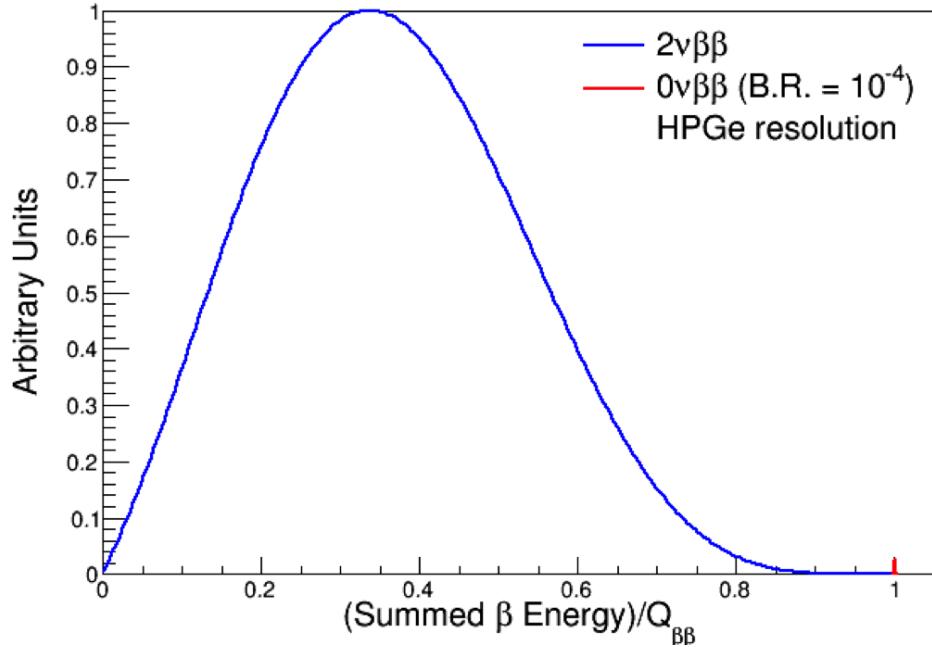


Figure 1.8: Simulated  $0\nu\beta\beta$  signal in a  $^{76}\text{Ge}$  detector. The  $0\nu\beta\beta$  signal is shown with an assumed half life of  $10^{25}$  [23] years.

For a substantial probability of discovery, experiments that seek  $0\nu\beta\beta$  need to have a source as large as possible with extremely low backgrounds. At that sensitivity level, natural isotopes such as  $^{222}\text{Rn}$ ,  $^{232}\text{Th}$ , and  $^{232}\text{U}$  can introduce background events. Cosmic rays are another potential source, along with

the decay of  $2\nu\beta\beta$  depending on the energy resolution of the experiment. Most of the experiments are thus housed underground to decrease the rate of cosmic-ray-induced backgrounds and cosmogenic activation of detector materials. They are also heavily shielded, and the experimental parts are ultra-clean. The experimental programs also need to perform radioactive assay studies to understand the backgrounds of their experiments.

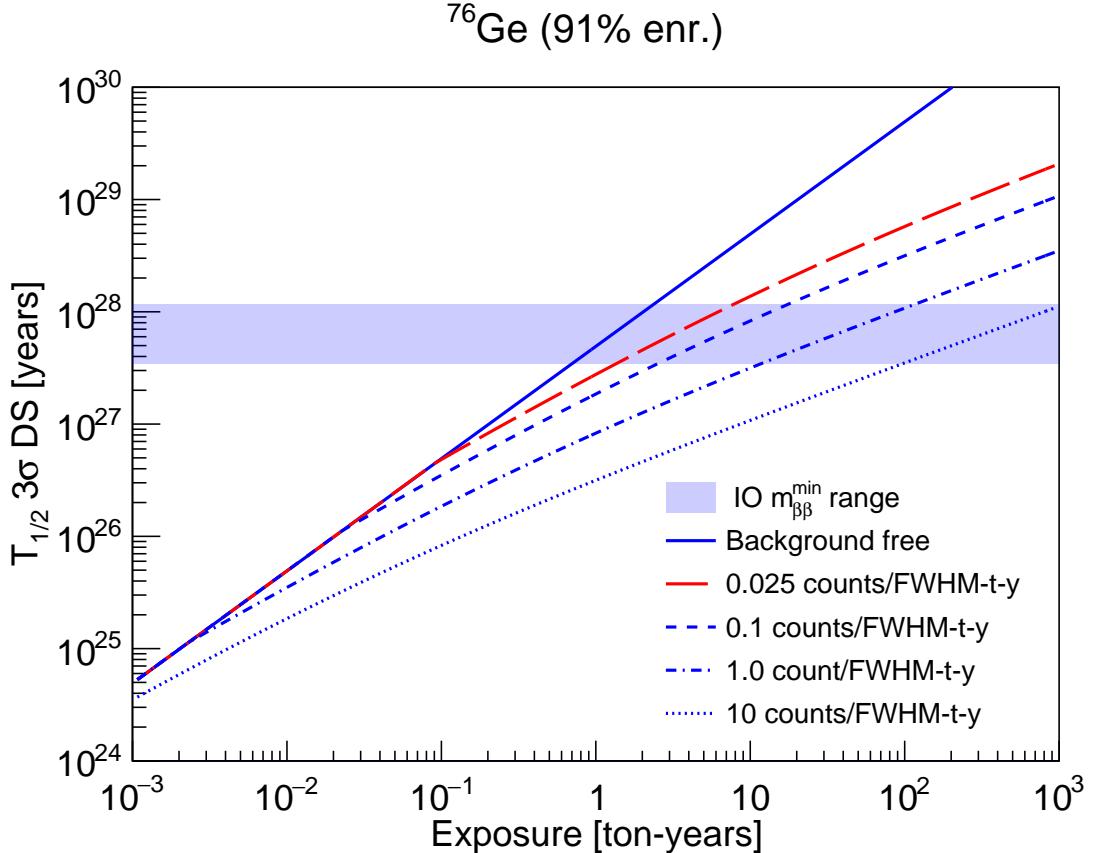


Figure 1.9: The exposure needed for  $3\sigma$  confidence level discovery of neutrinoless double beta decay for different half life limit, given the inverted hierarchy. Various lines represent scenarios with different background levels in ROI. Red line indicates the LEGEND-1000 background goal. The blue band shows the bottom of the inverted ordering.

The  $\beta\beta$  isotope used should ideally have a high Q value and a low double beta decay rate. It must also be able to turn into large detectors to increase the probability of discovery. Several isotopes are used in experiments searching for neutrinoless double beta decay, including  $^{76}\text{Ge}$  (GERDA [24], MAJORANA DEMONSTRATOR [25]),  $^{136}\text{Xe}$  (KamLAND-Zen [26], nEXO [27]), and  $^{130}\text{Te}$  (CUORE [28], SNO + [29]). KamLAND-Zen established the leading half-life limit in  $^{136}\text{Xe}$  with  $T_{1/2}^{0\nu} > 2.23 \times 10^{26}$  years [26].

In the coming years, the next generation of experiments will attempt to explore the entire inverted hierarchy region of figure 1.7. Figure 1.9 shows the exposure needed to achieve the discovery of  $3\sigma$  at the bottom of the inverted hierarchy region in  $^{76}\text{Ge}$ . The different lines represent background levels in counts per Region of Interest (ROI). The different background scenarios underscore the importance for experiments to reduce or eliminate all of their background in ROI and follow the solid blue line. As the experiments increase their exposure time, the sensitivity will increase and reach the bottom of the inverted hierarchy band. The bottom of the band is shown with the shaded blue box, which is calculated by using various calculations of nuclear matrix elements. Next-generation experiments would test an inverted hierarchy space and provide new insight into the nature of neutrinos.

Thus, the next decade would be an exciting time for research in neutrino physics. With ton-scale experiments such as LEGEND-1000 [30], there might be major revelations about the properties of the neutrino. We might find that neutrinos are their antiparticles, or at least that inverted ordering is not true. In that case, new experiments and possible theories must be postulated to explain the neutrino mass and its interaction. Neutrinoless double beta decay will inform a new Standard Model of physics. In the next section, we explore  $^{76}\text{Ge}$  as a choice of detector and the LEGEND experiment's search for neutrinoless double-beta decay.

## CHAPTER 2: HIGH PURITY GERMANIUM DETECTORS

### 2.1 Introduction

$^{76}\text{Ge}$  double-beta decays into  $^{76}\text{Se}$ , represented by the equation 2.1. It has several advantages as a source for  $0\nu\beta\beta$  decay experiments. As a semiconductor, it can be made into detectors with excellent energy resolution, which is critical because energy is the only observable that is both necessary and sufficient for a discovery. Its Q value is 2.039 MeV, which is above the energies of most gamma events. The crystal growth process for Germanium has purification stages that can greatly reduce internal backgrounds such as  $^{238}\text{U}$  and  $^{232}\text{Th}$  chain contaminants. Germanium is abundant in nature and can be easily enriched, allowing for large-scale experiments. Signal formation inside Germanium is well understood and enables precise topology-based discrimination between background and signal events through pulse shape discrimination (PSD) techniques. High-purity germanium (HPGe) detectors thus have a long history in nuclear physics and rare event searches.



Figure 2.1 shows a timeline of HPGe-based experiments. Over the years, research and development have resulted in experiments that continuously provide a larger bound on the half-life of the reaction while reducing the background.

### 2.2 Signal formation in Germanium

#### 2.2.1 Germanium properties

The lattice structure of crystalline metals results in allowed energy bands for the electrons. The outermost bound state is termed the valence band, and the unbound state is termed the conduction band. The difference between the valence band and the minimal conduction band energy is called the band gap. Electrons of all lattice atoms can move freely in the conduction band. In metals, the band gap is

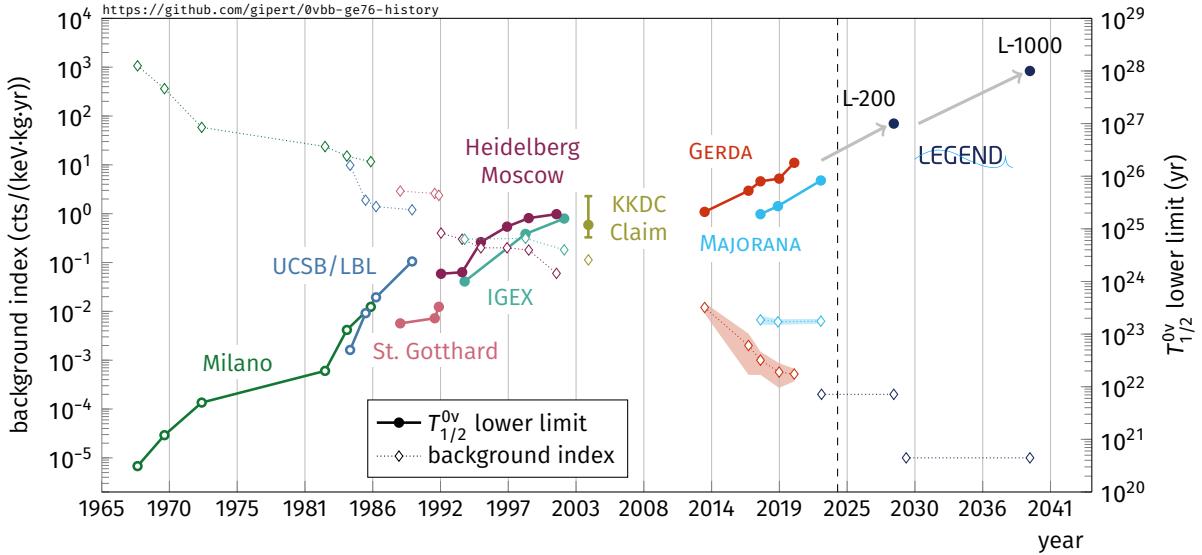


Figure 2.1: A plot showing a timeline of past and future HPGe  $0\nu\beta\beta$  decay experiments. The half-lives probed by them is shown along with their background index. Next generation experiments aim to continue improve half life limit while having the lowest background index. Picture credit: Luigi Pertoldi.

nonexistent, with the valence and conduction bands overlapping, making metals very good conductors. In insulators, the band gap is large, making them poor electrical conductors.

In semiconductors, the band gap is intermediate (2.96 eV for Ge), which makes it possible to use semiconductor crystals as a radiation detector. Due to the ideal band gap, radiation excites multiple electrons to the conduction band, leaving behind a positively charged atom (hole). Both charge carriers can drift along the electric field and can be measured at a readout electrode. In reality, the holes do not drift. An electron from an adjacent atom fills the hole, another electron fills the new hole, and the process continues until the last hole is filled by the cathode. The process gives the illusion that the hole is drifting towards the negative terminal, and we will treat it as such.

### 2.2.2 P-type Point Contact Germanium Detectors

A HPGe detector measures the total energy deposited by measuring the current induced by the drifting charge carriers on the readout electrodes. This is achieved by applying an external electric field to drift the charges towards the electrodes. There is still a steady leakage current as a result of the thermal excitation of the electrons. To reduce the leakage current, detectors are operated at low temperatures, but low electric fields make it difficult to efficiently detect radiation with just pure Germanium. Impurities are added to improve detector performance.

The impurities come in two forms: electron donors and receivers. The donors provide excess loosely bound electrons in the semiconductor, which is called the n-type, whereas the receivers increase the net number of holes, which is called the p-type. Impurities can be intentionally added to semiconductors to control the charge carriers and increase the conductivity by a process called doping. For example, doping Germanium with a group 15 element such as Arsenic will make it n-type, since Arsenic has 5 valence electrons compared to four in Germanium. Similarly, doping Germanium with a group 13 element such as Aluminum (3 valence electrons) would make it p-type [31].

During the manufacturing of an HPGe crystal, impurities are added to a molten bath, and the crystal is pulled via the Czochralski method. The concentration of impurities is not uniform throughout the crystal. In Germanium, this depends on the solubility ratio of the impurities in the solid and liquid phases, known as the segregation coefficient. A segregation coefficient close to 1, such as that for Aluminum, will result in it being uniformly distributed in the crystal. For several n-type impurities such as Phosphorus and Oxygen, the segregation coefficient is less than 1, and the impurities would prefer to be in a liquid state. As the crystal is pulled from its melt, the impurities with low segregation coefficients prefer the liquid phase and get concentrated in the bath. This will result in the tail having a higher impurity concentration. The net effect is an impurity gradient, which is important to consider while simulating signal formation inside the detector.

A diode is created when a p-type semiconductor is joined with an n-type semiconductor. The excess charges diffuse into each other, creating a region with no free charge carriers, called the depletion region. The process produces a net positive charge on the n side of the material and an equal-magnitude negative charge on the p side. This results in a potential difference across the material, termed the contact potential. If an external potential is applied opposite to the contact potential (positive potential on the p side, termed forward bias), there will be a free flow of current; however, if the external potential difference is applied opposite to the contact potential (positive potential on the n side, termed reverse bias), more charge carriers will diffuse, resulting in a larger depletion region and lower leakage current. This depletion region is ideal for detecting incident radiation, as the electrons ionized by the radiation are drifted by the bulk electric fields, creating a signal that is read out at the electrodes.

The depletion region can be inferred from the capacitance of the detector. The capacitance will continue to decrease until a certain potential, called the depletion voltage, is reached. The depletion voltage is where the biggest possible depletion region is achieved. Reducing the capacitance of HPGe

detectors decreases the inherent noise in the readout circuit, thereby improving the energy resolution. This can be achieved by using a point-like central contact for the electrodes. In general, holes are less sensitive to charge trapping, and thus P-type detectors are preferred over N-type detectors for point-contact detector geometries.

### 2.2.3 Shockley-Ramo Theorem

Energy deposition in the detector will create charge clouds of electrons and holes. These charges induce an image charge on the point contact which is kept at zero potential. As these charge clouds drift under the electric field inside the detector, the image charges change, inducing a current at the point contact. The electrons and holes drift in opposite directions, resulting in a net negative induced signal on the read-out electrode. The Shockley–Ramo theorem gives the induced charge  $Q(t)$  on an electrode caused by a moving charge  $q$ :

$$Q(t) = q\Delta\phi_0(t) \quad (2.2)$$

The weighting potential  $\phi_0(t)$  is the solution of the Laplace equation with the boundary conditions of  $\phi_0 = 1$  for the readout electrode and  $\phi_0 = 0$  for all other electrodes. Figure 2.2 left shows the calculated weighting potential inside a point contact detector used in LEGEND-200. Note that the weighting potential is close to zero for most of the bulk. Thus, most energy depositions create charge clouds in the detector region with a low weighting potential. The holes travel to the  $p^+$  point contact at the bottom and the electrons travel to the nearest  $n^+$  region which surrounds the detector. Thus, holes undergo a larger change in the weighting potential than electrons since the point contact has a weighting potential of 1. This bigger change means that the hole contribution dominates the signal. The holes encounter most of this potential difference in the region near the point contact, which results in a step-shaped signal shown by a sample waveform from a  $^{76}\text{Ge}$  detector in figure 2.2.

The pulse shape, along with a long drift time and a short collection time, provides an opportunity to veto multi-site events from gamma rays using analysis cuts. As the drift time varies with the path taken, multiple steps can often be identified in waveforms associated with multi-site events.

The MAJORANA, GERDA and LEGEND collaborations are experiments with extensive experience in the research and development of HPGe detectors for  $0\nu\beta\beta$  searches.

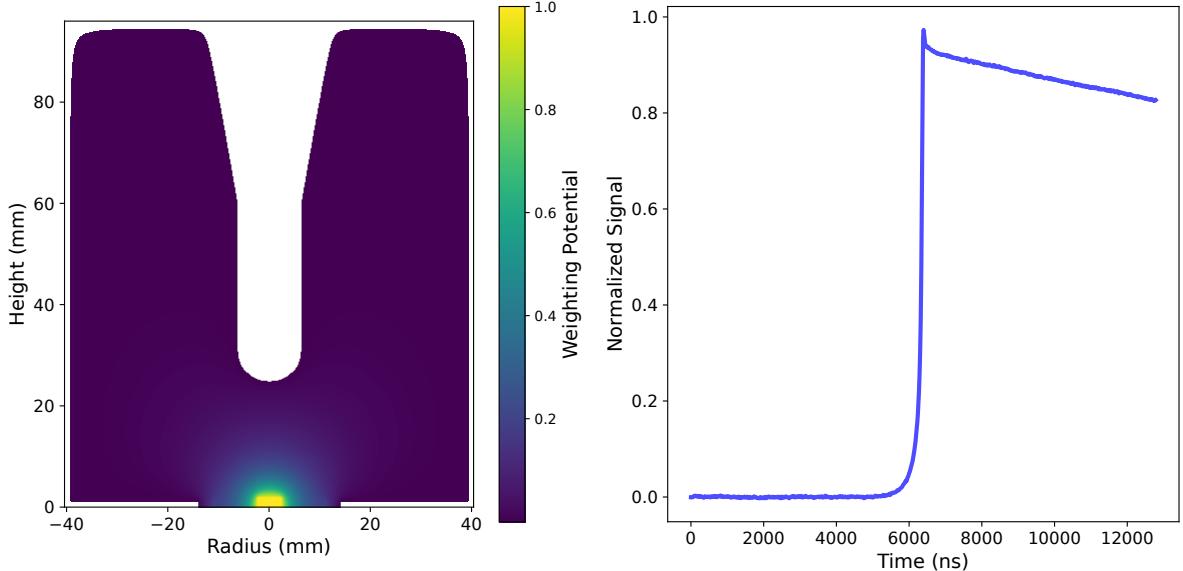


Figure 2.2: Left: Simulated weighting potential inside a LEGEND ICPC style point contact detector. Right: Collected signal from a detector showing sharp step shaped waveform.

### 2.3 The MAJORANA DEMONSTRATOR

The MAJORANA DEMONSTRATOR (MJD), located at the Sanford Underground Research Facility (SURF) in the USA, operated a 30 kg array of enriched point-contact p-type (PPC) detectors developed by ORTEC. The detectors were divided between two vacuum-insulated cryostats housed within a low background shield and operated in a vacuum at liquid nitrogen (LN) temperature, as shown in figure 2.3.

MJD pioneered the development of underground electroformed copper (EFCu), where the purest commercially available Cu is obtained and electroformed underground to reduce contamination from cosmogenically produced  $^{60}\text{Co}$ . This also further reduced the concentration of natural contaminants  $^{238}\text{U}$  and  $^{232}\text{Th}$  by a factor of approximately 30 [32]. Another key innovation was the use of a low-mass front-end (LMFE) electronics board. The LMFE was designed to be highly radiopure and hence reduced the background close to the detector. It also reduced the noise in the signal, as the first stage of amplification occurred very close to the detector. With its ultra-clean procedures, underground electroformed Copper and advances in front-end electronics, MJD achieved the lowest energy resolution in any large scale  $0\nu\beta\beta$  decay experiment of 2.52 keV (0.12%) FWHM at  $Q_{\beta\beta}$  when combining all detectors, and set a half-life limit  $8.3 \times 10^{25}$  years (90% C.L.). MAJORANA DEMONSTRATOR had the second-lowest background index of  $6.6 \times 10^{-3}$  cts/(keV kg yr). [25].

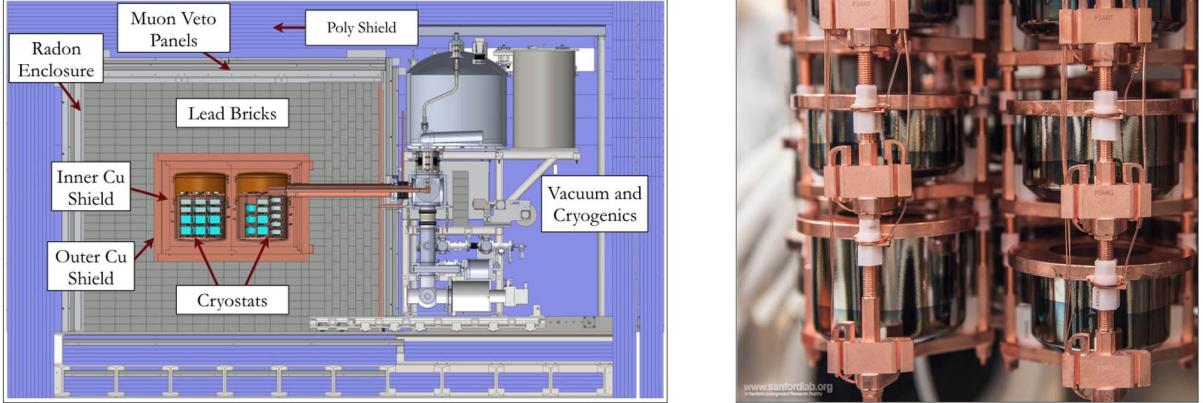


Figure 2.3: The MAJORANA Demonstrator setup (left) and its Ge detector arrays (right). Source: MAJORANA Collaboration.

## 2.4 The GERmanium Detector Array

The GERmanium Detector Array (GERDA) collaboration was another enriched Germanium detector-based experiment, located at the Laboratori Nazionali del Gran Sasso (LNGS) in Italy. GERDA consisted of an array of 20 kg of customized enriched versions of broad-energy Ge (BEGe) detectors and 15.6 kg of enriched semicoaxial detectors developed by Mirion, as shown in figure 2.4. In its last year of data collection, GERDA deployed an additional 9.6 kg of the new p-type inverted coaxial point contact (ICPC) detectors to verify their technical maturity for future LEGEND experiments.

GERDA took a unique approach to reduce background: operating the detectors in liquid Argon. The detector strings submerged in liquid Argon were surrounded by a shroud of wavelength-shifting fibers, and the entire LAr cryostat was surrounded by water that enabled a muon veto using Cherenkov radiation. The fibers converted Ar 128 nm scintillation photons into green photons, which were observed using silicon photomultipliers (SiPM) attached to the ends of the fibers. This enabled a coincidence combining events detected in the Germanium detectors and the scintillation light read by SiPMs. The LAr veto allowed GERDA to distinguish gamma backgrounds from  $0\nu\beta\beta$  signal candidate events with high efficiency, achieving a background rate of  $5.2 \times 10^{-4}$  cts/(keV kg yr) and reaching a half-life sensitivity greater than  $10^{26}$  years [24].



Figure 2.4: GERDA setup (left) with Ge detectors (right) arranged in vertical strings. Source: GERDA Collaboration.

## 2.5 The LEGEND experiment

The Large Enriched Germanium Experiment for Neutrinoless  $\beta\beta$  Decay (LEGEND) combines the best techniques and resources from GERDA and MJD. In both previous experiments, the maximum mass of the detectors was constrained by the electrode geometry to  $\sim 1$  kg. Higher mass results in a lower surface-to-volume ratio, which reduces the per-channel background and decreases the number of channels needed, thus reducing per-kg background from cables, connectors, and detector mounts. A major innovation for LEGEND has been the development of p-type inverted-coaxial point contact (ICPC) detectors, which can be manufactured with masses of around 2 – 4 kg [33]. LEGEND is taking a phased approach to a double-beta decay experimental program.

### 2.5.1 LEGEND-200

The first phase currently underway consists of 200 kg of Germanium detectors housed in an upgrade of the GERDA infrastructure at LNGS. In the first configuration, the MJD and GERDA detectors constitute 55.8 kg of the source mass, and the ICPC detectors 86.7 kg. The setup is illustrated in figure 2.5.

MJD achieved excellent energy resolution, primarily due to its low noise and low background electronics. The background model spectral fits indicated that the dominant background source in MJD was not from  $^{208}\text{Tl}$  sources close to the detectors. [34]. On the other hand, GERDA achieved a lower background index with the help of its efficient LAr veto. The prominent features of its background

model were the  $^{210}\text{Po}$  on the surface of the Ge-detector p<sup>+</sup> electrode,  $^{232}\text{Th}$  and  $^{238}\text{U}$  decay chains, and  $^{42}\text{Ar}$  progeny. GERDA's background model fits of GERDA suggest that the dominant background contributions [24] were from components close to the detector array.

LEGEND-200 inherits the best of both experiments. It uses MJD's low background materials and electronics to counter the background GERDA observed closer to the detector and uses a LAr veto with improved scintillation light readout to reduce the distance-component backgrounds faced by MJD. The introduction of ICPC detectors also results in a lower surface-to-volume ratio, allowing for a reduction in net radioactive decays of alpha and beta particles near the surface. Higher-mass detectors also reduce the need for detector support materials and electronics that can be sources of background.

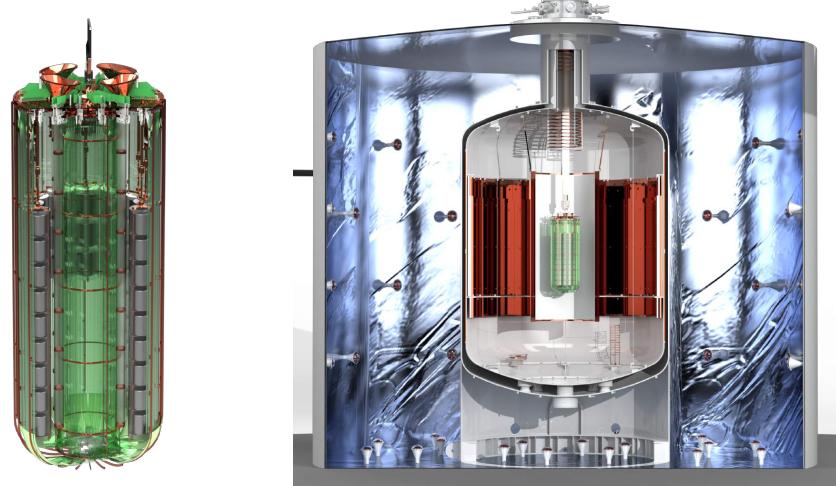


Figure 2.5: Left: LEGEND-200 Ge detectors strings surrounded by wavelength-shifting fibers for LAr veto. Right: The L-200 design. The detector array is mounted in the center of the LAr cryostat. Muon veto is achieved by placing the cryostat in a water tank and detecting Cherenkov radiation with photo-multipliers. Source: LEGEND Collaboration.

The three types of point contact detector geometries that LEGEND uses are illustrated in Figure 2.6. The p<sup>+</sup> contact is created by implantation of boron at the crystal end with the highest impurity. The n<sup>+</sup> region is created by diffusion of lithium atoms onto the detector surface. To isolate the two electrodes from each other, the surface region in between is passivated, usually with amorphous germanium. The BeGe and ICPC detectors also have a passivated ditch. The electric field is carefully designed by controlling the geometry and impurity gradient in the crystals.

LEGEND-200 took physics data in a 142 kg source configuration from March 2023 until February 2024. The first data set was unblinded on 13 June 2024. After applying analysis cuts, a total of seven

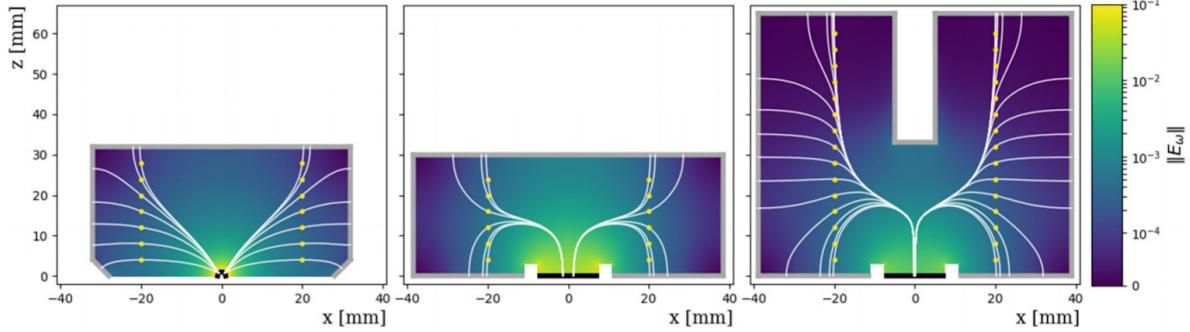


Figure 2.6: The three point contact detector geometries used in LEGEND: MAJORANA DEMONSTRATOR PPC detectors (left), GERDA BEGe detectors (middle), and newly developed ICPC detectors (right). The color bar depicts weighting potential within each detector’s cross section. The thick black and gray lines are the  $p^+$  and  $n^+$  electrode, respectively. The yellow points are locations of simulated energy depositions. The holes travel along the white lines to the  $p^+$  electrode, while electrons travel along white line to the  $n^+$  electrode. Figure from [35].

events were observed in the region of interest, corresponding to a background index of  $(5.3 \pm 2.2) \times 10^{-4}$  cts/(keV·kg·yr). In a combined fit of GERDA, MAJORANA DEMONSTRATOR, and LEGEND-200, the lower limit on the  $0\nu\beta\beta$  half-life was set at  $T_{1/2}^{0\nu} > 1.9 \times 10^{26}$  yr (90% C.L.) with an expected sensitivity of  $2.8 \times 10^{26}$  years. [36]

### 2.5.2 LEGEND-1000

The subsequent phase of LEGEND is a proposed ton-scale experiment that will probe the entire inverted ordering mass hierarchy, with the capability to make an unambiguous discovery of  $0\nu\beta\beta$  with  $3\sigma$  certainty for the  $m_{\beta\beta}$  range of 9–21 meV. It would also cover half of the normal ordering parameter space, under certain assumptions [37]. LEGEND-1000 builds on the innovation and research done in LEGEND-200. It will use only ICPC detectors, which have demonstrated lower surface backgrounds, excellent energy resolution, and precise pulse shape discrimination. The experiment also plans to use underground-sourced liquid Ar, which would greatly reduce the  $^{42}\text{K}$  background. LEGEND-1000 would probe a half-life of  $10^{28}$  years after 10 years of live time. The preconceptual design report for LEGEND-1000, with a detailed design description and information about the discovery potential, can be found at [37]. LEGEND-200 will be the main scope of this thesis, with the aim of developing techniques for LEGEND-1000.

## 2.6 LEGEND-200 background model

LEGEND-200 has very low backgrounds. Figure 2.7 shows the expected background contribution in LEGEND-200 at  $Q_{\beta\beta}$  after all analysis cuts. These contributions were calculated using radiopurity assays of selected materials and Monte Carlo simulations.

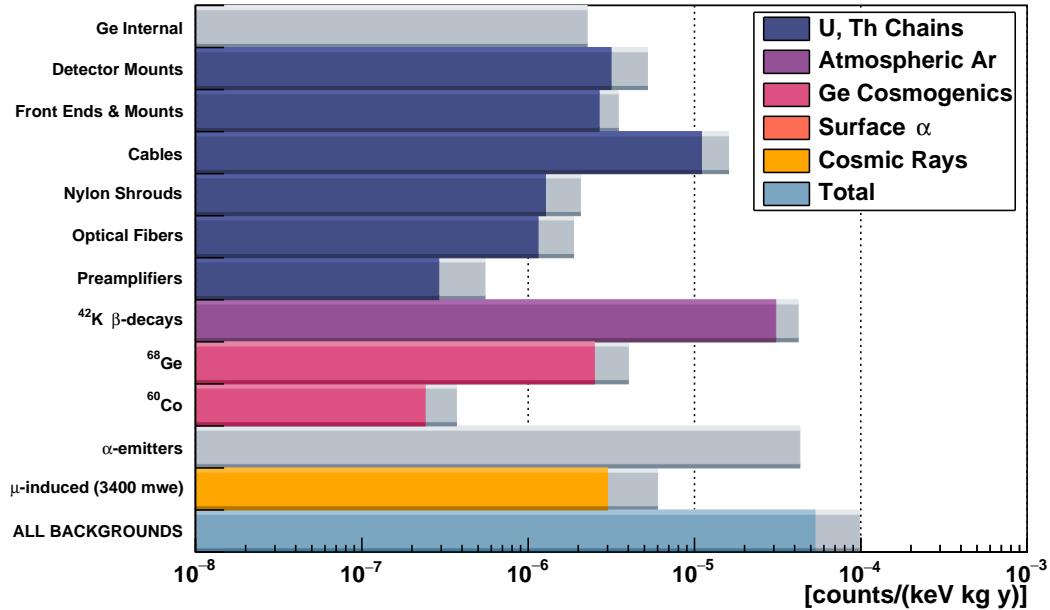


Figure 2.7: LEGEND-200 Projected background contributions near  $Q_{\beta\beta}$  after all analysis cuts. Grey bands indicate uncertainties in assays and background rejection.

The decays of long-lived isotopes  $^{238}\text{U}$  and  $^{232}\text{Th}$ , as well as their short-lived progeny, are the main backgrounds of component materials. They are primarily reduced by using higher purity materials and by reducing materials near the detectors.

Cosmogenic backgrounds from  $^{68}\text{Ge}$  and  $^{60}\text{Co}$  isotopes are produced when the HPGe crystal is exposed to neutrons produced by cosmic rays. This background can be reduced by controlling the cosmic-ray exposure during detector production and increasing the underground cool-down period for the HPGe detectors prior to the experiment.

$0\nu\beta\beta$  events are inherently single-site. Most background events in HPGE deposit energy in multiple locations. During the analysis, LEGEND approaches these backgrounds using two broad techniques: Pulse shape discrimination (PSD) and Anti-Coincidence (AC) tagging.

The HPGe point contact configuration enables identification of the unique drift of the charge clouds for each location. PSD techniques take advantage of this ability to reject multi-site and surface backgrounds. An example of a PSD technique is the use of the A/E discriminator, which relies on the maximum value of the derivative of the waveform, the current amplitude. Figure 2.8 shows the current amplitude for four different events in the experiment:  $n^+$  surface beta event,  $p^+$  surface alpha event, a multi-site gamma event and a single-site signal event, such as a  $0\nu\beta\beta$  or  $2\nu\beta\beta$  event. Each event has a unique current signal, and the maximum value of the current provides an effective way to accept or reject them. The cuts are very effective at rejecting backgrounds, but their precise level of effectiveness is difficult to measure, particularly for surface backgrounds, as we do not have access to a large or pure sample of this category of events. Knowing this rejection efficiency helps us improve PSD cuts and is needed to predict background in future experiments, like LEGEND-1000.

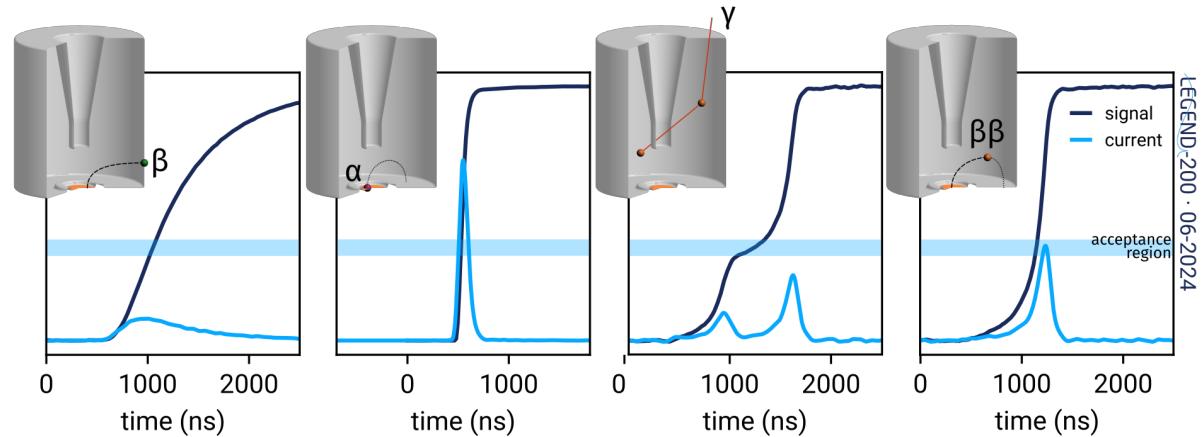


Figure 2.8: Signal (black) and current (blue) for four different types of events in LEGEND-200. The point contact geometry enables long and unique drift times that allow for pulse shape discrimination. The maximum of the current signal can be used to reject the three background events and accept the  $2\nu\beta\beta$  events. Credit: The LEGEND Collaboration.

The AC cuts are used to detect background depositing energies at multiple locations. The liquid argon cut rejects events that happened in the germanium detector while producing scintillation in the LAr volume. Similarly, the Muon veto rejects events that produce Cherenkov light in the water tank. Finally, the granularity cut removes events producing energy depositions in multiple HPGe detectors.

