HIGH PRECISION MODELING OF GERMANIUM DETECTOR WAVEFORMS USING BAYESIAN MACHINE LEARNING

Benjamin E. Shanks

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.

Chapel Hill
2017

Approved by:

John F. Wilkerson

Joaquin Drut

Matthew P. Green

Reyco Henning

David C. Radford

ABSTRACT

Benjamin E. Shanks: High precision modeling of germanium detector waveforms using
Bayesian machine learning
(Under the direction of John F. Wilkerson)

The universe as we see it today is dominated by matter, but the Standard Model of
particle physics cannot explain why so little antimatter remains. If the neutrino is its own
antiparticle – a so-called Majorana particle – lepton number must be violated, which is a key
component of theories that explain the observed matter-antimatter asymmetry. Neutrinoless
double-beta decay ($0\nu\beta\beta$), a hypothetical radioactive decay in certain nuclei, is the only
experimentally accessible signature that can prove if neutrinos are Majorana in nature. But
if it exists, $0\nu\beta\beta$ must be exceedingly rare, with current half-life limits over 10^{25} years.
Measuring a process with such a faint signal requires extraordinary efforts to eliminate
backgrounds. The MAJORANA DEMONSTRATOR is a search for $0\nu\beta\beta$ of germanium-76 in
an array of germanium detectors, with the goal of "demonstrating" backgrounds low enough
to justify building a larger experiment with ~ 1 tonne of isotope.

Reducing backgrounds even further will be critical to the discovery potential of a tonne
scale experiment. One powerful method to reject background is pulse shape discrimination,
which uses the shape of measured detector signals to differentiate between background and
candidate $0\nu\beta\beta$ events. With a better understanding of pulse shapes from our detectors,
we may be able to improve the discrimination efficiency. We have developed a detailed
model of signal formation in germanium detectors, where the shape depends sensitively
on characteristics specific to each individual detector crystal. To train the parameters for
specific crystals in the DEMONSTRATOR, we have implemented a Bayesian machine learn-
ing algorithm which is able to infer detector characteristics using only standard calibration
waveforms. This model is accurate to the level of parts per thousand of the signal amplitude,

is able to discriminate against common background events, and has even shown some ability to estimate the position of origin for signals inside the detector.

ACKNOWLEDGEMENTS

There are countless people whose support has made this dissertation possible. Thanks to my advisor, John Wilkerson, who allowed me the freedom to explore new ideas, the direction to bring them to fruition and a steady hand when things seemed hopeless. Chris O'Shaughnessy was the earliest and strongest supporter of this work, and it would never would have materialized without his encouragement and input. I've relied extensively on the work and germanium detector expertise of David Radford. Reyco Henning provided valuable advice and suggestions. Matt Green taught me how to be useful around a vacuum system. I have been fortunate to be surrounded many excellent MAJORANA collaborators, both to learn from and to help pass the time in South Dakota.

Before coming back to physics, I spent a few years at General Electric. Mike Durling's backing and advice came at a critical point in my career. Mauricio Castillo-Effen took me under his wing and introduced me to MCMC. Their influence on this work and my career can't be understated.

Finally, thanks to my parents for a lifetime of encouragement and support, to my friends for keeping me grounded, and to Carly for inspiring and sharing the journey alongside me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

$0\nu\beta\beta$ Neutrinoless Double-beta Decay

$2\nu\beta\beta$ Double-beta Decay

A/E   Current pulse amplitude to energy ratio

BEGe   Broad Energy Germanium

DEP   Double Escape Peak

DNS   Diffusive Nested Sampling

FET   Field Effect Transistor

HPGe   High Purity Germanium

LMFE   Low Mass Front End

MCMC   Markov Chain Monte Carlo

PPC   P-type Point Contact

PSD   Pulse Shape Discrimination

Q End-point Energy

ROI   Region of Interest

SEP   Single Escape Peak

SURF   Sanford Underground Research Laboratory

CHAPTER 1:   Introduction

Section 1.1:   Neutrinoless double beta decay

Determining the Majorana or Dirac nature of the neutrino is one of most important open challenges in neutrino physics. Unfortunately, any effect that would distinguish between them appears only at $\mathcal{O}(m_\nu/E)$. Typical experiments fall in the regime $m_\nu/E \ll 10^{-6}$ [1], making a measurement of the neutrino nature extremely difficult. The only signature of a Majorana mass currently considered feasible to detect is neutrinoless double-beta decay.

Two-neutrino double beta decay ($2\nu\beta\beta$) is a second order standard-model-allowed process observed in certain isotopes which have even numbers of both protons and neutrons. These "even-even" nuclei are sometimes more stable than neighboring odd-odd isotopes due to spin-pairing, which occurs independently for both protons and neutrons. In such an instance normal β-decay is energetically forbidden, but Goeppert-Mayer [2] suggested in 1935 that a double beta decay $((A, Z) \rightarrow (A, Z+2) + 2e^- + 2\bar{v}_e)$ might occur as a second order process with an extremely long half-life ($> 10^{17}$ yr). A half century passed before the first direct laboratory measurement by Elliott $et.$ $al.$ of $2\nu\beta\beta$ in ^{82}Se [3], with a half life of 1.08×10^{20} years.

In 1937, Ettore Majorana suggested that neutral fermions could be their own antiparticles [4], a description now known as Majorana fermions. Four years after Goeppert-Mayer proposed $2\nu\beta\beta$, Furry [5] showed that the existence of a Majorana neutrino would imply an additional double decay mode in which no neutrino is emitted. In the modern formulation of this process, the right-handed antineutrino emitted from one decaying neutron can develop a left-handed component and be absorbed by a neighboring neutron. This is not possible for a Dirac neutrino, since the neutrino and antineutrino are distinct particles. The

Majorana nature of the neutrino therefore permits virtual neutrino exchange to mediate the $0\nu\beta\beta$ decay $(A, Z) \rightarrow (A, Z + 2) + 2e^-$, violating lepton number by two units.

There are several different lepton number violating mechanisms which could mediate $0\nu\beta\beta$. The scenario considered most likely is "light neutrino mediated" decay. If we assume the weak charged current is purely left handed and that either no unknown particles exist, or any unknown particles are too massive to contribute significantly to the $0\nu\beta\beta$ process, then the process is mediated only by the three light neutrinos. To first order, the half life of the light neutrino exchange $0\nu\beta\beta$ process is then given by [6]

$$\left(T_{1/2}^{0\nu}\right)^{-1} = G_{0\nu} \left|M_{0\nu}\right|^2 \langle m_{\beta\beta}\rangle^2 \tag{1.1}$$

where $G_{0\nu}$ is a phase space factor, $\left|M_{0\nu}\right|^2$ the nuclear matrix element and $\langle m_{\beta\beta}\rangle^2$ the effective Majorana mass, which is essentially an average of the mass states weighted by their mixing in the electron neutrino:

$$\langle m_{\beta\beta}\rangle \equiv \left|\sum_i U_{ei}^2 m_i\right| = \left|c_{12}^2 c_{13}^2 m_1 + c_{13}^2 s_{12}^2 m_2 e^{i2\phi_2} + s_{13}^2 m_3 e^{i2\phi_3}\right| \tag{1.2}$$

Here, U_{ei} are the component of the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) neutrino mixing matrix [7] that describes the composition of the electron neutrino in terms of the mass state m_i. The Majorana phases, ϕ_1 and ϕ_2, are possible CP violating terms in the PMNS matrix for Majorana neutrinos (Dirac neutrinos have only one such phase). The terms s_{ij} and c_{ij} are the sine and cosine, respectively, of the mixing angles θ_{ij} of the PMNS matrix, as measured by neutrino oscillation experiments. The relationship between the lightest neutrino mass and $m_{\beta\beta}$ is shown in Figure 1.1. The effective Majorana mass, and therefore the half-life of $0\nu\beta\beta$, depends sensitively on the neutrino masses, and especially on the neutrino mass ordering. A measurement of the half-life could, in turn, provide a measurement of the neutrino mass.

The nuclear matrix element present in the half-life expression Equation 1.1 encapsulates

2
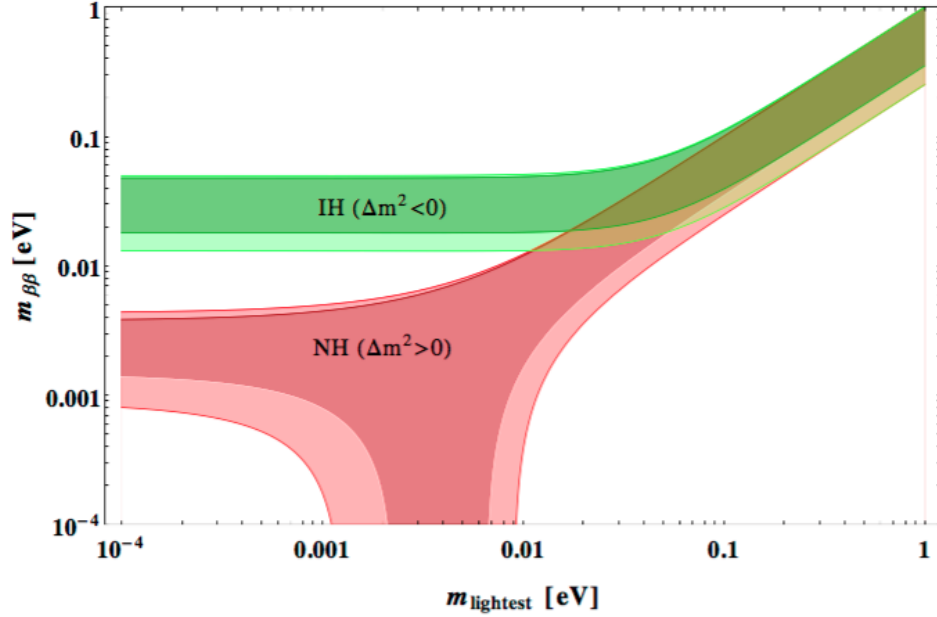
Figure 1.1: Effective Majorana mass plotted as a function of lightest neutrino mass for the cases of both normal (red) and inverted (green) mass ordering. The solid lines outline the allowed parameter space allowed from neutrino oscillation experiments, with uncertainties from neutrino mixing angles shown in the lighter color bands. Figure from [8].

the nuclear physics of the transition between the initial and daughter nucleii, and is not exactly calculable. The term can be written

$$M_{0\nu} = g_A^2 M^{(0\nu)}, \tag{1.3}$$

where g_A is the axial vector coupling constant, and $M^{(0\nu)}$ contains the contributions from the Fermi, Gamow-Teller and tensor operators for the transition [8].

The value of g_A is renormalized during calculations of nuclear models, and so can take on an effective value to describe the observed transition strength. In general, limitations in the model cause the the effective value of g_A in model calculations to be reduced, or "quenched," from the value for quarks of 1.269 [9]. Because the half life term depends on the effective value as g_A^4, the presence of quenching will affect the half life with fourth power dependence. However, the strength of the quenching effect is currently not well understood and is the

3

focus of current theoretical efforts [10].

Multiple approximation method have been employed to calculate $M^{(0\nu)}$, as reviewed in [10], notably the quasi-random phase approximation (QRPA), the interacting boson model (IBM) and the interacting shell model (ISM). However, there are a large number of virtual intermediate states available to the nucleus during the decay which increase computational complexity of the calculation. For a given $0\nu\beta\beta$ candidate nucleus, disagreement between calculations can vary by up to a factor of three – creating an order of magnitude uncertainty in the half-life, and further complicating comparison between experiments in different isotopes.

The experimental signature of $0\nu\beta\beta$ is a peak in the energy spectrum at the endpoint energy of the decay ($Q_{\beta\beta}$), as all the energy is shared between the pair of detected electrons. The $2\nu\beta\beta$ mode creates a continuous spectrum varying up to the endpoint of the decay. A simulated spectrum is shown in Figure 1.2. Typical measured half-lives for the $2\nu\beta\beta$ mode are around 10^{21} years, while the best limits for $0\nu\beta\beta$ have extended to 10^{25} years [11].

Given the extremely rare nature of the decay, $0\nu\beta\beta$ experiments must be constructed with a number of common design goals in mind. First, the largest possible mass of candidate isotope should be present. Many experiments are designed such that the source material forms the bulk of the detector. Backgrounds must be suppressed through radiopure construction materials, analytic signal rejection or both. In order to distinguish remaining backgrounds (including the irreducible $2\nu\beta\beta$ background) from signal, it is desirable to have good energy resolution in the region of interest around $Q_{\beta\beta}$. A summary of experimental considerations is found in [6].

The search for neutrinoless double beta decay is an active field, and there are current experiments in a number of isotopes, including ^{76}Ge, ^{130}Te, and ^{136}Xe. Table 1.1 summarizes the limits from major experiments. In order to cover the parameter space preferred for inverted ordering neutrinos, an experiment must reach a sensitivity of $m_{\beta\beta} \approx 20$ meV (see Figure 1.1), requiring tens of tonne-years of exposure. For this reason, most collaborations intend to scale their current design into a next-generation detector with a tonne or more of
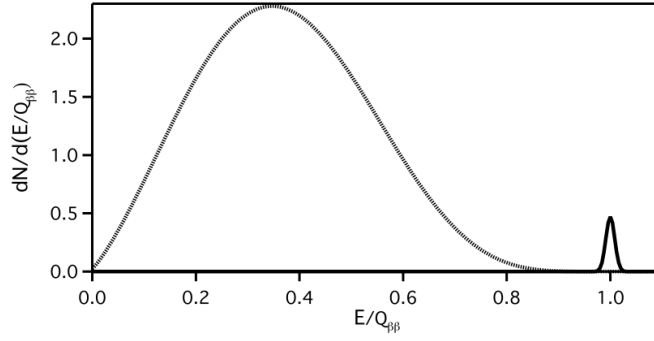
Figure 1.2: Spectra for $2\nu\beta\beta$ (dashed line) and $0\nu\beta\beta$ (solid line). Curves are drawn assuming the decay rate of $0\nu\beta\beta$ is 1% that of $2\nu\beta\beta$, with a 2% resolution. Figure from [6].

Experiment Name	Isotope	Half life limit ($\times 10^{25}$ y)
GERDA Phase II [12]	^{76}Ge	4.0
CUORE-0 [13]	^{130}Te	0.4
EXO-200 [14]	^{136}Xe	1.1
KamLAND-Zen [15]	^{136}Xe	10.7

Table 1.1: Summary of current $0\nu\beta\beta$ half life limits.

isotopic mass.

Section 1.2: The Majorana Demonstrator

The Majorana Demonstrator [16] is a search for neutrinoless double beta decay in the isotope ^{76}Ge. The experiment consists of an array of high purity germanium (HPGe) detectors of the p-type point contact (PPC) geometry. In total, there are 44.1 kg of detectors, of which 29.7 kg are 88% enriched in ^{76}Ge. The main goal of the Demonstrator is to prove a germanium detector array can be constructed with sufficiently low background to justify investment in a tonne-scale germanium-76 $0\nu\beta\beta$ experiment. Using the nuclear matrix element calculated with QRPA[17], a ^{76}Ge experiment must be sensitive to a half-life $T^{0\nu}_{1/2} \approx 5 \cdot 10^{27}$ years in order to cover the inverted ordering at $m_{\beta\beta} \approx 20$ meV. With backgrounds < 1 count/tonne-year, 10 tonne-years of exposure are required to reach this sensitivity at 3σ confidence. The exposure required increases significantly at higher background levels.