<sup>42</sup> K betas are among the largest contributions to the background. These backgrounds come from betas that interact on all surfaces of the detectors. At the  $n^+$  surface, the thick dead layer blocks some of the betas from reaching the active volume, but PSD is the most important tool in reducing these

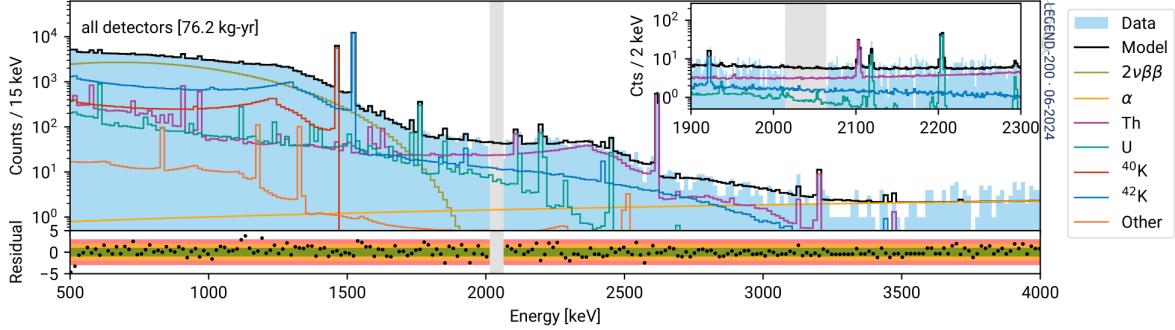


Figure 2.9: LEGEND-200 background contributions near  $Q_{\beta\beta}$  after all analysis cuts. Credit: The LEGEND Collaboration.

backgrounds. The use of liquid Argon introduces two short-lived isotopes,  $^{39}\text{Ar}$  and  $^{42}\text{Ar}$ .  $^{42}\text{Ar}$  decays into  $^{42}\text{K}$  which is positively charged and can drift with the applied electric field to the detector surface. Then it can undergo a beta decay with a  $Q$  value of 3525 keV that overlaps with the region of interest.  $^{42}\text{Ar}$  is formed by cosmogenic exposure to atmospheric Ar, and the best way to mitigate its background is to use underground-sourced Argon, as proposed for LEGEND-1000. LEGEND-200 uses nylon shrouds that inhibit the drift of  $^{42}\text{K}$  ions. AC and PSD cuts further help to identify the beta background.

Alpha backgrounds are  $\text{p}^+$  and passivated surface events that are derived primarily from surface contaminants such as  $^{210}\text{Po}$  and  $^{210}\text{Pb}$ . These originate from the decay of the natural  $^{222}\text{Rn}$  gas that is adsorbed on the surface during the fabrication, storage, and assembly processes of the detectors. Components can also develop static charges that can increase adsorption. Alpha backgrounds can be reduced by handling the detectors and surrounding materials with care in ultraclean and Rn-reduced environments. In LEGEND, support materials are etched using acids or leached in Nitric acid to reduce the adsorption of radon gas, and then the detectors are assembled and installed into cryostats using a glovebox with a radon-reduced dry Nitrogen environment. The use of liquid Argon reduces the alpha background due to the gaseous  $^{222}\text{Rn}$  previously observed in MJD.

The alpha background contribution in figure 2.7 is estimated based on the rate of high-energy background events and their survival after cuts, as observed in GERDA. This is both a small sample size and an extrapolation to the  $0\nu\beta\beta$  region of interest and, therefore, has a high uncertainty. These uncertainties are driven by uncertainty in the pulse shape rejection for background components in the  $0\nu\beta\beta$  region of interest. Since there is no first-principles background model for surface events, it is hard to estimate the fraction of background events that leak through the cuts. Passivated surface effects can

also affect signal efficiency. If an event occurs near the passivated surface, the surface effects may lead the detector to measure a lower energy than the actual deposited event. Thus, the event might not be registered as  $0\nu\beta\beta$ , which then reduces the active detector volume for  $0\nu\beta\beta$  events and distorts the shape of the  $2\nu\beta\beta$  spectrum.

Figure 2.9 shows the fit of the background model to the data in LEGEND-200. It is constructed using a global binned likelihood fit using measured spectra, radio assay data, and Monte Carlo simulations. The alpha background was modeled using an analytic spectral shape model developed by the GERDA Collaboration. [38]. GERDA used multiple PDFs corresponding to different  $p^+$  contact thicknesses to model partial energy deposits of alpha decays on the passivated surface. This is combined with a linear continuum to handle highly degraded events. Although the overall background model accurately describes the gamma-induced background, the alpha background component at  $Q_{\beta\beta}$  is mostly flat and remains poorly constrained due to modeling challenges. The GERDA model is not effective for MAJORANA PPCs, which have a larger passivated surface.

Pulse Shape Simulations (PSS) model how the charge carriers move and induce signals in the HPGe detectors. PSS enable improved detector design, position reconstruction, and improved pulse shape discrimination by accurately modeling the geometry and electric fields of each detector. They are also key to estimating which background events remain after analysis cuts. Thus, PSS are critical for identifying the sources of background in current experiments and projecting background levels in future searches. Two longstanding obstacles have hindered reliable PSS: incomplete models of charge collection from surface event interactions and the complexity of replicating the detector’s electronic response.

In this work, we focus on methods to improve PSS to model the backgrounds. Chapter 3 focuses on the challenges involved in modeling alpha interactions on the passivated surfaces. The chapter introduces the EH-Drift simulation, which can model surface events using charge diffusion, self-repulsion, and reduced mobility on the surface. Chapter 4 addresses the computational challenges posed by the EH-Drift by showing how the simulation can be accelerated by using GPU computing techniques, enabling large-scale parallelism. Chapter 5 presents the results of EH-Drift simulations. It shows how EH-Drift can reproduce data from test stands, estimate efficiency losses in the region of interest for  $0\nu\beta\beta$  events, and generate detector-physics-driven spectra for surface backgrounds.

Chapter 6 transitions to the challenge of replicating the detector’s electronic response. It describes how the readout chain modifies waveforms, why first-principles approaches are difficult, and how

CPU-Net, a deep learning framework, can improve electronics modeling. Chapter 7 explains the data preparation, simulation pipelines, and training procedures for the neural network–based electronics model. Chapter 8 presents the outcomes of CPU-Net, showing how simulated pulses can be translated into realistic detector-like waveforms and analyzes network performance on critical pulse-shape parameters.

## CHAPTER 3: SURFACE EVENT MODELING AND PULSE SHAPE SIMULATIONS

### 3.1 Introduction

The events on passivated surfaces create a unique modeling challenge in HPGe detectors. These interactions are best illustrated with alphas, since they deposit energy over short distances on the surface, creating large charge clouds with complex drift and collection processes, making it quite challenging to understand. We therefore study the events in terms of alpha interactions, but the same techniques can be applied to simulations of other passivated surface backgrounds such as beta and low-energy gamma interactions. This chapter begins by reviewing alpha backgrounds, then discusses the challenges in modeling surface backgrounds, and introduces a new model to address these challenges.

### 3.2 Properties of Alpha Interactions

Alpha particles are composed of two protons and two neutrons tightly bound together. They are identical to ionized Helium atoms with a charge of  $+2e$ . The energies of alpha particles are correlated with the half-life of the parent isotope, such that the ones with the highest energies are from parent atoms with the shortest half-lives. The energies of the alpha particles are usually between about 4 and 6 MeV. For a given material, the linear stopping power  $S$  is defined as  $-\frac{dE}{dx}$ . The energy loss of charged particles in a material is then given by the Bethe formula [39]:

$$-\frac{dE}{dx} = \frac{4\pi e^4 z^2}{m_0 \nu^2} N B \quad (3.1)$$

such that

$$B = Z \left[ \ln \frac{2m_0 \nu^2}{I} - \ln \left( 1 - \frac{\nu^2}{c^2} \right) - \frac{\nu^2}{c^2} \right] \quad (3.2)$$

Here  $\nu$  and  $ze$  are the charges of the given particle.  $N$  and  $Z$  are the number density and atomic number of the absorber atoms, and  $m_0$  and  $e$  are the rest mass and charge of the electron, respectively. Parameter

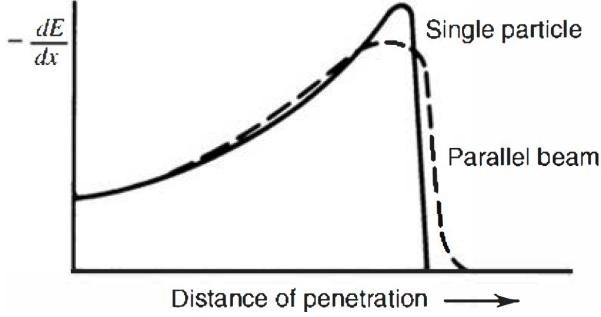


Figure 3.1: The specific energy loss in a material for an alpha particle with several MeV initial energy. Plots are shown for a single alpha particle track and for the average behavior of a parallel beam of alpha particles. Alphas lose most of their energies in a small region. Picture from [31]

I represents the average excitation and ionization potential of the absorber and is normally determined experimentally.

Equation 3.1 suggests that the energy loss for non-relativistic particles is proportional to  $1/\nu^2$ , or the kinetic energy. Hence, the slower the velocity, the longer the particles spend in the vicinity of electrons of the material's atoms, and the higher the energy loss. Equation 3.1 is proportional to  $z^2$  for a constant velocity. Thus, the heavier particles will have more energy loss for a given velocity. This results in alphas having a very low penetration depth, usually around 10 micrometers.

The energy loss of the alpha particles can be best understood by plotting the specific energy loss along the track. The plot known as the Bragg curve is shown in figure 3.1. During most of the track, the energy loss increases roughly as  $1/E$ , as predicted by equation 3.1. As the particles slow down, electron pickup reduces the charge and the curve falls off. As a result, alpha deposits a large amount of energy in a very localized area. For Germanium, this penetration depth of alpha particles is about 17.6 - 20.0 microns [31]. The n+ electrode is too thick for the alphas to penetrate and deposit energy. The p+ electrode in LEGEND detectors does have alpha events, but the charge collection is well understood, as the events are collected immediately and thus can be rejected easily using PSD techniques. Between these contacts is a thin insulating passivated surface, where alphas can penetrate the active volume of the detector, and the charge collection is more complex. The modeling technique developed here is focused on this surface.

### 3.3 Alpha Background Observations

The alpha background can vary between different detector environments and experiments. The MAJORANA DEMONSTRATOR and GERDA experiments had different spectral shapes for alpha backgrounds. In the MAJORANA DEMONSTRATOR, alpha events degraded more substantially in the  $0\nu\beta\beta$  ROI, while in GERDA, prominent peaks of the  $^{210}\text{Po}$  and the  $^{226}\text{Ra}$  chains were identified above 5 MeV, along with lower energy tails of partially collected surface alphas. Factors such as the cryostat environment, detector geometry, and material purity can contribute to the alpha background at  $Q_{\beta\beta}$ .

This can also be seen in dedicated test stands that were developed to study alpha backgrounds. Two scanning systems were developed to study alpha interactions in HPGe detectors: TUBE [40] and GALATEA [41]. These scanners yielded conflicting dependencies of the energy loss as a function of radial position. TUBE reported that the energy degraded inversely with the detector radius, whereas GALATEA reported the opposite trend. Figure 3.2 shows a known alpha waveform in the TUBE data. The alpha waveform, being close to the surface, has a much sharper rise than the bulk waveform. However, the tail of the alpha waveform has an upward slope towards the end. This is due to slow collection of charges on the passivated surface. We call this effect the Delayed Charge Recovery Effect (DCR) [42]. This feature and others like it are used to tag and reject alphas. Such cuts have been highly effective, but it is difficult to know which events are not tagged without modeling the signal.

Figure 3.3 shows the LEGEND-200 physics spectrum. Muon and multiplicity cuts remove backgrounds such as muons and gamma rays, and the events left are primarily surface events shown by the unfilled histogram. As shown in green, Pulse Shape Discrimination (PSD) cuts are highly effective in removing these backgrounds, but the efficiency of those cuts cannot be determined due to the lack of a model for surface events.

### 3.4 Challenges in Modeling Surface Events

An alpha particle deposits an energy of approximately 4 - 6 MeV with a penetration depth of approximately 17.6 - 20.0 microns on the passivated surface. This creates a lot of charge carriers and produces a dense charge cloud. In such cases, effects such as diffusion of the charges and self-repulsion among charges become significant. These effects could also push the charge onto the passivated surface.

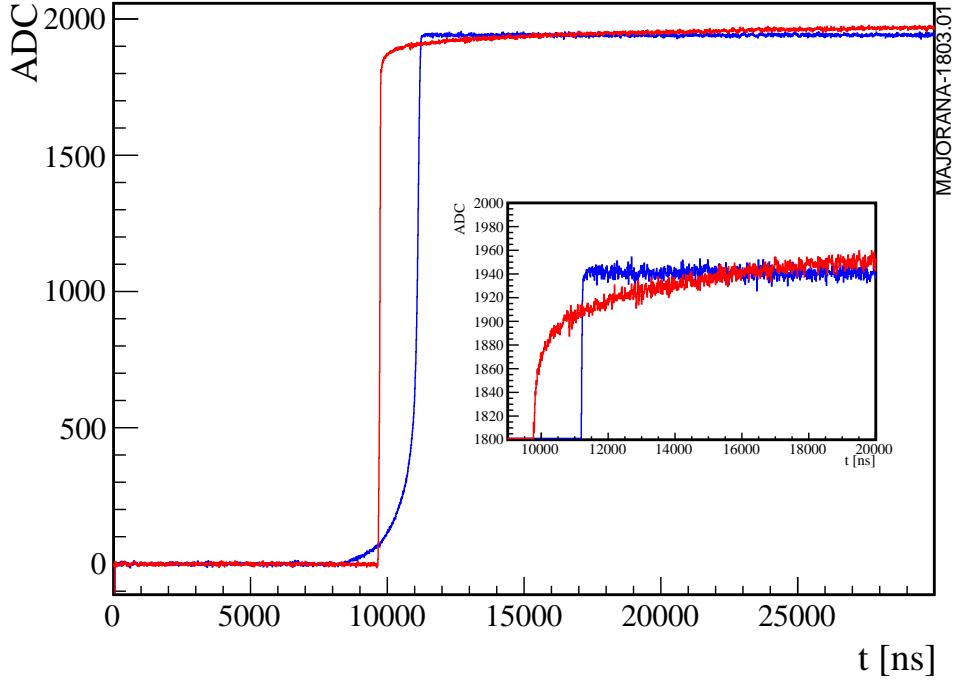


Figure 3.2: An alpha waveform (red) compared to a bulk waveform (blue) in the MJD PPC detector. The DCR effect results in a slow charge collection after the fast rise of the signal, as shown by the insert plot in the box. [40]

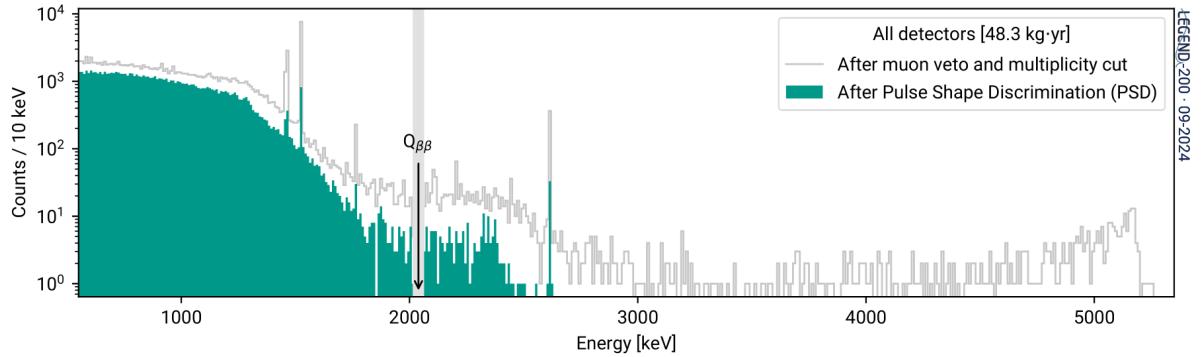


Figure 3.3: LEGEND-200 physics spectrum. The events above 3000 keV after the muon and multiplicity cut are primarily alphas. These events can be effectively removed using PSD cuts as shown in green. Credit: LEGEND collaboration

The charges on the passivated surface move at slower speeds than in the detector bulk. This is mainly due to additional scattering mechanisms and modified band structure on the surface, described in [43]. In bulk, the hole and electron transport velocities can be approximated using a simple warped-band approximation. However, on the passivated surface, the band structure is altered by electrical fields

normal to that on the surface. This leads to a higher probability of scattering of charges and surface-related effects such as roughness scattering. Simulations in Ref. [43] estimate that the drift speed on the surface is 40 to 50 times slower than in the bulk.

Alpha events create a highly localized, dense charge cloud close to the surface. Because the charge that ends up on the surface drifts slower than the bulk, it could get stretched into a non-spherical cloud. Thus, spherical charge cloud approximation cannot be used for such events.

### 3.5 Surface Charge Effect

In operation, the HPGe detectors can accumulate static charge on the passivated surface, which affects the charge collection for surface events. The sign and magnitude of the accumulated charges can vary depending on the operating conditions. Figure 3.4 shows how the presence of a surface charge alters the electric field close to the passivated surface. When there is no surface charge (top), the field lines are parallel to the surface. The presence of a negative surface charge (middle) attracts the field lines towards the surface. The positive surface charge (bottom) repels the field lines from the surface. The negative surface charge can pull the holes to the surface, where they can drift slowly. Similarly, positive surface charges can pull the electrons to the surface.

Since the surface charge changes the overall electric field of the detector, it could also alter the depletion voltage for the detector. Figure 3.5 shows how the depletion voltage of a LEGEND PPC detector changes with surface charge. Typically, the detector's operational voltage is higher than the depletion voltage, but if the surface charges are not properly accounted for, a detector could become undepleted after being deployed in the experiment.

Pulse shape simulations can help model these effects accurately and build a background model for surface events. A simulation to model surface events should allow for a non-spherical charge cloud while incorporating surface drift, diffusion, and self-repulsion. It should also properly account for surface charge effects. Next, we look at the current pulse-shape simulation frameworks that are being used by LEGEND.

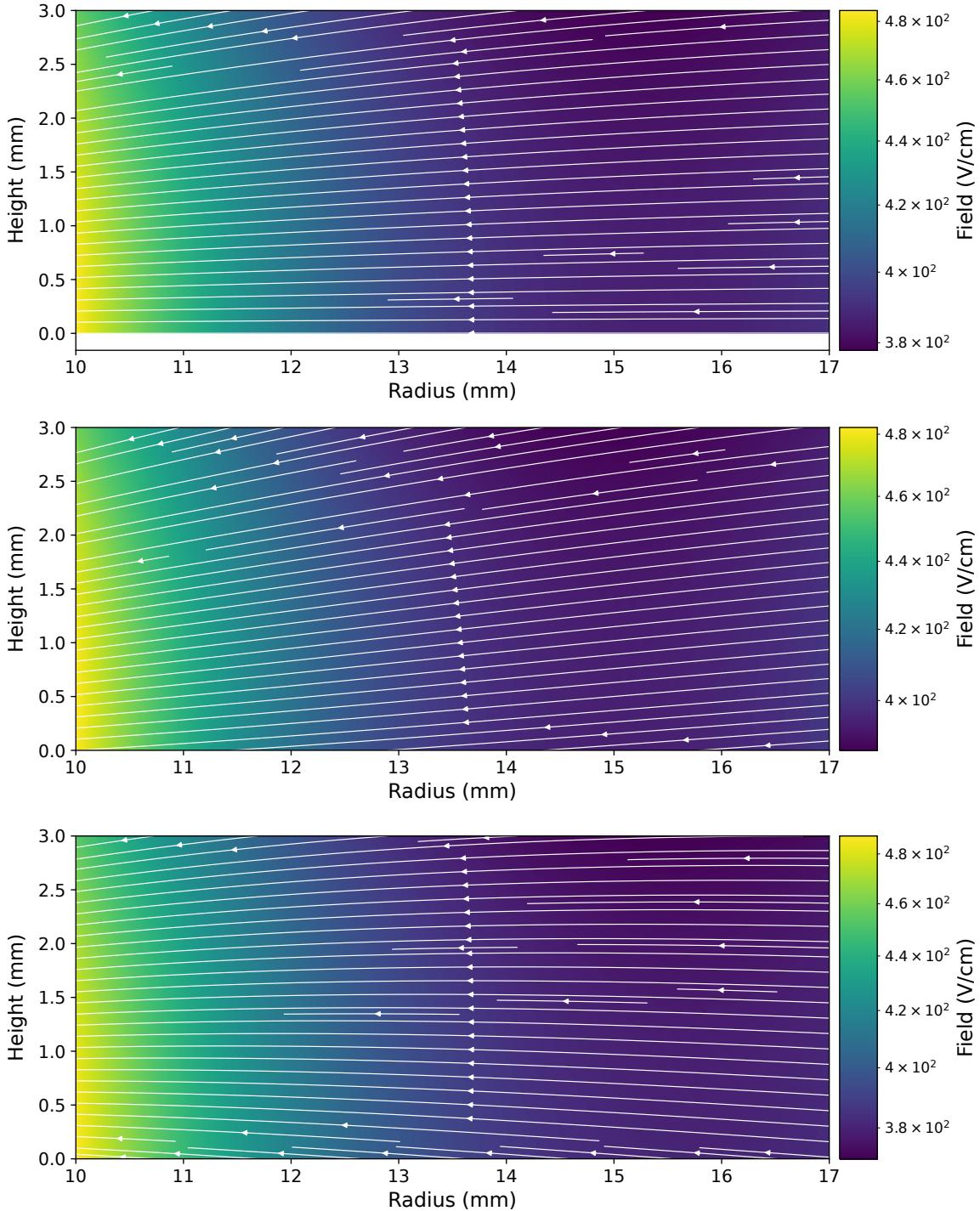


Figure 3.4: Electric field magnitude and field lines in a region near the passivated surface of the PONaMa-1 PPC detector with different surface charges. The white arrows shows the direction of electric field lines. Geometry information for the detector can be found in table 5.1. Charges on the passivated surface can pull electrons and holes to the surface and cause energy loss and slow charge collection. Top figure has no surface charge and the field lines are parallel to surface. Middle figure has a surface charge of  $-0.5 \times 10^{-10} \text{ e}/\text{cm}^2$  which pulls the filed lines to the surface. Bottom figure has surface charge of  $0.5 \times 10^{-10} \text{ e}/\text{cm}^2$  that repels the field lines form the surface.

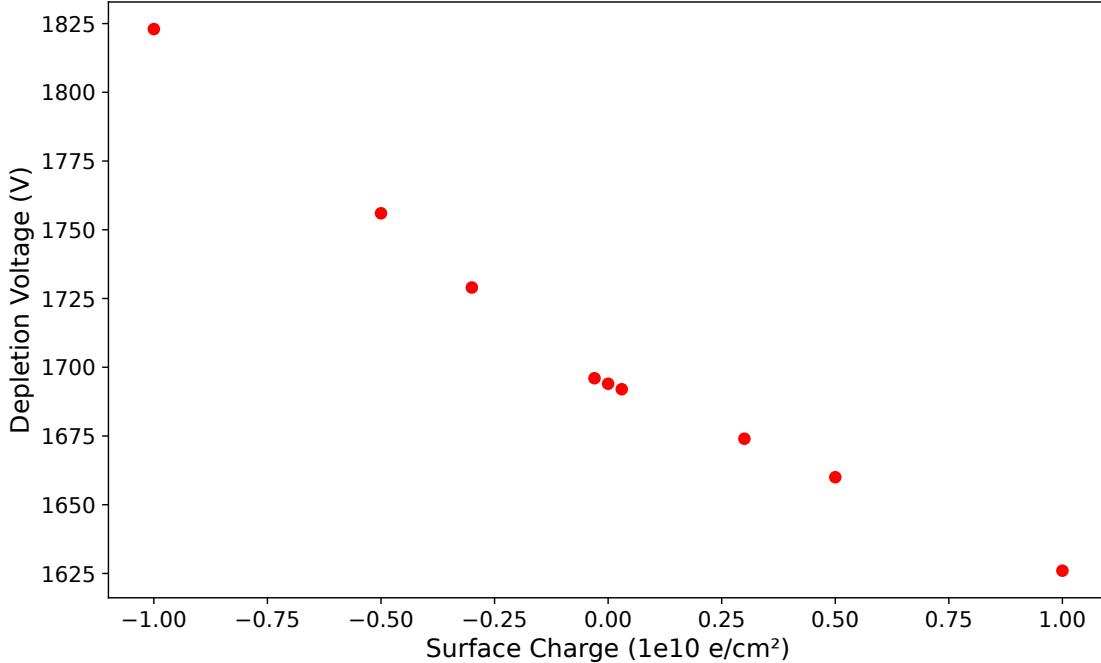


Figure 3.5: Effect of surface charge on depletion voltage for LEGEND PPC detector P00661C. Depletion voltages calculated using `siggen` software. The nominal depletion voltage is 1694 V at zero surface charge. Surface charge can alter the depletion voltage and potentially cause a detector to become undepleted after being deployed.

### 3.6 Current Status in Pulse Shape Simulations

Signal formation in HPGe detectors is modeled using the Shockley-Ramo theorem. This enables the building of accurate simulations that can model the waveforms. The LEGEND Collaboration uses two HPGe signal modeling software packages, both developed in part by members of the Collaboration: `siggen` and `SSD` simulations.

`siggen` is a C-based program developed by David Radford at Oak Ridge National Laboratory [44]. It consists of two components: `fieldgen` and `siggen`. `fieldgen` is used to calculate the electric potential and weighting potential of point-contact detectors in two dimensions. `siggen` then uses the weighting potential calculated from `fieldgen` to generate signals for a charge originating at a given point in the detector using the Shockley–Ramo theorem. `fieldgen` simulations can be used to calculate the depletion voltage, the volume of the depletion region, and the capacitance of the detector. Diffusion in `siggen` is approximated using a Gaussian convolution, and there is no mechanism to model nonspherical charge clouds since `siggen` uses point charges to represent the entire charge cloud.

`siggen` simulations have been crucial in the design and manufacturing of the LEGEND detectors, as they allow precise modeling of electric fields and projected PSD performance.

`SolidStateDetectors.jl` (SSD) is a software package developed by the Abt group at the Max Planck Institute in Munich [45]. It is written in Julia and can perform calculations in full 3-dimensions. SSD can calculate electric fields and potentials outside the detectors. The SSD enables full 3-D diffusion and models self-repulsion in the signal. It does this by using a limited number of charges, each of which represents many real charges. Charges are individually tracked using their position and velocity. The electric field calculation can be performed on GPUs.

### 3.7 EH-Drift

`EH-Drift` is a newly developed method to simulate surface events while directly simulating diffusion and self-repulsion. It is a specialized model designed to simulate passivated surface interactions. The model uses 2-D approximations to optimize run-time while adjusting for 3-D effects. The original software version was developed in C by David Radford and built as an extension to `siggen`. As charges drift through the detectors, the software keeps track of charge densities at a pixel-by-pixel level, which allows for nonspherical charge clouds and for extremely large variations in charge density at different positions in the detector. The charges that end up on the surface have their velocities reduced by a predetermined factor, which is set based on dedicated computational modeling of passivated surfaces in HPGe detectors. `EH-Drift` also enables simulation of the effects of surface charges. The simulation is performed in the  $r$  and  $z$  directions while assuming  $\phi$  symmetry; therefore, the charge cloud is simulated by a ring of charges, and the calculations account for the 3-D effects using analytic approximations. This approach does not take into account the varying mobilities as a function of the crystal axis because this is a small effect for events dominated by surface drift.

#### 3.7.1 Charge Transport Equation

To model the time evolution of electrons and holes in a detector, we used the continuity equation added together with drift and diffusion terms. Let  $\rho(r, t)$  be the charge density of one carrier (electron or hole) at position  $r$  and time  $t$ . The total current density,  $J$ , has two contributions:

- Diffusion current  $-D\nabla\rho$ , where  $D$  is the diffusion coefficient,

- Drift current  $\mu \rho E$ , where  $\mu$  is the mobility and  $E$  is the electric field.

Combining these into the continuity equation for charge conservation gives:

$$\frac{\partial \rho(r, t)}{\partial t} = \nabla \cdot (D \nabla \rho(r, t)) - \nabla \cdot (\mu \rho(r, t) E(r, t)) \quad (3.3)$$

The first term on the right-hand side describes how thermal diffusion spreads the charge carriers from regions of higher density to lower density. The second term captures the drift of charges under the influence of the electric field. To solve this numerically in discrete time steps, we write  $t^{n+1} = t^n + \Delta t$ . Using a first-order approximation, we get:

$$\rho^{n+1}(r) = \rho^n(r) + \Delta t \left[ \nabla \cdot (D \nabla \rho^n(r)) - \nabla \cdot (\mu \rho^n(r) E^n(r)) \right] \quad (3.4)$$

where  $\rho^n(r)$  and  $E^n(r)$  denote the density and electric field at the  $n$ -th time step, respectively. After discretization, we update both the drift and diffusion terms in a single time step  $\Delta t$ . Although drift and diffusion have different characteristic timescales, we adopt the one that is more stringent. Typically, drift in an electric field changes more quickly. The time step is thus set using this assumption by calculating the Courant number, discussed in Section 3.7.7. To solve this equation, EH-Drift takes the Lagrangian splitting approach. This means that each step, such as diffusion and charge drift, is solved independently, and then the net effect is calculated by combining them. Figure 3.7 shows how the EH-Drift program works.

### 3.7.2 Initial Setup

During the initial setup, the detector is divided into a grid. The coordinates of this grid are shown in figure 3.6.

Each point on the grid has a value for the potential, weighting potential, electron and hole densities, and impurity. The boundary conditions are then set according to the detector's input geometry, surface charge, and bias voltage. Next, the electric potential and weighting potential are calculated using an over-relaxation algorithm. The capacitance and depletion are estimated to understand the detector's

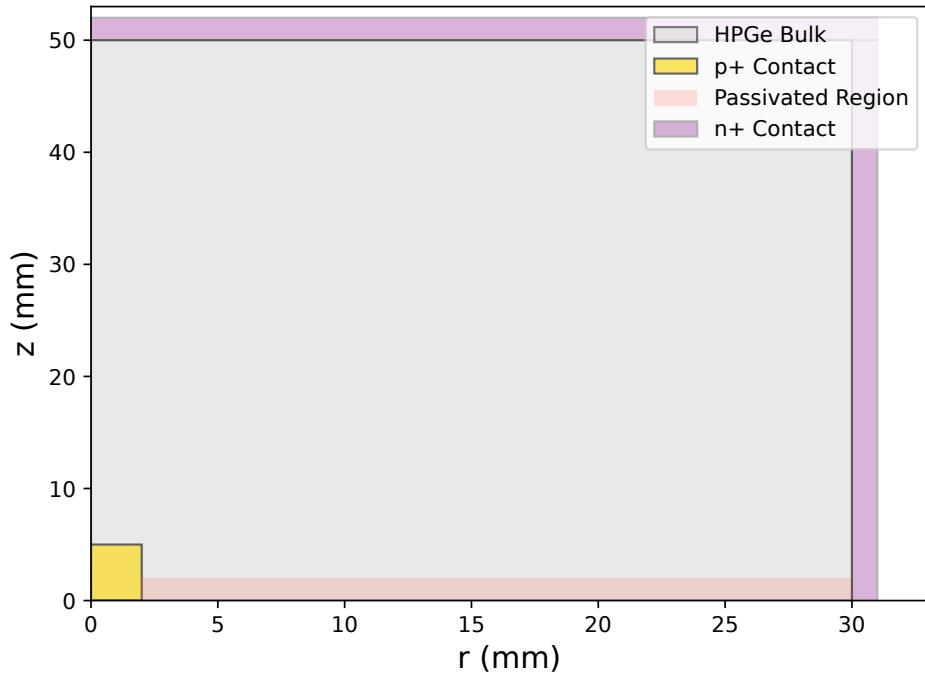


Figure 3.6: Diagram showing a simple setup and coordinates in EH-Drift for a PPC detector. The n<sup>+</sup> electrode is kept at high voltage while p<sup>+</sup> is kept at zero. Passivated region is the region in between the two electrodes.

behavior. The capacitance is calculated by relating two equations for the energy stored in an electric field:

$$\frac{1}{2}CV^2 = \frac{1}{2}\epsilon \int E^2 dV \quad (3.5)$$

$$C = \frac{\epsilon \int E^2 dV}{V^2} \quad (3.6)$$

where C is the capacitance of the detector,  $\epsilon$  is the permittivity of the medium, E is the electric field magnitude, and V is the bias voltage. The depletion voltage is calculated by finding the minimum bias required to fully deplete the detector. Then the software checks to find local minima in the potential to help identify any pinch-off depletion region. The initial charge densities are distributed at a given location on the basis of the position, usually over two adjacent grid points. The initial densities are determined using the following:

$$\rho_E \left( 10^{10} \frac{\text{pairs}}{\text{cm}^3} \right) = \rho_H = \frac{E(\text{keV}) \times (1 \text{ keV} = 1000 \text{ eV}) \times 10^{-10}}{(3 \text{ eV/pair}) \times (10^{-3} \text{ cm}^3 = 1 \text{ mm}^3) \times [\text{grid (mm)}]^3} \quad (3.7)$$

where  $\rho$  represents the density of the charge (electrons or holes) in units of  $10^{10}/\text{cm}^3$ . The energy deposited  $E$  is the energy deposited in keV, which is first converted to eV. The factor of  $10^{-10}$  accounts for the final units. The approximate energy required to generate an electron-hole pair in Germanium is 3 eV. Finally, dividing by  $\text{grid}^3$  in mm normalizes the number of charge pairs generated by the volume of the grid. The densities for both holes and electrons are deposited equally in points  $(z,r)$  and  $(z+1,r)$ . The density is then multiplied by a calibration factor that is tuned such that the 2-D simulation of the self-repulsion effect matches the 3-D model. The calibration factor we found in dedicated studies was a factor of 20 for all detector geometries.

After the initial setup, the program steps through a series of time steps. The charges are allowed to diffuse and then drift in the electric field created by the combination of the intrinsic detector field and the self-repulsion of charges. The electric field is updated at each time step by solving Poisson's equation with the net charge density, including both fixed impurities and free charges, to account for effects like self-repulsion. The calculation of the electric field is discussed in Section 4.2. In the following sections, we discuss the rest of the critical components of EH-Drift.

### 3.7.3 Diffusion

Diffusion arises from the random thermal motion of electrons and holes, which causes them to spread out from regions of high concentration into regions of lower concentration. In thermal equilibrium, the diffusion coefficient  $D$  and the mobility  $\mu$  are related by the Einstein relation:

$$D = \mu \frac{k_B T}{q}, \quad (3.8)$$

where  $k_B$  is Boltzmann's constant,  $T$  is the absolute temperature, and  $q$  is the elementary charge. Thus, if the mobility  $\mu$  is known or has been fitted to experimental data,  $D$  can be determined for a specified temperature. In germanium at cryogenic temperatures (e.g.  $\sim 77$  K), the mobilities are sufficiently large

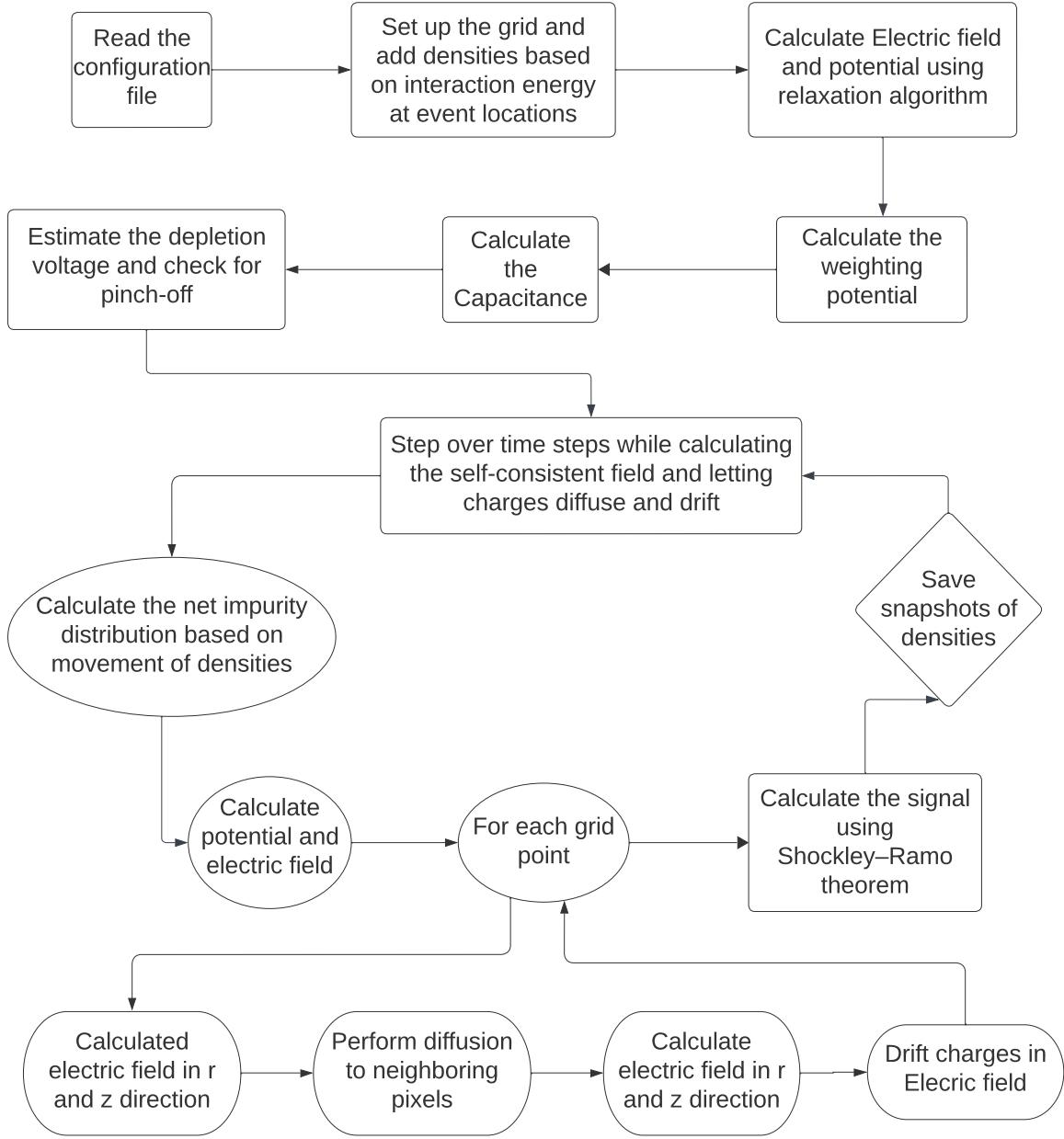


Figure 3.7: A flow chart illustration the EH-Drift steps. The initial steps to calculate intrinsic parameters of the detector (e.g. weighting potential, capacitance, etc.) are performed once. Then subsequent loops step through each time step and each grid point to track the movement of charge clouds.

that diffusion can still have a noticeable effect on the final spread of a charge cloud, although it is smaller than under room temperature conditions.

In EH-Drift the diffusion is simulated by redistributing the charge among neighboring grid cells in the ( $r$ ) and ( $z$ ) directions using the diffusion coefficients. During each time step  $\Delta t$ , a fraction of the carriers in the cell  $(z, r)$  diffuse to adjacent cells  $(z \pm 1, r \pm 1)$  according to two quantities:  $\delta_z$  and  $\delta_r$ .

These fractions are computed from an approximate diffusion parameter  $f$  and the local drift velocity magnitudes  $|\mathbf{v}_z|$ ,  $|\mathbf{v}_r|$  as follows:

$$\delta_z = \frac{\Delta x \ v_z f}{E_z} \quad (3.9)$$

$$\delta_r = \frac{\Delta x \ v_r f}{E_r} \quad (3.10)$$

where  $\Delta x$  is the grid size,  $v_z$  and  $v_r$  are the drift velocities,  $E_z$  and  $E_r$  are the local field components, and  $f$  incorporates the diffusion coefficient  $D$  and corrections for grid sizes. If  $E_z$  or  $E_r$  is below 1 V/cm, the program sets  $\delta_z = 0$  or  $\delta_r = 0$  assuming that the low field regions produce negligible drift velocity.

### Volume Correction

In cylindrical coordinates  $(r, z)$ , each radial ring has a different physical area. The simulation tracks these scaling factors with arrays:

$$s1[r] = 1 + \frac{0.5}{r - 1}, \quad s2[r] = 1 - \frac{0.5}{r - 1},$$

For  $r=1$  and  $r=0$ , the code uses a special fixed weighting factor to handle the geometry. These scalars adjust the amount of charge that is diffusing into or out of the neighboring ring. Specifically, if  $\delta_r$  is the fraction of carriers leaving cell  $(r)$  in the  $+r$  direction, the actual increase in the neighbor cell  $(r + 1)$  is multiplied by  $s1[r]$ , while the old cell is reduced by the same fraction multiplied by  $s1[r]$  to balance volume differences. Similarly, diffusion to cell  $(r - 1)$  uses  $s2[r]$ . The same volume correction is used in the field calculation.

## Diffusion Density Update

After computing the diffusion fractions  $\delta_z$  and  $\delta_r$ , the code subtracts those amounts from the point's density and adds them to the four neighbors:

$$\rho^{\text{new}}(z, r+1) += \rho^{\text{old}}(z, r) \times \delta_r \times s_1(r) \times \frac{r-1}{r} \quad (3.11)$$

$$\rho^{\text{new}}(z-1, r) += \rho^{\text{old}}(z, r) \times \delta_z \quad (3.12)$$

$$\rho^{\text{new}}(z+1, r) += \rho^{\text{old}}(z, r) \times \delta_z \quad (3.13)$$

$$\rho^{\text{new}}(z, r-1) += \rho^{\text{old}}(z, r) \times \delta_r \times s_2(r) \times \frac{r-1}{r-2}. \quad (3.14)$$

The offsets  $r$  and  $r - 2$  are due to a one-based indexing in the code.

### 3.7.4 Drift in Electric Field

Germanium has a diamond-cubic lattice structure, where each lattice point hosts a two-atom basis. This is shown in 3.8. The crystal axes are shown along with the three principal crystallographic basis axes  $\langle 100 \rangle$ ,  $\langle 110 \rangle$ , and  $\langle 111 \rangle$ . Charges in the three basis axes travel at different velocities. Mobility relates the velocity of the charges to the electric field. At low fields, the relationship is Ohmic:

$$\vec{v} = \mu_0 \vec{E} \quad (3.15)$$

At higher fields, the scattering with the crystal lattice results in a nonlinear relationship. This scattering causes the velocity to saturate at higher values. The relationship in a higher field can be modeled using [48]:

$$v(E) = \frac{\mu_0 E}{(1 + (E/E_0)^\beta)^{1/\beta}} \quad (3.16)$$

$\beta$  is the empirical parameter that shapes the curve and typically falls between 0.2 and 2.

In EH-Drift, we calculate the electric field in the  $r$  and  $z$  directions at each point of the grid. We use values from [49], which gives  $v(E)$  from the local electric field  $E$ . We use a piecewise linear interpolation to find  $v(E)$ . To do this, we define an array of electric field points, `drift_E`, which partitions the field range into intervals. For each interval  $[E_i, E_{i+1}]$ , we store two quantities: a drift offset,

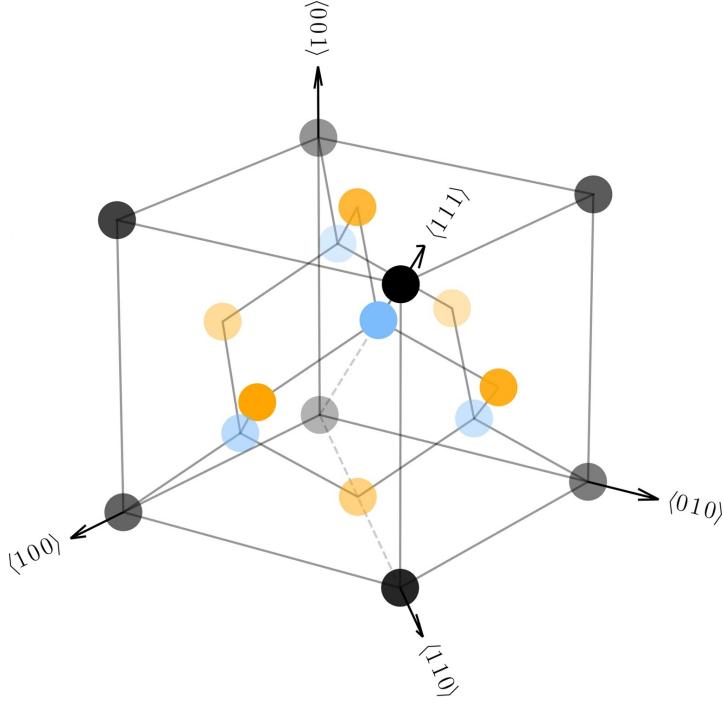


Figure 3.8: Germanium crystal unit cell, showing the crystal axis and the crystallographic basis axes. The atoms on the corners, faces and inside the cube are shown in black, orange and blue circles respectively.  $\langle 001 \rangle$ ,  $\langle 010 \rangle$ , and  $\langle 100 \rangle$  are the axes of Ge crystal unit cell. And  $\langle 100 \rangle$ ,  $\langle 110 \rangle$ , and  $\langle 111 \rangle$  are the crystallographic basis axes. The crystallographic basis axes are the direction used to describe physical properties such as mobility and energy band structure. Source: [46, 47]

corresponding to the drift velocity at  $E = E_i$ , and a drift slope, which is the rate of change of the drift velocity with respect to the electric field in that interval. The drift velocity is then computed by:

$$v(E) = \text{drift offset}[i] + \text{drift slope}[i] (E - E_i), \quad (3.17)$$

with  $E_i \leq E < E_{i+1}$ . Figure 3.9 shows the relation between the drift velocity and the electric field. The drift velocity is then multiplied by the time step to find where the charges will drift, and then the charges are moved to the new location. For simplicity, we only used the  $\langle 100 \rangle$  direction velocities in the current model. As shown in figure 3.9, for a given electric field magnitude, the variation in mobility with the crystal axis is expected to be 20 – 40%, while the surface drift effects we are studying originate from drift velocities that are up to 1000 times slower.

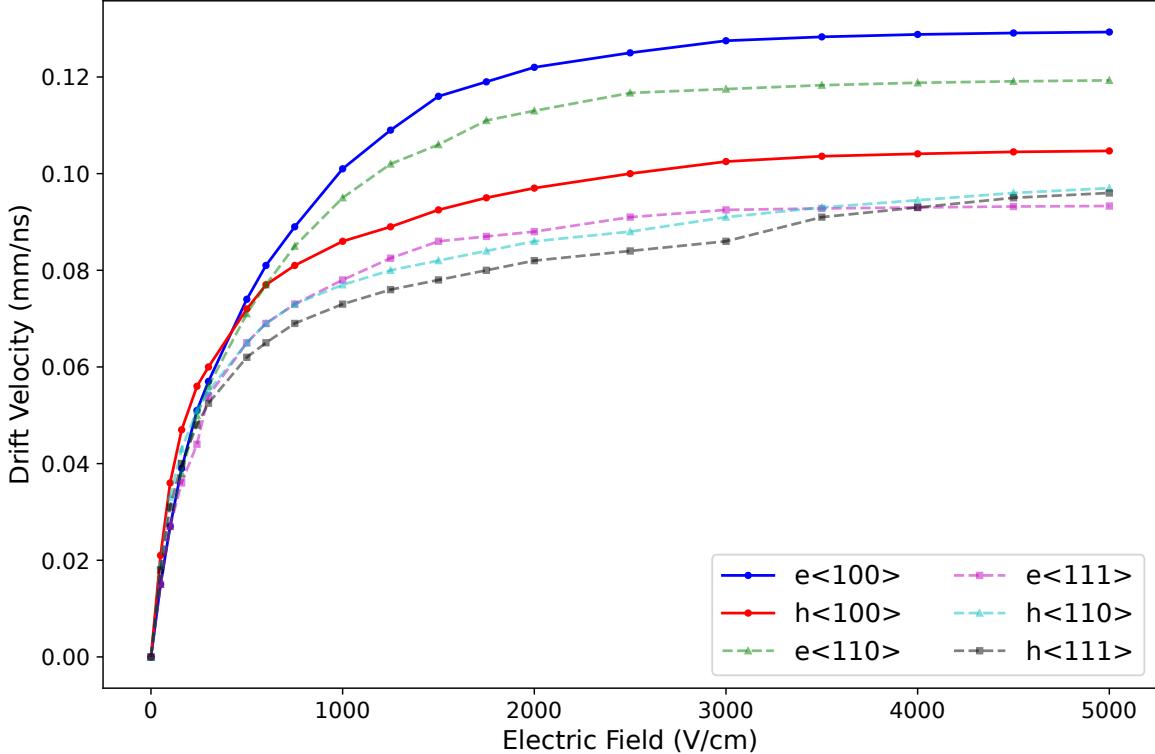


Figure 3.9: Relation of drift velocity versus electric field in EH-Drift. The points shows the experimental values adapted from [49]. To show how EH-Drift estimates the values we create 500 samples between 0 and 5000 V/cm and performed the piecewise linear interpolation between the points. Only  $<100>$  direction velocities are used in the current model.

### Fraction Splitting

We use a splitting approach to move charges to the new cell. Since we are using a grid, we want to account for the discretization by splitting the charges between two grid points. Suppose that after a time step  $\Delta t$ , the density in the grid cell  $(z, r)$  moves  $\Delta z$  and  $\Delta r$  to a new position  $(k, i)$ . We use the following equation to define  $(k, i)$  and fractional parts  $f_z$  and  $f_r$ :

$$k = z + \Delta z, \quad f_z = \lceil \Delta z \rceil - \Delta z \quad (3.18)$$

$$i = r + \Delta r, \quad f_r = \lceil \Delta r \rceil - \Delta r \quad (3.19)$$

where  $\lceil \Delta z \rceil$  means the floor of  $\Delta z$ . The charges are split using these fractions:

$$\text{fraction in } (k, i) = f_r \times f_z \quad (3.20)$$

$$\text{fraction in } (k, i + 1) = (1 - f_r) \times f_z \quad (3.21)$$

$$\text{fraction in } (k + 1, i) = f_r \times (1 - f_z) \quad (3.22)$$

$$\text{fraction in } (k + 1, i + 1) = (1 - f_r) \times (1 - f_z). \quad (3.23)$$

Finally, the densities are updated using:

$$\rho^{\text{new}}(k, i) += \rho^{\text{old}}(z, r) \times [f_r \times f_z] \times G_{r,z} \quad (3.24)$$

$$\rho^{\text{new}}(k, i + 1) += \rho^{\text{old}}(z, r) \times [(1 - f_r) f_z] \times H_{r,z} \quad (3.25)$$

$$\rho^{\text{new}}(k + 1, i) += \rho^{\text{old}}(z, r) \times [f_r (1 - f_z)] \times G_{r,z} \quad (3.26)$$

$$\rho^{\text{new}}(k + 1, i + 1) += \rho^{\text{old}}(z, r) \times [(1 - f_r) (1 - f_z)] \times H_{r,z}. \quad (3.27)$$

### 3.7.5 Geometric Factors

In cylindrical coordinates, each cell at index  $r$  corresponds to an annular region of approximate circumference  $2\pi(r \Delta r)$  and thickness  $\Delta z$ . When the charge moves from cell  $(z, r)$  to cell  $(k, i)$ , we apply the factors:

$$G_{r,z} \equiv \frac{(r - 1)}{(i - 1)} \quad H_{r,z} \equiv \frac{(r - 1)}{(i)}$$

reflecting the difference in volumes at radii  $r$  and  $i$ . The offsets  $(r - 1)$  and  $(i - 1)$  are for one-based indexing in the code. For points near  $r = 0$  or  $i = 0$ , the code uses a modified factor that preserves volume scaling.

### 3.7.6 Surface Drift

When charge carriers drift within the detector, they may encounter the passivated surface, which is handled separately. To model the surface, the lowest set of grid points is split into two regions: one modeling the passivated surface of a fixed depth (set by the user and taken by default to be  $2\mu$ ), and one modeling the remaining bulk region, of grid size minus the surface depth. We store the charges on the surface in a special row.

If the result of the charge drift calculation places a set of charges below the bottom of the detector, they are fully added to the surface row. If the charge enters the passivated layer, we split it up using the fraction splitting used in charge drift. Performing the diffusion on the last grid point in equation 3.12, the bottom z-point is considered to be the passivated surface.

We employ the same methods described above for drifting and diffusing the charges present on the surface, but the drift and diffusion are suppressed by the surface drift velocity factor. This factor can be set by the user in a configuration file; it is set to 0.001 by default. It represents the ratio of speed on the surface to that in the bulk and is derived from experimental measurements or computational models of surface drifts. The charges from the surface can drift to the bulk or to other points on the surface. Points on the surface can diffuse to other points on the surface.

### 3.7.7 Courant Number and Adaptive Time Step

The Courant–Friedrichs–Lowy (CFL) condition is a requirement for numerical solutions of partial differential equations involving moving particles [50]. It is commonly used in fluid dynamics to ensure that particles travel only to adjacent grid points. We introduce the CFL condition in the EH–Drift to ensure that the time step is well matched to the velocities of the charges. This helps maintain consistency for different input grid sizes. During each simulation time step, we calculate the local Courant number at every point on the grid using the following.

$$C(z, r) = \max \left( \frac{v_r \Delta t}{\Delta x}, \frac{v_z \Delta t}{\Delta x} \right), \quad (3.28)$$

where  $v_r$  and  $v_z$  are the drift velocities in r and z directions, respectively, and  $\Delta x$  is the grid spacing. The time step is updated using the largest Courant number on the grid :

$$\Delta t = \frac{1}{\max(C(z, r))}. \quad (3.29)$$

The use of adaptive time-stepping based on the Courant number minimizes errors from the discretization process while reducing unnecessary computation during periods of the simulation when only slow-moving surface charges are present.

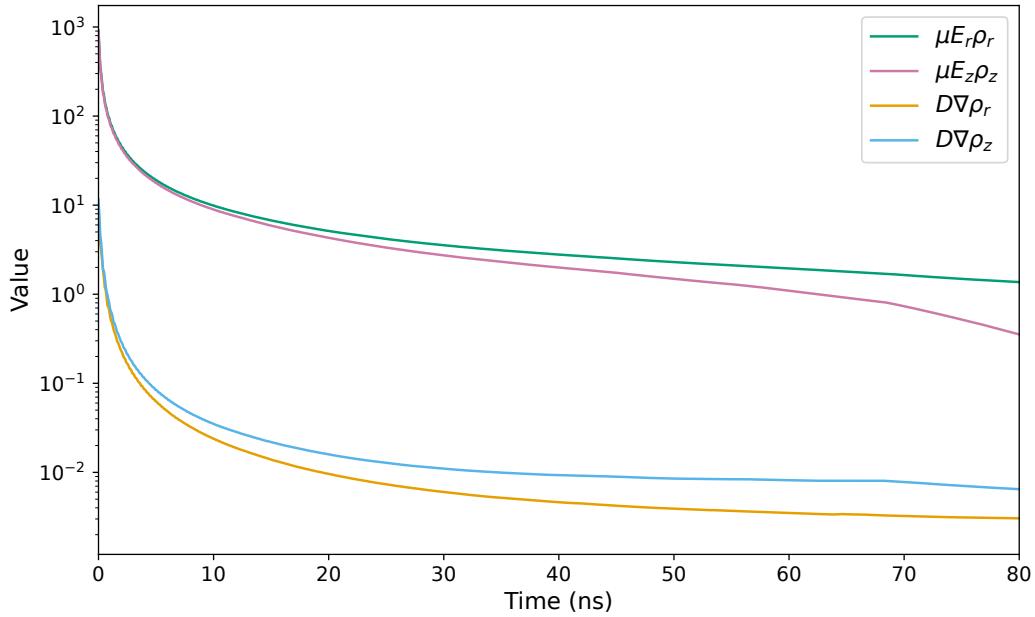


Figure 3.10: Maximum values of terms in the charge transport equation 3.3 at a given time step for an event in EH-Drift. The drift terms are about two orders of magnitude large than the diffusion terms throughout the time evolution of the charge cloud. This justifies the use of drift velocity to define the timestep used for Courant number calculation.

During this calculation, we make an inherent assumption that the drift component in equation 3.4 has the larger contribution, and thus use it to set the time step. This can be verified by plotting the drift terms ( $\mu E_r \rho_r$  and  $\mu E_z \rho_z$ ) against the diffusion terms ( $D \nabla \rho_r$  and  $D \nabla \rho_z$ ) as shown in figure 3.10. The drift components are an order of magnitude larger than the diffusion components throughout the simulations, which justifies the assumption.

### 3.7.8 Impurity Correction

Surface events occur in a detector region with low electric fields, and high-energy alpha events, in particular, leave densely ionized tracks. The high density of charge carriers creates a significant local electric field that must be taken into account to accurately calculate the electric potential. At a given time, the impurity correction is given by:

$$I_t(z, r) = I_0(z, r) + (\rho_h(z, r) - \rho_e(z, r)) \times \frac{e}{\epsilon} \times \frac{(\Delta x)^2}{2}. \quad (3.30)$$

The  $I_0$  is the bulk impurity level of the crystal, measured at production.  $\rho_h(z, r)$  and  $\rho_e(z, r)$  are the hole and electron densities, respectively.  $\frac{e}{\epsilon}$  is the conversion factor that relates the charge density to the resulting electric field, derived from Gauss's law in Germanium with  $\epsilon = 16\epsilon_0$ .  $\frac{e}{\epsilon}$  is 11.310 in charge units of  $10^{10} \frac{e}{cm^3}$ .  $(\Delta x)^2$  is the area of the grid that converts the density to charge, and then the conversion factor converts it into  $10^{10} \frac{e}{cm^3}$  units to match the units of impurity used.

### 3.7.9 Signal Calculation

In cylindrical coordinates, each grid cell with radius  $r$  and height  $z$  has an area proportional to  $(r - 1)$ . Once we read the electron density  $\rho_e(r, z)$  and hole density  $\rho_h(r, z)$  at time  $t$ , the weighted sums are defined as:

$$S_e(t) = \sum_{z=1}^{L-1} \sum_{r=1}^{R-1} \rho_e(r, z) (r - 1) \text{wpot}[r - 1][z - 1] \quad (3.31)$$

$$S_h(t) = \sum_{z=1}^{L-1} \sum_{r=1}^{R-1} \rho_h(r, z) (r - 1) \text{wpot}[r - 1][z - 1], \quad (3.32)$$

where  $\text{wpot}[r - 1][z - 1]$  is the weighting potential at that point in the grid. These  $S_e(t)$  and  $S_h(t)$  represent the induced signal contributions of electrons and holes, respectively, using the Shockley-Ramo theorem.

The initial unweighted sums are defined as:

$$R_e(0) = \sum_{z=1}^{L-1} \sum_{r=1}^{R-1} \rho_e(r, z) (r - 1), \quad (3.33)$$

$$R_h(0) = \sum_{z=1}^{L-1} \sum_{r=1}^{R-1} \rho_h(r, z) (r - 1), \quad (3.34)$$

Combining these quantities, the induced signal at time  $t$  is computed using:

$$\text{Signal}[t] = \frac{S_h(t) - S_h(0)}{R_h(0)} - \frac{S_e(t) - S_e(0)}{R_e(0)} \quad (3.35)$$

Since the electrons' contribution would be negative, we subtract it from holes to get the net induced signal. Normalization by  $R_e(0)$  and  $R_h(0)$  ensures that the signal is expressed as a fraction of the original total electron/hole count, so it starts near zero at  $t = 0$  and approaches 1 as all charges arrive at the contacts.

### 3.7.10 Density Snapshots

The ability to store density snapshots at each time step provides detailed information about how the charges drift in the detector. Figure 3.11 shows a snapshot at time=80ns in an EH-Drift simulation for a 5 MeV energy deposition close to the surface without any surface charge. The event started at  $r=15$  mm. The electrons drift towards the  $n^+$  contact, which is located to the right at  $r = 34.7$ mm, and the holes drift towards the  $p^+$  contact, located to the left starting at  $r = 1.5$ mm. A 5 MeV energy deposition can create a charge cloud 1.5 mm high. Charges in such a large charge cloud drift and diffuse to the surface, even in the absence of surface charges. The projected density (bottom panel of figure 3.11) shows the density of the charge cloud as a function of radius. The peak of each distribution is created by the charges traveling normally through the bulk of the detector. The broad tail of the projected density has an additional secondary peak, seen at  $r = 15$  mm, which is due to the charges that were initially pushed to the surface by self-repulsion when the charge clouds were at their highest density. The points in between the two peaks are charges that drifted or diffused to the surface in intermediate steps.

Figure 3.12 shows the same event at the same time, but with a negative charge on the surface. The negative charges alter the field in such a way that holes are pulled onto the surface and electrons are repelled. Thus, the holes have a significant fraction of their charge attracted to the surface, as seen in the projected density plot. Electrons, being repelled from the surface, have no slow-moving component.

Similarly, Figure 3.13 shows the same event but with a positive charge on the surface. In this case, the electrons are pulled onto the surface and drift slowly, while the holes are repelled and do not have surface drift.

### 3.7.11 Model Parameters

In EH-Drift, certain inputs can be adjusted to match the observed data. These parameters fall into three categories: event-to-event quantities, detector-level constants, and operating conditions.

Event-to-event quantities such as the initial energy and event position are known or fixed by the source configuration, such as a mono-energetic alpha beam or a background location from Monte Carlo

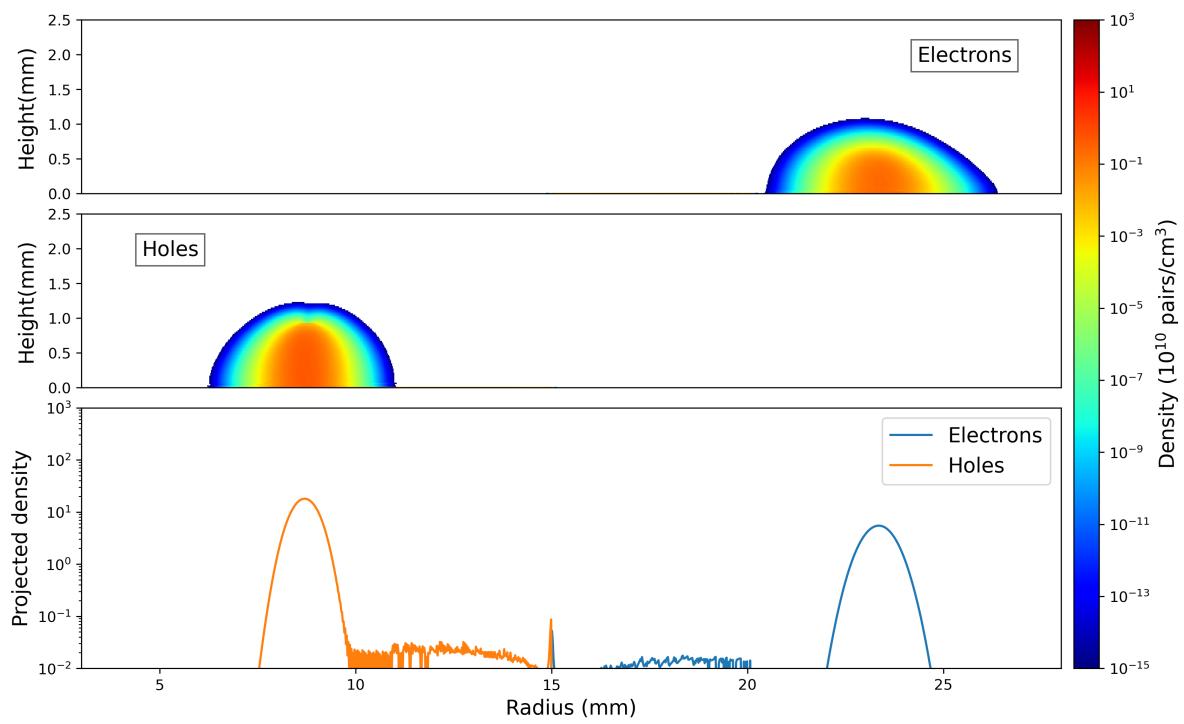


Figure 3.11: Drift of electron and hole charge clouds in EH-Drift with zero surface charge. The projected densities show how the charges are distributed along the radius. The densities have two peaks, one due to the fast moving component in bulk and another due to the slow moving component on the passivated surface.

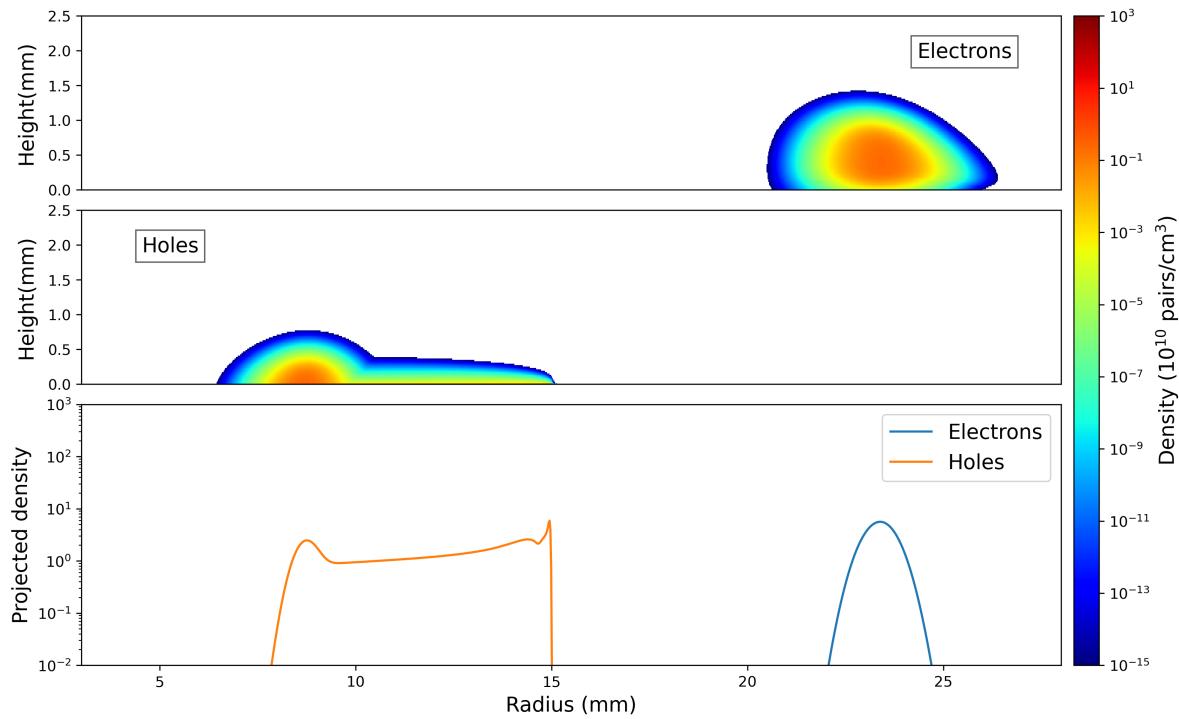


Figure 3.12: Drift of electron and hole charge clouds in EH-Drift with negative surface charge. The negative surface charge pulls the holes onto the surface, which then move at a slower speed.

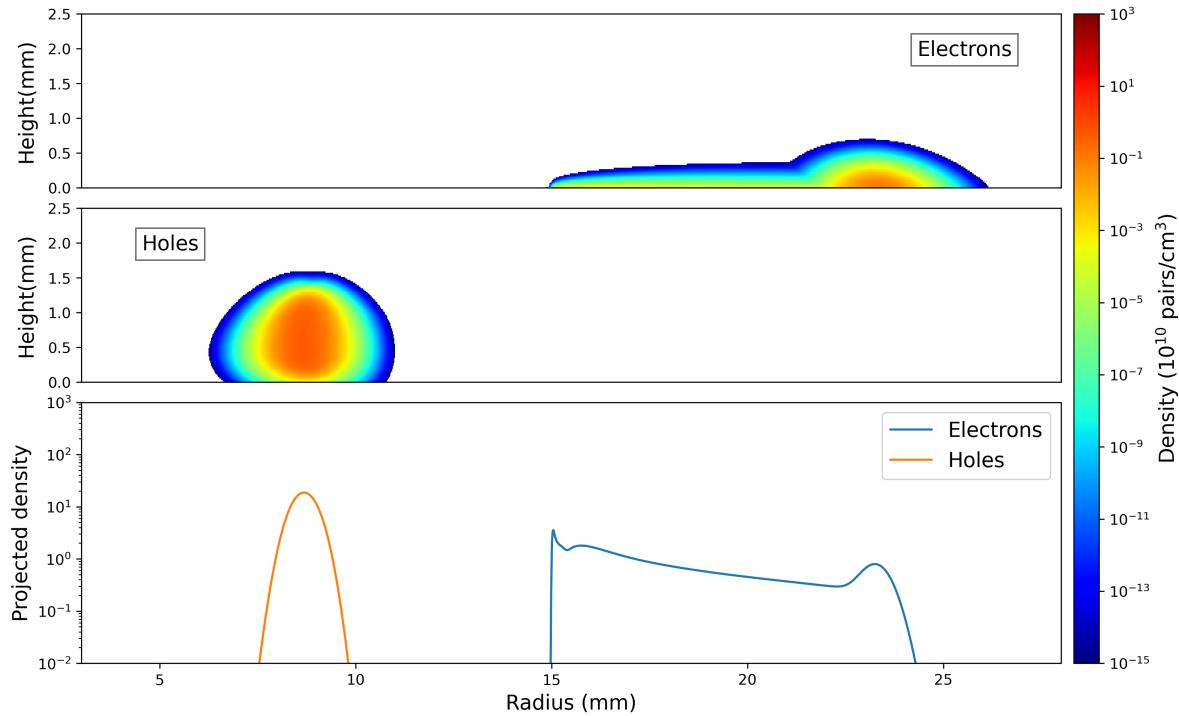


Figure 3.13: Drift of electron and hole charge clouds in EH-Drift with positive surface charge. The positive surface charge pulls the electrons onto the surface which move at a slower speed.

simulations. These would not usually vary to match data, but rather set based on the known test stand or source energy.

Detector-level constants, such as passivated surface thickness or the ratio of surface-to-bulk drift speed, are physically determined by the detector fabrication and inherent surface conditions. Although these remain constant for a single detector, they may not be well measured and thus serve as unknown but fixed parameters. Tuning them helps reproduce the shape of waveforms. Figure 3.14 illustrates how the shape of the resulting waveform changes as surface-to-bulk drift and surface charge vary. The high magnitude of the surface charge means that more charges will be pulled onto the surface, and thus the sharply rising part of the waveform, which is due to bulk charge collection, will have a lower magnitude. A faster surface-to-bulk ratio means that the charges on the surface will be collected more quickly, and therefore the tail shape will be different.

The deployment or operating conditions, such as surface charge density or bias voltage, can change between different experimental setups or over time. The surface charge, for instance, might build up differently after each deployment, while the operating voltage is controlled by the detector configuration and the conditions of the experiment.

### 3.7.12 Input and Output

`EH-Drift` is compiled using the GCC compiler. It utilizes a configuration file that is the same as `siggen`. The configuration file contains information on the detector geometry, and an example can be found in [51]. In addition to the configuration file, we provide flags that can be passed to the executable. Table 3.1 shows the input flags that allow one to set multiple parameters related to the event.

The simulation output is stored in an HDF5 file, a hierarchical data format optimized for handling large structured datasets. Each simulated event is recorded in the ‘event data’ dataset as a compound data type containing energy, radius and height positions, surface charge, surface drift velocity factor, and the signal for the event. The dataset is dynamically extendable to allow new events to be appended without rewriting existing data. The file includes grid size, passivated surface thickness, self-repulsion flag, and detector name as attributes. Attributes are stored as scalars at the file root level for efficient retrieval without redundancy. This storage is critical for High Performance Computing (HPC) when we simulate thousands of waveforms for multiple detectors.

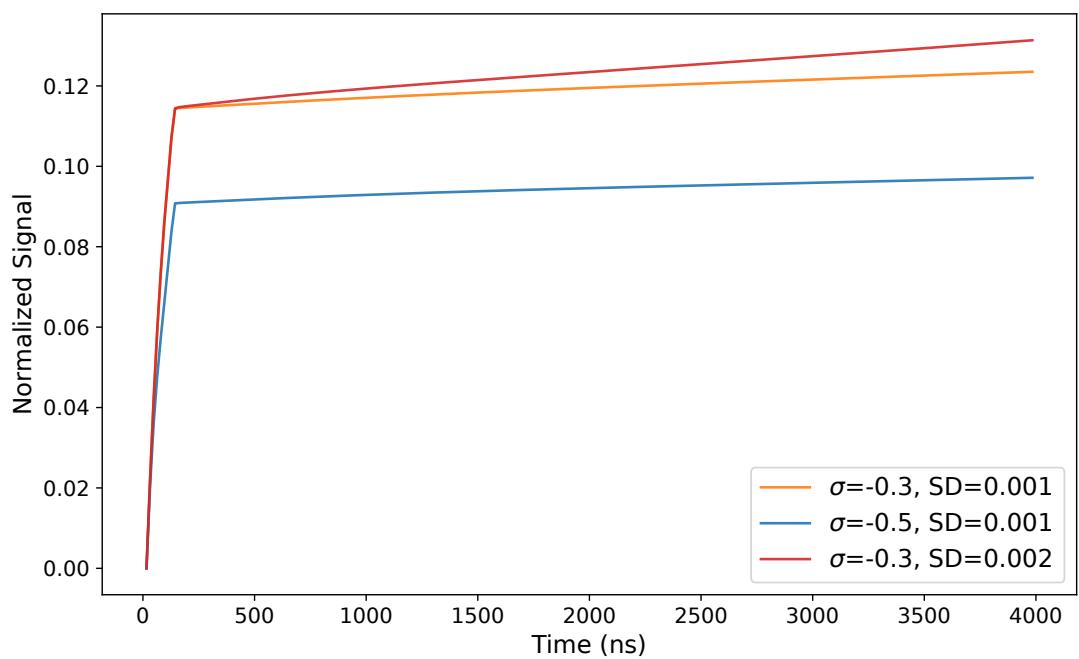


Figure 3.14: A comparison of waveforms generated by EH-Drift for a LEGEND-200 PPC detector as surface charge ( $\sigma$ ) and relative surface drift velocity (SD) are varied. The events were at  $r=15$  mm  $z=0.02$  mm.

<b>Flag</b>	<b>Description</b>	<b>Example</b>
-r	Set the radial position of the event in mm.	15.00
-z	Set the axial position of the event in mm.	0.50
-g	Specify the detector name.	P42575A
-s	Set the surface charge density in $10^{10} e/cm^2$ .	-0.50
-e	Input the interaction energy in keV.	5000
-v	Choose whether to write density files (0 = no, 1 = yes).	1
-f	Choose whether to recalculate the electric field at each time step (0 = no, 1 = yes).	1
-w	Choose whether to save the electric field at the end of the program (0 = no, 1 = yes).	1
-d	Choose whether to save the depletion surface (0 = no, 1 = yes).	1
-p	Choose whether to write the weighting potential (0 = no, 1 = yes).	1
-b	Set the bias voltage in volts.	3500
-h	Specify the grid size in mm.	0.0200
-m	Define the passivated surface depth in mm.	0.00020
-c	Set the velocity of surface charges compared to bulk.	0.001
-a	Input a custom impurity density profile file.	filename.dat
-t	Define the waveform length to be simulated, in ns.	16000
-u	Set the frequency of output signal saving in ns.	16

Table 3.1: Input parameters to EH-Drift

## CHAPTER 4: GPU ACCELERATION

### 4.1 Introduction

The EH-Drift approach better models surface alpha waveforms by incorporating diffusion, self-repulsion, and field recalculations. However, it is computationally expensive. A simulation of a 800 ns event for a surface alpha event with a 10-micrometer grid can take up to 7 hours to run using a non-parallelized CPU-based implementation on a single compute server in the Longleaf cluster at the University of North Carolina at Chapel Hill. This is primarily due to the electric potential being recalculated at every time step. In the program, the Poisson equation is solved numerically using a relaxation algorithm, where the program goes through each point averaging around the neighbors. This can be computationally expensive since the program has to update every grid point hundreds of times until convergence is achieved.