The DEMONSTRATOR aims to achieve < 3 counts/tonne-year in a 4-keV-wide region of interest (ROI) around $Q_{\beta\beta}$, which is 2039 keV for ^{76}Ge. Due to additional background reduction from self-shielding and granularity cuts, this scales to < 1 count/tonne-year for a tonne-scale experiment. The current design goal for a tonne scale experiment is 0.1 count/tonne-year in the ROI, which has been revised downward to ensure the inverted ordering region is covered for the worst-case nuclear matrix element calculations. In addition to showing the feasibility of a tonne-scale experiment, the DEMONSTRATOR will have a sensitivity to $T_{1/2}^{0\nu} \approx 10^{26}$ yr. Using the same matrix element as above, this corresponds to $m_{\beta\beta} \approx 100$ meV.

The DEMONSTRATOR was designed and constructed to achieve the lowest possible background. Cosmic muon-induced signals are diminished by 4850' of rock overburden above the experiment site at the Davis Campus of Sanford Underground Research Facility (SURF) in Lead, SD. An underground cleanroom lab facility was established with < 500 particles/ft^3. Parts within the cryostat, which therefore have a direct shine path to detectors, were maintained in dry boxes purged with liquid nitrogen boil-off to minimize radon plate out. All detector manipulation took place in a ntirogen-purged glovebox with a particle count < 10 particles/ft^3. Radon levels and particle counts in the lab were carefully monitored before sensitive work proceeded.

Construction materials were chosen to ensure radiopurity. The majority of the cryostat material is composed of copper, which was manufactured with excellent chemical purity via electroforming. However, copper is cosmogenically activated, inducing long-lived radioactive isotopes, mainly ^{57}Co and ^{60}Co. To avoid cosmogenics, the MAJORANA collaboration built an underground electroforming facility where copper was grown starting in 2011. The copper was machined into parts in an underground shop staffed with a dedicated machinist.

A six-layer compact shield was designed with both active [18] and passive components. On the outer layer, borated plastic moderates neutrons. Scintillator panels actively reject muon events. A nearly airtight enclosure is filled with nitrogen purge gas to mitigate radon.

The remaining layers are composed of lead and copper, with electroformed copper used for the innermost layer. A schematic of the shield is shown in Figure 1.3.

The MAJORANA detector unit design is shown in Figure 1.4. The germanium crystal is electrically isolated using plastic standoffs, which double as the thermal path for cooling. High voltage is applied via a copper contact ring, and signals are read out with a copper contact pin. A novel low noise low-mass front-end (LMFE) board [19] was designed to provide a resistive feedback readout. Since it is low mass and made of ultra pure materials, the LMFE can be placed near the detector, which minimizes stray capacitance. Three to five detectors are stacked into a "string" which shares a common thermal connection to a coldplate above. Each MAJORANA cryostat maintains cryogenic temperature and ultrahigh vacuum for up to seven strings. Figure 1.5 contains an image of a fully loaded cryostat.

1.2.1: Initial results of the MAJORANA DEMONSTRATOR

The DEMONSTRATOR is currently taking data with both modules in-shield. The collaboration has analyzed the first 32.4 days of data in this final configuration, taken from August 25 to September 27, 2016, and constituting 1.39 years of isotopic exposure. The spectrum after preliminary analysis cuts is shown in Figure 1.6. For energies greater than ∼500 keV, the two neutrino spectrum dominates the spectrum. Given the low exposure of this dataset, we evaluate the background in a 400 keV wide window centered around $Q_{\beta\beta}$. After cuts, one background event remains in this window, which projects to $5.1^{+8.9}_{-3.2}$ counts / (ROI · t · y) for a 2.9 keV ROI for module one detectors, and 2.6 keV ROI for module two detectors (where each ROI was chosen to optimize sensitivity given resolution at $Q_{\beta\beta}$). This corresponds to a background index of 1.8×10^{-3} counts / (keV · t · y).

Section 1.3: Signal modeling for germanium $0\nu\beta\beta$ experiments

The DEMONSTRATOR is meant to show that we can achieve backgrounds low enough to justify building a tonne scale experiment that could cover the allowed parameter space for
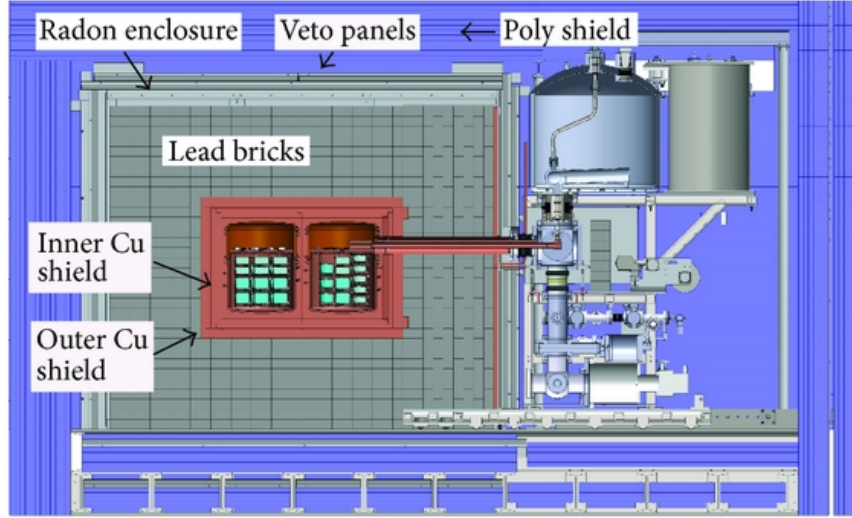
Figure 1.3: (a) A schematic of the DEMONSTRATOR shield. (b) A picture of the shield components up to the Radon enclosure. A cryostat is visible inside the shield. The removable "blank" monolith seen here pulled away from the shield has been replaced with a second cryostat, and both modules are currently taking data.

(a)