Figure 4.1 shows the run time for a 5 MeV event in EH-Drift for different grid sizes in the non-parallelized CPU implementation. In general, the smaller the grid, the more points are needed to update, and the longer the program takes. As discussed in chapter 3.3, alphas have a penetration depth of less than  $20\mu m$  in Germanium, so, ideally, the grid size should be at most  $10\mu m$ . Moreover, to correctly model alpha backgrounds, many simulations will need to be run involving different detectors, incident alpha positions, radii, heights, surface drift speed, and detector surface charges. The computation time scales quickly for sampling these parameters.

One way to speed up EH-Drift is by using parallel calculations with Graphic Processing Units (GPUs). We can transform the entire detector into the GPU memory so that each GPU kernel represents a grid point. Then we can perform the calculations at each grid point in parallel, and instead of going through the program one grid point at a time, the points can be updated simultaneously. We executed this strategy using the NVIDIA CUDA C++ framework. Of all the steps executed by EH-Drift, it is expected that performing the relaxation algorithm on the GPU would have the greatest improvement in run time. We thus begin with a discussion of 2-D Poisson equation solvers based on iterative relaxation methods, then explain how we adapted them to GPU architectures.

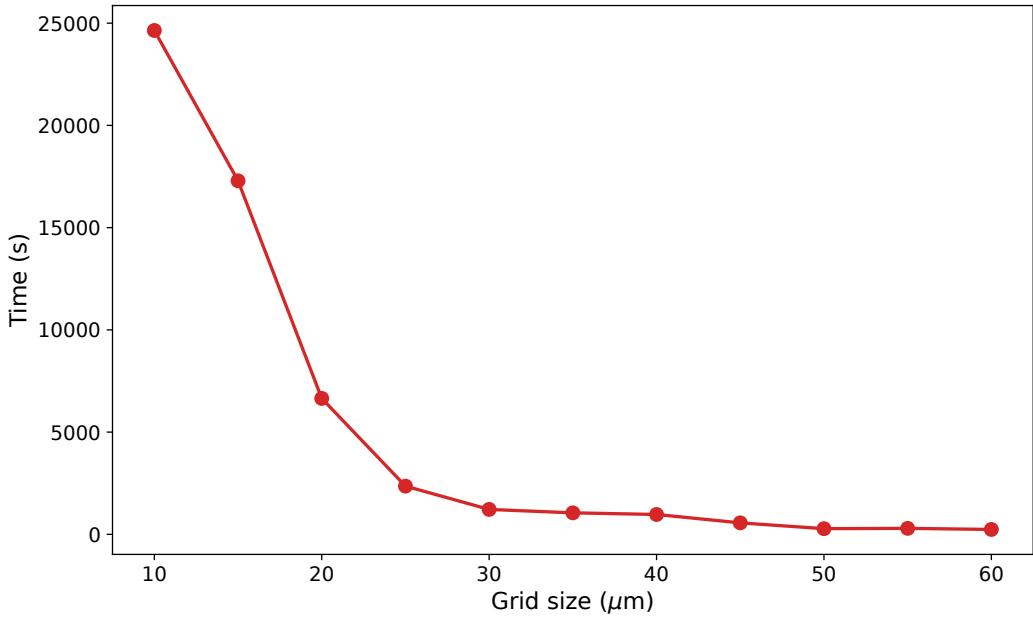


Figure 4.1: Run time for a single EH-Drift event on CPU. Run time based on a single compute server in the Longleaf cluster at the University of North Carolina at Chapel Hill.

## 4.2 Iterative Relaxation Algorithms for Solving 2-D Poisson Equation

The electric potential inside the detector can be solved by solving the Poisson equation:

$$\nabla^2 \phi = -\frac{\rho}{\epsilon}, \quad (4.1)$$

where  $\phi$  is the electrostatic potential,  $\epsilon$  is the permittivity of the material, and  $\rho$  is the charge density. An analytic solution to the Poisson equation does not exist for most geometries, and we rely on numerical techniques to approximate a solution. To solve for the potential in the EH-Drift, the detector is divided into grid points and assigned the following classification: Point Contact (PC), High Voltage Contact (HVC), Inside the detector, On the Passivated Surface (Passive), Pinchoff points, Ditch, Ditch Edge, and Contact Edge. We treat each point type separately in solving for the potential. The boundary conditions are determined by the geometry of the detector and the impurity concentration. For the passivated surface, the boundary condition is a reflection symmetry about that surface. The impurity concentration on the passivated surface is modified to include the effects of surface charge if present. In the absence of surface charge, this means that the electric field lines at the passivated surface are parallel to the surface. Points

on the high-voltage (HV) and point-contact (PC) surfaces have Dirichlet boundary conditions fixed at the contact voltages.

### Jacobi method

A simple way to solve for the potential is to set the boundary condition, run through all the points in the grid, and estimate the solution at a point in the grid by the average of four points around it. The code continues to iterate through the detector, updating all the points until the difference  $\phi_{i,j}^n - \phi_{i,j}^{n-1}$  at  $n^{\text{th}}$  iteration at all points is less than the convergence threshold. This method is called the Jacobi method, and the points are updated using equation 4.2 [52].

In this method, the iteration order through the points must be fixed: the method begins with the bottom leftmost point. Then it progresses through the row from left to right and repeats, moving up a row each time.

$$\phi_{i,j}^n = \frac{1}{4} \left( \phi_{i+1,j}^{n-1} + \phi_{i-1,j}^{n-1} + \phi_{i,j+1}^{n-1} + \phi_{i,j-1}^{n-1} \right) \quad (4.2)$$

Here the  $n$  denotes the iteration number,  $i$  and  $j$  are the radii and height of the grid point under consideration.

### Gauss-Seidel Method

Although the Jacobi method works well to obtain a numerical solution, it is slow. One way to speed up the method is to use the newly calculated grid points to update the current grid. This is represented by equation 4.3 and is termed the Gauss-Seidel method [52].

$$\phi_{i,j}^n = \frac{1}{4} \left( \phi_{i+1,j}^{n-1} + \phi_{i-1,j}^n + \phi_{i,j+1}^{n-1} + \phi_{i,j-1}^n \right) \quad (4.3)$$

By incorporating updated values as soon as they are available, Gauss–Seidel converges faster than the Jacobi method.

### Successive over-relaxation

A way to achieve faster convergence with the Gauss-Seidel method is to use a relaxation factor  $\omega$  to extrapolate the solution and accelerate the convergence rate. The successive over-relaxation method

(SOR) was developed by David Young [53]. It is represented by equation 4.4 and helps to reach convergence quickly. It is the method used in `siggen` simulations to calculate the electric potential. The value of  $\omega$  can be between 1 and 2, and there is an optimal value that gives the fastest convergence. For most PDE problems,  $\omega$  is typically slightly less than 2. Equation 4.4 was empirically tuned to approach a value of less than 2 for large detectors while remaining stable for smaller detectors.

$$\phi_{i,j}^n = (1 - \omega)\phi_{i,j}^{n-1} + \frac{\omega}{4} (\phi_{i+1,j}^{n-1} + \phi_{i-1,j}^n + \phi_{i,j+1}^{n-1} + \phi_{i,j-1}^n) \quad (4.4)$$

$$\omega = \frac{1.991 - 1500.0}{L * R} \quad (4.5)$$

The CPU version of EH-Drift was written using the SOR algorithm, which proved to be quite effective at reaching fast convergence; however, this cannot be parallelized on GPU. This is because when a grid point is updated, a combination of new values ( $\phi^n$ ) and old ( $\phi^{n-1}$ ) values is used, as shown in figure 4.2. This cannot be implemented in parallel, as the order of update is significant for the SOR method, and thus cannot be parallelized.

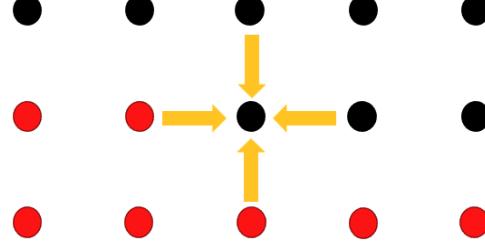


Figure 4.2: Updating grid points in the SOR method. Red grid points represent all the points that have already been updated. A given grid point is updated using a mixture of old and new values.

### Red-Black Successive Over-Relaxation

To achieve parallelism, we can use a modified SOR algorithm termed Red-Black Successive Over-Relaxation (RB-SOR) [54]. In RB-SOR, we assign a point as red if  $(i + j)$  is even and black if  $(i + j)$  is odd. We update all the red points using old values and then use the newly updated red values to update the black nodes.

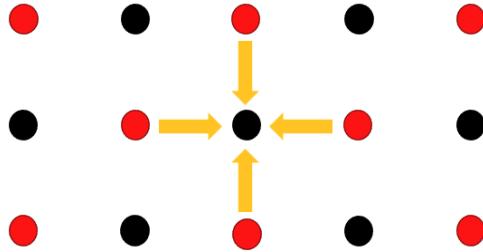


Figure 4.3: Updating grid points in Red-Black SOR. All the red grid points are updated in parallel using old values, and the black grid points are updated using the new red grid points.

Figure 4.3 shows how RB-SOR differs from traditional SOR. This method reaches convergence in about the same number of iterations as SOR, but the update at each iteration can now be done in parallel using GPUs. Parallel reduction techniques can be used to verify convergence by finding the maximum difference per point between the  $n$  and  $n - 1$  iterations. The update equations of the RB-SOR method are:

$$\phi_{i,j}^n = (1 - \omega)\phi_{i,j}^{n-1} + \frac{\omega}{4} (\phi_{i+1,j}^{n-1} + \phi_{i-1,j}^{n-1} + \phi_{i,j+1}^{n-1} + \phi_{i,j-1}^{n-1}) \quad i+j \text{ is even} \quad (4.6)$$

$$\phi_{i,j}^n = (1 - \omega)\phi_{i,j}^{n-1} + \frac{\omega}{4} (\phi_{i+1,j}^n + \phi_{i-1,j}^n + \phi_{i,j+1}^n + \phi_{i,j-1}^n) \quad i+j \text{ is odd} \quad (4.7)$$

Figure 4.4 shows a solution to the Poisson equation in the program. With the framework set for implementing the relaxation algorithm on the GPU, we are now ready to discuss the programming model used and the parallel computing techniques.

### 4.3 GPU Programming with CUDA C++

A Central Processing Unit (CPU) and a Graphics Processing Unit (GPU) are designed with different computations in mind. A CPU is designed to execute a sequence of operations as fast as possible, one at a time, whereas a GPU is designed to execute tens of thousands of operations in parallel. A CPU can perform a single task significantly faster than a GPU, but GPUs dramatically reduce run time in problems where parallelization can be established, such as adding vectors, manipulating matrices, finding gradients, etc. Thus, GPUs are extensively used in areas such as computer vision, gaming, and machine learning. A GPU always needs a CPU to drive the workflow.

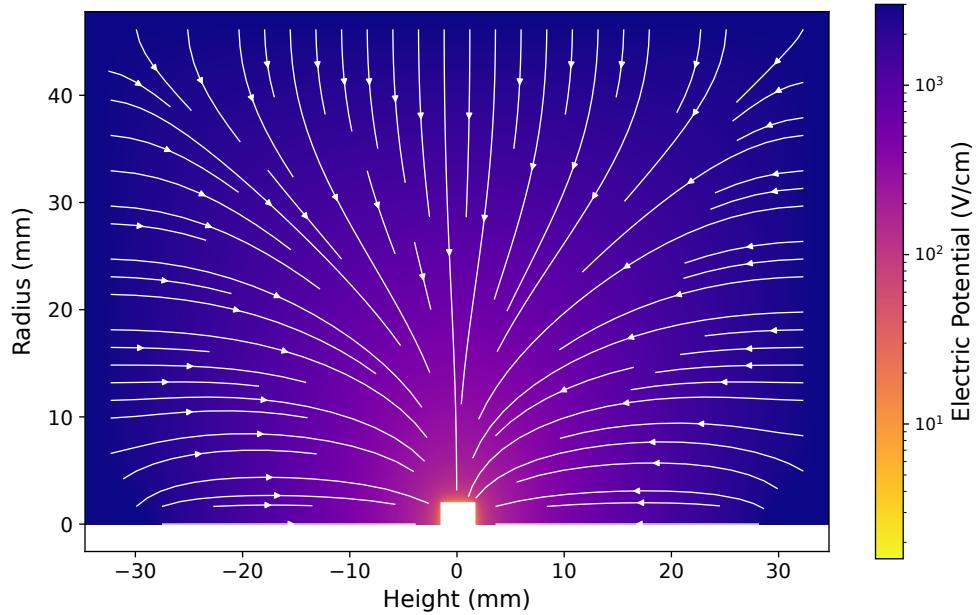
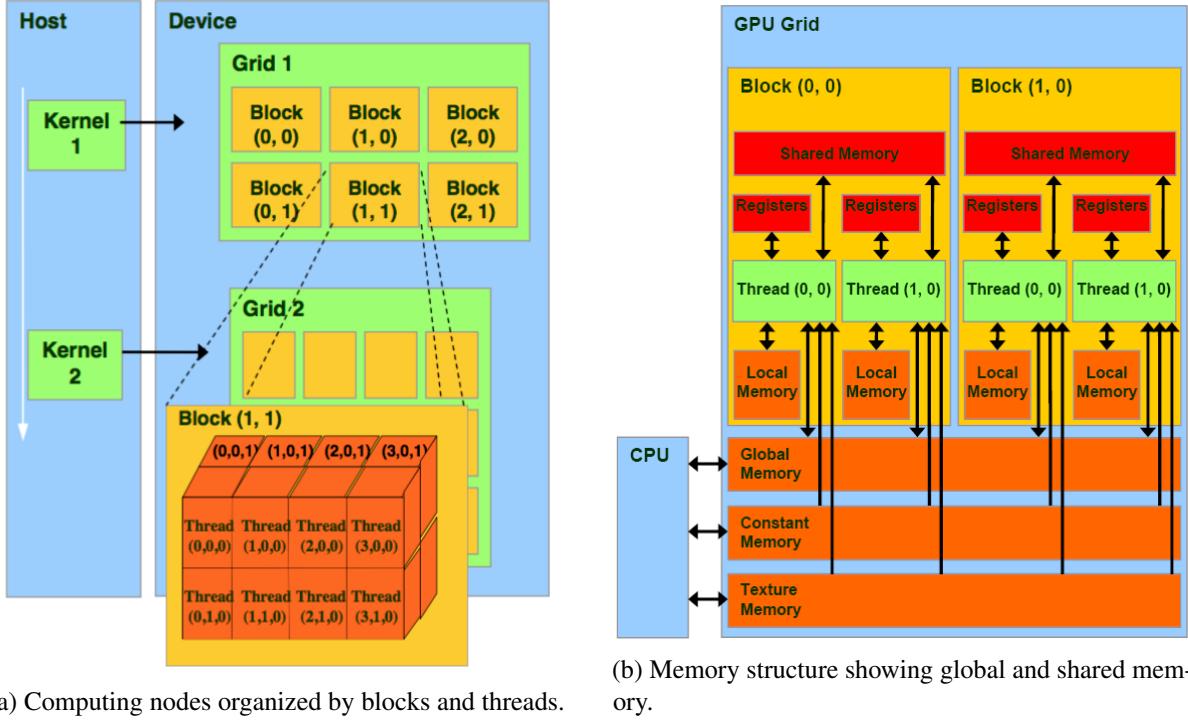


Figure 4.4: A solution to the Electric Potential using Red-Black SOR in EH-Drift. The white lines show electric field lines.

For our purpose, we used NVIDIA GPUs. The Compute Unified Device Architecture (CUDA) is a programming environment that allows writing C++ code on GPUs. CUDA extends the standard C++ environment to enable functions, called kernels, to be executed in parallel by many threads. The GPU performs calculations in the kernel, which contains the sets of instructions each thread has to follow. The kernel execution is divided into blocks, each with a given number of threads, shown in figure 4.5a. To enable parallelism, we need to break down the problem onto the GPU grid and then write instructions for threads and blocks to follow.

In CUDA, the maximum number of blocks that can be called is 65535 and the maximum number of threads per block is 1024. We thus want to find a way to split our detector into blocks and threads and index it appropriately for each grid point  $(r, z)$ . Our first attempt was to set the height  $z$  as the block index and the radius  $r$  as the grid index such that  $z=10, r=15$  corresponds to block 10 and thread 15. The  $z$  and  $r$  values used are in grid units, so we never run into decimal point issues with this approach; however, we quickly ran out of threads on smaller grid sizes because there can be at most 1024 in a given block. To mitigate the problem, we use modular arithmetic to slice the detector of radius  $R$  and height  $L$  into rectangles of width  $R$  and length = max threads, such that the total number of blocks and grids needed is:



(a) Computing nodes organized by blocks and threads.

(b) Memory structure showing global and shared memory.

Figure 4.5: A depiction of the GPU architecture. Source: NVIDIA CUDA C++ Programming Guide

$$\text{Number of blocks} = R \times \text{ceiling} \left( \frac{L}{\text{max threads}} \right) \quad (4.8)$$

$$\text{Number of threads per block} = \text{max threads} \quad (4.9)$$

Here  $R$  and  $L$  are in terms of the number of grid units. Then for a given (block index) and (thread index), the corresponding  $r$  and  $z$  are:

$$r = (\text{blockIdx}) \bmod R \quad (4.10)$$

$$z = \text{floor} \left( \frac{\text{blockIdx}}{R} \right) \times \text{max threads} + \text{threadIdx} \quad (4.11)$$

With the blocks and grid framework in place, we can focus on memory management. The GPU memory exists in a different space than the CPU memory. To perform any calculation, we need to copy the data to GPU memory, perform the task with GPUs, and copy the data back to CPU memory. This

copying operation is slow, so we aim to minimize the number of copying steps as much as possible. We assigned constants such as grid size, radius, height, etc. to the small shared memory such that both the GPU and CPU can access them simultaneously. Larger arrays such as potentials, point types, and impurities were copied to the large GPU global memory. The GPU also does not perform well on multidimensional arrays, so we flattened them using the formula:

$$\text{Array}[i][j][k] = \text{flat array}[(i \times (L+1) \times (R+1)) + ((R+1) \times j) + k] \quad (4.12)$$

The nvcc library was used to compile and link the GPU C++ code with the EH-Drift C code. Using GPUs only for the relaxation algorithm, we achieved a 13x speed improvement over the CPU-based program. In the new program, the most computationally expensive task was copying between CPU and GPU memory, which was performed every time step after field recalculation. The run time could be reduced by copying all the data needed at the beginning of the program, performing all the necessary calculations on the GPU, and copying back only the required data. This required the implementation of the diffusion and charge drift components and signal calculation on the GPU. Running on the GPU does not result in significantly faster run time for these components, but it does reduce the memory transfer at each step, significantly decreasing the run time.

Performing diffusion and self-repulsion in parallel requires one to address one issue: the interference of threads with each other. When densities are allowed to drift and diffuse, two densities could end up on the same grid point. This would not be an issue on the CPU since points are addressed sequentially, but on the GPU, this would mean that two threads are trying to update the same memory location. This was enough to cause disagreement between CPU and GPU results. CUDA enables the implementation of ‘atomic operations’, which allow a thread to perform a task that is guaranteed to be performed without interference from other threads. Programming the densities to update using atomic operations fixed the discrepancies between the GPU and CPU with no noticeable time loss.

Warp divergence occurs when threads in the GPU attempt to perform different operations. This can lead to divergent results and can also cause performance problems. To avoid this, we modularized the GPU kernels into smaller executions and added device synchronization between each kernel. This made sure that all threads performed the calculation before moving to a new kernel.

Having diffusion and charge drift running on the GPU enabled the entire time step loop in figure 3.7 to run on the GPU without copying back to CPU memory. This required careful tracking of the pointers in GPU memory using a C struct to store the locations of all GPU pointers. This struct is initialized in the time step loop by allocating space in GPU memory and copying the required data into it. The calculations are then performed in GPU memory, and only the required data is transferred to CPU memory.

A final improvement in the run time was achieved by generating the signal on the GPU without the need to write the charge cloud densities to the file. In the initial implementation, the signal was calculated by saving snapshots of the densities during the program, and the snapshots were read later to generate the signal. A faster approach is to calculate the signal collected during each time step on the GPU and skip writing the snapshots to the file. Signal calculation entails summing up the charges collected and multiplying them by the weighting potential.

On the GPU, such operations can be performed using the parallel reduction technique. In this technique, two elements of the array are grouped in pairs. Each pair is computed in parallel with the others, halving the overall array size in one step. This reduction is continued until there is one final number: the sum in our case. We used the NVIDIA Thrust library to perform sums on the GPU using parallel reduction techniques and generate the signal on the GPU. Now, the only memory that is transferred to the CPU is the final simulated waveform.

#### 4.4 GPU Performance Comparisons and Benchmarking

The GPU program accurately reproduced the CPU results with less than 0.69 % difference in the waveforms. Figure 4.6 shows a comparison between the waveforms of the two programs. The difference is primarily due to the different methods of calculating the electric potential: SOR in the CPU versus RB-SOR in the GPU. As shown in figure 4.7, it reduced to 0.045% when both simulations were allowed to relax to a threshold of numerical convergence at the double precision level. The default convergence threshold was set to  $8 \times 10^{-4}$  by studying the maximum error and run time for different convergence thresholds.

However, the GPU program was extremely fast, with nearly grid-size-independent execution time. We primarily used GPUs on the Perlmutter supercomputer at the National Energy Research Scientific Computing Center (NERSC). Figure 4.8 shows the time taken by each program. For a grid size of

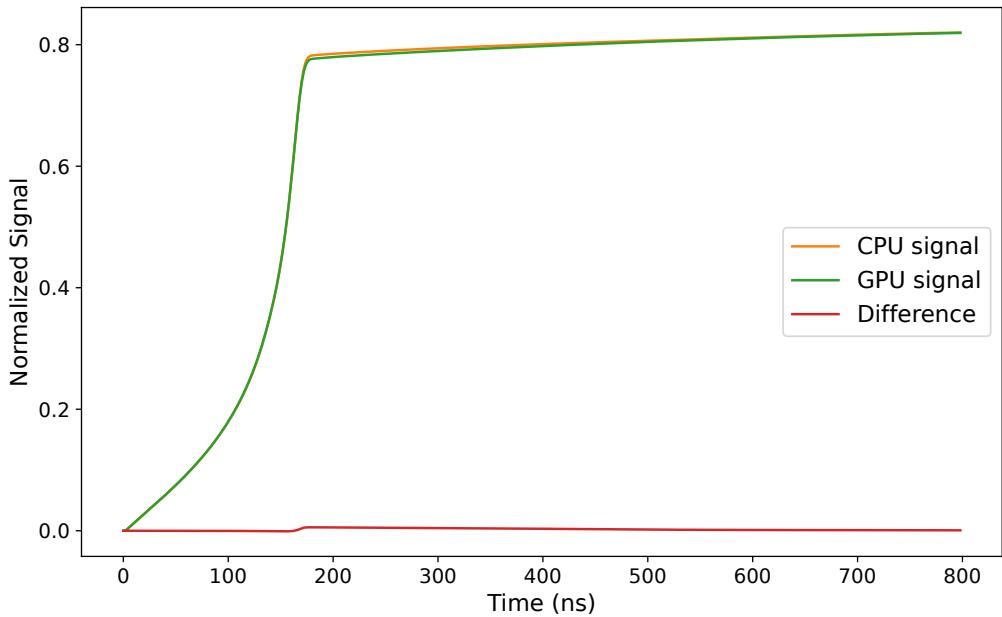


Figure 4.6: Comparison of waveform generated by CPU program versus GPU program in EH-Drift at the default converge threshold. The event was at  $r=15\text{mm}$  and  $z=0.10\text{mm}$ .

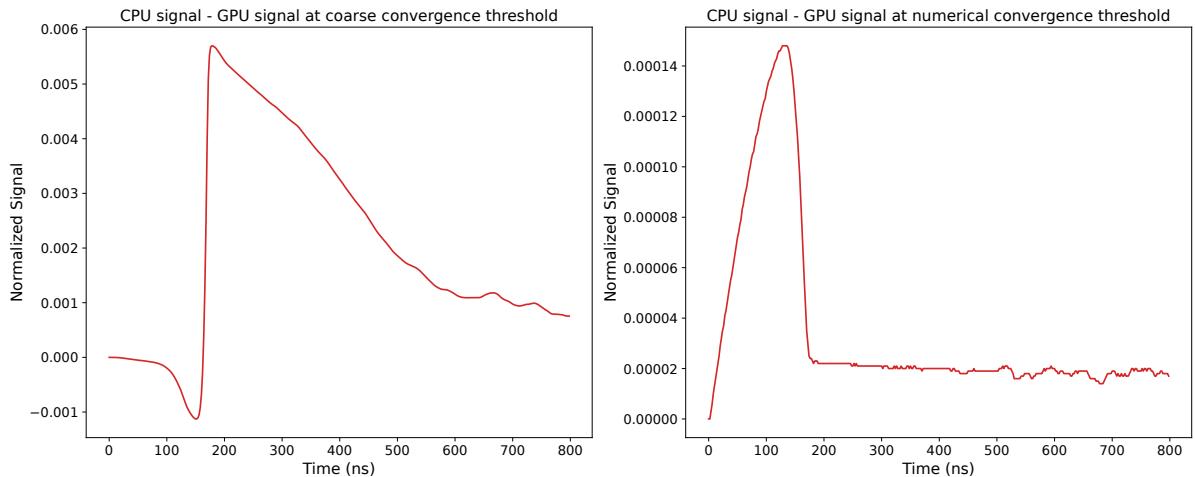


Figure 4.7: Difference in waveforms produced by the CPU and GPU programs. Left plots shows the difference at default converge threshold for RB-SOR algorithm. Right plot shows the difference at numerical convergence which is set to  $8 \times 10^{-8}$ .

10 microns, the CPU program took 24642 seconds ( $\sim 7$  hours) to run, while the GPU program took 394 seconds ( $\sim 7$  min) to run, representing a speed increase of 62.5x. The biggest contribution to the speed-up was the implementation of the electric potential calculation on the GPU, which was also grid-size independent. Other contributions were the fact that the densities did not have to be written to a

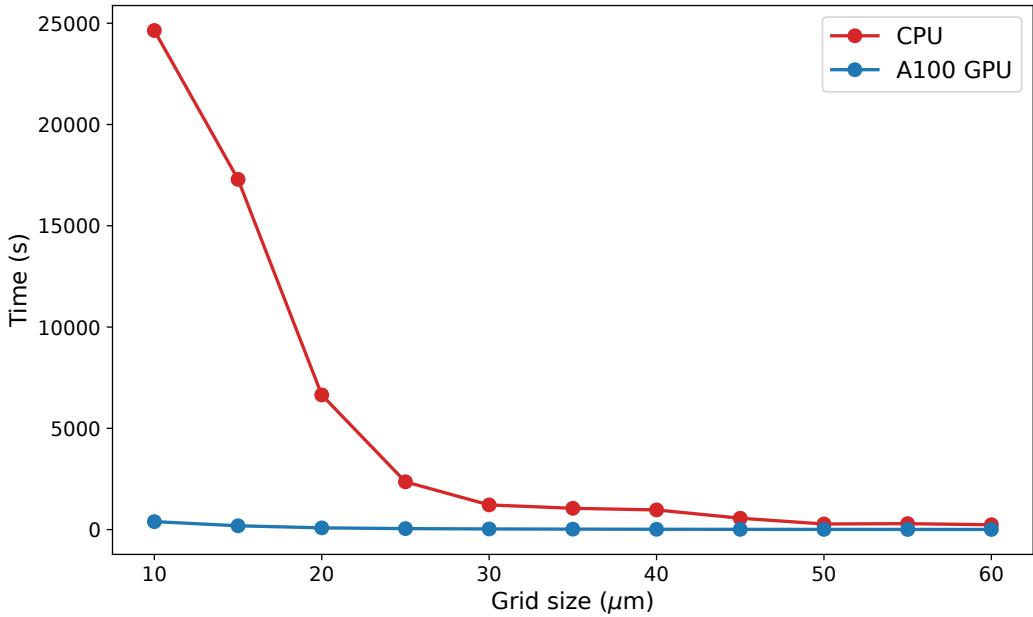


Figure 4.8: Comparison of run time for an EH-Drift event on GPU and CPU. Run times are based on a single CPU node on the Longleaf cluster at UNC-Chapel Hill and an NVIDIA A100 node of the Perlmutter supercomputer at the NERSC.

file to generate the signal and that there was no need for a transfer of data from GPU memory to CPU memory at each time step. This speed-up will allow for running thousands of simulations and building a waveform library with different configurations.

Figure 4.9 shows how the run time varies for different NVIDIA GPUs at the two clusters. The faster the GPU, the faster the program runs, suggesting that we have achieved parallelization using sufficient utilization of GPU resources. Table 4.2 shows the profiling results of GPU kernels that did not use the standard library. The kernel name is listed alongside its contribution to the total execution time as a percentage time (%), its cumulative execution time in nanoseconds Total Time ( $ns$ ), and the number of times it was executed during profiling (Instances). The table also includes the average execution time per invocation (Avg ( $ns$ )), which indicates the computational efficiency of each kernel. Furthermore, the dimensions of the thread blocks (blocks ( $x, y, z$ )) and the configuration of the grid (grid ( $x, y, z$ )) are reported. The most computationally intensive kernel was the one in which self-repulsion was performed. This is where we determine where the charges drift to, which can contain many conditional statements and for loops that can slow down GPU performance. This section of the code is an excellent target for

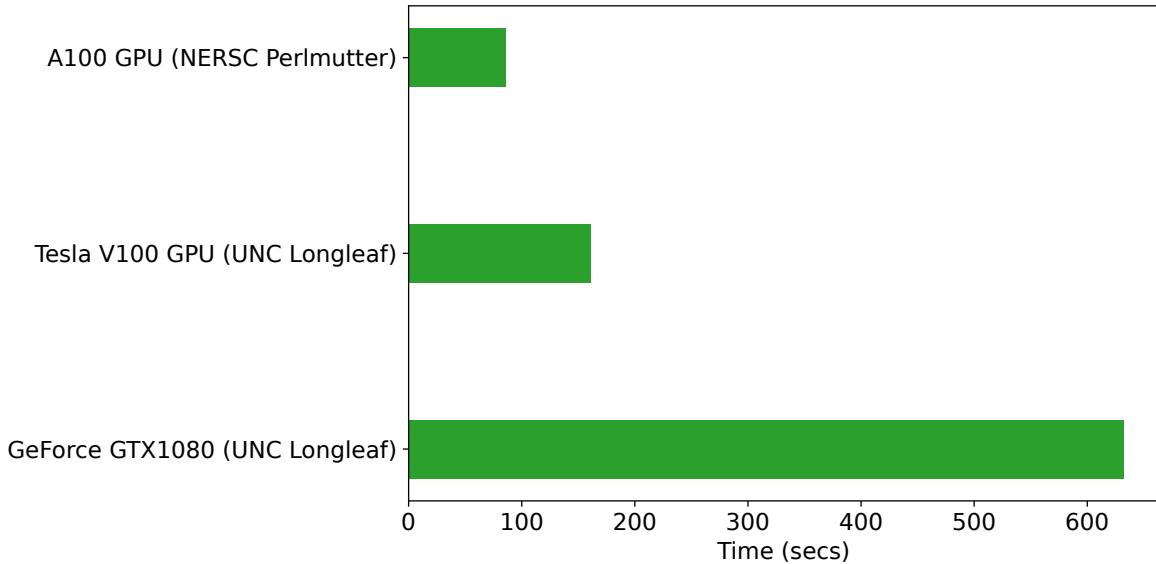


Figure 4.9: Program run time on various GPU models for a  $20\mu$  grid. The faster GPU results in lower run time.

further optimization. The kernel that was called the most was the relax step of the RB-SOR algorithm. This is expected as, for a given time step, the program needs to iterate several times to reach convergence. The Block and Grids usage column suggests that we are using x components of the block and grid. This can be improved in the future to use an entire grid, although that would be a computational challenge requiring significant changes to the program. In the next section, we describe some of the results from EH-Drift to show how it can be used to model background events from the passivated surface.

Kernel <i>Name</i>	Function	Time (%)	Total Time (ns)	Insta -nce	Avg (ns)	Blocks (x, y, z)	Grids (x, y, z)
gpu_self_repulsion	Finds location where the charges will drift	22	$3.43 \times 10^{10}$	18594	$1.85 \times 10^6$	(1024,1,1)	(5211,1,1)
diff_update	Diffuse charges to the new location	16	$2.42 \times 10^{10}$	18594	$1.30 \times 10^6$	(1024,1,1)	(5211,1,1)
gpu_sr_update	Drift the charges to new location	15	$2.35 \times 10^{10}$	18594	$1.26 \times 10^6$	(1024,1,1)	(5211,1,1)
gpu_diffusion	Calculates the diffusion rate to neighbors	11	$1.71 \times 10^{10}$	18594	$9.22 \times 10^5$	(1024,1,1)	(5211,1,1)
relax_step	Performs relaxation step in R-B SOR	9	$1.43 \times 10^{10}$	51138	$2.79 \times 10^5$	(1024,1,1)	(5211,1,1)
reset_rho	Resets charge density values	5	$7.49 \times 10^9$	18594	$4.03 \times 10^5$	(1024,1,1)	(5211,1,1)
set_rho_zero	Sets charge density to zero	3	$4.56 \times 10^9$	9297	$4.90 \times 10^5$	(1024,1,1)	(5211,1,1)
update_impurities	Updates impurity concentrations	3	$4.46 \times 10^9$	9297	$4.79 \times 10^5$	(1024,1,1)	(5211,1,1)
cal_esum1	Calculates $S_e(t)$	2	$3.00 \times 10^9$	10898	$2.75 \times 10^5$	(1024,1,1)	(5211,1,1)
surface_drift	Drifts charges on the surface	1	$2.61 \times 10^9$	18594	$1.40 \times 10^5$	(1024,1,1)	(5211,1,1)
cal_hsum1	Calculates $S_h(t)$	1	$2.58 \times 10^9$	10898	$2.37 \times 10^5$	(1024,1,1)	(5211,1,1)
hvc_modification	Modifies high voltage points	1	$2.14 \times 10^9$	18594	$1.15 \times 10^5$	(1024,1,1)	(5211,1,1)
cal_hsum2	Calculates $R_h(0)$	1	$1.98 \times 10^9$	10898	$1.82 \times 10^5$	(1024,1,1)	(5211,1,1)
cal_esum2	Calculates $R_e(0)$	1	$1.97 \times 10^9$	10898	$1.80 \times 10^5$	(1024,1,1)	(5211,1,1)
surface_drift_calc	Calculates where the density on surface will drift	1	$1.74 \times 10^9$	18594	$9.34 \times 10^4$	(1024,1,1)	(5211,1,1)
z_reflection_set	Sets initial reflection symmetry in z-direction	0	$1.56 \times 10^8$	25569	$6.10 \times 10^3$	(1,1,1)	(1736,1,1)
reflection_symmetry	Sets reflection symmetry during simulation	0	$1.47 \times 10^8$	25569	$5.74 \times 10^3$	(1,1,1)	(2543,1,1)
relax_step_init	Relax step for initial R-B SOR	0	$6.26 \times 10^7$	1332	$4.70 \times 10^4$	(1024,1,1)	(580,1,1)
set_passivated_imp	Sets impurity for passivated surface	0	$1.17 \times 10^7$	2295	$5.08 \times 10^3$	(1,1,1)	(1737,1,1)

Table 4.2: Profiling results of custom GPU kernels during simulation.

## CHAPTER 5: SURFACE SIMULATIONS RESULTS

### 5.1 Introduction

In the previous chapters, we introduced EH-Drift as a new method to model passivated surfaces in germanium detectors and showed that we can achieve a significant runtime improvement using GPUs. In this chapter, we present the first results from the simulations. The chapter begins with an overview of preprocessing the output waveform, then shows how EH-Drift can be used to simulate varying conditions and match the experimental data from the previous scanners. We then show how the model can be used to estimate the collection efficiency of  $0\nu\beta\beta$  events and finally demonstrate how it can be used to model background source spectral components with two examples.

### 5.2 Post-Processing and Activeness Maps

In HPGe detectors, digital filtering is used to determine the energy of each event. The use of a filter can reduce the electronic noise, and thus improve the energy resolution. We applied a trapezoidal filter to the pulse to match one of the techniques used to determine the event energy in the experimental data in LEGEND. Figure 5.1 shows the waveform and trap filter output. Then the energy is picked from the indicated time on the flat top, taken a fixed time after the start time of the pulse's rise. We call the output of the filter the activeness value. Activeness represents the fraction of the initially deposited energy that is collected. Since EH-Drift waveforms are normalized, the activeness values range between 0 and 1, with 0 representing no collection and 1 representing full collection.

Although many results in this chapter focus on maps at a single initial energy, the same detector location can exhibit a different partial charge at different energies. The PONaMa-1 detector (PPC from ORTEC made from Natural Material) is a research and development PPC detector manufactured by ORTEC using natural abundance Germanium. Its properties are given in table 5.1. Figure 5.2 shows this variation of activity with initial energy in PONaMa-1 detector for various surface charges. The measured activeness increases with radius. One possible explanation is that, for higher energy, larger charge clouds

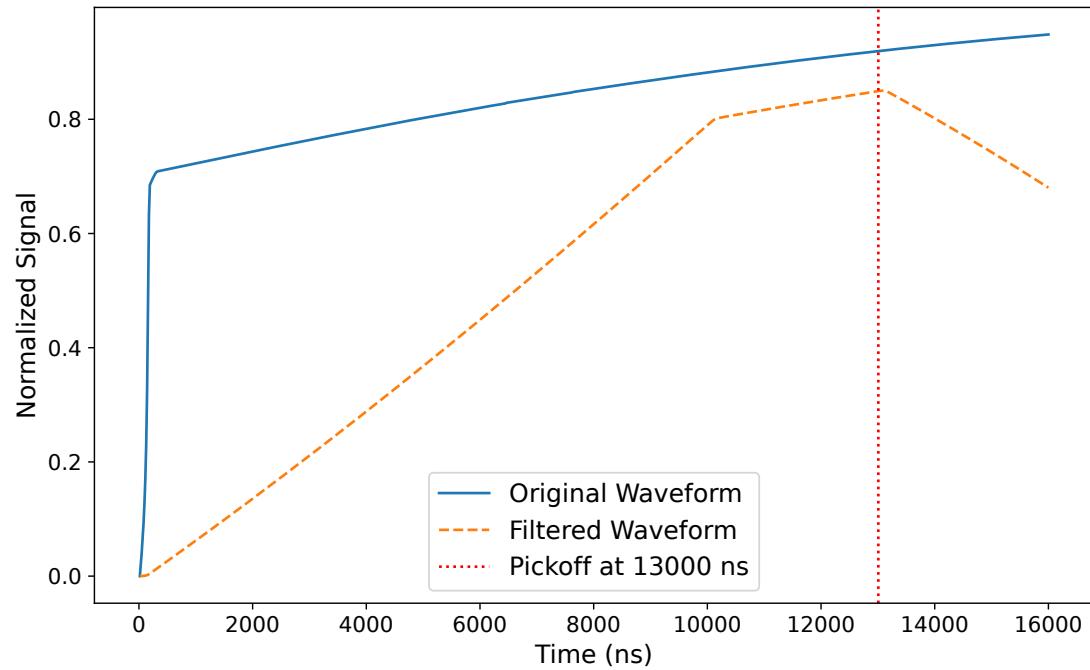


Figure 5.1: A simulated waveform and the resulting trapezoidal filtered waveform for a highly energy degraded surface waveform from EH-Drift simulations. The filter had a rise time of  $10\mu s$ , flat time of  $3\mu s$ , and the energy estimate is based on the value of the filtered waveform  $13\mu s$  after the start of the pulse rise.

are more susceptible to self-repulsion, which can push the charges away from the surface. This means that a smaller fraction of charges is pulled onto the surface in relation to their total number of carriers. Therefore, to accurately model background processes that span a range of energies, we would have to create activeness maps with different positions and initial energies.

Detector Geometry	Value
Height	50.5 mm
Radius	34.5 mm
Point-contact height	2.1 mm
Point-contact radius	1.4 mm
Taper length	4.5 mm
Capacitance	1.8 pF
Depletion voltage	1850 V
Impurity at $z=0$	$-0.318 \times 10^{10} \text{ e/cm}^3$
Impurity gradient	$0.025 \times 10^{10} \text{ e/cm}^4$

Table 5.1: Key geometry dimensions for the PONaMa-1 detector.

To understand the charge collection at different points on the detector, we map the detector using about 100 points, as shown in figure 5.3. The varying vertical spacing of the points was chosen to accurately capture the behavior in the rapidly varying region nearest to the surface. These points can be used to generate an activeness map for a detector such as in figure 5.4. This map is produced by interpolating using cubic interpolation. The activeness map gives an understanding of how the charge collection efficiency works in the detectors. Points at the same height but larger radii have a lower activeness, as holes must travel a longer distance, increasing the likelihood of being pulled onto the surface by the negative surface charge. Since each event takes a few minutes to run, it is important to know the optimum number of events needed to sample the map enough for interpolation. We found that approximately 100 points sampled in total in the  $r$  and  $z$  directions were sufficient.

Figure 5.5 illustrates the activeness maps without any surface charge. In this case, there is not much reduction in activeness.

Figure 5.6 shows the activeness map for positive surface charge. The loss of activeness is smaller than that for negative surface charges, since electrons are the minority charge carriers. Points at the same height but lower radii have lower activeness because electrons must travel a longer distance, increasing the probability that they are pulled onto the surface by a negative surface charge.

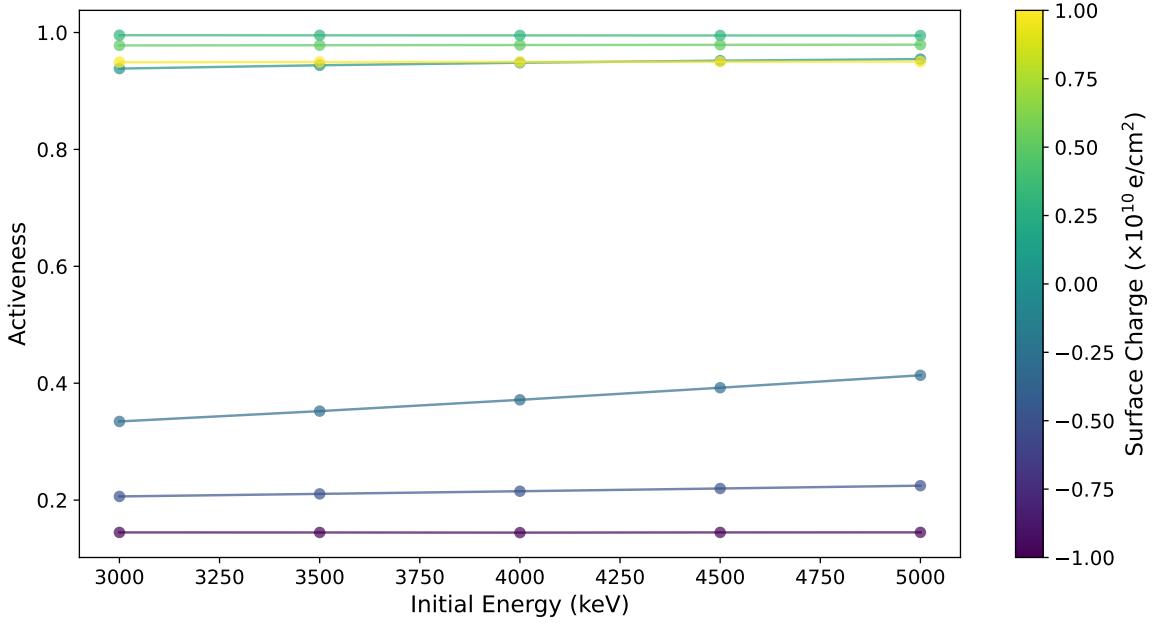


Figure 5.2: Activeness versus initial energy for event at  $r=14\text{mm}$  and  $z=1.00\text{mm}$  in PONaMa-1 detector. The lower activeness loss for higher energy could be due to self-repulsion pushing charges away from surface thus leading to smaller fraction of total charges on surface.

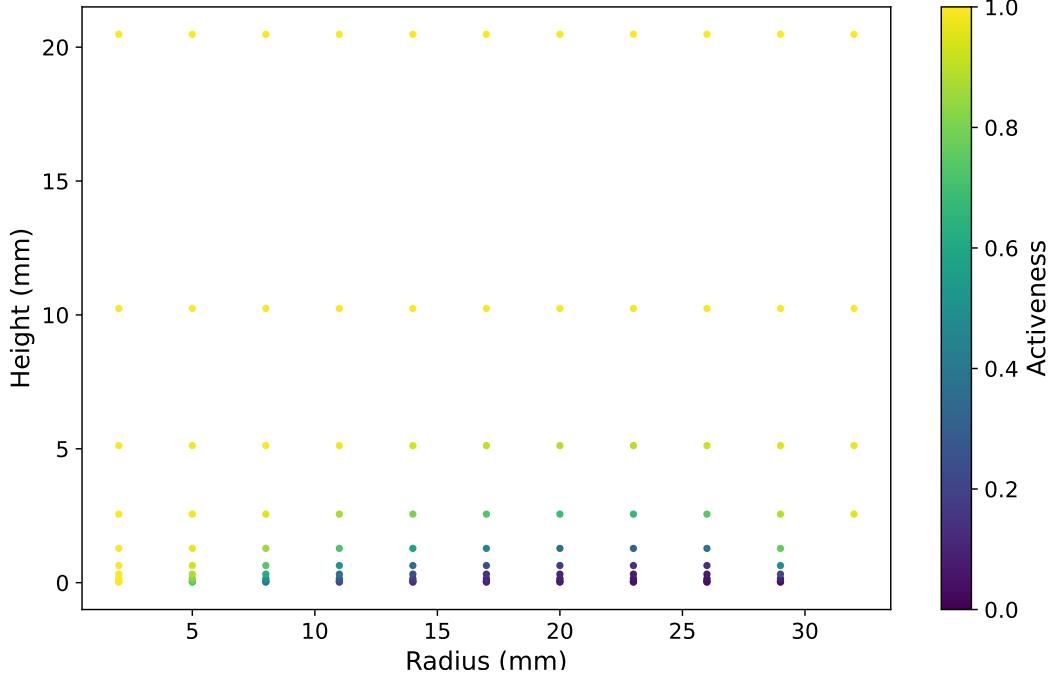


Figure 5.3: Activeness values to map a PPC detector EH-Drift with an initial energy 5000 keV and negative surface charge  $(-0.30 \times 10^{-10} \text{e}/\text{cm}^2)$ . The radius points were uniformly spaced while the height points were densely sampled close to the surface.

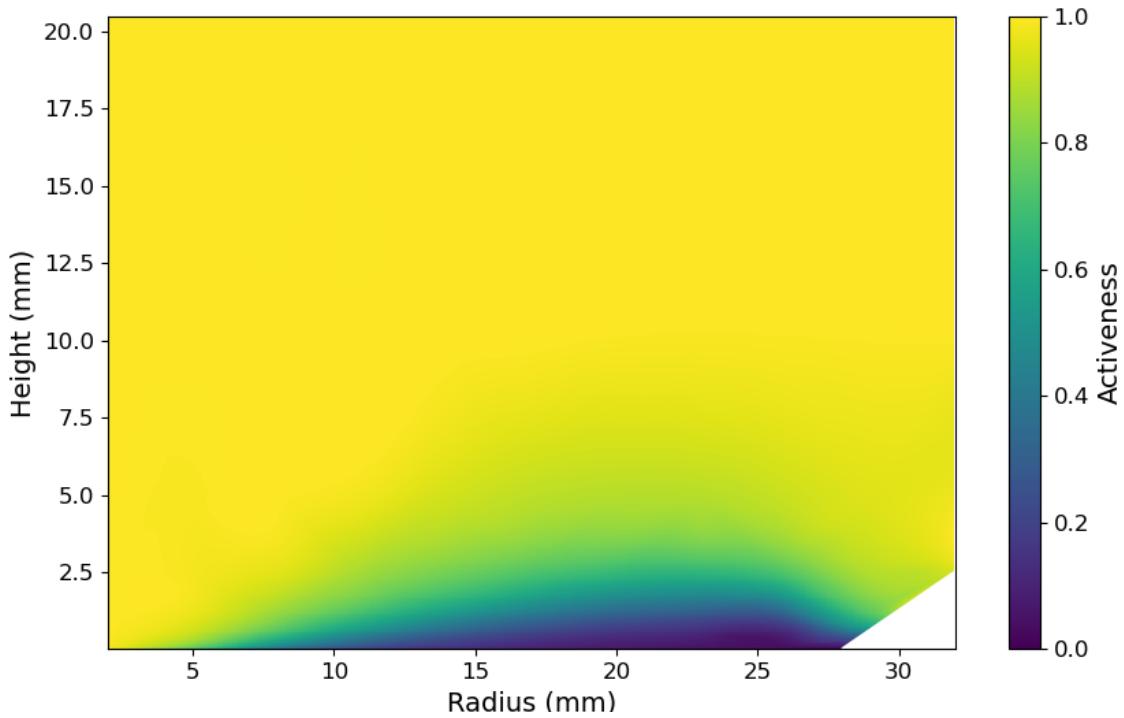


Figure 5.4: Interpolated activeness map for a PPC detector using EH-Drift with negative surface charge ( $-0.30 \times 10^{-10} \text{ e/cm}^2$ ). For a larger radius events there is a higher chance of having the hole component pulled onto the surface, causing higher loss of activeness.

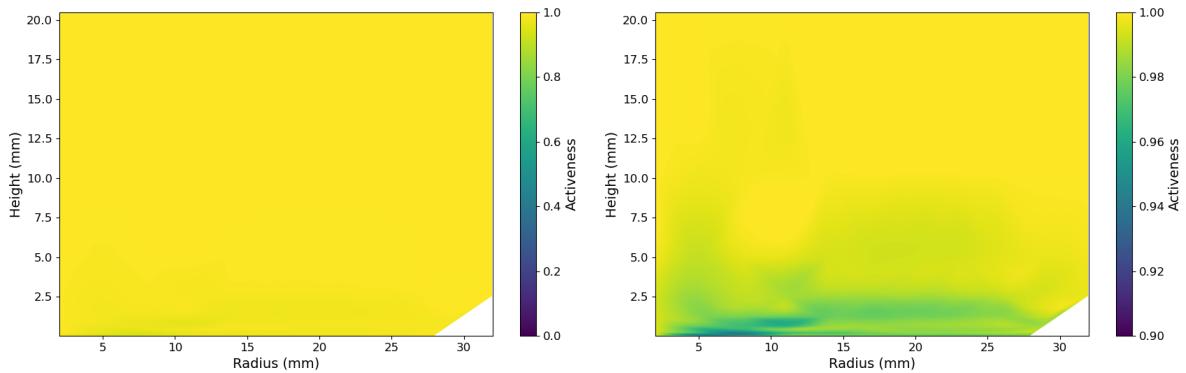


Figure 5.5: Interpolated activeness map for a PPC detector using EH-Drift with zero surface charge. Without surface charge there is nearly full collection. Right plot shows the same map with using narrow color scale.

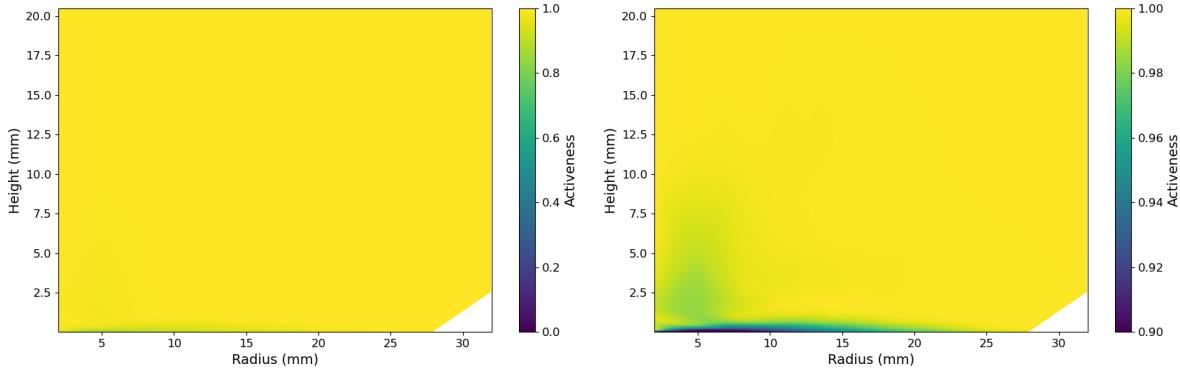


Figure 5.6: Interpolated activeness map for a PPC detector using EH-Drift with positive surface charge ( $0.30 \times 10^{-10} \text{ e/cm}^2$ ). The right plot is shown using narrow color scale. At a smaller radius, there is a higher chance of the electron component being pulled onto the surface causing higher loss of activeness. The activeness loss is smaller than negative surface charge since electrons are minority signal component.

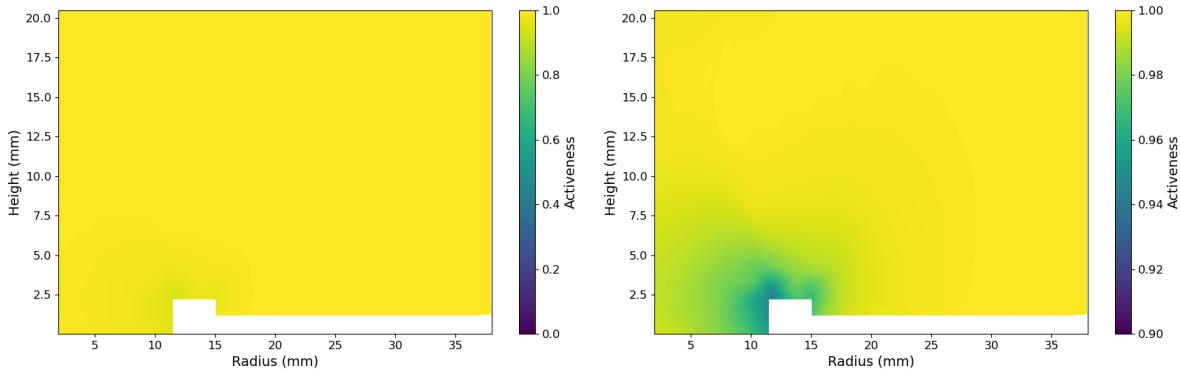


Figure 5.7: Interpolated activeness map for a Mirion ICPC detector using EH-Drift with negative charge ( $-0.30 \times 10^{-10} \text{ e/cm}^2$ ). The right plot is shown in a narrow color scale. The ICPC detector has significant less reduction in activeness than PPC detectors. The activeness loss also does not depend on the surface charge and the same activeness is found for positive and negative surface charge.

Figure 5.7 shows the activeness maps for the Mirion ICPC detectors. The ICPC detectors have a smaller passivated surface and thus have a lower loss in activeness compared to the PPC detectors. In addition, these detectors have a ditch feature at the passivated surface location, increasing the vertical component of the electric field in this region. This decreases the likelihood that charges are pushed to the surface of the detector in this remaining passivated surface region. Figure 5.12 shows the electric field near the detector ditch. The field lines suggest that this loss is primarily due to electrons being pulled in the ditch.

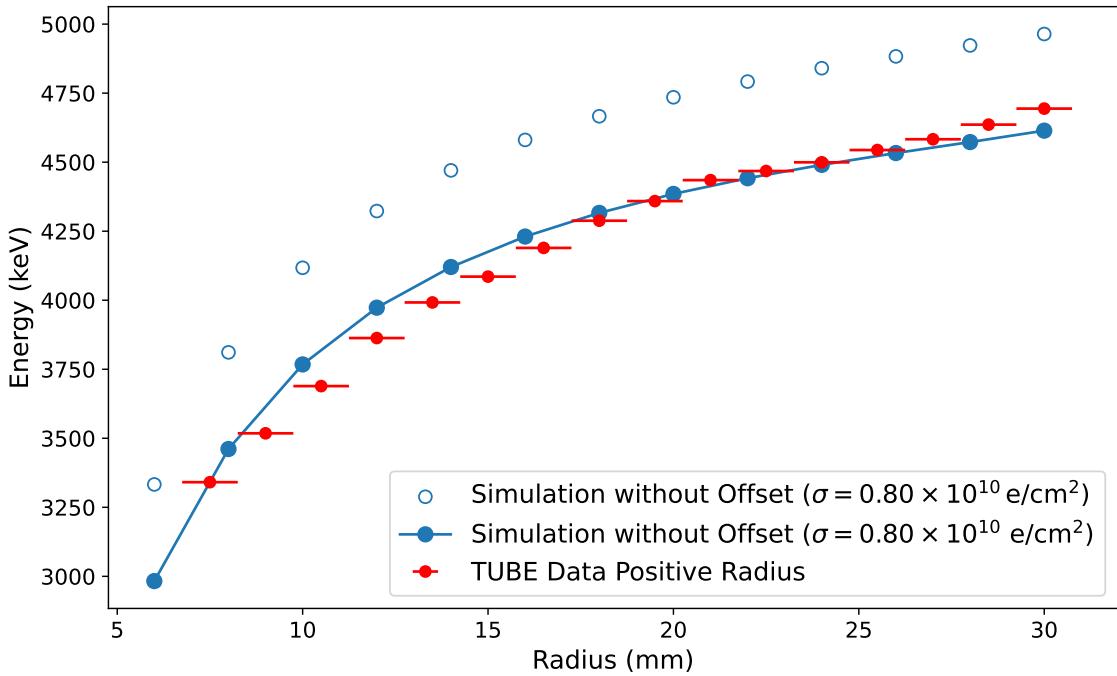


Figure 5.8: Fitting the activeness in the TUBE scanner with EH-Drift. A positive surface charge can help explain the increase in collected energy with radius. A 350 keV energy offset and 2.0 mm radial offset is needed to match the data, which could occur if there was a thin fully dead region either in the source or at the detector surface. Experimental values from [40]

### 5.3 Reproduction of Results from Established Test Stands

To validate EH-Drift against existing measurements, we modeled alpha events studied for PONaMa-1 detector operated in the TUBE [40] and GALATEA [41] scanning test stands. The teams that took these measurements hypothesized that the difference in behavior observed between them was due to opposite-sign charge build-up on the detector’s passivated surface. If this hypothesis is correct and the EH-Drift code is correctly modeling the charge collection behavior in this region, varying the surface charge in the model should allow for an accurate fit to both sets of data.

We simulated events using various surface charges in the PONaMa-1 detector. We then fit those different curves to the data to estimate the surface charge needed to match the behavior observed in each measurement.

We found a positive surface charge that explains the behavior observed in TUBE. Figure 5.8 compares the TUBE scanner data with EH-Drift simulations with a positive surface charge. A positive surface

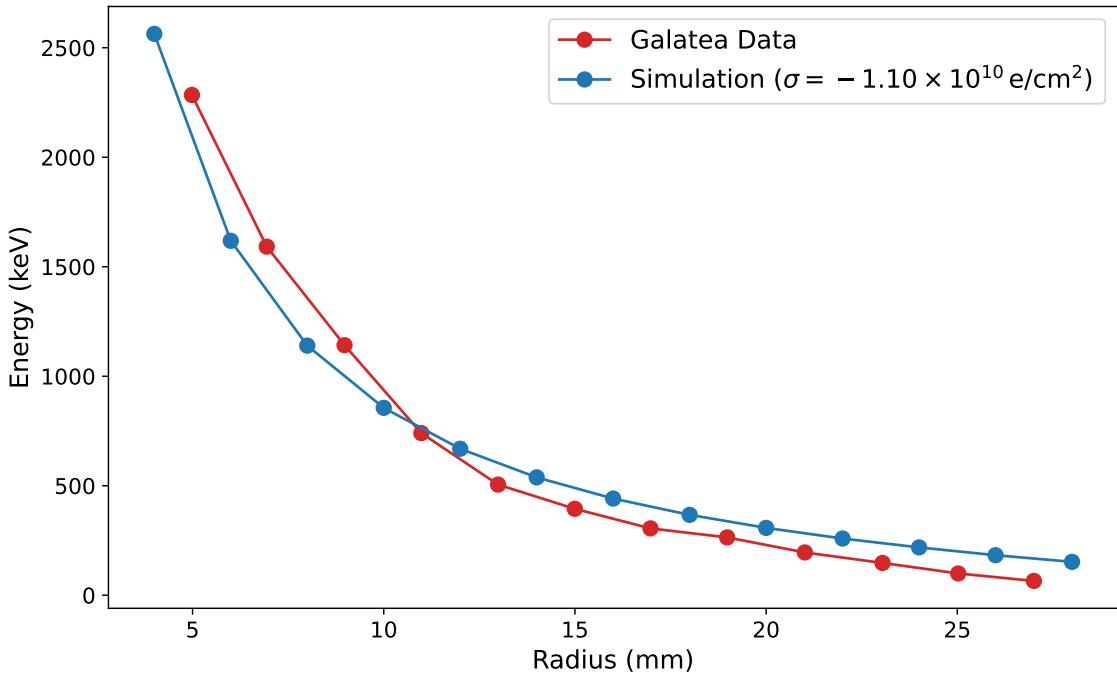


Figure 5.9: Fitting the activeness in GALATEA scanner with EH-Drift. A negative surface charge can help explain the decrease in collected energy with radius. Experimental values from [41]

charge would attract electrons on the surface, resulting in energy degradation. In this scenario, electrons are collected at the n+ surface, so the closer the points are to the origin, the more chance there is for the electrons to get trapped onto the surface and the more energy degradation there is. Thus, the energy collected is proportional to the radius. Although the simulations have matched the trend well, an offset of 350 keV in energy and 2.0 mm in radius offset is required to match the data. This could be because the source beam used in TUBE had an uncertainty of about 0.75 mm in radial position. There could also be a thin dead layer either within the source used or at the detector surface, before the passivated surface. We added an energy offset to account for this offset.

In the GALATEA scanner, the significantly different behavior can be explained using a negative surface charge as shown in figure 5.9. The negative surface attracts holes to the surface, and since the holes are collected at the point contact, located at  $r=0$ , the higher the radius, the more charges end up on the surface. Thus, the collected energy falls inversely with the radius.

## 5.4 Efficiency Estimations in the Region of Interest

We can use the maps created by EH-Drift to estimate the efficiency of  $0\nu\beta\beta$  event collection. LEGEND’s search for  $0\nu\beta\beta$  is conducted by fitting a peak to events at  $Q_{\beta\beta}$ , the shape of which is determined from the energy resolution of the detector. Only events falling in this narrow peak would be counted as  $0\nu\beta\beta$  signal events. However, if an event occurs close to the passivated surface, there is a chance that the collected energy would be degraded enough to not register as an event in the Region of Interest (ROI). Standard calculations of the detector’s active volume that account for the loss of efficiency from regions of partial charge collection do not include any passivated surface effects.

To understand this effect, we created activeness maps for detectors with a 2039 keV initial energy, which is the Q value of  $0\nu\beta\beta$  in Germanium. We did this for various surface charges. We then randomly sampled 20,000 points uniformly in the detector and estimated their activeness using the interpolated map. A threshold of 3 keV is defined for the ROI, which means that if the activeness at a point reduces the measured energy of the simulated  $0\nu\beta\beta$  event below 2036 keV, that event is not registered as a  $0\nu\beta\beta$  event. For this initial study, this threshold 3 keV is chosen as an approximation for the peak width used to fit  $0\nu\beta\beta$  events, since the average FWHM of the detectors in LEGEND is roughly 2.5 keV. Each point is assigned a value of 1 if an event at that location is counted as  $0\nu\beta\beta$ , or 0 otherwise. Figure 5.10 shows this binary activeness for a PPC and Mirion ICPC detector with negative surface charge. The overall efficiency is then calculated as the sum of activeness values divided by the total number of points, while factoring in the azimuthal angle dependence.

Surface Charge ( $\times 10^{-10}$ e/cm $^2$ )	V07647A (%)	PONaMa-1 (%)	P00698A (%)
-0.30	99.82	83.80	95.21
-0.15	99.82	87.78	97.25
-0.03	99.82	95.34	98.85
0.00	99.82	99.02	99.91
0.03	99.82	99.98	99.98
0.15	99.82	99.88	99.90
0.30	99.82	99.76	99.71

Table 5.2: Estimated  $0\nu\beta\beta$  efficiency for three different germanium detectors at different surface charges.

Table 5.2 summarizes the results of the study for the PONaMa-1 detector, a LEGEND PPC detector, and a Mirion ICPC detector. Figure 5.11 plots the efficiency against the surface charge. The efficiency

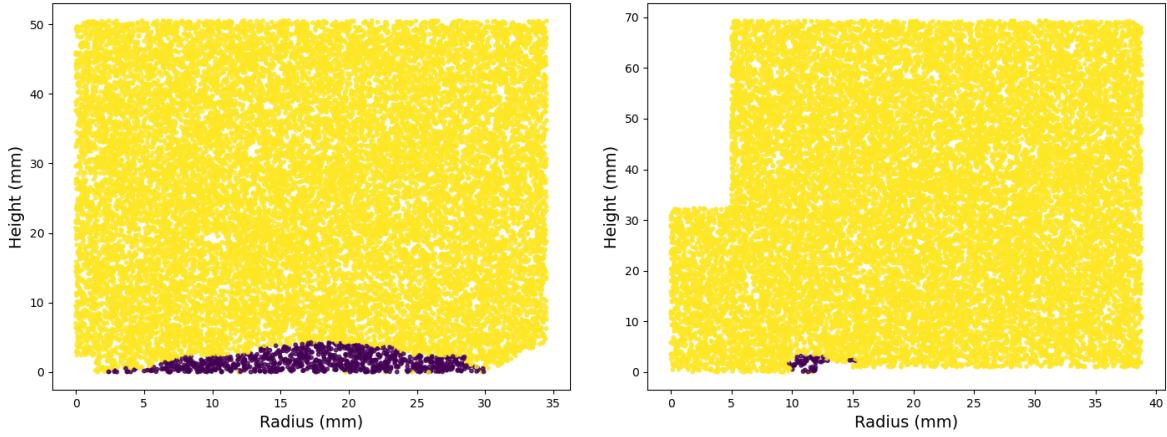


Figure 5.10: Binary activeness for PONaMa-1 detector and Mirion ICPC detector with negative surface charge. A 3 keV energy resolution threshold is defined. If energy is degraded above it, it is labeled zero, else 1. The purple dots show where  $0\nu\beta\beta$  events would be not recorded for the detectors.

drops rapidly for negative surface charges in both PPC detectors; however, it decreases most sharply for PONaMa-1, which has a lower depletion voltage and thus a weaker electric field near the surface. The LEGEND PPC detector also loses efficiency quickly due to its larger passivated surface, but its higher electric field moderates the overall loss more than in PONaMa-1.

The Mirion-style ICPC detector exhibits virtually no reduction in activeness, primarily because it has a much smaller passivated surface. The collection efficiency is 99.82%. To understand this slight reduction in efficiency, one can look at the electric field in the ditch of the detector shown in figure 5.12. The field lines indicate that the electron component, which drifts opposite to the field lines, near the surface will be pulled onto the ditch, which is passivated. This loss in activeness is much smaller than in other types of detectors. Since LEGEND-1000 plans to use only Mirion-style ICPC detectors, surface backgrounds in that experiment are expected to be significantly reduced.

## 5.5 Generating Background Spectra

`Geant 4` is a Monte Carlo-based toolkit that can simulate the passage of particles through matter. To demonstrate how EH-Drift can be used to create spectral components for background modeling, we performed a `Geant 4` simulation of the PONaMa-1 detector exposed to a planar 5 MeV alpha source. We recorded the resulting hits from alpha particles entering the detector. We then used the activeness map to estimate the energy that would be collected from surface effects. Figure 5.13 shows what the collected energy spectrum would look like for various surface charge scenarios. In particular, the spectrum for the

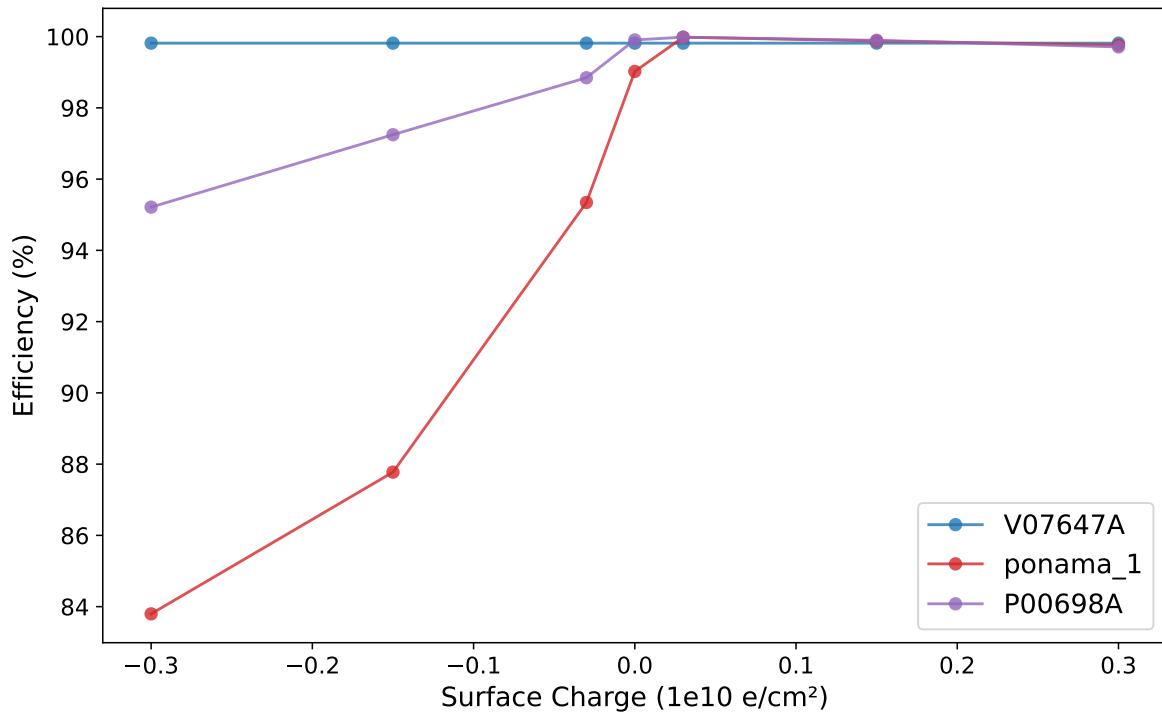


Figure 5.11: Efficiency in the ROI versus surface charge calculated using the binary activeness.

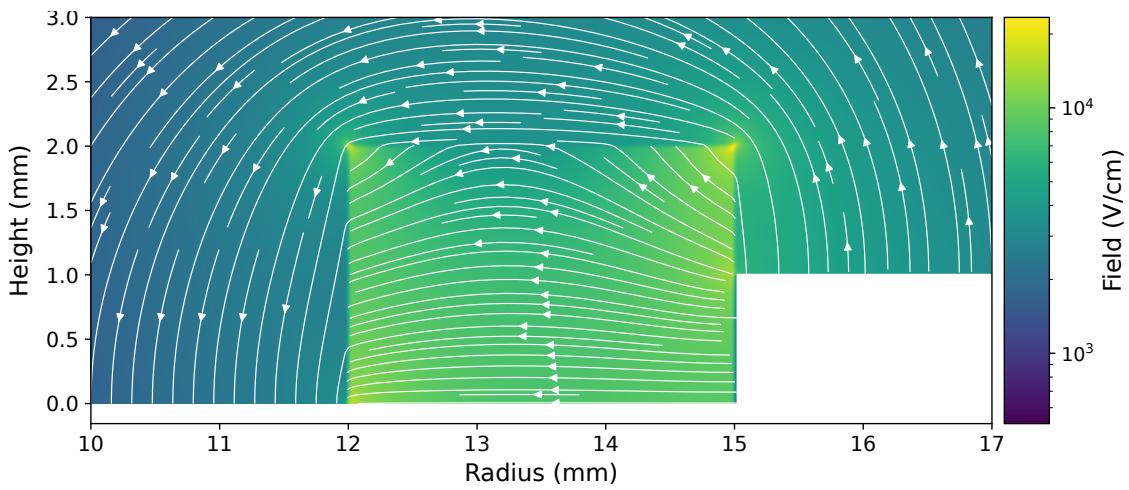


Figure 5.12: Electric magnitude and field lines near the ditch of a Mirion ICPC detector with negative surface charge. In Mirion ICPC, the flat  $p^+$  point contact, where the field lines end, extend to the start of ditch at  $r=12$ . The  $n^+$  contact starts at end of ditch at  $r=15$ , shown by the origin of field lines. The field lines attract electron component to the surface which drift slowly. The reduction in activeness is fraction of other detector geometries.

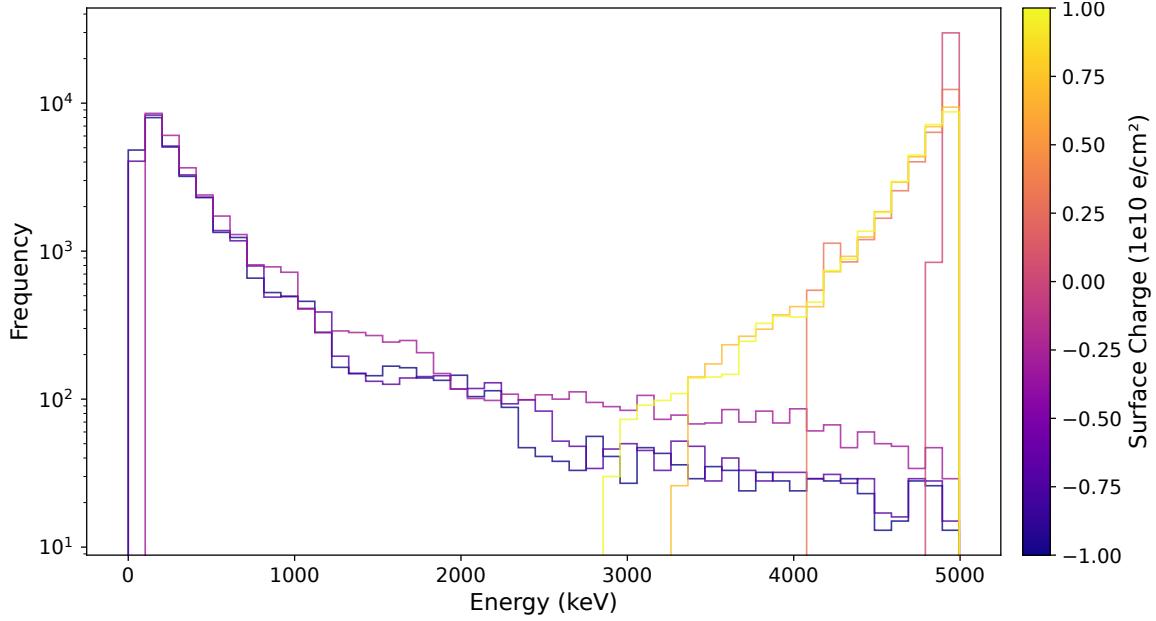


Figure 5.13: Degradation of the alpha-particle energy spectrum under various surface charges. The location of hits of alpha particles were generated using a Geant 4 simulation of planar alpha source of 5 MeV on PONaMa-1 detector. Energy degradation for an alpha event was calculated from interpolated map for each surface charge. The unit of surface charge is in  $10^{-10} \text{ e/cm}^2$ .

negative surface charge closely matches the alpha spectra observed in the MAJORANA DEMONSTRATOR experimental data and initial background modeling for PPC detectors in the LEGEND data.

To understand the effect of the passivated surface on beta decays such as those that form a significant background in LEGEND, we simulated the spectrum of  $^{42}\text{K}$  using Geant 4 for PONaMa-1 detector. The output energy spectrum is shown by the blue histogram in figure 5.14. The spectrum has a prominent gamma line at 1500 keV and a broad beta continuum extending to 3500 keV, which is the Q-value of the decay. Using EH-Drift we estimate the activeness of the passivated surface assuming a 2039 keV energy map and a surface charge of  $-0.5 \times 10^{-10} \text{ e/cm}^2$ . The resulting spectrum is shown by the yellow curve.

These two preliminary results, along with the dramatic active-volume effects for  $0\nu\beta\beta$  signal efficiency, show the importance of incorporating the results of the EH-Drift simulations into the LEGEND background model.