(b)
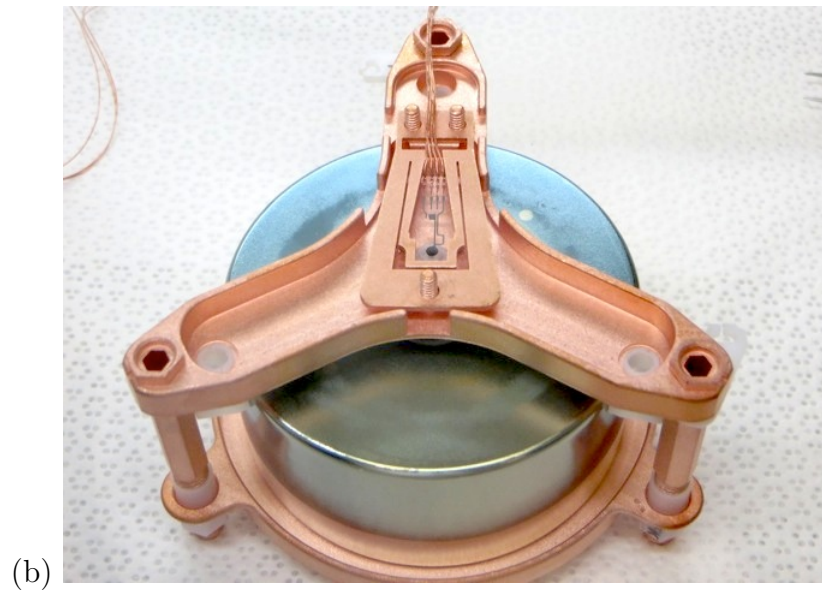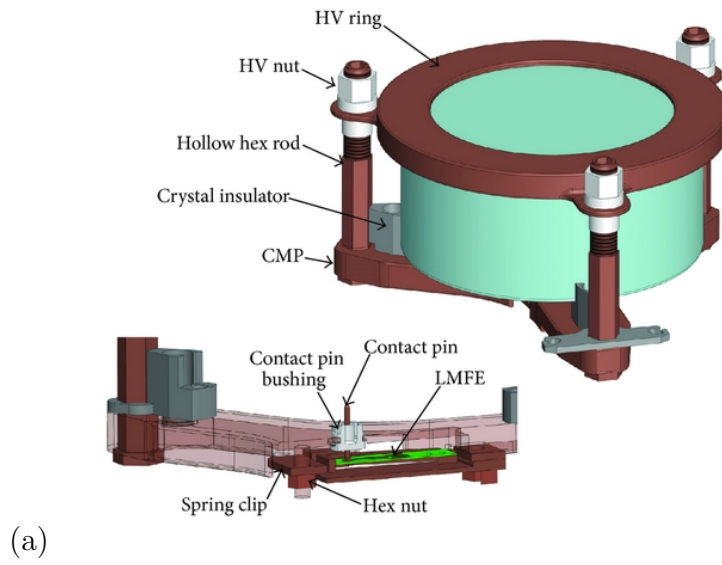
Figure 1.4: (a) A conceptual drawing of the DEMONSTRATOR detector unit. (b) Picture of an assembled detector unit with LMFE visible on top.
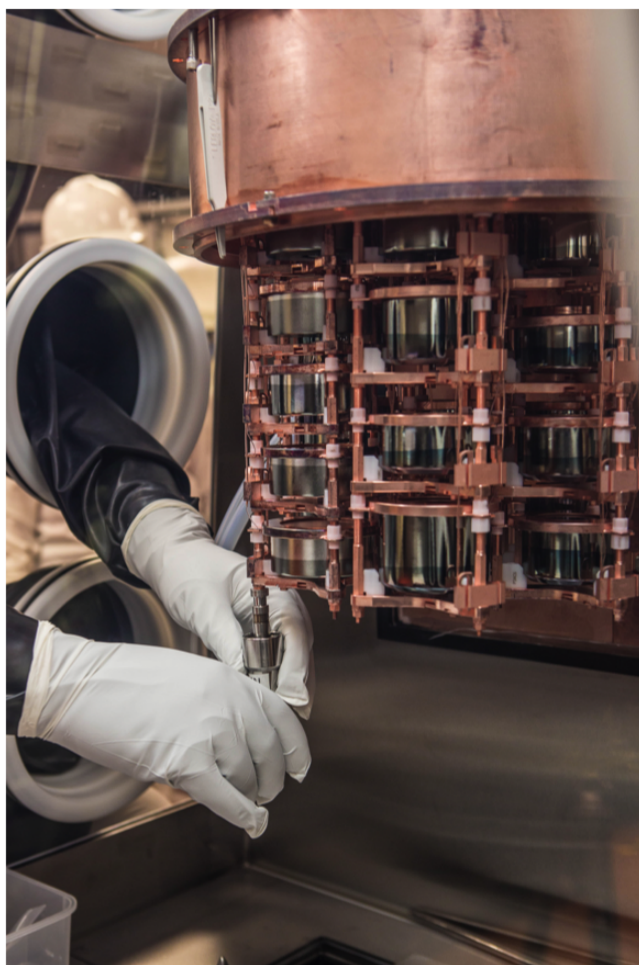
Figure 1.5: Image of strings hanging from a DEMONSTRATOR cryostat. Seven strings hang from a copper cold plate, obscured in the photo by the cryostat exterior.
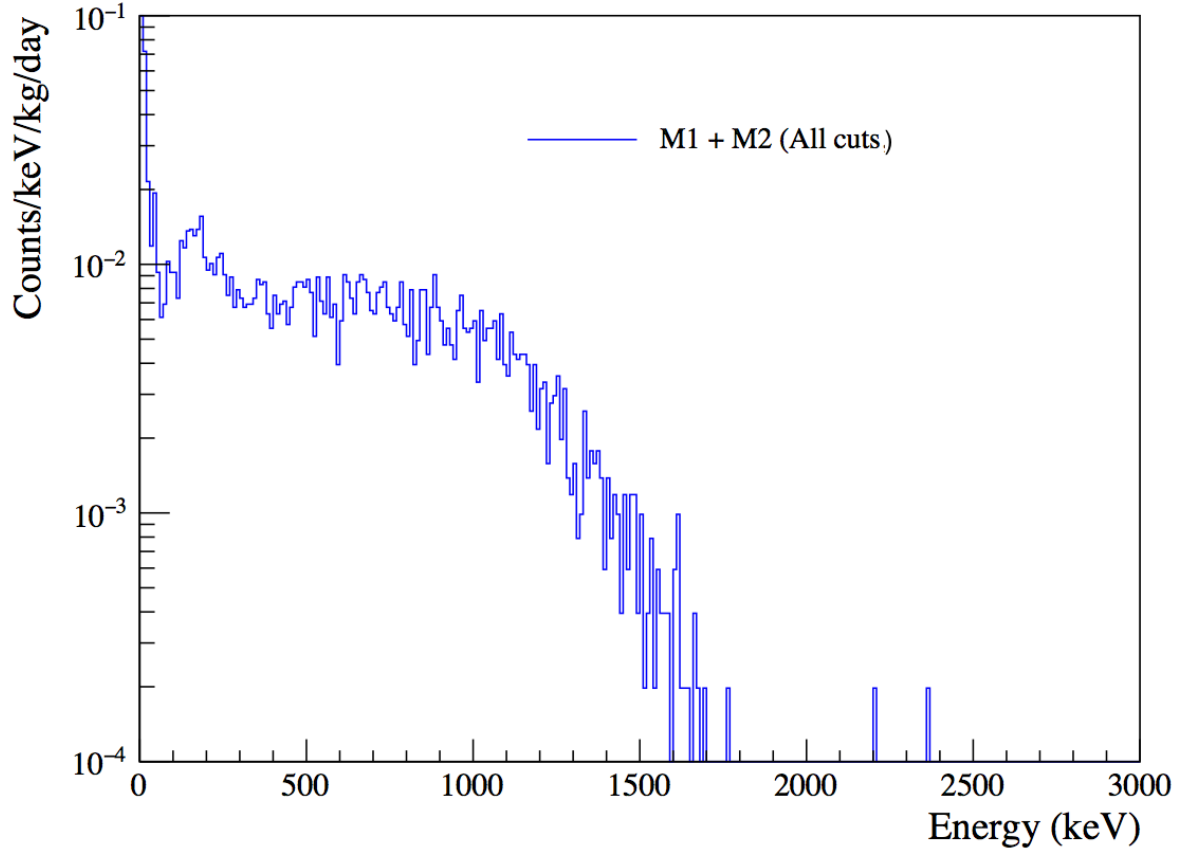
Figure 1.6: Preliminary spectrum of initial DEMONSTRATOR data. The spectrum encompasses the first 1.39 kg·y of enriched detector exposure for the DEMONSTRATOR in its final configuration. The spectrum is shown after analysis cuts are applied. $2\nu\beta\beta$ of ^{76}Ge is the dominant spectral feature in the region above 500 keV. $Q_{\beta\beta}$ for ^{76}Ge is 2039 keV.

$0\nu\beta\beta$ in the inverted ordering. In the case of a background-free experiment, the half-life sensitivity level $S(_{T_{1/2}})$ scales linearly with exposure,

$$S(_{T_{1/2}}) \propto M \cdot t \tag{1.4}$$

where M is the isotopic mass and t the total time over which data are collected. In the presence of background, due to poisson statistics, the sensitivity instead scales as a square root of exposure

$$S(_{T_{1/2}}) \propto \sqrt{\frac{M \cdot t}{b \cdot \Delta E}} \tag{1.5}$$

where b is the background index and ΔE the size of the ROI. Figure 1.7 plots the relationship between exposure and sensitivity for an enriched germanium, both for the case of a background-free experiment and for several background indexes. As we approach the $\sim$ 10 tonne-years of exposure necessary to cover the inverted ordering region, the background dramatically impacts the sensitivity level of the experiment. Given this reality, it is especially important that the design of next generation experiments maximize the capability to discriminate against background with analysis cuts.