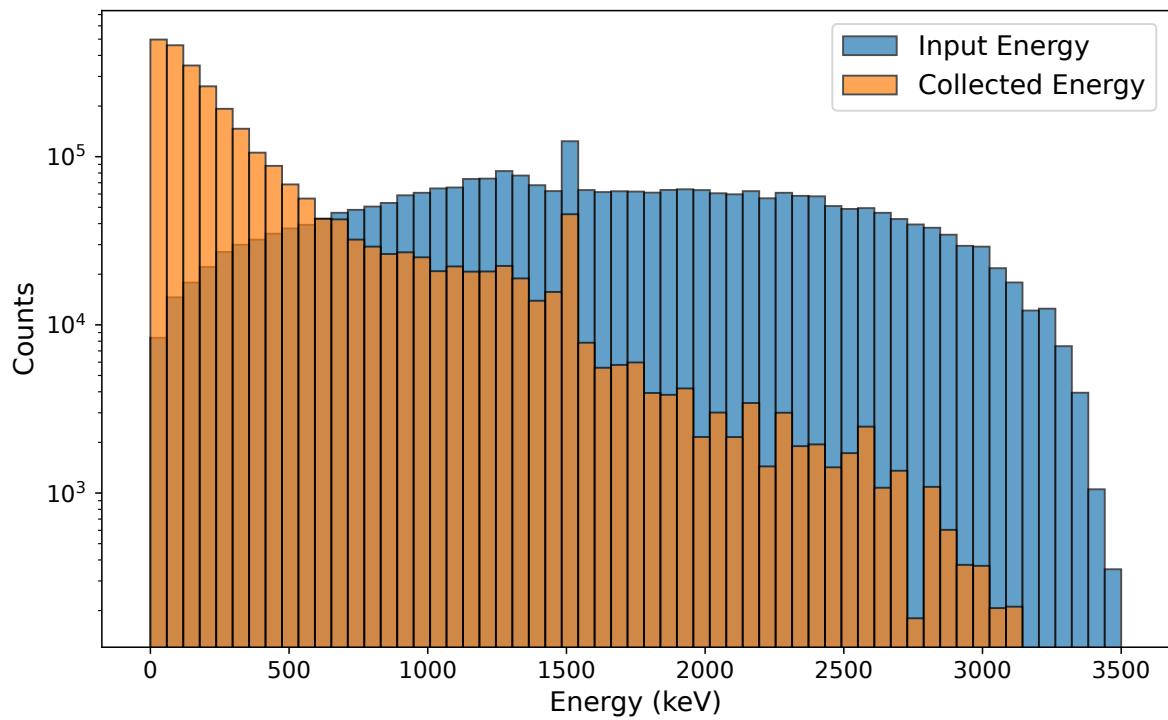


Figure 5.14: Degradation of the  $^{42}\text{K}$  energy spectrum in PONaMa-1 detector. The detector surface had a surface charge of  $-0.5 \times 10^{-10} \text{ e/cm}^2$ . The activeness was calculated assuming 2039 keV energy map. This result does not include a model for the dead layer created by the lithiated n<sup>+</sup> contact.

## CHAPTER 6: NEURAL NET BASED ELECTRONICS RESPONSE MODELING

### 6.1 Introduction

As discussed in chapter 2, HPGe detectors provide excellent energy resolution and good waveform-shape-based event identification, making them ideal for low-background physics searches such as  $0\nu\beta\beta$  searches. However, the signal from the HPGe detector must pass through an electronics readout chain consisting of components such as a preamplifier, filtering stages, and digitization. These electronics transform the signal in ways that can distort waveform shapes, alter the rise time, and add overshoots/undershoots or baseline shifts. Accurate modeling of this transformation is crucial for producing realistic pulse shape simulations.

This chapter begins with an overview of typical HPGe detector readout electronics. It is followed by an overview of LEGEND-200 electronics and how electronics introduce distortions in waveforms, called the electronics response. This is followed by challenges in modeling the electronics response and a neural network approach to address them.

### 6.2 LEGEND-200 Readout Electronics

LEGEND-200 uses a resistive feedback charge sensitive amplifier (CSA) to minimize noise and achieve a low radioactive background, as shown in figure 6.1. The first stage is termed the Low Mass Front End (LMFE). The LMFE is placed inside the liquid argon cryostat of the experiment, only a few millimeters from the signal contact of each HPGe detector [55]. The LMFE is therefore constructed with high radiopure materials with a tolerance typically on the order of  $1\mu\text{ Bq}$  per channel. The LMFE builds on the one developed by MAJORANA and is produced using Suprasil substrates, titanium-gold (TiAu) traces, an in-die junction field effect transistor (JFET, Moxtek MX11), and thin-film amorphous germanium feedback resistors that can operate at cryogenic temperatures. The first stage is connected to the second stage using four ultra-pure Axon pico-coaxial cables.

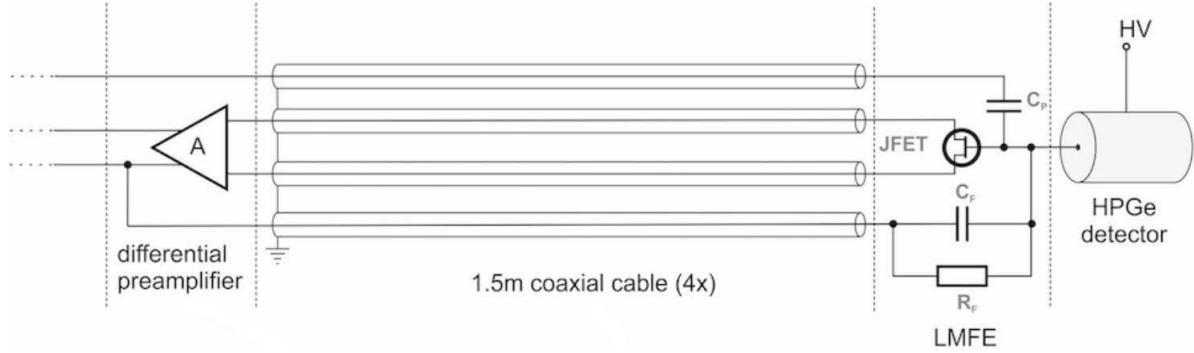


Figure 6.1: A schematic of the LEGEND-200 signal readout chain. [55]



Figure 6.2: Pictures of the readout electronics of LEGEND-200. Left shows a close up of the detector with the LMFE, middle shows the detector strings and front end electronics, with the CC4s located at the top of the array. Below them are the connectors for the high voltage cable. Right shows the Flashcam crates in the DAQ rack. Photo: Michael Willers

The second amplifier is based on a folded cascode arrangement and is located 30 -150 cm from the detector to provide additional gain and convert the signal to a differential output. This stage can tolerate slightly higher radioactivity levels of  $50\mu\text{Bq}$  per channel and is built using commercial surface mount components on Kapton circuit boards called the CC4. The second stage is connected to the data acquisition system using a  $\sim 10\text{ m}$  long transmission line. An active buffer called the Head Electronics, located outside the cryostat, receives the signals from the second stage. It adjusts offsets and gain while providing impedance matching. Finally, the signals are fed into a FlashCam digitizer, where the signal is digitized and stored for analysis.

Together, these design choices enable sub-3 keV FWHM energy resolution in the ROI at  $Q_{\beta\beta}$  2039 keV while minimally contributing to the overall radioactive background. They also enable a fast rise time of  $\leq 100$  ns, which allows for pulse shape discrimination of the signals. In addition, they have a high range of up to 10 MeV that enables the measurement of high-energy alpha decays, which would provide additional information for background modeling.

### 6.3 Electronics Influence on Waveforms

The LEGEND-200 electronics chain inevitably modifies the raw current signals from the detector. Figure 6.3 shows some of these effects by comparing a simulated waveform (blue) with a corresponding real data waveform (red).

The baseline is the region before  $t_0$ , the start of the waveform. Ideally, this baseline should remain near a stable DC offset determined by the quiescent operating point of the preamplifier. Two factors can shift or tilt the baseline. First, a slow high-pass filter effect (on the order of milliseconds) arises from the capacitive coupling between the first and second preamplifier stages. Second, if events occur in rapid succession, the tail of one waveform may not fully decay before the next arrives, causing the baseline to overshoot or undershoot. These shifts become especially noticeable at higher rates, leading to soft pile-up where partial overlap deforms the apparent starting level of subsequent waveforms.

The rising edge, shown in the orange region, results from the movement of holes and electrons in the bulk of the detector, from the initial energy deposition to the end of normal charge collection. The long cable and folded-cascode amplifier stages introduce a finite bandwidth, effectively acting as a low-pass filter that damps the highest-frequency components. As a result, the observed rising edge in real data tends to be broader and smoother than in simulations.

The RC decay tail (light blue region) develops from the CSA feedback network in the LMFE, where the detector current is integrated on a feedback capacitor and slowly discharged by the feedback resistor. The second stage of amplification has a shorter, differentiator-like time constant, further modifying the decay shape seen in the data. Together, these processes create a waveform with multi-exponential decay.

In general, the three waveform regions, the baseline, the rising edge, and the tail, are each influenced by different parts of the LEGEND-200 electronics chain. Circuit simulations, such as SPICE, can

reproduce some of the bulk features, but subtle responses, such as parasitic inductance, JFET nonlinearity, or thermal drifts, are hard to model.

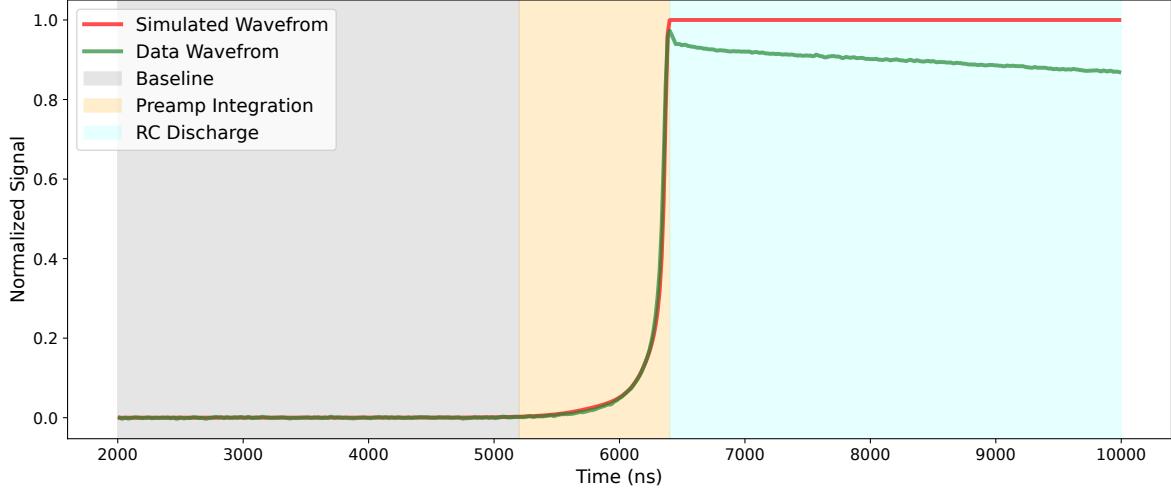


Figure 6.3: Comparison of a simulated waveform with a real data waveform. Both pulses were randomly sampled from datasets and are aligned for clarity. Three components of the waveforms are labeled. The data pulses contain electronics effects such as RC decay tail, electronics noise and preamplifier integration.

#### 6.4 Challenges in Modeling Electronics Response

Pulse shape simulation aims to generate waveforms that are similar to the actual detector waveforms. Having a direct mapping between a simulated waveform and its corresponding incident particle would enable applying the same cuts to the simulations as to the data. This would help us understand the efficiency of cuts such as multi-site cuts [56].

A frequency-dependent response function must be added to the simulations to reproduce the detector response. The electronics transfer function describes the series of linear transformations on the detector output signal by the readout electronics components. The electronics response is typically derived from the circuit response to a step function input. However, the experimental conditions commonly used in low-background and cryogenic experiments, such as long cable runs and extended amplification chains, make it difficult to generate step-like waveforms close to the detector. This poses significant challenges in measuring the electronics response directly. Figure 6.2 shows the front-end electronics of LEGEND-200.

This contrasts with figure 6.1, which shows how, in practice, the realization of the electronics chain can be quite complicated.

Although simulations can account for some of these effects, they are inherently limited by assumptions. These assumptions often fail to capture the complexities of real-world behavior, such as detector-specific anomalies, nonlinearity in the amplification chain, or subtle effects from experimental conditions. Furthermore, inaccuracies in simulation assumptions, such as incorrect drift velocity calculations in the electric field or oversimplified charge cloud dynamics, can exacerbate discrepancies between simulated and real detector waveforms. The result is a persistent mismatch between the simulated and measured waveforms, necessitating ad hoc corrections.

Heuristic methods are often used to approximate the electronics response. Current experiments avoid detailed electronics modeling by calculating reconstruction features directly from Monte Carlo particle-interaction simulations, relying on heuristic methods to meet most of their simulation needs [57, 58]. These methods fall short in their ability to account for detector-by-detector variations and operational changes in experimental conditions, which can evolve over time. This lack of adaptability introduces systematic errors and limits the fidelity of pulse shape simulations.

The electronics response is a series of frequency-dependent transformations applied to each waveform, and deep learning could provide a method to learn these transformations. Transfer learning is an unsupervised learning approach to learn the translations between the simulated waveform and the corresponding detector waveform. In the next section, we present a new neural network model called Cyclic Positional U-Net (CPU-Net) that allows learning the translations from the data and applying them to simulations without explicit programming. To correctly model the electronics transformation, a transfer learning model needs to meet two requirements. First, it must preserve underlying physical features like multi-site or single-site waveform. Thus, a single-site waveform must be translated to a single-site waveform, and a multi-site waveform must be translated to a multi-site waveform. Second, it must reproduce ensemble distributions of key pulse shape features, such as tail slope and current amplitude distribution. In the next sections, we discuss two neural networks that we designed for this purpose.

## 6.5 Deep Learning Approach to Electronics Modeling

### 6.5.1 Mathematical Formalism

The following formalism was developed by Aobo Li in Reference [59]. The task of learning the transfer function of electronics from data and adding its effects to the simulation can be reformulated under the transfer learning framework. The goal is to translate simulated waveforms into real waveforms by learning the differences between them.

In this approach, the source domain  $\mathcal{D}_{Source}$  represents the simulated waveforms. It includes features of the simulated waveforms, such as energy, maximal current amplitude, tail slope, etc. These features form a feature space ( $\mathcal{Y}$ ) and follow a probability distribution ( $P(Y)$ ). The source task refers to the relationship between the reconstruction features ( $Y$ ) in the source domain and the simulated waveforms ( $\mathcal{X}$ ). This is captured by the conditional distribution  $P(X|Y)$ , which describes how each parameter set ( $Y$ ) is assigned to a specific simulated waveform ( $X$ ). For instance, if we know the energy and rise time of a waveform, the source task models how these features produce a simulated waveform. Mathematically, we write the source domain as follows:

$$\mathcal{D}_{Source} = \{\mathcal{Y}, P(Y)\} \quad (6.1)$$

Then the reconstruction features are written as:

$$Y = \{E, I_{max}, c_{tail}\dots\} \in \mathcal{Y} \quad (6.2)$$

The source task is written as:

$$\mathcal{T}_{Source} = \{\mathcal{X}, P(X|\mathcal{Y})\} \quad X \in \mathcal{X} \quad (6.3)$$

Similarly, target domain  $\mathcal{D}_{Target}$  represents the detector data waveforms. It contains the feature space  $\mathcal{Y}'$ , whose elements follow the probability distribution  $P(Y')$ , and  $Y'$  are the reconstruction features of the detector waveforms ( $X'$ ).

$$\mathcal{D}_{Target} = \{\mathcal{Y}', P(Y')\} \quad (6.4)$$

$$\mathcal{T}_{Target} = \{\mathcal{X}', P(X'|\mathcal{Y}')\} \quad (6.5)$$

Pulse shape simulation typically learns  $P(X'|Y')$  and applies it to an arbitrary  $Y$  in  $\mathcal{D}_{Source}$  so that the generated simulated waveform ( $\mathcal{X}$ ) is similar to the data waveforms  $\mathcal{X}'$ . For example, one can fit the tail slope of the detector waveforms to determine the decay constant and add a pole with that time decay to the simulation to replicate the data.

This method requires a complicated collection of modeling and characterization data, along with computationally expensive fitting procedures in a highly degenerate high-dimensional parameter space [57, 58]. Instead, we can avoid the direct modification of  $P(X|Y)$  by introducing a simulation translator. We call it the Ad-hoc Translation Network (ATN) represented by  $\Lambda$ :

$$\Lambda = \{\hat{\mathcal{X}}, P(\hat{X} | X)\} \quad \hat{X} \in \hat{\mathcal{X}} \quad (6.6)$$

ATN accepts an input waveform  $X$  and translates it to an output waveform  $\hat{X}$ . This transformation is learned from a large sample of data waveforms. The collection of transformed output  $\hat{\mathcal{X}}$  should be very similar to  $\mathcal{X}'$  after training, so that by combining the ATN and  $\mathcal{T}_{Source}$ , we can replicate  $\mathcal{T}_{Target}$ :

$$\mathcal{T}_{Target} = \Lambda \mathcal{T}_{Source}, \quad (6.7)$$

Similarly, we define a data translator called the Inverse Ad-hoc Translation Network (IATN) represented by  $\bar{\Lambda}$ .

$$\mathcal{T}_{Source} = \bar{\Lambda} \mathcal{T}_{Target}, \quad (6.8)$$

which allows us to learn the features of  $\mathcal{T}_{Target}$  without explicit programming. For a model to be deemed accurate, it must fulfill two criteria: it must replicate the ensemble distributions of the dataset accurately and maintain the integral detector physics embodied within each waveform. In the next section, we discuss two neural networks that will help us train the ATN.

### 6.5.2 Positional U-Net

We explored a U-Net [60], a convolutional neural network initially developed for biomedical image segmentation, as a baseline model when designing the ATN. U-Net contains contracting (encoding)

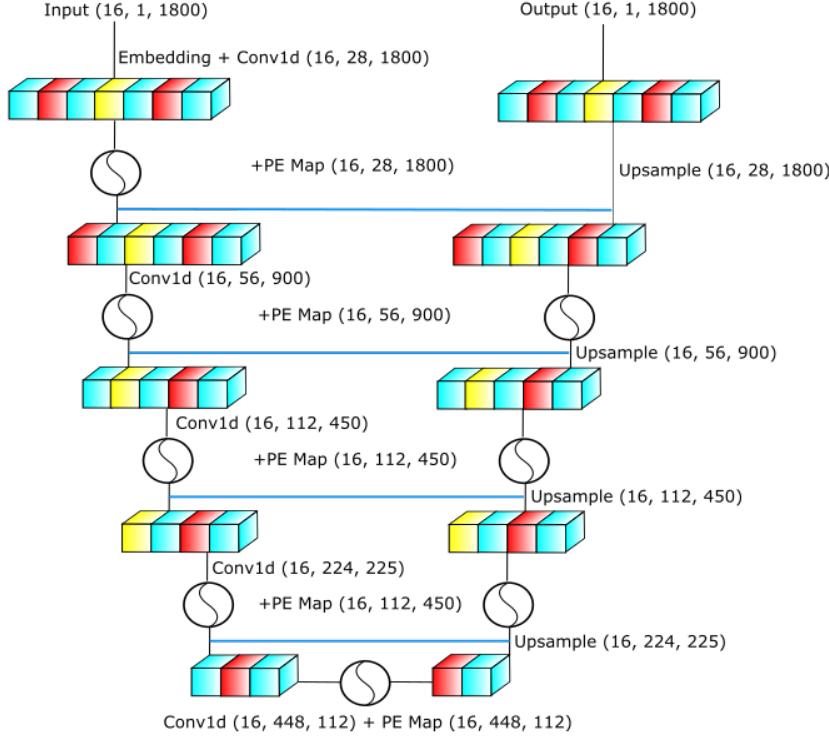


Figure 6.4: Layer-wise breakdown of the Positional U-Net. The blue lines represent the skip connections between contracting and expanding paths of the U-Net. Positional encoding layers are also shown for all levels. Credit: Aobo Li

and expansive (decoding) channels. The contracting path contains  $n$  convolutional layers to encode a waveform into a characteristic vector to capture contextual information and reduce the spatial resolution of the input. The expanding path contains  $n$  upsample layers to decode the feature vector back to an output with the same length. The network structure also allows information to flow at different levels to the decoding part to provide maximum reconstruction efficiency. The U-Net structure is depicted in figure 6.4, where the tensor shape is also denoted at each stage. The Conv1d module in figure 6.4 is a series of layers, whose breakdown is shown in figure 6.5. Within this module, the kernel size is an important hyperparameter to control the receptive field of CNN layers, and padding is added to guarantee the same input and output shapes. Max pooling is used in all Conv1d modules except for the first layer. Using max pooling reduces the feature map size, which increases the effective receptive field by focusing on the most essential features in the signal.

During initial testing, it was observed that the traditional U-Net does not reproduce the tail of the waveform because of the lack of positional information in intermediate layer outputs. The Transformer



Figure 6.5: Components of the Conv1d Module in the PU-Net from figure 6.4. Credit: Aobo Li

model [61] solves a similar problem by adding positional encoding  $\mathcal{M}_{\text{position}}$ , which contains sine and cosine functions with different frequencies to the output. The positional encoding in the Transformer model is given by:

$$PE_{(p,2i)} = \sin\left(\frac{p}{10000^{\frac{2i}{d_{\text{ch}}}}}\right) \quad (6.9)$$

$$PE_{(p,2i+1)} = \cos\left(\frac{p}{10000^{\frac{2i+1}{d_{\text{ch}}}}}\right) \quad (6.10)$$

where  $p$  is the position index of the token in the sequence,  $i$  is the dimension index, and  $d_{\text{ch}}$  is the dimension of input and output tensors. 10000 was the default value used in the transformer paper [61]. We added the position encoding to the layers of U-Net to preserve the positional information at each stage. Since each U-Net layer outputs a tensor with a different shape,  $\mathcal{M}_{\text{position}}$  must be generated separately for each layer and added to the layer output, as shown in figure 6.4.

Finally, we added a reparameterization trick to the bottom of PU-Net. The trick was developed in the Variational Autoencoder paper [62] to sample a random space while preserving the gradient flow. Reference [63] pointed out that this trick has the ability to increase the stochasticity of machine learning models. Based on our experiments, increased stochasticity helps the model learn the distributions of the reconstruction features better. We call this model Positional U-Net (PU-Net). Together, the PU-Net model has 7,213,781 trainable parameters.

### 6.5.3 RNN with Attention

RNN is an ideal model for tasks involving time series data because the position is intrinsically enforced by RNN. [64] An RNN consists of recurrent neurons in which each neuron has an output and a hidden state, which are passed to the next-step neuron. Thus, the RNN remembers the previous time state and uses it for the current output; however, the RNN can suffer from the vanishing gradient problem, where the gradients used to update the network become very small and do not allow it to learn long-range dependencies well across time steps. A Gated Recurrent Unit can help address this challenge by using "gates" that regulate the flow of information and gradients [65]. We developed a bidirectional RNN that adopts a Gated Recurrent Unit [65] as its internal structure.

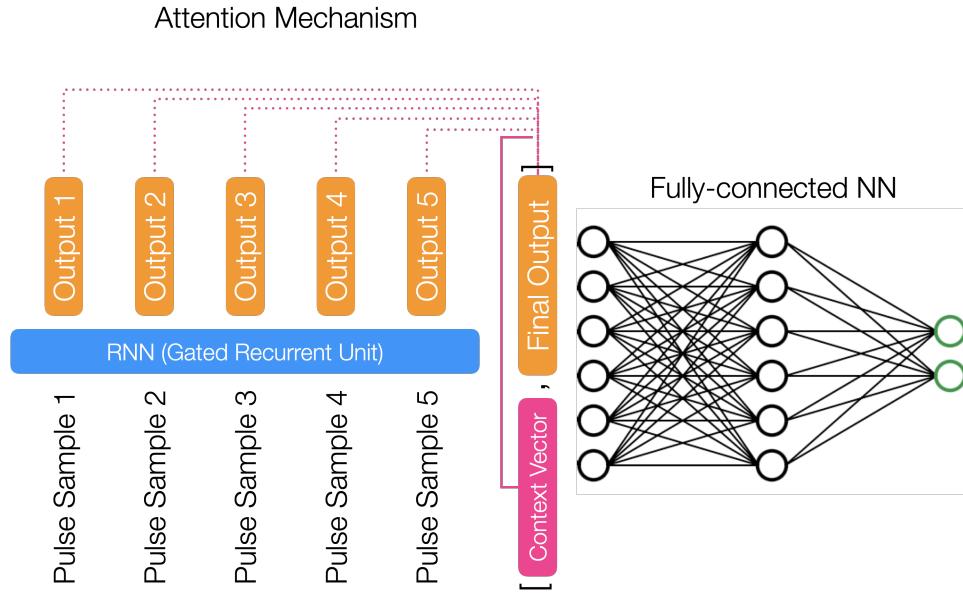


Figure 6.6: The attention coupled RNN discriminator. Waveforms are embedded into 128 dimension space and then passed to a GRU unit. The outputs of the GRU are used to calculate the attention score. Credit: Aobo Li

The RNN model is shown in figure 6.6. During training, the raw input waveforms are first embedded in the  $m = 128$  space. Embedding is done by initializing a lookup table that maps integer indices to dense vectors of fixed size. These vector representations are learned during training iterations to capture meaningful waveform features. We used an embedding trick that optimizes computation by directly retrieving the corresponding row from the embedding matrix instead of performing unnecessary matrix multiplications.

The embedded waveform is fed into a GRU that processes it sequentially while maintaining hidden states to capture temporal dependencies. This yields a 64-dimensional output  $\vec{I}(t)$  at each intermediate step  $t$  as well as a final output  $\vec{F}$ . We then use an attention mechanism [66] which can boost the performance of RNNs by allowing it to focus on different parts of the waveforms, such as the rising edge during different steps of the training. The attention mechanism contains an attention matrix  $A$  of dimension (64,64), which is used to calculate the attention score between  $\vec{F}$  and each  $\vec{I}(t)$ . This calculates how much each time step contributes to the final output:

$$s(t) = \text{Softmax}[\vec{I}(t)A\vec{F}] \quad (6.11)$$

A context vector is produced by summing  $\vec{I}(t)$  with the weight  $s(t)$  at each  $t$ . Finally, the context vector and the final output vector are concatenated and fed into a fully connected layer that produces a single scalar output. This model has 130,817 trainable parameters.

## CHAPTER 7: DATA PREPARATION & MODEL TRAINING

### 7.1 Introduction

The PU-Net and RNN layers provide a powerful way to learn the electronics response. Ideally, the ATN should be trained with paired waveforms of simulations and data. In reality, collecting such a paired dataset ranges from a major challenge to functionally impossible, depending on the desired population of events. Although we can simulate  $\mathcal{T}_{Source}$  from an arbitrary  $\mathcal{D}_{Source}$ , obtaining a corresponding  $\mathcal{T}_{Target}$  such that  $\mathcal{D}_{Source} = \mathcal{D}_{Target}$  is difficult without precise knowledge of  $P(X'|Y')$ . Consequently, training must be performed on unpaired datasets, where  $\mathcal{D}_{Source} \neq \mathcal{D}_{Target}$ . The CycleGAN framework [67] provides an unsupervised learning approach to train the ATN using adversarial losses. We call this model, which combines all the features, the Cyclic Positional U-Net (CPU-Net). CPU-Net was initially developed by Aobo Li and has been refined and improved upon for this thesis.

In this chapter, we describe the CPU-Net training process. First, we outline the detector data collection and processing. Then, we discuss how we obtained the simulation dataset. Finally, we detail the training of the network using CycleGAN.

### 7.2 Data selection

The LEGEND Collaboration characterizes a subset of newly produced HPGe detectors at Oak Ridge National Laboratory (ORNL). We used characterization data from a LEGEND Inverted-Coaxial Point-Contact (ICPC) detector, V06643A, manufactured by ORTEC. These characterization measurements were made by Morgan Clark as part of her dissertation [68]. The detector was mounted in a PopTop configuration, surrounded by lead shielding, and placed in a concrete alcove to reduce the external background. The data were taken with a flood Thorium-232 source placed on top of the detector. The detector signals were digitized using a FlashCam digitizer, and the pygama package [69] was used to convert the FlashCam output to LEGEND HDF5 files. The event energies were calibrated to convert

the ADC units to keV. The energies were then corrected for charge trapping. Figure 7.2 shows the final energy spectrum obtained from the detector.

The escape peaks in the spectrum are created from pair production events by a high-energy photon. In this process, the incident photon generates an electron-positron pair. The electron is collected by the detector, while the positron quickly annihilates, producing two additional gamma-ray photons. If all secondary gamma rays are fully absorbed within the detector, their combined energy yields a characterized peak in the energy spectrum, the Full Energy Peak (FEP).

If one of the annihilation gamma rays leaves the detector without depositing its energy, the spectrum shows a Single Escape Peak (SEP). This scenario results in a multi-site event because the signal comprises two distinct interactions within the detector: the collection of the pair-produced electron and the absorption of one of the two gamma rays. If both gamma rays escape the detector without interaction, a Double Escape Peak (DEP) is observed. This situation corresponds to a single-site event since the only interaction that contributes to the signal is the collection of the pair-produced electron. Tl-208 is part of the decay chain of  $^{228}\text{Th}$  and produces an FEP at 2614.53 keV. Given that the mass of the electron is 511 keV/c<sup>2</sup>, the SEP peak occurs at 2103.53 keV/c<sup>2</sup> and the DEP peak at 1592.53 keV/c<sup>2</sup>.

The FEP peak is an excellent region for training, given that it has a lot of events and contains a mixture of single- and multi-site events. SEP and DEP peaks can then be used as a validation data set, evaluating the model's response to multi-site and single-site events, respectively.

### 7.3 Simulations

We designed a simple Geant 4 setup geometry to simulate events in the ORNL characterization setup. The setup in Geant 4 is shown in figure 7.1. This includes a Germanium detector, a radioactive source, an Aluminum PopTop cryostat holder, surrounded by lead shielding. We simulated 100 million  $^{228}\text{Th}$  decay events originating from the source and recorded their energy depositions within the germanium detector. Given the location of energy deposition within the detector, we used `siggen` simulations to generate waveforms for hits in the DEP, SEP, and FEP energies. Figure 7.2 also shows the simulated spectrum. The simulated energy spectrum was obtained by summing the energy depositions for a given event. The `siggen` simulations used do not include any passivated surface effects. The `siggen` simulations used treat every energy deposit as a single point charge and heuristically account for diffusion

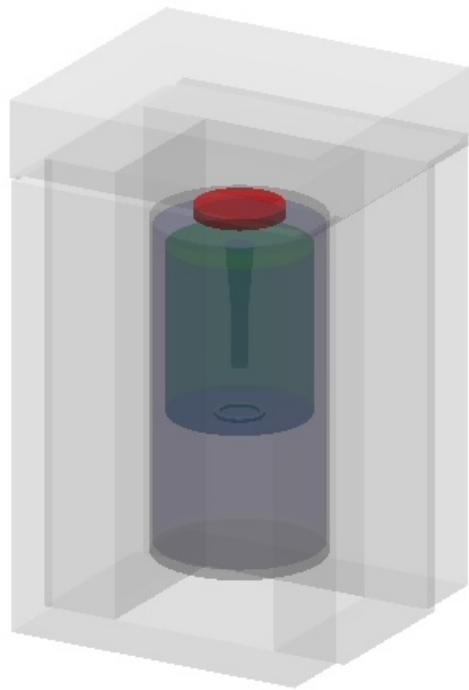


Figure 7.1: The simulation geometry of the ORNL characterization setup, showing the detector in green, the source in red, the aluminum holder in dark grey, and the lead shielding in light grey.

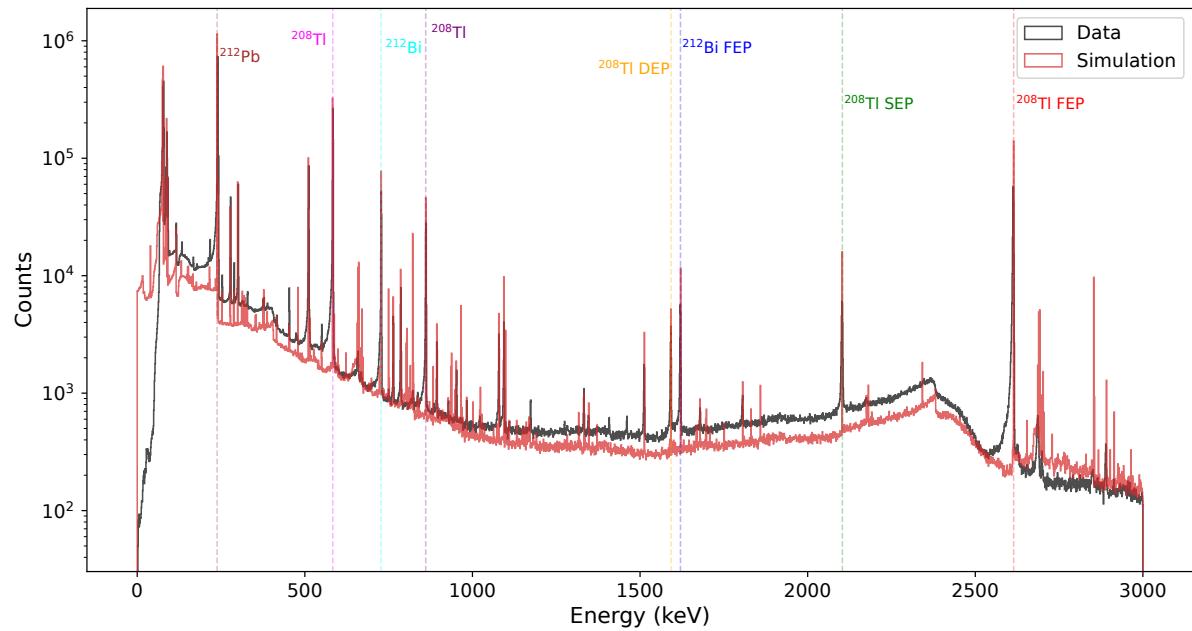


Figure 7.2: Calibrated energy spectrum from a  $^{228}\text{Th}$  source at ORNL compared to the Monte Carlo simulation results. Key peaks from  $^{228}\text{Th}$  source are labeled.

and self-repulsion by convolving the output signal with a 0.1 mm FWHM Gaussian. We disabled the preamplifier integration by setting the shaping constant  $\tau$  to zero.

Waveforms for multi-site events were created by energy-weighted summing of waveforms from individual hits. Figure 7.3 shows the simulation process for a few events. The left panel depicts the hit locations of the particles for the events. Event 3 and Event 4 are effectively single-site, as the energy depositions occur very close to each other and thus produce single-site events. In Event 1, the energy depositions result in two primary sites, giving rise to a two-site waveform. Event 2 features a trace of energy deposition localized to three regions, which is classified as a three-site event. Events 3 and 4 have different drift times because they are deposited at different distances from the point contact.

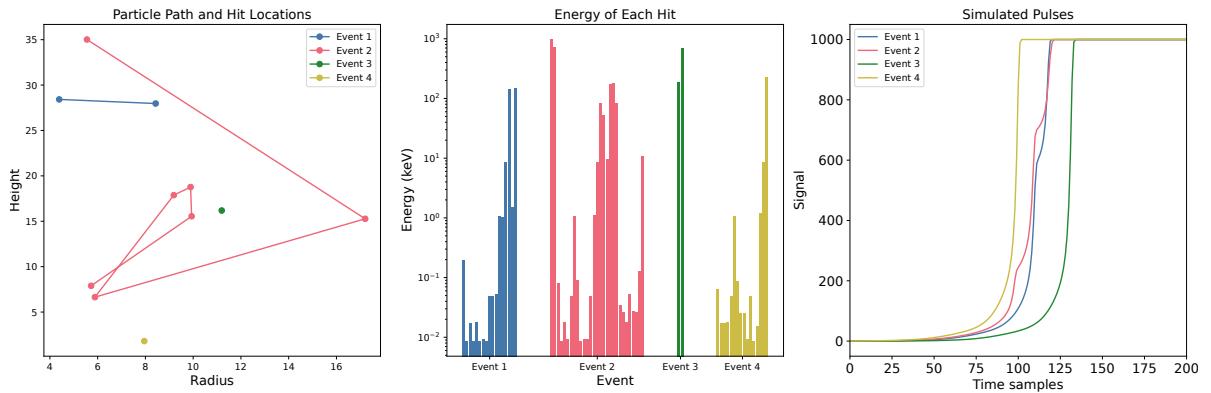


Figure 7.3: Production of energy weighted waveform in simulations for four events. Left figures show the location of hit locations for each event. The magnitude of energy deposited for hits in each event is shown in middle figure. Right plot show the energy-weighted waveforms for the corresponding events. Hits produced using Geant 4 simulations and waveforms are generated using `siggen` software.

## 7.4 Post Processing

Normalizing and aligning the waveforms is crucial for training the network. The simulated waveforms are already normalized between 0 and 1. The raw data waveforms are normalized by dividing by the 80% of the average of the last five samples. Both waveforms are then vertically shifted so that the average of the first 200 samples is zero. This ensures that all waveforms have their RC decay tails and baseline aligned, which can help the model learn the features better. Then we used the 99.9% rise time to horizontally align the waveform. We found that this method was more effective for training the network than aligning by the zero time point, as the waveform shape at the end of the rise time is far more consistent between events than the shape corresponding to the beginning of the charge drift, which

depends on the position of the energy deposition. The simulated waveforms are padded to ensure that there are 400 samples on both sides of the 99.9% rise time. The tail slope  $\tau$  is calculated by the slope of a linear fit of the logarithm of the last 300 waveform samples. Events with poor fit quality  $\chi^2$  or anomalous tail slope  $\tau$  are used to identify pile-up waveforms and remove them from training. Figure 7.4 shows the processed data and simulated waveforms that are fed into the neural networks.

## 7.5 Network Training

The training and validation are performed in PyTorch [70]. We construct two networks: an ATN  $\Lambda$  and an inverse ATN  $\bar{\Lambda}$ , both with the PU-Net structure. We then construct two RNN discriminator networks: a simulation discriminator  $\delta_S$  and a data discriminator  $\delta_T$  for the source and target waveforms, respectively. Figure 7.5 shows the overall training process.