By developing an accurate model of signal formation in germanium detectors, we have two different paths to achieving lower background in a tonne scale experiment. During the design phase, we can use the model to create accurate simulated waveforms. These waveforms can be used to realistically evaluate the efficiency of analysis cuts against simulated backgrounds, and inform detector design to achieve maximal efficiency.

In addition to being used for design simulations, it is possible to use the model to discriminate against background events. To first order, it can be used to identify which waveforms originate from the deposition of energy in the crystal bulk, like $\beta\beta$ events, thereby discriminating against backgrounds which scatter at multiple locations in the detector, or against events that are distorted due to edge effects in the detector fields. Because of the unique two-electron $\beta\beta$ signature, an extremely sensitive model could conceivably discriminate between
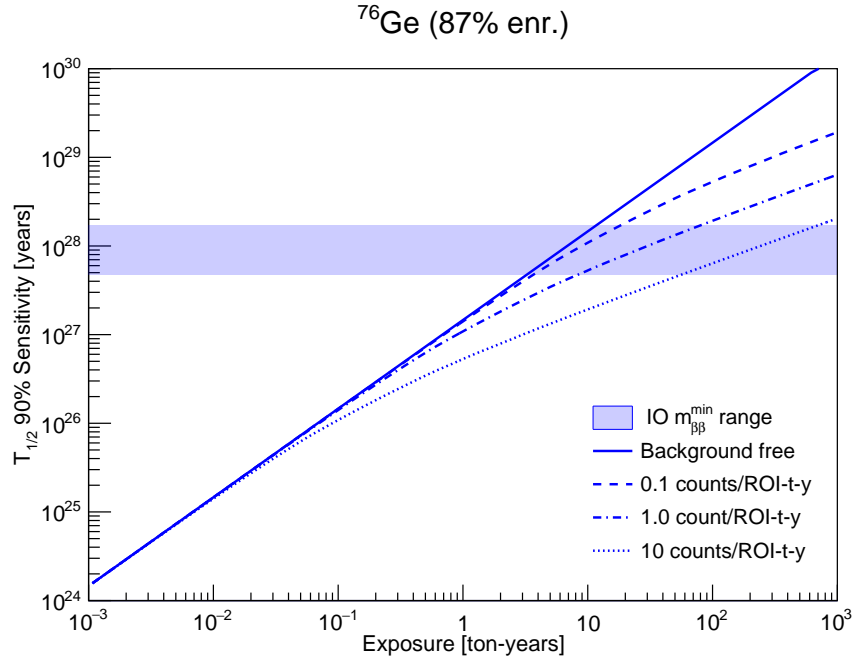
Figure 1.7: $0\nu\beta\beta$ 90% sensitivity as a function of background in the region of interest (ROI) for an enriched germanium-76 experiment. The horizontal blue band corresponds to the $m_{\beta\beta}$ at the bottom of the inverted ordering region, with uncertainty based on the difference between nuclear matrix element calculations, and assuming no quenching of g_A. Figure from Jason Detwiler.

$\beta\beta$ events and single-site γ events.

Any analysis based on comparing waveforms to a model should, in principle, extract the maximal amount of available information about that waveform (assuming the model is accurate and exhaustive), optimized for the physical characteristics of specific detectors. When applied to background discrimination, this means a sufficiently accurate model should outperform simpler pulse shape heuristics. Furthermore, additional interesting information can be ascertained about waveforms, including the location of energy deposition within the detector volume.

In Chapter 2, we describe a model of signal formation in germanium detectors, and extend it specifically for the unique circumstances of the DEMONSTRATOR detectors. Chapter 3 describes an algorithm which we have developed to optimize the parameters of the model to match observed waveforms. Finally, we evaluate the ability of the model to describe waveforms using DEMONSTRATOR data in Chapter 5.

CHAPTER 2:   Modeling signal formation in germanium detectors

Section 2.1:   Operating Principles

2.1.1:   Semiconductor detector design

The interaction of radiation within a semiconductor material transfers energy from the incident particle to electrons bound in the atomic valence band. When the energy absorbed by an electron exceeds a threshold, the electron enters the conduction band and leaves behind a vacancy in the valence band, called a hole. Each hole is defined by the absence of an electron and therefore carries positive effective charge. Because the energy required to create an electron–hole pair in a semiconductor is very small ($\sim$3 eV) [20], a large number of electron–hole pairs will be freed due to the interaction, creating a "cloud" of charge carriers. Both carrier types are free to drift through the material under the influence of an electric field, but in opposite directions due to their opposing charge. This feature allows the device to serve as a detector of the ionized signal.

A semiconductor detector measures the total energy deposited in an interaction by collecting the charge carriers at a readout electrode. Applying an external voltage across the detector creates an internal electric field which sweeps the carriers to the electrode. However, if the voltage difference also produces too large a steady-state current, it will overwhelm the signal from charge carriers produced from radiation. We therefore require a detector design that can hold an external voltage difference without generating a large current. In a sample of pure semiconductor, only electrons which are excited beyond the electron–hole creation threshold energy are mobile and able to contribute to the flow of current. If a perfect semiconductor were sufficiently cooled to mitigate thermal creation of electron–hole pairs, it could function as a detector simply by applying a voltage difference across the crystal and

measuring current spikes.

However, in a real world semiconductor sample, some population of elemental impurities will always occupy sites in the crystal lattice. Due to characteristics of their valence shell, some elements are able to donate an electron to the conduction band with an energy requirement significantly less than the band gap. These impurities are called electron "donors". Other impurities are prone to "accepting" electrons, thereby creating a hole in the conduction band. If there is a net surplus of donor over acceptor impurities, the concentration of electrons at finite temperature naturally contains a surplus of electrons over holes, and is considered "n-type." If the majority impurity type is acceptor, there are a surplus of holes, and the material is "p-type". In either case, the contribution from impurities increases the net number of mobile charge carriers, and thereby lowers the resistivity by several orders of magnitude. Because of the low resistivity, if a voltage difference is applied across conducting contacts in the simple detector design described above, a large steady-state current will flow between electrodes.

To reduce the current, we can make use of the convenient property that semiconductors can be easily manufactured into diodes. A diode is "rectifying," which means current flows freely when voltage is applied in the "forward" direction, but when "reverse" biased, almost no current flows. Diodes naturally form at the boundary between n and p type regions within a semiconductor. If this junction is reverse biased, it can be operated as a detector with very little current flow.

The overabundance of holes on the p side of the junction causes net diffusion of holes into the n material. Once across the junction, the hole recombines with a donor impurity, leaving an immobile positive ion in the n material. Similarly, electrons on the n side diffuse into the p, and are captured by acceptors. The net effect produces a positive space charge on the n side of the material, and an equal magnitude negative charge accumulation on the p side. The junction will therefore exhibit a natural steady-state potential difference, called the "contact potential." The region over which the impurities are ionized and space charge

16

exists is "depleted" of free charge carriers, because its majority carriers have recombined with stationary impurities.

If an external voltage is applied across the junction in a direction which reinforces the contact potential, it will attract electrons across the junction from the p region, and holes from the n. In both cases, these are the minority carriers, and are very low in concentration. The material therefore exhibits very low conductivity and nearly no current will flow. On the other hand, if a voltage is applied opposite and greater than the contact potential, the majority carriers will again cross the junction, and current will flow freely. This p-n junction therefore behaves as a rectifying element. Holding the p side of the diode at positive voltage to the n side constitutes forward biasing, and the opposite is reverse biasing. Increasing the reverse bias causes the depletion region to grow outward from the junction and deeper into the material. If the reverse bias reaches a critical level, the depleted region will break down and current will again flow rapidly.