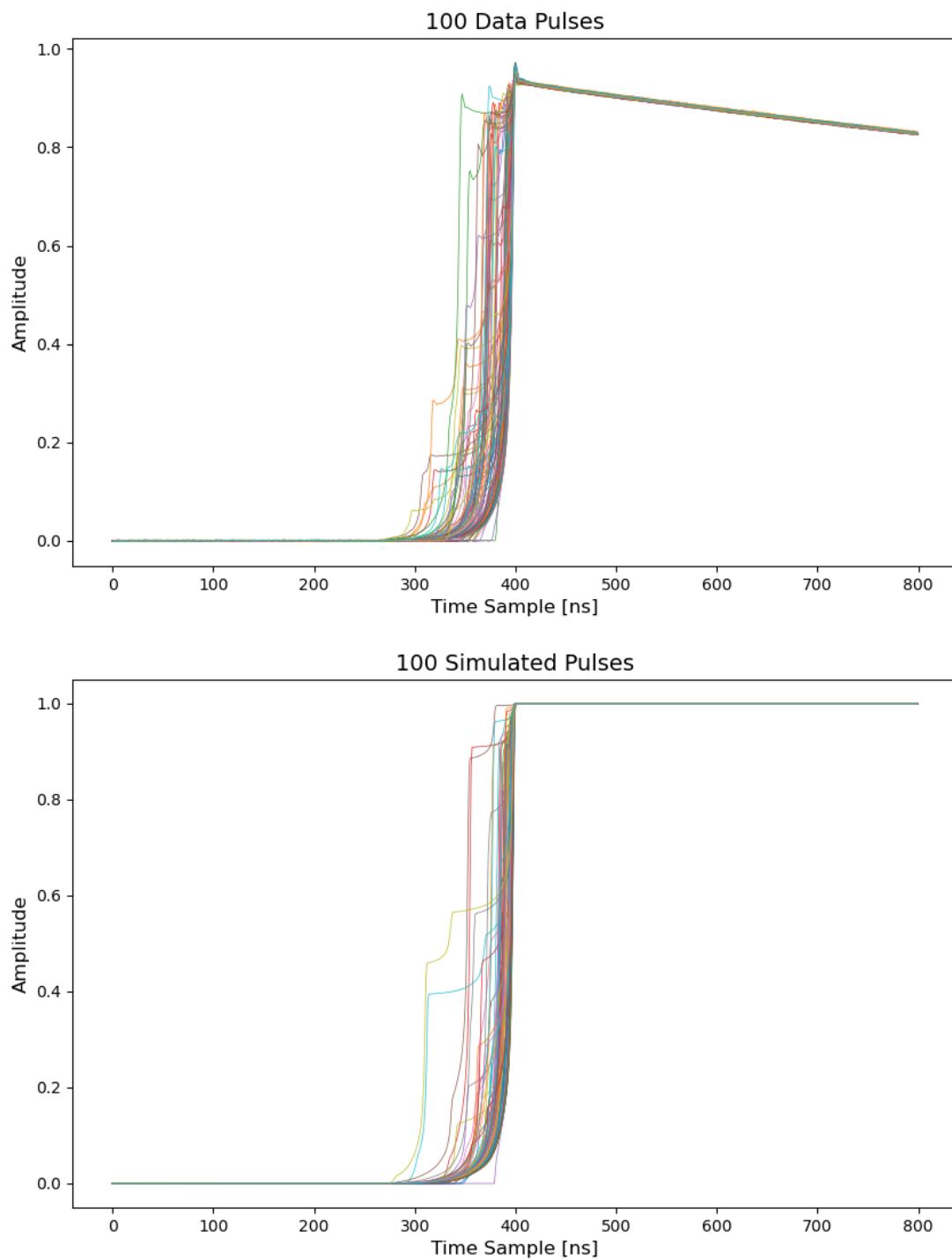


Figure 7.4: Input data and simulated waveforms. The waveforms are aligned by 99.9% rise time. Data pulses are normalized by aligning the tail and baselines.

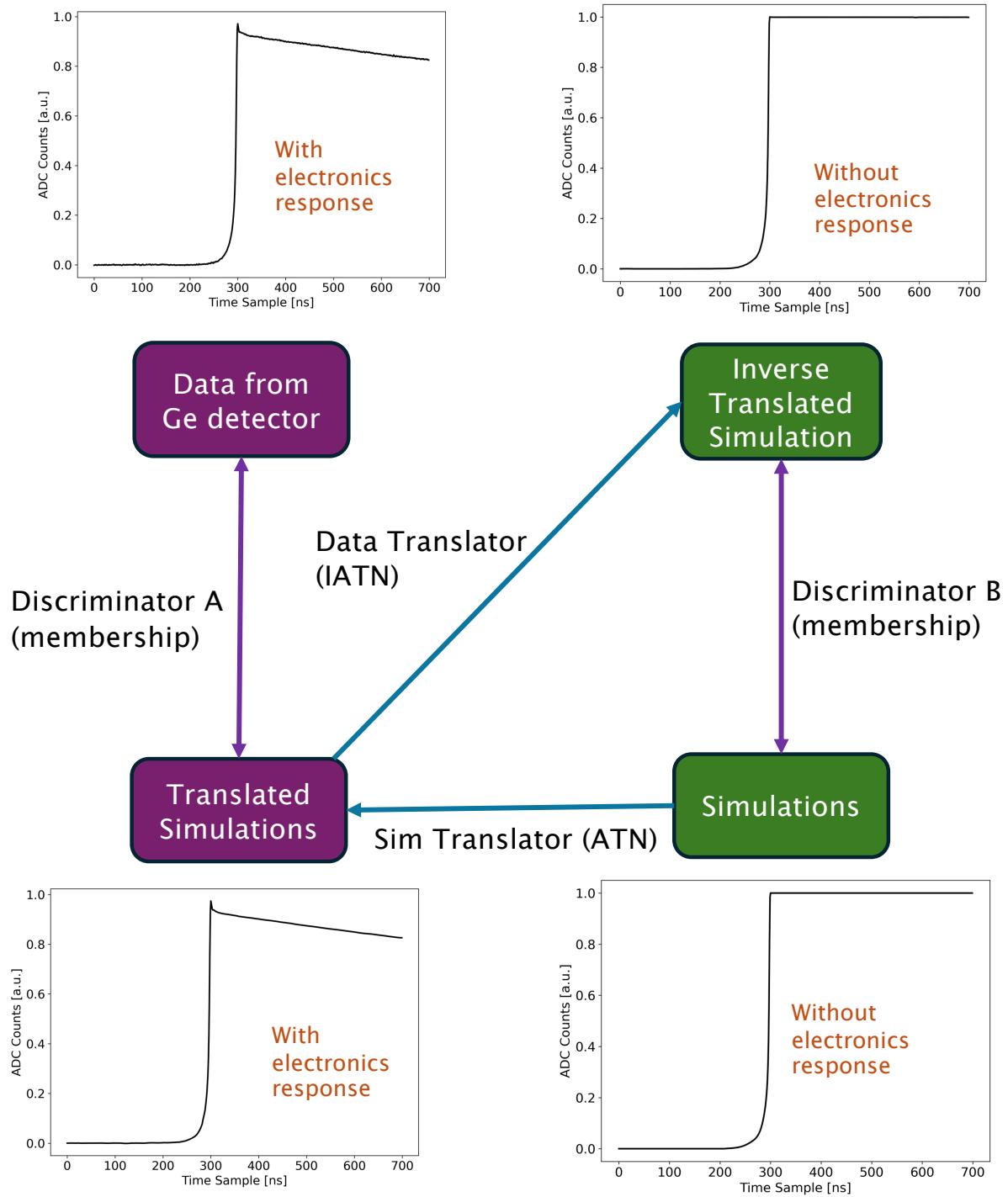


Figure 7.5: Training in CPU-Net. ATN transfers simulated waveforms to data-like waveforms, while the IATN transfers data waveforms to simulation-like waveforms. Discriminator A differentiates data-like waveforms outputs from the ATN from true data. Discriminator B differentiates simulation-like waveforms outputs from IATN from true simulations.

During training, a simulated waveform  $X$  is first fed to  $\Lambda$  to produce a translated waveform  $\Lambda(X)$ . The discriminator  $\delta_T$  attempts to distinguish  $\Lambda(X)$  from real data waveforms, while  $\Lambda(X)$  attempts to ‘fool’  $\delta_T$ . Then  $\Lambda(X)$  is fed to  $\bar{\Lambda}$  to translate back to  $\mathcal{X}$  space by  $\bar{\Lambda}(\Lambda(X))$ . A second discriminator  $\delta_S$  attempts to distinguish  $\bar{\Lambda}(\Lambda(X))$  from  $\mathcal{X}$ , while  $\bar{\Lambda}$  attempts to ‘fool’  $\delta_S$ . The  $\mathcal{X} \rightarrow \Lambda(\mathcal{X}) \rightarrow \bar{\Lambda}(\Lambda(\mathcal{X}))$  translation path is termed the forward cycle. The same process is performed in the other direction, starting with the detector waveform  $\mathcal{X}' \rightarrow \bar{\Lambda}(\mathcal{X}') \rightarrow \Lambda(\bar{\Lambda}(\mathcal{X}'))$  and called the backward cycle.

## 7.6 Loss Functions

Since we do not have a labeled training sample of matched data and simulation waveforms, we cannot use a direct comparison loss. Instead, we define three losses that are optimized during training to help the network learn the translations.  $L_{\text{Identity}}$  ensures that the identity relationship such that when a real detector waveform  $X'$  is fed into  $\Lambda$ , whose goal is to produce a detector-like waveform, the waveform should not be modified by  $\Lambda$ . This is shown in the equation 7.1.

$$L_{\text{Identity}} = |X' - \Lambda(X')| \quad (7.1)$$

The cycle-consistent loss in equation 7.2 ensures that the circular translation path preserves the original waveform shape. The forward cycle is when the waveform is fed into ATN to produce a detector-like waveform, which is then passed through IATN to produce a simulation-like waveform. This should return the original simulated waveform, and the cycle loss calculates the loss associated with this whole cycle. Similarly, the backward cycle loss is when data is passed through IATN to get a simulation-like waveform, which is passed through ATN to get the original data waveform back. The cycle-consistent loss acts in place of a direct comparison loss to ensure that the underlying physical features that are encoded in a waveform (e.g. multi-site vs. single-site, or energy deposition location) are not destroyed by the translation networks.

$$L_{\text{Cycle}} = |X - \bar{\Lambda}(\Lambda(X))| \quad (7.2)$$

In both cycles and identities, the waveform is compared with another waveform. For such a comparison, we define a specialized mean absolute error (L1 loss) that emphasizes different parts of the

Loss	Calculation	Type of Loss	Optimizer
Identity Losses	Data - ATN(Data)	Custom L1 Loss	Optimizer 1
	Sim - IATN(Data)		
Cycle Losses	Data – ATN(IATN(Data))		
	Sim – IATN(ATN(Sim))		
Generator Losses	Disc B (IATN(Data)) Disc A (ATN(Sim))	Binary Cross Entropy	
Discriminator A	[Disc A(Data) - 1] + [Disc A(Sim) - 0]	Binary Cross Entropy	Optimizers 2
Discriminator B	[Disc B(Sim) - 1] + [Disc B(Data) - 0]	Binary Cross Entropy	Optimizers 3

Table 7.1: Overview of Loss Calculations, Loss Types, and Optimizers in CPU-Net training.

waveform by assigning them varying weights. It is designed to give more importance to the rising and falling edges of the waveform, which are critical for accurate pulse shape analysis.

The adversarial loss calculates the loss associated with the generator and discriminator ‘fooling’ each other. The discriminator takes in the waveform from the generator and outputs a single value between 0 and 1. A value close to 1 means that the discriminator classifies it as real data, while a value close to zero means that the discriminator classifies it as simulated waveform.

Suppose we start with a simulated waveform  $X$ , and let  $X' = \Lambda(X)$  be the output of the generator  $\Lambda$ . The adversarial loss for the generator is equal to  $\delta_T(X')$ , so that it is encouraged to produce a waveform that the discriminator classifies as data. The discriminator loss for  $\delta_T$  is computed as  $1 - \delta_T(X')$ , encouraging it to classify it as simulation. Both losses are quantified using the binary cross-entropy loss, as shown in equation 7.3:

$$L_{\text{Adversarial}} = E_{X'} \log(\delta(X')) - E_{\Lambda(X)} \log(1 - \delta(\Lambda(X))) \quad (7.3)$$

An additional three complementary losses are defined for the detector waveform translation path. Therefore, a total of eight losses are optimized during training: six are associated with the generators and two are associated with the discriminators. The loss calculations are summarized in table 7.1.

AdamW [71] optimizers are used for all losses. Compared to the traditional Adam optimizer, AdamW applies weight decay as a separate step during gradient descent optimization. This avoids interference with the learning rate schedule and helps stabilize the training process. In total, we define three optimizers,

Hyperparameter	Value	Description
batch_size	32	Number of pulses used in one training iteration.
baseline_len	200	Number of samples assigned to baseline portion of the waveform.
rising_edge_len	250	Number of samples assigned to the rising edge of the waveform.
tail_len	350	Number of samples assigned to the RC decay tail of the waveform.
baseline_weight	3.0	Weight given to baseline portion of the waveform in the loss.
ris_edge_weight	10.0	Weight given to rising edge portion of the waveform in the loss.
tail_weight	7.0	Weight given to RC decay tail portion of the waveform in the loss.
iters	7000	Maximum number of iterations for training.
decay	1000	Iteration at which learning rate starts to decay.
lrate_gen	$1 \times 10^{-3}$	Learning rate for the generator networks.
lrate_disc	$1 \times 10^{-3}$	Learning rate for the discriminator networks.
cyc_loss_weight	20	Weight of the cycle consistency losses.
iden_loss_weight	5	Weight of the identity loss.
gan_loss_weight	9	Weight of the generator loss.
max_grad_norm	100	Maximum gradient norm for gradient clipping.
w_decay	$1 \times 10^{-4}$	Weight decay in the optimizers.
n_disc_iters	30	Number of iterations after which the discriminators are updated.

Table 7.2: Hyperparameters used for CPU-Net training.

as shown in table 7.1. The six losses for ATN and IATN are optimized together, and the discriminators each have their own optimizers.

## 7.7 Hyperparameter Tuning

Achieving stability in CycleGAN training can be challenging. This is because the loss optimization process is complex as there is not one metric such as mean squared difference being optimized. Training must balance the learning progress of generators and discriminators while preventing gradients from exploding or imploding to zero. To improve training, we introduced hyperparameters at different levels of the model that are summarized in table 7.2.

It was found that the discriminator typically overpowers the generator, since the generator has a more complex task of generating waveforms while maintaining cycle and identity consistency and also fooling the discriminator. To balance this, we update the weights of the generators more frequently than the discriminators. We introduced a hyperparameter for the number of intervals after which the discriminator is updated. This allowed the generator enough steps to adapt to changes in the discriminator without destabilizing the adversarial process.

The weighting of the adversarial, cycle-consistency, and identity losses enables fine-tuning the learning of the generators. At some point, the generator will have learned just enough to fool the discriminator, and further learning might be hampered. The loss weights are used to optimize the generator to learn translation, such as the cycle and identity consistency even if it has been successful in overpowering the discriminator. These hyperparameters are carefully tuned to achieve a balance, as values that are too high will cause the gradient to explode; too low will cause the model not to learn much.

We also introduced a learning rate decay for the optimizers. In the beginning, the learning rate is intentionally kept high for the model to explore the entire parameter space. Then, the learning rate decays linearly to help it converge in the right direction. This is particularly important for the ATN optimizer, as it is optimizing six losses, and this ensures that all the space is explored.

To prevent overfitting during training, weight decay is applied to the optimizers so that it penalizes large weights and ensures generalization among all parameters. Gradient clipping is applied, limiting the norm of gradients during training and preventing exploding gradients, particularly in the layers of the U-Net. The threshold for clipping the magnitude of gradients is kept high since the loss is multiplied by different loss weights multiple times, which can increase the magnitude.

Together, these parameters enable precise fine-tuning CPU-Net training to obtain the right balance during training. A single training takes about 1 GPU hour on an NVIDIA A100 GPU. The trained CPU-Net generates both an ATN and an inverse ATN, enabling bidirectional translation between the simulation and data domains. The ATN is the primary focus of this work, but the reverse ATN can also enhance analysis by refining the waveform reconstruction. In the next chapter, we discuss some results from the ATN output.

## CHAPTER 8: CPU–NET RESULTS

### 8.1 Introduction

In order to have accurate simulations, the ATN needs to reproduce the correct ensemble distribution of the parameters used in PSD. Since  $0\nu\beta\beta$  events are single-site events, LEGEND seeks to distinguish between single-site and multi-site events using PSD techniques. We trained CPU–Net on 110,000 FEP waveforms and then used 1,200 SEP events and 3,000 DEP events for validation. In this chapter, we present the results of ATN’s ability to reproduce detector waveforms using some critical parameters used in PSD.

### 8.2 Training progression

The progression of training losses is shown in figure 8.1. The cycle-consistent and identity losses converge rapidly towards zero. The adversarial training is evident in the losses for the discriminators and generators. During training, the discriminator initially wins, and its loss decreases while the generator’s loss increases. As training continues, the generator catches up, and the networks approach an equilibrium in which each attempts to out-compete the other. Thus, the loss functions fluctuate as they continuously adapt and improve.

### 8.3 Waveform Translation

The translation of the simulated waveform is shown in figure 8.2. The ATN translates the simulated waveform to the ATN output by smoothing the sharp turning edge in the orange region. This is a consequence of the non-zero integration time in detector waveforms, which is set to 0 in `siggen`. The RC discharge effect of the electronics readout system is also set to 0 in `siggen`, leading to a tail slope of 0 in the simulated waveforms. The ATN learns to translate the flat tail in the cyan region into an exponential decay.

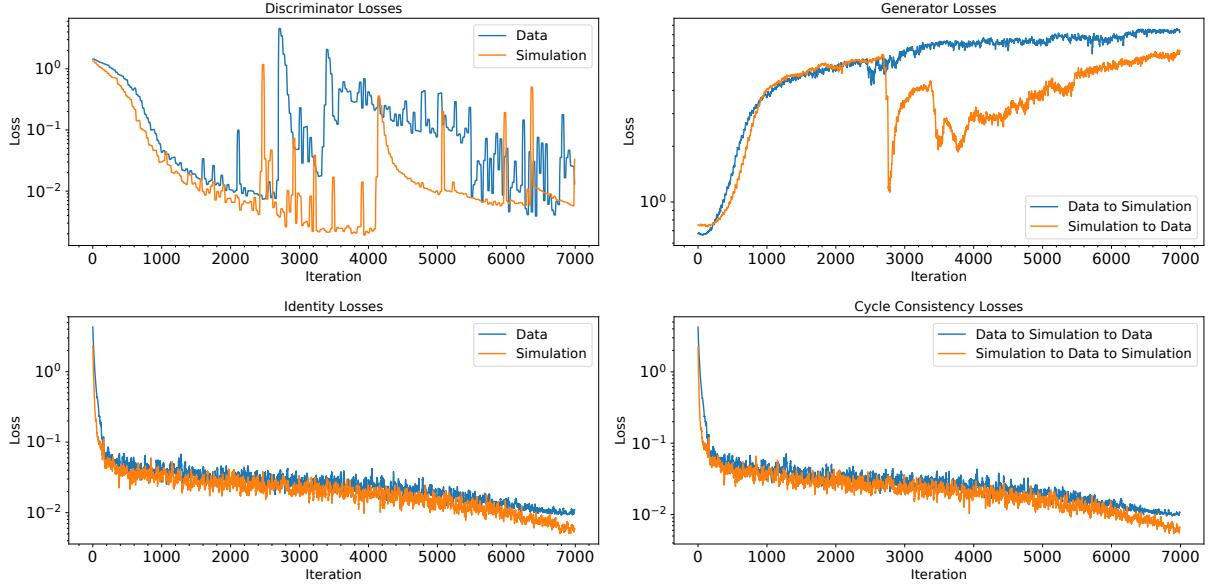


Figure 8.1: Training losses for CPU-Net. Curves are smoothed using a moving average of 10 samples for clarity. The identity and cycle losses rapidly converged to zero while the generator and discriminator losses fluctuate.

Figures 8.3 and 8.4 show the SEP waveforms as they progress through the cycle in CPU-Net. The network effectively translates the waveforms in forward and reverse translations. In the `siggen` simulations, the RC discharge effect is not modeled, resulting in a flat tail slope of zero. ATN learns to transform this flat tail into an exponentially decaying one while including more subtle effects like overshoot and a faster decay early in the tail, matching the observed behavior in real detector data, and IATN learns to transform it back to a zero slope, matching the simulations.

#### 8.4 Validation of Key Waveform Parameters

The accuracy of translations can be assessed at the distribution level by plotting the histogram of key reconstruction parameters. We use the Intersection over Union (IoU) metric, also known as the Jaccard index, to measure the overlap between distributions  $A$  and  $B$  [72, 73]. We first calculate the histogram of the distribution being compared using the same binning. The *intersection* is  $\sum_i \min(\text{bin}_i^A, \text{bin}_i^B)$ , the *union* is  $\sum_i \max(\text{bin}_i^A, \text{bin}_i^B)$ , and

$$\text{IoU} = \frac{\sum_i \min(\text{bin}_i^A, \text{bin}_i^B)}{\sum_i \max(\text{bin}_i^A, \text{bin}_i^B)}. \quad (8.1)$$

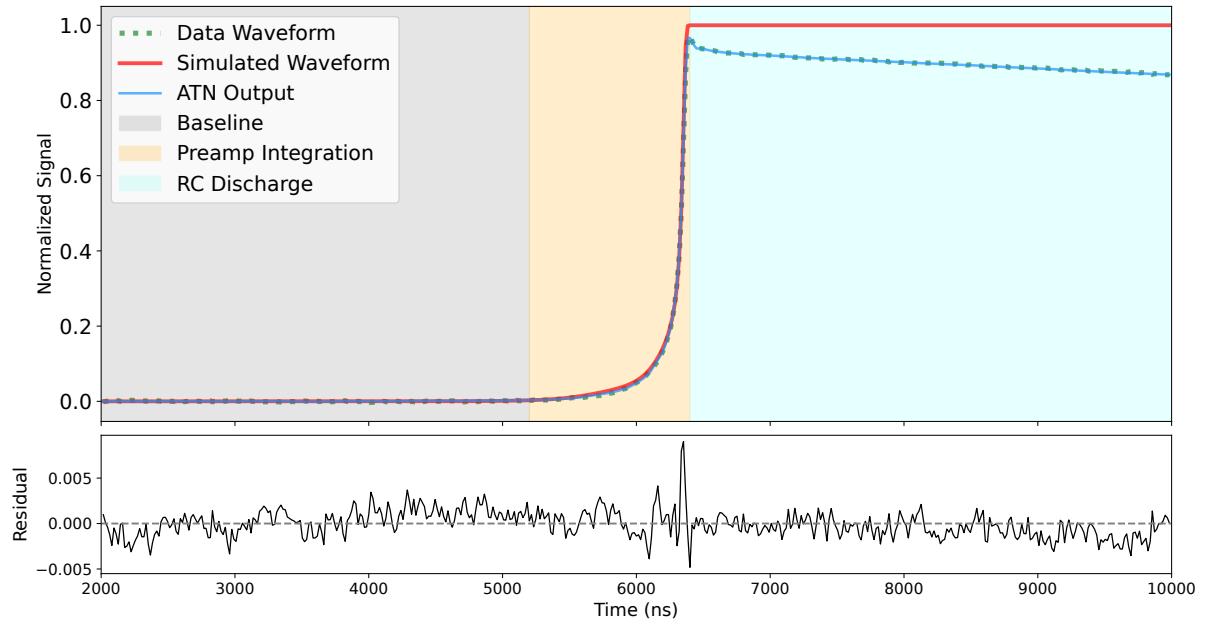


Figure 8.2: An example ATN output (blue) generated by the input simulated waveform (red). A detector waveform (dotted green), manually selected from data to resemble the ATN output, is included for reference. The orange region depicts the area where the preamplifier integration effect is most visible. The blue region shows the impact of the RC circuit discharge effect. Grey region shows the baseline region with electronics noise. A residual panel comparing the ATN output to the reference waveform is included below. A 0 residue is not expected, as the reference waveform is not a one-to-one match with the simulated waveform.

We express the result in percentages. An IoU of 100% indicates perfect agreement, and 0% indicates no agreement between the histograms.

#### 8.4.1 Drift Time Distribution

The electric field of the HPGe detector allows events at different locations to have unique waveform shapes, allowing reconstruction of the event topology. As shown in figure 7.3, each location has a different drift time, making it a crucial parameter for waveform analysis. Overall, data waveforms have a longer drift time than simulations because of the integration of the preamplifier. To calculate the drift time, we define two time points shown in figure 8.5. The metric  $tp_1$  is defined as the time the waveform reached 1% of its maximum value, and  $tp_{100}$  represents the time the waveform reached its maximum value. The time points are found by first finding the maximum time point and then searching backward to

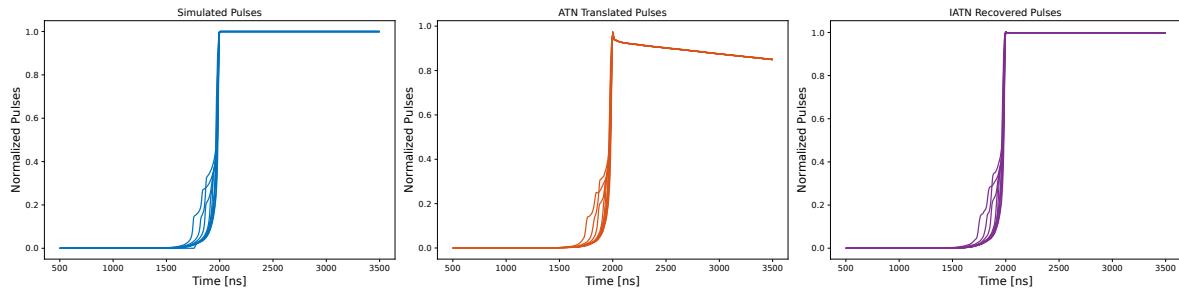


Figure 8.3: Cycle consistency in waveform translation in the forward direction. Starting with 10 simulated waveforms, the ATN first translates them into detector-like waveforms, and then IATN translates them back to the simulation like waveforms.

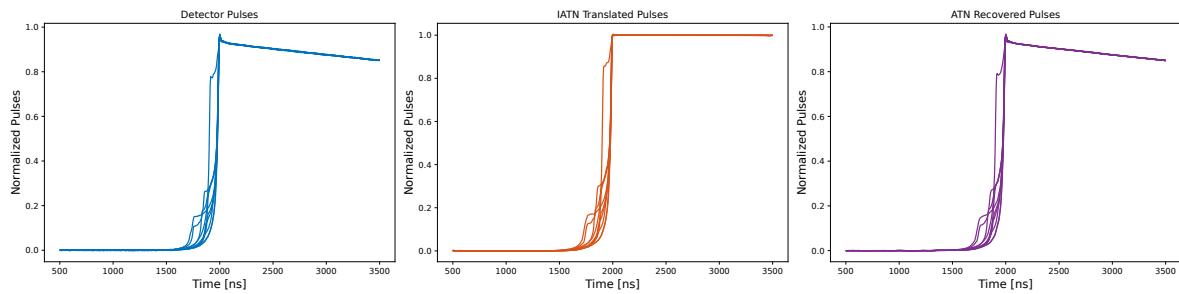


Figure 8.4: Cycle consistency in waveform translation in the backward direction. Beginning with 10 detector waveforms, the IATN translates them into simulated-like waveforms, and then ATN translates back to detector-like waveforms.

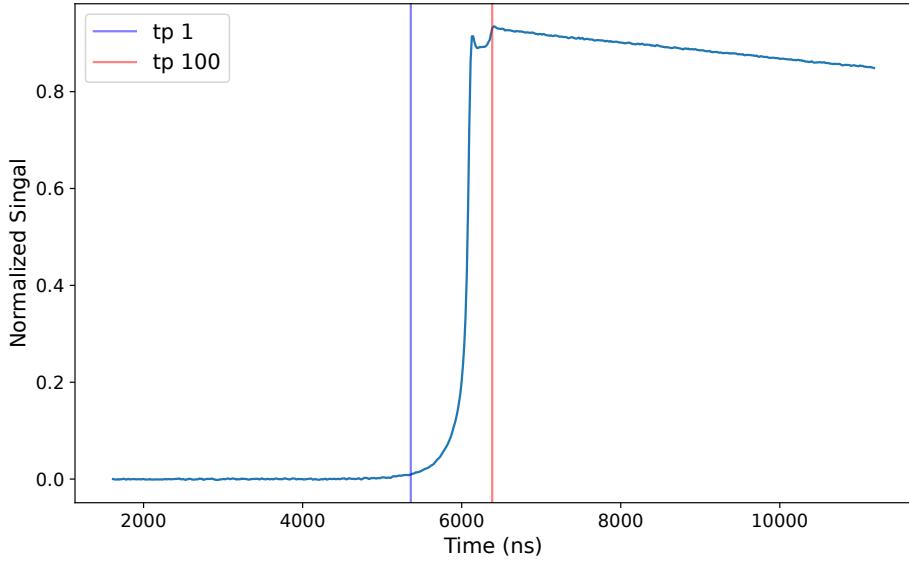


Figure 8.5: Calculating the time points of the waveform.  $T_{Drift}$  of the waveform is the time between the two points.

find the point when the waveform first crosses that amplitude. The drift time  $T_{Drift}$  of the waveform is the time between the two points.

Figures 8.6 and 8.7 illustrate the distribution of  $T_{Drift}$  in the SEP and DEP validation datasets. ATN nets learn to slow the drift time of the simulated waveforms to match the distribution of the data. The SEP  $T_{Drift}$  IoU increases to 62.4% from 39.5%. The DEP IoU increases to 22.5% from 5.4%.

#### 8.4.2 Current Amplitude

For a given event energy, single-site events produce a localized energy deposition, resulting in a sharper and faster increase in  $I_{max}$ . In contrast, multi-site events, characterized by energy deposition at multiple locations within the detector, yielding the same current spread over multiple peaks, decreasing the magnitude of the maximum current value. This distinction makes  $I_{max}$  highly effective in differentiating single-site from multi-site events, thereby establishing it as a crucial parameter in waveform shape simulations [74].

The maximum current amplitude  $I_{max}$  is determined by differentiating the waveform and identifying the maximum value of its derivative. Figure 8.8 shows the steps taken to calculate the current amplitude. The waveform example used here is from an event that deposited energy in two locations in the detector.

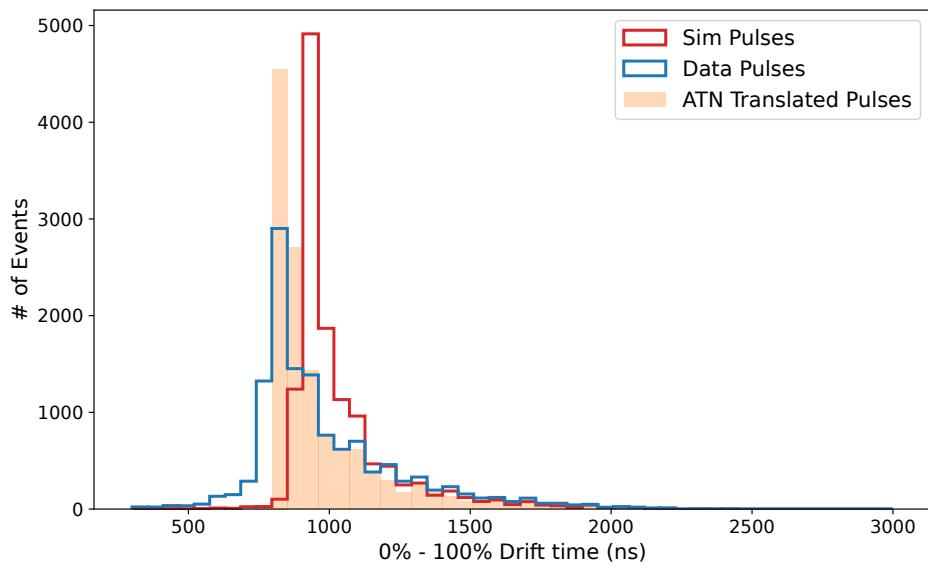


Figure 8.6: The distribution of the  $T_{Drift}$  on SEP dataset.

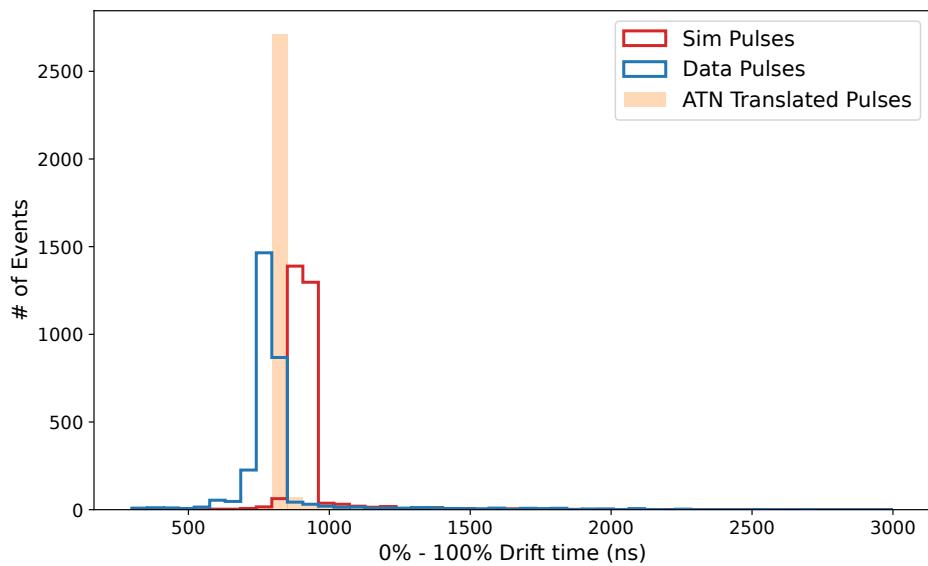


Figure 8.7: The distribution of the  $T_{Drift}$  on DEP dataset.

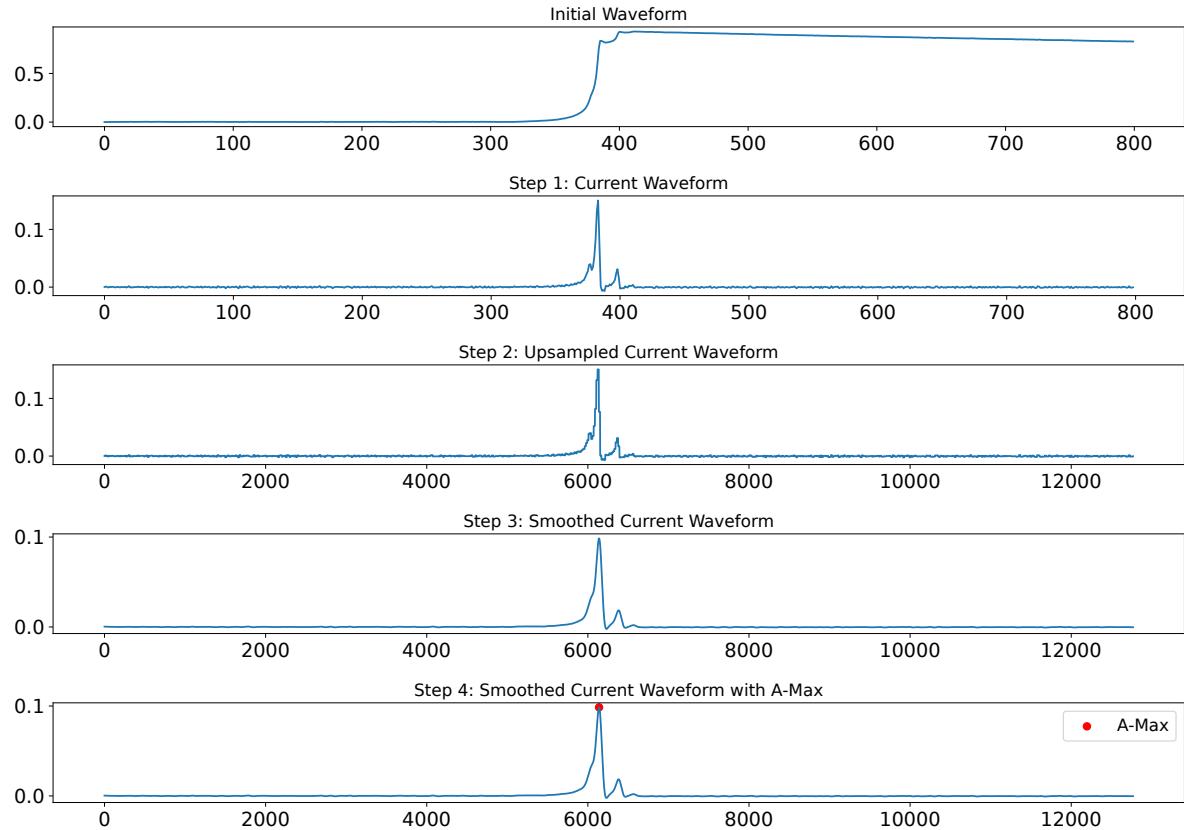


Figure 8.8: Steps involved in calculating the maximum current amplitude of the waveform. First the discrete derivative of the waveform is calculated which is then up-sampled by a factor of 16 to extract finer detail. Then the current is smoothed to reduce the noise. The maximum value is  $I_{max}$ . The ratio of the  $I_{max}$ , to energy is used as a PSD parameter.

The current amplitude thus has two peaks. Then the maximum current would be lower than one for a single-site event, which would only have a single peak and thus higher  $I_{max}$ . This is why current amplitude is highly accurate in distinguishing single- and multi-site events.

Figures 8.9 and 8.10 illustrate the distribution of  $I_{max}$  for both validation datasets. The SEP dataset distributions have two peaks, the higher of which corresponds to single-site events. The peak at a lower value of  $I_{max}$  corresponds to multi-site events where one of the sites has an energy of 511 keV. DEP has only one peak for single-site events. ATN learns to correctly slow the current amplitude of the simulated waveforms to align with the data. The SEP  $I_{max}$  increases to the IoU of 63.71% from 27.53%. The DEP IoU increases to 15.5% from 4.2%.