The reverse biased diode is able to function as a detector because incident radiation can still generate new electron–hole pairs within the depleted region. In this case, the internal electric field produced by applied voltage and the impurity sites in the crystal, which have been ionized due to the external reverse bias, will force the minority carrier towards the junction. The depleted region therefore constitutes the active volume in which energy deposited by radiation can be collected.

A p-n junction is usually created by doping two halves of a crystal with opposite impurities. If the both regions have equal impurity concentrations, the depleted zone will extend to equal depth on each side. However, if one region has a much greater impurity concentration, its carriers will need to travel further across the junction to find a recombination site, thereby creating a deeper depletion layer in the lower concentration material. In a detector, it is desirable to maximize the volume that can be depleted before reaching the breakdown voltage. Semiconductor detectors are therefore created using a sample of material with low net impurity concentration, and the junction is formed by doping a thin surface layer with

extremely high concentration [21]. This high concentration layer is referred to as either $p+$ or $n+$ type, and generally serves as one of the detector electrodes. A detector manufactured from n-type crystal will have a thin $p+$ rectifying electrode, and one from a p-type crystal will have a thin $n+$ electrode.

Because the detector bulk is very high purity, an appreciable number of minority carriers will remain after depletion. To reduce steady-state leakage current from these carriers, the readout contact is highly doped with the impurities of the same type as the bulk crystal material. This highly doped region will have very low concentrations of the minority carrier. This is called a "noninjecting" electrode. As a junction of two similar-typed semiconductors, it does not rectify, and is an "ohmic" contact [22].

2.1.2: P-type point contact detectors

The P-type point contact (PPC) germanium detector geometry was developed to maintain a low detector capacitance in a relatively large-volume design [23]. Detector capacitance is coupled to the electronic noise generated on the observed signal. Therefore, a low capacitance detector is able to achieve low thresholds and retain high resolution at low energies. The detectors were originally designed to increase sensitivity to the low energy nuclear recoil from dark matter scattering, and PPC detector experiments have produced leading exclusion limits for low dark matter masses [24, 25]. The possibility of measuring coherent elastic neutrino–nucleus scattering has also been explored [26]. Stable operation has been achieved with thresholds at or below ~ 500 eV [24, 27].

The approximate geometry of a PPC detector is diagrammed in Figure 2.1, and a picture shown in Figure 2.2. Using a small, shallow $p+$ "point" contact reduces the capacitance compared to traditional "coaxial" geometries, which use a long central contact bored several centimeters into the detector diagrammed in Figure 2.3. The $p+$ layer is doped via ion implantation, often with boron, and is ~ 0.3 μm thick. The $n+$ layer is created by drifting lithium into the crystal lattice, and can be up to a mm thick [28]. A proprietary passivation

18
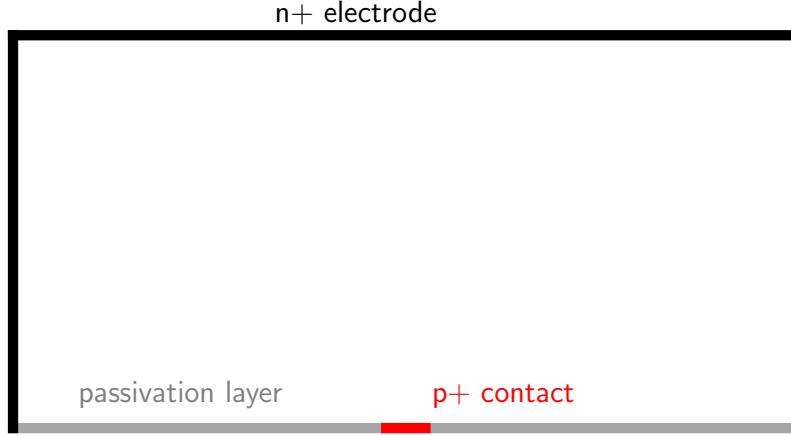
Figure 2.1: Cross-sectional schematic of the PPC detector geometry. The detectors are cylindrically symmetric, and produced with diameters around 6-7 cm, and axial lengths between roughly 3-5 cm. The $p+$ "point" contact is $\sim$2 mm in diameter. A positive bias of a few kilovolts is placed on the outer $n+$ electrode, while the point contact is held at ground. The passivation layer reduces the flow of current along the detector surface.

treatment on the surface between the outer and point contact reduces current flow across the detector surface.

An extremely low net impurity concentration is required to create depletion layer thickness on the order of centimeters with external bias of a few kilovolts. The net number density of charge carriers in the crystal bulk cannot exceed a few times 10^{10} cm^{-3}, which corresponds to less than one part per trillion germanium atoms. This level of purity can be achieved in germanium using sophisticated manufacturing techniques [21]. Beginning with a bar of raw germanium, the material undergoes the zone refinement process, in which one end of the crystal is melted, then the molten zone is slowly swept to the opposite end. Because some impurities have different solubility in liquid than solid germanium, repeated application of this process segregates the impurities and can improve the purity by two or three orders of magnitude. A "boule" of crystal then is formed by Czochralski growth in a pure hydrogen
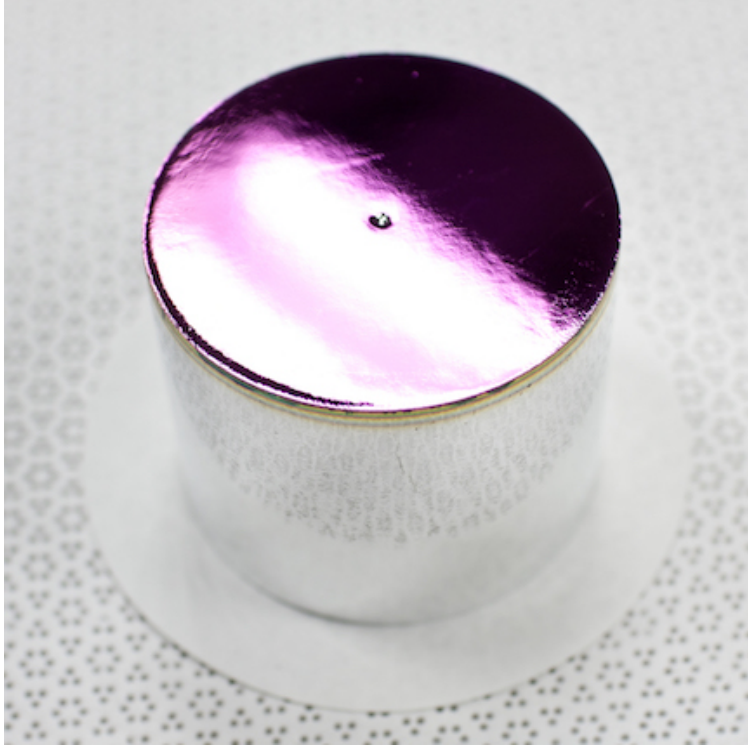
Figure 2.2: Picture of a PPC detector. The point contact is visible at the center of the top surface.

atmosphere, which further increases the material purity. The boule is then manufactured into detectors.