Figure 8.9 plots the  $I_{max}$  of the ATN against  $I_{max}$  of the simulated pulses. This shows that the ATN performs the translation while maintaining the relative order of the events. The waveforms are

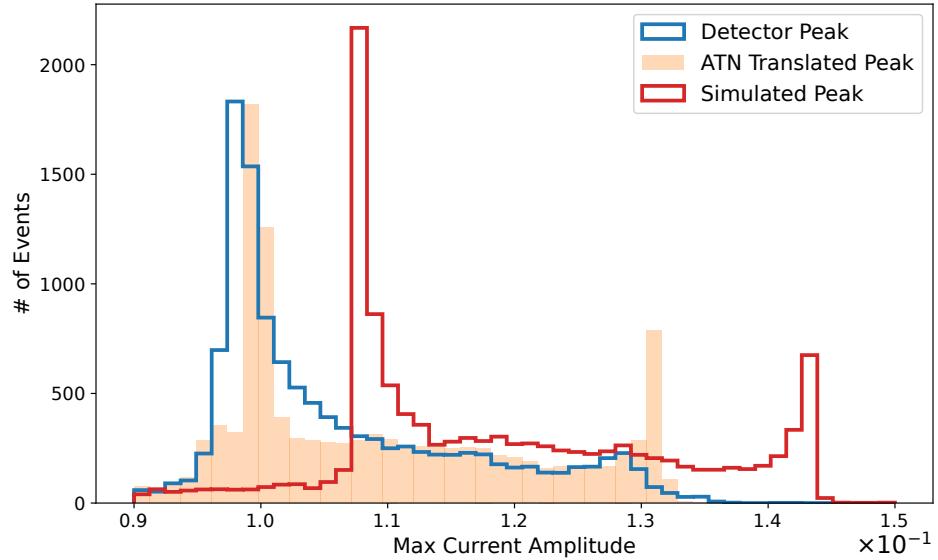


Figure 8.9: Distribution of maximum current amplitude ( $I_{max}$ ) on SEP validation datasets.

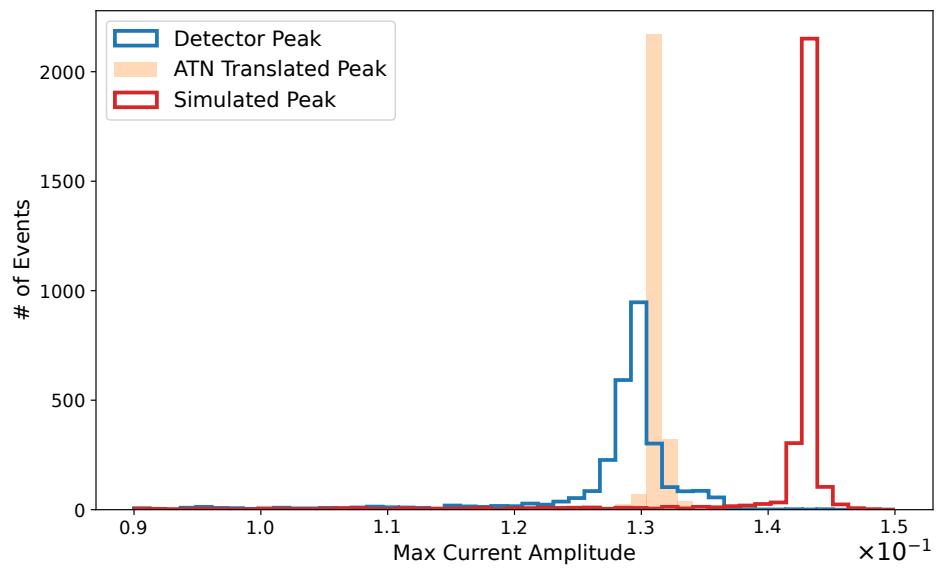


Figure 8.10: Distribution of maximum current amplitude ( $I_{max}$ ) on DEP validation datasets.

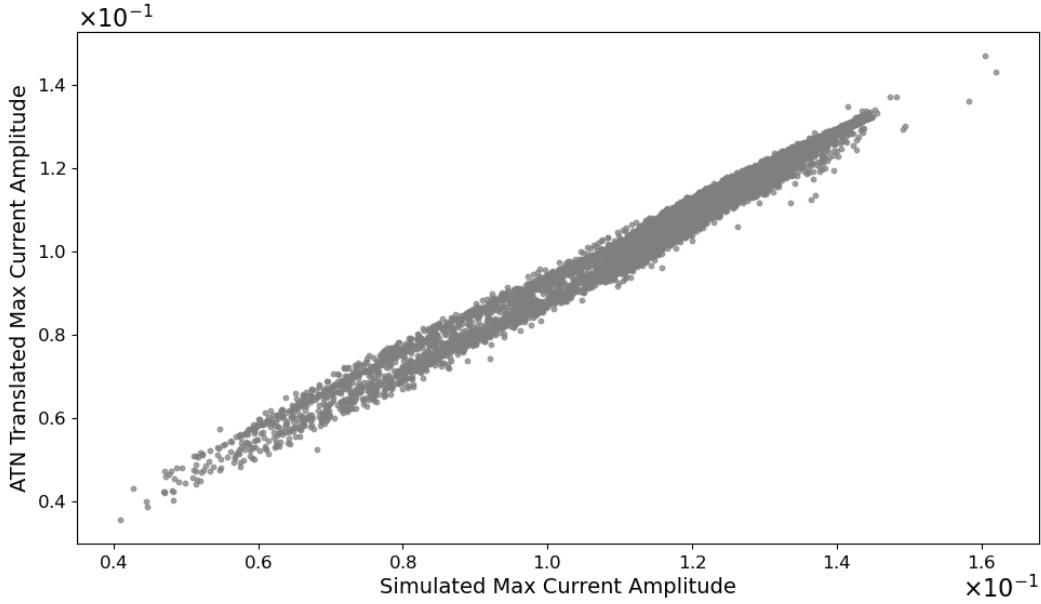


Figure 8.11: Scatter plot between  $I_{max}$  of simulated waveforms and ATN translated waveforms. ATN shifts the amplitude distribution to align with data while maintaining the relative ordering of individual events.

translated in a consistent way: waveforms that had the larger  $I_{max}$  in the simulations have larger  $I_{max}$  after translation. In other words, it is shifting the histogram instead of recreating it and reorganizing the events.

CPU-Net provides significant improvement in agreement between translated and data waveforms for drift time and current amplitude distributions. These agreements could likely be increased by using improved input simulations. The `siggen` simulation includes heuristic models for diffusion and self-repulsion, but it treats each energy deposition as a single point charge, which can lead to overly similar shapes of single-site waveforms. Including actual charge cloud effects in the simulation can help incorporate effects due to cloud evolution during drift in the input simulations and improve the performance of the model.

### 8.4.3 Tail Slope

The strength of the RC decay can be measured by the mean tail slope parameter  $c_{tail}$ . Since the RC decay is an exponential decay,  $c_{tail}$  was calculated by a linear fit of the logarithm of 300 samples of the waveform as shown in figure 8.12. The simulation waveforms do not have an RC decay and thus have a

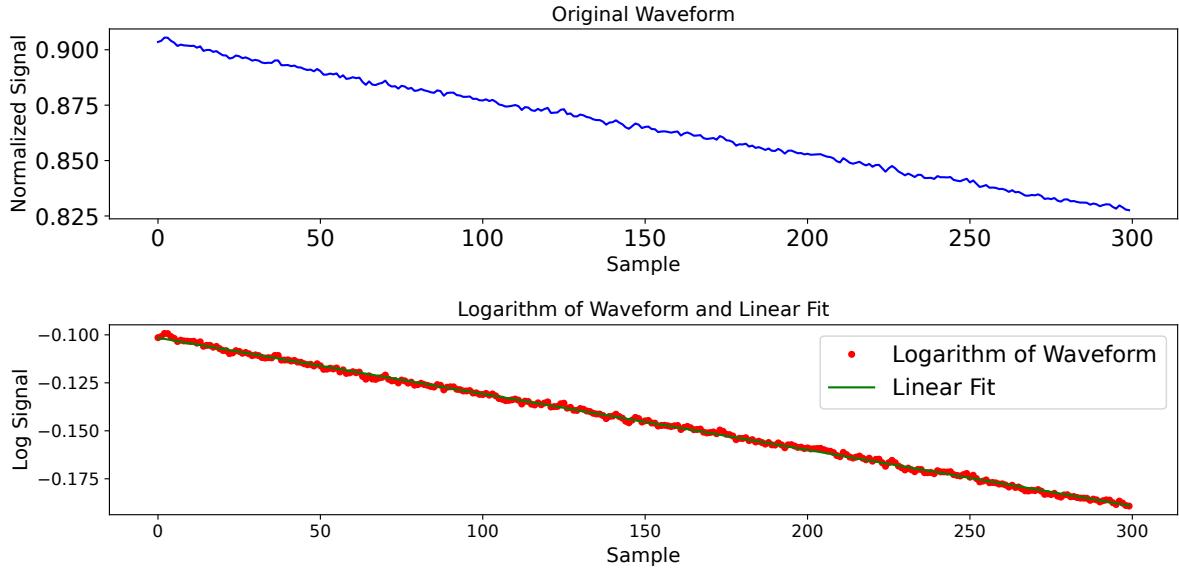


Figure 8.12: Calculating the tail slope ( $c_{tail}$ ) of the waveform. We take the logarithm of last 300 samples of the waveform and linear fit it to find the slope.

mean tail slope of zero. The data waveform was found to have a mean  $c_{tail}$  of  $-1.8605 \times 10^{-4}$  ADC/ns which is equal to a decay constant  $\tau = 1/|c_{tail}|$  of  $54.60\mu s$ , while the ATN translated waveform had a mean  $c_{tail}$  of  $-2.9769 \times 10^{-4}$  ADC/ns, which is equal to a  $\tau$  of  $53.75\mu s$ . This is a difference of 1.58% from the actual value. Figure 8.13 shows this relative magnitude of  $c_{tail}$  for simulations, data, and ATN translated waveforms. Relative to the simulations, the distribution of ATN translated waveforms lies close to the data waveforms.

Figure 8.14 shows the comparison of the  $c_{tail}$  distribution between the data and the ATN translated waveforms. The distributions do not match perfectly, suggesting that the model is not learning this parameter distribution well. This could be a numerical precision issue since the tail slope magnitude is quite small. It could also be that at some step in training the generator determined that the RC decay tail it was adding to the simulated waveforms was enough to fool the discriminator and thus does not improve further to match the distribution of the data. In the next chapter, we suggest some ways to improve the model.

Although our main focus in this work is on translating simulations to resemble measured data, CPU-Net's bidirectional nature also makes it a powerful tool for denoising real detector waveforms. By transforming noisy detector signals into cleaner, simulation-like waveforms, CPU-Net can help

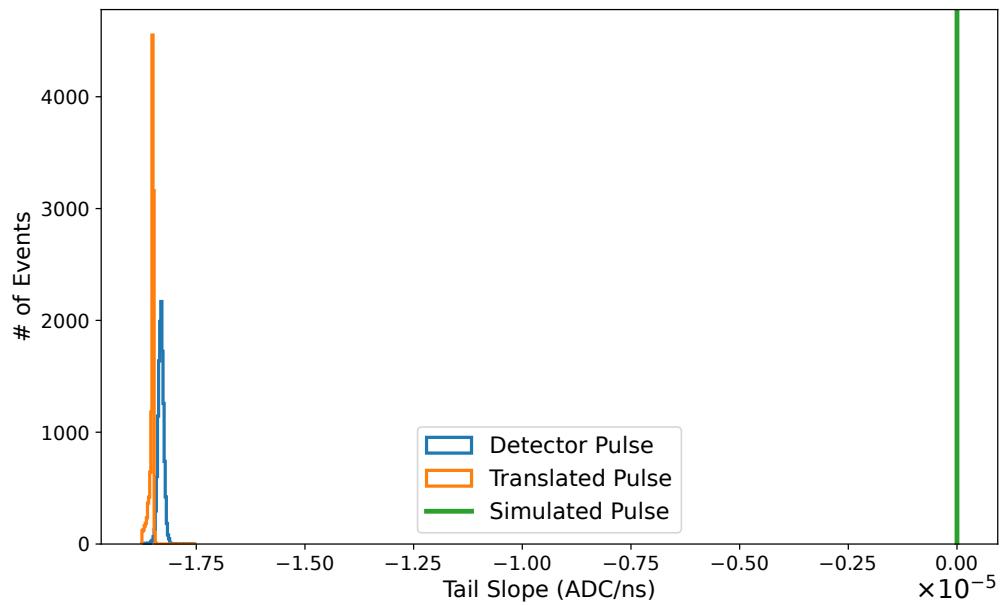


Figure 8.13: Distribution of  $c_{tail}$  for simulations, data, and ATN translated waveforms.

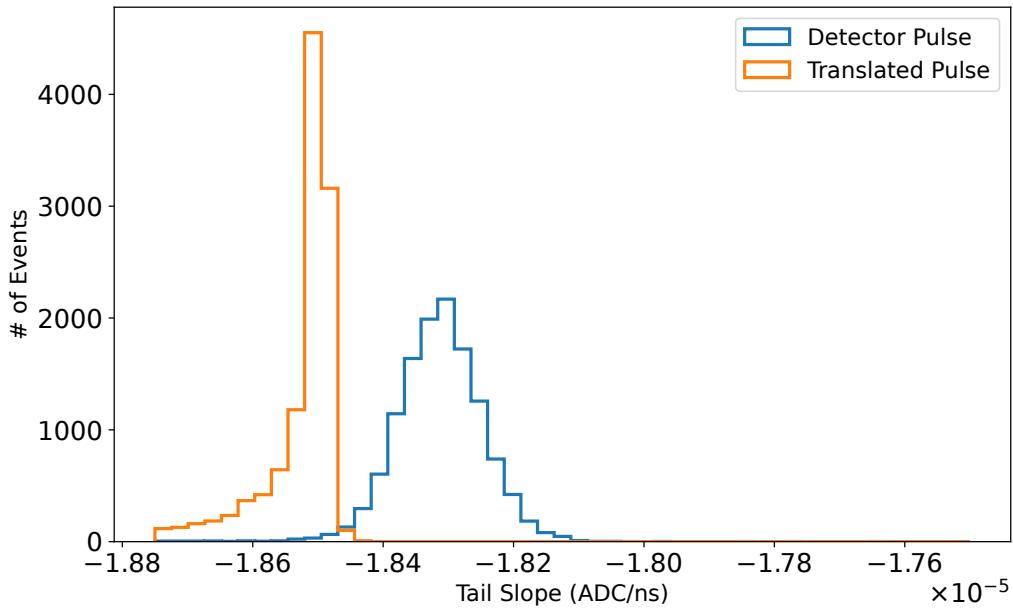


Figure 8.14: Distribution of tail slope ( $c_{tail}$ ) between data and ATN translated waveforms.

isolate and identify subtle features in the event topology. This capability could ultimately improve spatial reconstruction and improve our understanding of particle interactions within the detector.

Electronics response modeling is necessary for accurate pulse shape simulations and is challenging to determine from first principles. CPU-Net is a deep learning framework that is showing good performance in accurately translating simulated waveforms to match the data with the electronics response. The translations preserve the necessary physics and have shown success in reproducing waveform parameter distributions.

## CHAPTER 9: CONCLUSION

### 9.1 Summary of Results

Neutrinos are fundamental to understanding the matter-antimatter asymmetry in the universe. The theory of how neutrinos get their mass suggests that neutrinos could be Majorana particles. To experimentally verify this fact, one can search for neutrinoless double beta decay, which can occur if neutrinos are Majorana particles. HPGe detectors provide a compelling approach for looking for neutrinoless double-beta decay, since they have excellent energy resolution and high internal purity. Signal formation in these detectors is very well understood and can be modeled using simulations. The LEGEND Collaboration is seeking  $0\nu\beta\beta$  in Ge-76 using a phased approach and with the goal of investigating the full inverted hierarchy parameter space of  $m_{\beta\beta}$ . Pulse shape simulations that model signal formation in the detector help us to understand the background model components and simulate the statistical methods used to reject background such as multi-site and surface events. Reliable simulations have been hindered by a lack of modeling of surface events and replication of the effects of the experiment’s electronics response.

Surface backgrounds are potentially the largest components for the future LEGEND-1000 experiment and a significant source of uncertainty due to the lack of proper modeling of charge collection. We developed a new method that can model these challenging backgrounds on the passivated surfaces of detectors. EH-Drift extends the capabilities of existing HPGe detector simulations by incorporating detailed processes such as diffusion, self-repulsion, surface charge, and reduced surface mobility. Its ability to model large charge cloud densities and surface drift makes it a valuable tool for understanding detector response to surface events and improving background modeling in experiments. Using GPU computing, the simulations can be significantly accelerated, allowing for a large-scale background modeling study. Using activeness maps created using EH-Drift we were able to reproduce the measurements made by previous test stands that measured the detector response to alphas at varying positions on the detector surface. We also showed that maps can be used to calculate the active volume for  $0\nu\beta\beta$  events efficiency and to create spectral models for alpha and beta events incident on the passivated surface of HPGe detectors.

Another challenge in accurately simulating HPGe detector waveforms is to correctly model the effects of the readout electronics. It is quite challenging to use first-principles corrections to fully account for the electronics response. We developed a CycleGAN network that learns to translate simulated pulse shapes into data-like signals without explicitly modeling any of the electronics. The model combines a positional U-Net generator with an RNN-based discriminator. This model trains on unpaired data sets of simulated and detector pulses that are collected during routine calibration measurements. CPU-Net successfully translates simulated pulses to output waveforms that closely resemble actual detector pulses by applying corrections using a neural network. We show that CPU-Net preserves the necessary physics in the waveforms and correctly reproduces the critical pulse shape parameters.

## 9.2 Future Directions

This work has laid the foundation for improving pulse shape simulations in LEGEND by addressing some longstanding challenges in modeling surface events and the detector electronics response. Looking ahead, there are several avenues for improvement in the two models. The first step would be a detailed validation of the results, which would allow for refining the models.

### 9.2.1 Validation of Results

To validate the results of EH-Drift, a direct comparison of the surface event waveforms with EH-Drift output can be performed. This will help validate the results while providing inputs to better determine tunable parameters such as surface drift velocity, passivated surface depth, etc. A surface scanning cryostat is currently being commissioned at UNC and will provide additional data needed to validate the simulations. The scanner aims to perform multidetector alpha measurements with a  $^{241}\text{Am}$  source and a monoenergetic beta study with a  $^{137}\text{Cs}$  source. Waveforms from the scanner can be used to validate the simulations as the energy and location of events are known from the source type and location. The ability of the scanner to perform rapid scanning of different detectors will help us determine the detector-to-detector variation in the surface effects. This will also help us understand how these values depend on event location, energy deposition, alpha incidence angle and detector temperature.

The Compton scanner in TUM provides an opportunity to validate the CPU-Net results [75]. The scanner operates by irradiating the detector with gamma rays and then collecting the scattered gamma

rays with pixelated detectors. Using the Compton camera technique, these measurements can be used to create a library of data waveforms with known positions. We can use this paired dataset to train and validate CPU-Net since we can now perform a direct comparison of ATN output with data waveforms.

### 9.2.2 Improving EH-Drift Activeness Maps

Activeness maps in EH-Drift have proven useful for modeling surface effects while reducing simulation runtime by interpolating between points. Until now, we have been using cubic interpolation to build activeness maps in EH-Drift. Although this method is a good approximation, it does not always capture the sudden changes in activeness near the passivated surface. This is because the charge-collecting efficiency varies much more rapidly in the near-surface areas where surface effects dominate than in the bulk of the detector. Thus, the cubic interpolation results tend to have too sharp transitions close to the surface. One way to address this is to adopt interpolation that accounts for local gradients in the activeness function. Techniques like radial basis function interpolation could be used to refine the interpolation in high-variation regions while maintaining a coarser interpolation in the bulk. Another possibility is to perform future studies on the activeness to determine the optimal sampling of the detector. The electric field near the surface with detector type, energy, and surface charge can give a relation for how the activeness is going to change, and this can be used to sample each detector optimally.

### 9.2.3 Future for Pulse Shape Simulations

The future development in pulse shape simulations for Germanium detectors will be centered around SSD . Although EH-Drift successfully captures diffusion, self-repulsion, and surface charge effects, it also makes several assumptions in 2-D simulations of a 3-D problem that SSD can, in principle, avoid. The fully three-dimensional approach of SSD provides a more complete picture of electric fields and charge drift, and it also provides the ability to model the effects of the surroundings of a detector. The program is written in Julia which provides a natural way to parallelize the calculations.

The techniques developed in EH-Drift such as GPU-based drift and diffusion and surface drift can be used to improve SSD. Although SSD already uses GPUs for field calculations, performing drift and diffusion on GPUs will enable simulations using large charge cloud surface events more efficiently in SSD. In SSD the number of charges used is an important input. The higher the number of charges, the better the results are, but after a certain number, there is a diminishing return in improvement. A study

comparing EH-Drift results and SSD can help find this optimal point to get accurate simulations while optimizing runtime. We performed some initial SSD simulations with surface drift effects included, and the results looked promising.

#### 9.2.4 Challenges in CycleGAN based Model

The CPU-Net was quite effective at translating simulated waveforms to data-like pulses, while matching the current amplitude and drift time distribution. However, the CycleGAN training can be quite unstable and requires fine-tuning of several hyperparameters to get the right balance. This can be seen in figure 8.14, where the distribution of data and the ATN output waveform do not match. We also noticed that the noise added by CPU-Net is not always consistent with what we see in the data.

To test both effects, we added pole-zero decay and pink noise to the simulated waveform. This gave us a direct one-to-one data set for validation. We trained CPU-Net with simulations plus electronics effects as data. After training, we observed that while the ATN output waveform resembled the data waveform better, there was still a significant difference between them. This shows that the model learns the electronics response to an extent to fool the discriminator, but not necessarily to the high precision we would like to achieve.

#### 9.2.5 CPU-Net Model Improvements

One way to improve the model is to incorporate distribution-based loss metrics, such as the Wasserstein distance, into loss optimization [76]. This can be done for known parameters such as the tail slope, current amplitude, and drift time. This forces the model to learn the correct distribution during training. We would not know how the model works for unknown electronics parameters; however, proving the known parameters can give the model a first guess to improve upon. Combining the Wasserstein metrics with cycle consistency and identity consistency could help the generator learn the features more effectively. Input simulations can also be improved by using SSD instead of siggen, and using SSD's capability to model charge cloud effects more accurately. This will provide better signal modeling and initial approximation for drift time and current amplitude distributions.

Another approach would be to replace CycleGAN with an advanced generative model such as the diffusion model [77] or the Transformer model [61]. The diffusion model gradually adds Gaussian noise to the data pulse in a Markov chain. Then it trains a network to reverse the chain. This allows for finer

control over the learned distribution. The advantage of the method is that we can control the noise being encoded. Thus, using known waveform shape parameter priors, such as tail slope, current amplitude, and drift times, we could guide the model to learn the correct translations of simulated pulses.

### 9.3 Implications for the LEGEND Physics Program

Together, the two methods improve the accuracy of pulse shape simulations. This will enable a large-scale simulation study of LEGEND backgrounds, tests of new pulse shape discrimination strategies, and allow simulations to be adapted to reflect different detector configurations.

#### 9.3.1 Passivated Surface Background Modeling for LEGEND

Having an accurate model will enable us to create a background model for events on the passivated surfaces that correctly accounts for the anomalous charge collection effects in this region. This will impact the predicted spectral shape of all background components, particularly in detectors like PPCs that have large passivated surfaces, with particularly large impacts on the backgrounds associated with surface events such as alpha and beta interactions. Using Geant4 simulations, we can estimate the location and energy of interactions near the passivated surface from varying background sources. Then, using libraries of EH-Drift activeness maps with different surface charges, we can apply the charge collection efficiency determined for each location to estimate the energy spectrum for these depositions. Ultimately, this would enable us to create a background model that accurately describes the behavior of the passivated surface events. Knowing the fraction of those events appearing in the  $0\nu\beta\beta$  ROI would help reduce the uncertainty in the alpha background contribution in the LEGEND experiment, improve models for the shape of the K-42 spectrum, and improve estimates of the signal efficiency in detectors with large passivated surfaces.

#### 9.3.2 Beyond Standard Model Searches

The EH-Drift may also have an important application in low energy studies. Enriched detectors in LEGEND-1000 will have limited cosmogenic exposure on the surface and a very low  $^{39}\text{Ar}$  background event rate, resulting in ultra-low backgrounds at low energy. This will enable highly sensitive searches for other physics Beyond the Standard Model (BSM) such as signatures of light weakly interacting massive

particles (WIMP) and bosonic dark matter (DM). Low-energy BSM physics searches would require an accurate model of the passivated surface. This is because most background events at low energy happen on the surface. A model for these passivated surface interactions would directly improve the low-energy sensitivity in LEGEND-1000 by improving background modeling near threshold. BSM effects can alter the shape of  $2\nu\beta\beta$ , and studying the distribution can set limits or make a measurement of BSM processes like Majoron emission. EH-Drift will play a crucial role in modeling  $2\nu\beta\beta$  spectral distortions since such studies would require detailed modeling of the background of  $^{39}\text{Ar}$  beta, a passivated surface effect and of how bulk events in the vicinity of the passivated surface are degraded in energy.

### 9.3.3 LegendGeSim and CPU-Net Integration

The LEGEND Germanium Simulation Chain (LegendGeSim) is a simulation pipeline that can be used to simulate the entire electronics chain from geometry visualization to the simulation of raw files with realistic data that mimic the data. The goal is to produce a set of simulated waveforms that are compatible with tools used to process data waveforms. CPU-Net will be incorporated into the LegendGeSim raw tier to help add electronics effects to the data. This will enable more realistic pulse shape simulations and enable calculating the efficiency of cuts by directly applying them to the data.

## BIBLIOGRAPHY

- [1] Andrei D Sakharov. Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe. *Soviet Physics Uspekhi*, 34(5):392, May 1991.
- [2] F. Reines and C. L. Cowan. Detection of the free neutrino. *Phys. Rev.*, 92:830–831, Nov 1953.
- [3] B. T. Cleveland, Timothy Daily, Raymond Davis, Jr., James R. Distel, Kenneth Lande, C. K. Lee, Paul S. Wiedenhein, and Jack Ullman. Measurement of the solar electron neutrino flux with the Homestake chlorine detector. *Astrophys. J.*, 496:505–526, 1998.
- [4] B. Pontecorvo. Neutrino Experiments and the Problem of Conservation of Leptonic Charge. *Zh. Eksp. Teor. Fiz.*, 53:1717–1725, 1967.
- [5] Q. R. Ahmad et al. Measurement of the rate of  $\nu_e + d \rightarrow p + p + e^-$  interactions produced by  ${}^8\text{B}$  solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, 2001.
- [6] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.
- [7] J. L. Hewett, H. Weerts, et al. Fundamental Physics at the Intensity Frontier. Technical report, US Department of Energy, Office of High Energy Physics, 5 2012. Report of the Workshop held December 2011 in Rockville, MD.
- [8] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Physical Review*, 105(4):1413–1415, February 1957.
- [9] Ettore Majorana. Teoria simmetrica dell'elettrone e del positrone. *Nuovo Cim.*, 14:171–184, 1937.
- [10] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016.
- [11] M. et al. Aker. Direct neutrino-mass measurement based on 259 days of katrin data. *Science*, 388(6743):180–185, 2025.
- [12] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, 1962.
- [13] KAI ZUBER. Neutrinoless double beta decay. *Pramana*, 79(4):781–791, Oct 2012.
- [14] S. R. Elliott, A. A. Hahn, and M. K. Moe. Direct evidence for two-neutrino double-beta decay in  ${}^{82}\text{Se}$ . *Phys. Rev. Lett.*, 59:2020–2023, Nov 1987.
- [15] et al. Agostini. Final results of gerda on the search for neutrinoless double- $\beta$  decay. *Phys. Rev. Lett.*, 125:252502, Dec 2020.
- [16] N. et al. Ackerman. Observation of two-neutrino double-beta decay in  ${}^{136}\text{Xe}$  with the exo-200 detector. *Phys. Rev. Lett.*, 107:212501, Nov 2011.
- [17] C. Alduino et al. Measurement of the two-neutrino double-beta decay half-life of  ${}^{130}\text{Te}$  with the cuore-0 experiment. *European Physical Journal C*, 77:13, 2017.

- [18] Ruben Saakyan. Two-neutrino double-beta decay. *Annual Review of Nuclear and Particle Science*, 63(1):503–529, 2013.
- [19] J. Menéndez. Neutrinoless  $\beta\beta$  decay mediated by the exchange of light and heavy neutrinos: The role of nuclear structure correlations. *J. Phys. G*, 45(1):014003, 2018.
- [20] J. Schechter and J. W. F. Valle. Neutrinoless double- $\beta$  decay in  $su(2)\times u(1)$  theories. *Phys. Rev. D*, 25:2951–2954, Jun 1982.
- [21] Ivan Esteban, M. C. Gonzalez-Garcia, Michele Maltoni, Ivan Martinez-Soler, and Thomas Schwetz. Updated fit to three neutrino mixing: exploring the accelerator-reactor complementarity. *Journal of High Energy Physics*, 2017(1), Jan 2017.
- [22] Matteo Agostini, Giovanni Benato, and Jason A. Detwiler. Discovery probability of next-generation neutrinoless double- $\beta$  decay experiments. *Phys. Rev. D*, 96:053001, Sep 2017.
- [23] C. R. Haufe et al. Modeling backgrounds for the Majorana Demonstrator. *AIP Conf. Proc.*, 2908(1):100001, 2023.
- [24] M. Agostini et al. Final Results of GERDA on the Search for Neutrinoless Double- $\beta$  Decay. *Phys. Rev. Lett.*, 125(25):252502, 2020.
- [25] I. J. Arnquist et al. Final Result of the Majorana Demonstrator’s Search for Neutrinoless Double- $\beta$  Decay in Ge76. *Phys. Rev. Lett.*, 130(6):062501, 2023.
- [26] S. Abe et al. Search for Majorana Neutrinos with the Complete KamLAND-Zen Dataset. *arXiv*, 6 2024.
- [27] G. Adhikari et al. nEXO: neutrinoless double beta decay search beyond  $10^{28}$  year half-life sensitivity. *J. Phys. G*, 49(1):015104, 2022.
- [28] C. Arnaboldi et al. CUORE: A Cryogenic underground observatory for rare events. *Nucl. Instrum. Meth. A*, 518:775–798, 2004.
- [29] B. Aharmim et al. Combined Analysis of all Three Phases of Solar Neutrino Data from the Sudbury Neutrino Observatory. *Phys. Rev. C*, 88:025501, 2013.
- [30] N. Abgrall et al. The Large Enriched Germanium Experiment for Neutrinoless Double Beta Decay (LEGEND). *AIP Conf. Proc.*, 1894(1):020027, 2017.
- [31] Glenn F. Knoll. *Radiation detection and measurement*. John Wiley & Sons, Inc., 4th ed. edition, 2010.
- [32] N. Abgrall et al. The Majorana Demonstrator radioassay program. *Nucl. Instrum. Meth. A*, 828:22–36, 2016.
- [33] R.J. Cooper, D.C. Radford, P.A. Hausladen, and K. Lagergren. A novel hpge detector for gamma-ray tracking and imaging. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 665:25–32, 2011.
- [34] Micah James Noecker Buuck. *A Radiogenic Background Model for the MAJORANA DEMONSTRATOR*. PhD thesis, U. Washington, Seattle (main), 2019.

- [35] Tommaso Comellato, Matteo Agostini, and Stefan Schönert. Charge-carrier collective motion in germanium detectors for  $\beta\beta$ -decay searches. *Eur. Phys. J. C*, 81(1):76, 2021.
- [36] Luigi Pertoldi. The first year of legend-200 physics data in the quest for  $0\nu\beta\beta$  decay. In *Proceedings of the XXXI International Conference on Neutrino Physics and Astrophysics (Neutrino 2024)*, Milano, Italy, June 2024. University of Milano-Bicocca. Plenary talk S6: Neutrino properties 1.
- [37] N. Abgrall et al. LEGEND-1000 Preconceptual Design Report. *arXiv e-prints*, page arXiv:2107.11462, July 2021.
- [38] M. Agostini et al. Modeling of GERDA Phase II data. *JHEP*, 03:139, 2020.
- [39] H. Bethe and J. Ashkin. Passage of radiations through matter. In E. Segrè, editor, *Experimental Nuclear Physics*, page 253. John Wiley & Sons, New York, 1953.
- [40] I. J. Arnquist et al.  $\alpha$ -event characterization and rejection in point-contact HPGe detectors. *Eur. Phys. J. C*, 82(3):226, 2022.
- [41] Frank Edzards, Lukas Hauermann, Iris Abt, Chris Gooch, Björn Lehnert, Xiang Liu, Susanne Mertens, David C. Radford, Oliver Schulz, and Michael Willers. Surface Characterization of P-Type Point Contact Germanium Detectors. *Particles*, 4(4):489–511, 2021.
- [42] Julieta Gruszko. *Surface Alpha Interactions in P-Type Point-Contact HPGe Detectors: Maximizing Sensitivity of  $^{76}\text{Ge}$  Neutrinoless Double-Beta Decay Searches*. PhD thesis, U. Washington, Seattle (main), 2017.
- [43] Paul Mullowney, Ming-Chieh Lin, Kevin Paul, Cory Ahrens, Michael Loh, R.J. Cooper, D.C. Radford, and Ethan Hull. Computational models of germanium point contact detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 662(1):33–44, 2012.
- [44] S. Mertens, A. Hegai, D. C. Radford, N. Abgrall, Y. D. Chan, R. D. Martin, A. W. P. Poon, and C. Schmitt. Characterization of High Purity Germanium Point Contact Detectors with Low Net Impurity Concentration. *Nucl. Instrum. Meth. A*, 921:81–88, 2019.
- [45] I. Abt, F. Fischer, F. Hagemann, L. Hauermann, O. Schulz, M. Schuster, and A. J. Zsigmond. Simulation of semiconductor detectors in 3d with solidstatedetectors.jl. *Journal of Instrumentation*, 16(08):P08007, 8 2021.
- [46] David Alberto Hervas Aguilar. *Characterizing Bulk Signals in an Inverted Coaxial Point-Contact Detector to Inform Rare-Event Searches*. Ph.d. dissertation, University of North Carolina at Chapel Hill, Chapel Hill, NC, 2023. Advisor: John F. Wilkerson; Committee: Jonathan Engel, Charles R. Evans, Julieta Gruszko, Reyco Henning, David C. Radford.
- [47] Charles Kittel. *Introduction to Solid State Physics*. John Wiley & Sons, Inc., Hoboken, NJ, 8 edition, 2005.
- [48] D.M. Caughey and R.E. Thomas. Carrier mobilities in silicon empirically related to doping and field. *Proceedings of the IEEE*, 55(12):2192–2193, 1967.
- [49] M.Ali Omar and Lino Reggiani. Drift velocity and diffusivity of hot carriers in germanium: Model calculations. *Solid-State Electronics*, 30(12):1351–1354, 1987.

- [50] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen*, 100:32–74, January 1928.
- [51] Kevin Bhimani. Eh-drift: Charge transport simulation code, 2024. GitHub repository, accessed on February 2025.
- [52] Richard S. Varga. *Matrix Iterative Analysis*, volume 27 of *Springer Series in Computational Mathematics*. Springer-Verlag Berlin Heidelberg, 2 edition, 2000.
- [53] David M. Young. *Iterative Methods for Solving Partial Difference Equations of Elliptic Type*. PhD thesis, Harvard University, 1950.
- [54] G. Alefeld. On the convergence of the symmetric sor method for matrices with red-black ordering. *Numerische Mathematik*, 39:113–117, 1982.
- [55] Michael Willers and for the LEGEND collaboration. Signal readout electronics for legend-200. *Journal of Physics: Conference Series*, 1468(1):012113, feb 2020.
- [56] S. I. Alvis et al. Multisite event discrimination for the MAJORANA DEMONSTRATOR. *Phys. Rev. C*, 99(6):065501, 2019.
- [57] Benjamin Shanks. High precision modeling of germanium detector waveforms using bayesian machine learning. *PhD Thesis*, 2017.
- [58] Samuel Joseph Meijer. Precision modeling of germanium detector waveforms for rare event searches. *PhD Thesis*, 2019.
- [59] Aobo Li, Julieta Gruszko, Brady Bos, Thomas Caldwell, Esteban León, and John Wilkerson. Ad-hoc Pulse Shape Simulation using Cyclic Positional U-Net. In *36th Conference on Neural Information Processing Systems: Workshop on Machine Learning and the Physical Sciences*, 12 2022.
- [60] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [61] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [62] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [63] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.
- [64] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.
- [65] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [66] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

- [67] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [68] Morgan L. Clark. *Development of a Characterization Facility and Analysis Process for LEGEND-200 HPGe Detectors*. PhD thesis, University of North Carolina at Chapel Hill, 2023.
- [69] Matteo Agostini, Jason Detwiler, Luigi Pertoldi, Ian Guinn, George Marshall, Valerio D’Andrea, Patrick Krause, Grace Song, Erin Engelhardt, Sam Borden, Rosanna Deckert, David Sweigart, Andreas Zschocke, Clint Wiseman, Tim Mathew, Yoann Kermaïdic, and Ben Shanks. pygama, March 2024.
- [70] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- [71] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.
- [72] Allan H. Murphy. The finley affair: A signal event in the history of forecast verification. *Weather and Forecasting*, 11(1):3–20, 1996.
- [73] Paul Jaccard. Etude de la distribution florale dans une portion des alpes et du jura. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, 37:547–579, 01 1901.
- [74] Alvis et al. Multisite event discrimination for the majorana demonstrator. *Phys. Rev. C*, 99:065501, Jun 2019.
- [75] Iris Abt, Chris Gooch, Felix Hagemann, Lukas Hauertmann, David Hervas Aguilar, Xiang Liu, Oliver Schulz, Martin Schuster, and Anna Julia Zsigmond. A novel wide-angle Compton Scanner setup to study bulk events in germanium detectors. *Eur. Phys. J. C*, 82(10):936, 2022. [Erratum: Eur.Phys.J.C 82, 1004 (2022)].
- [76] Leonid N. Vaserstein. Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredači Informacii*, 5(3):64–72, 1969. English translation: *Problems of Information Transmission*, 5(3):47–52, 1969.
- [77] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. *arXiv e-prints*, page arXiv:2006.11239, June 2020.