Impurity segregation during the Czochralski process tends to introduce a gradient in impurity concentration along the axis of crystal growth [21]. In a PPC detector, the point contact is always manufactured on the higher impurity side of the detector. When depleted, the impurities in the p type material will leave stationary negative charges in the detector. By placing the point contact on the high impurity end, these impurity sites create an axial electric field which sweeps holes towards the $p+$ contact. Since the axial fields in a PPC detector can otherwise be quite low, the impurity gradient is important to ensure proper collection of hole carriers [23].
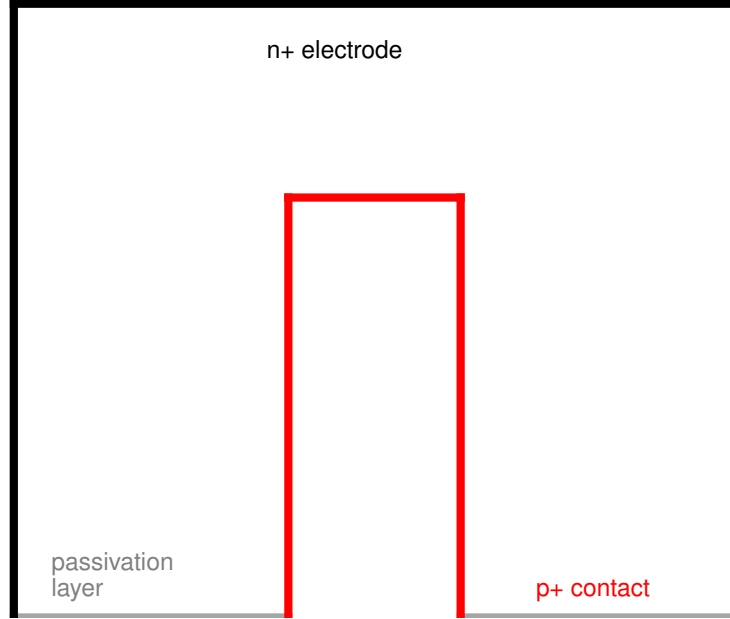
Figure 2.3: Cross-sectional schematic of the coaxial germanium detector geometry. The p+ contact is bored several cm into the detector, which is ∼5-7 cm long.

2.1.3: Signal Formation: The Shockley-Ramo Theorem

As soon as a cloud of electron–hole pairs is created in the depleted region of a detector, they begin to drift under the influence of the electric field. This creates a current which, in turn, induces a current at the readout electrode. The generated signal therefore persists over the entire drift time of *both* charge carriers. In a PPC detector, holes will drift to the $p+$ point contact, and electrons towards the $n+$ outer contact. Because they are opposite charges drifting in opposite directions, they will each induce a net positive signal on the readout electrode.

The magnitude of induced charge as a function of time depends on the positon of charges and the electrode geometry of the detector. It is possible to express this for each electrode in terms of a quantity called the "weighting potential," denoted φ_0 [29, 30]. The weighting potential is not the electric potential inside the detector, but instead the solution of the Laplace equation with the boundary conditions $\varphi_0{=}1$ for the readout electrode and $\varphi_0{=}0$ for
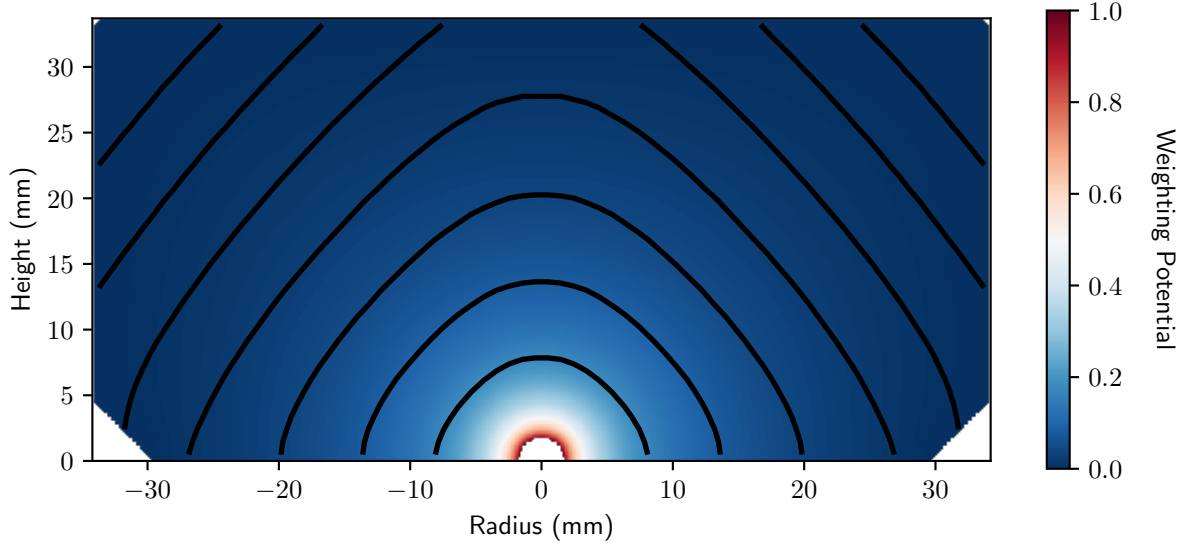
Figure 2.4: Weighting potential for the point contact in a PPC detector. The point contact is located at bottom center, in the "dimple." Black lines are "isochrones," or lines of equal drift time for holes to reach the point contact. Each isochrone is spaced by 100 ns. The range of isochrones in this plot is from 100-700 ns.

every other electrode. By this definition, the weighting potential will always carry a value between zero and one. The Shockley-Ramo theorem then relates the total charge induced on the electrode, Q, as a carrier with charge q drifts through the weighting potential as a function of time:

$$Q(t) = q\Delta\varphi_0(t) \tag{2.1}$$

Given a model for charge carrier migration, the Shockley-Ramo theorem can be used to calculate the signal produced at the readout electrode.

A calculated weighting potential for a PPC detector is shown in Figure 2.4. The PPC geometry characteristically produces a sharp peak in weighting potential in the millimeters directly surrounding the point contact. This weighting potential configuration has two important consequences for the signals produced in PPC detectors. First, since the high weighting potential region is a small fraction of the total detector volume, the vast majority of signals originate in an area with low initial weighting potential. For these events, holes drifting toward the point contact traverse a $\Delta\varphi_0$ of nearly unity, while the $\Delta\varphi_0$ of electrons
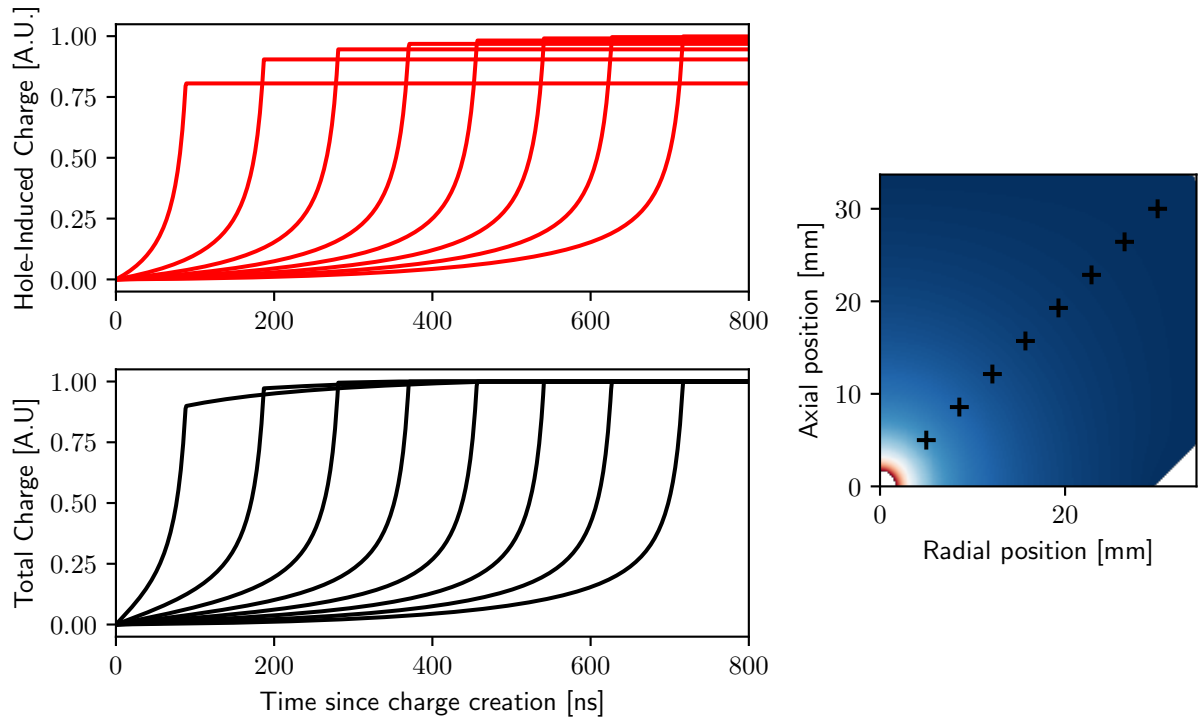
Figure 2.5: Simulated waveforms from eight points in a PPC detector. Positions are marked on the plot at right, which shows half a detector cross section. The shortest waveforms on left correspond to the marker closest to the point contact. The top left plot shows only the hole contribution, while the bottom shows the total contribution from both carriers.

is nearly zero. Therefore, the hole contribution dominates the total signal produced for most PPC waveforms. Second, given the low weighting potential in the bulk, relatively little signal forms until the holes reach the immediate vicinity of the point contact. The result is a waveform that sharply increases just before all the charge is collected. Both of these attributes are visible in the simulated waveforms shown in Figure 2.5. This figure together with the isochrones, or contours of equal drift time, in Figure 2.4, show that the drift time can vary by hundreds of nanoseconds, but no matter the point of origin, the majority of the signal is generated over the final few tens of nanoseconds.

2.1.4:   Pulse Shape Discrimination

For neutrinoless double beta decay experiments, an important benefit of characteristic PPC waveform shape is that it enables discrimination between single and multiple point interactions [23]. A multisite interaction occurs when incident radiation deposits energy at more than one location in the detector, such as high energy gamma ray Compton scattering. Since a $0\nu\beta\beta$ decay deposits its energy within $\sim$1 mm^3, rejecting multisite events can reduce background signals.

Due to the long drift time and short collection time of PPC waveforms, charges deposited at different isochrones (see Figure 2.4) will have clearly time-separated signal components. In terms of the time dependence of the collected charge, the arrival of each new charge cloud generates a new "step" to a higher total charge, seen in Figure 2.6. By differentiating the collected charge and looking at current, each time–separated charge cloud appears as a distinct peak. Methods of multisite pulse shape discrimination in $0\nu\beta\beta$ experiments include cuts based on the energy-normalized amplitude of the current pulse [31] and χ^2 comparison of each waveform to a set of known single site pulse shapes [32].

Section 2.2:   Signal Parameterization

The Shockley-Ramo theorem provides a framework to calculate the detector response for any arbitrary signal in the detector. Calculating the weighting potential requires understanding the detector geometry. Next, a model must be developed to simulate the position of charge carriers as a function of time. Understanding the drift path, in turn, requires a model for the electric fields in the detector and velocities of the charge carriers.

If we wish to compare simulated waveforms to data, we must account for the fact that the frequency response of electronics in the data acquisition system also affects the shape of the observed waveform. Finally, a high precision model requires adjustments for higher-order effects such as size and shape of the charge cloud.
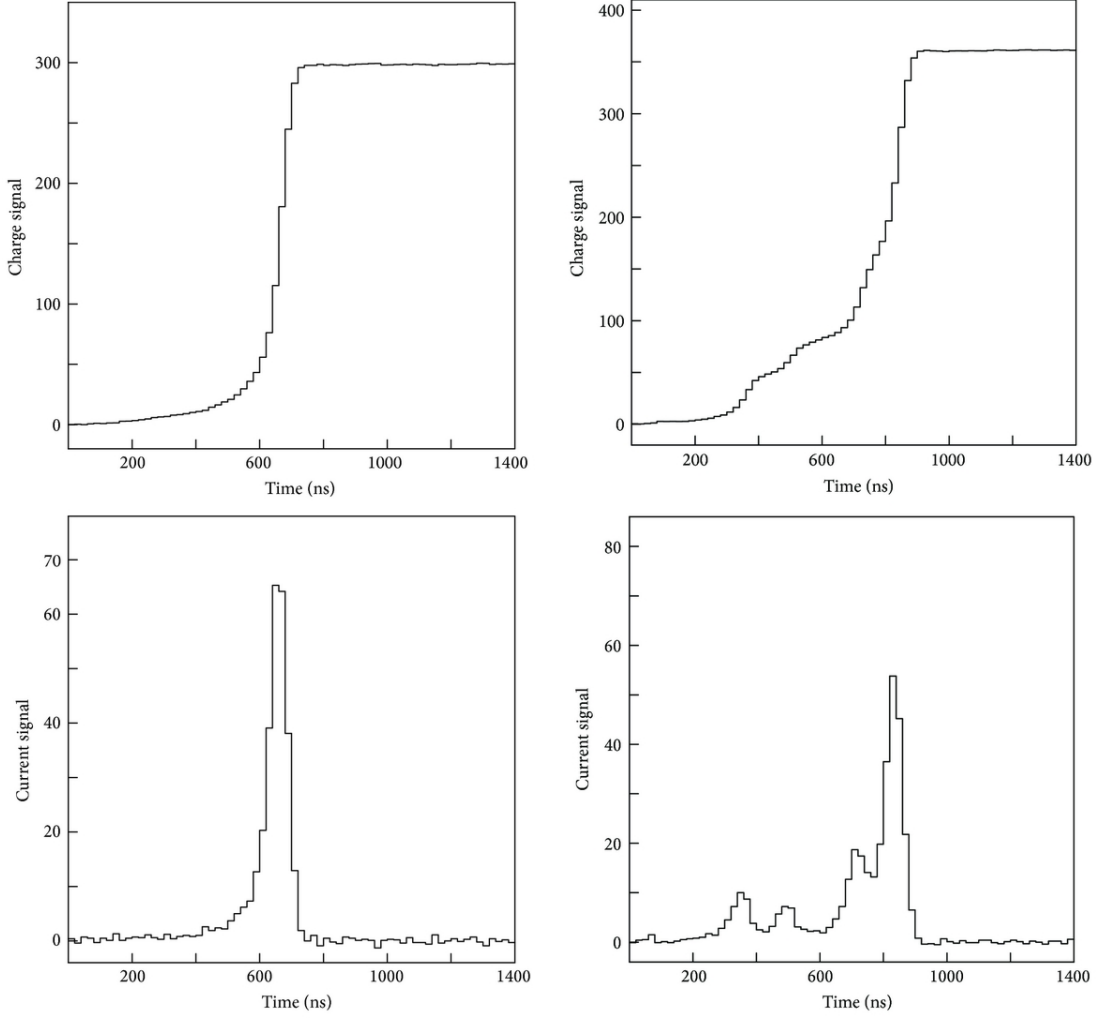
Figure 2.6: Charge (top) and current (bottom) signals formed by single and multisite interactions in a PPC detector. The left plots show a single site interaction with one smooth pulse, while the multisite waveform at right has "steps" in the charge pulse, and multiple peaks in the current pulse. Figure from [33].

This section describes the overall model used in this work. Parameters are sorted by three categories:

1. Detector Parameters: Affect the signal created at the electrode for every waveform simulated in a detector. Examples include detector dimension or drift mobility.

2. Waveform Parameters: Specific to each individual waveform. Includes the location of energy deposition and charge cloud size.

3. Electronics Parameters: Describe the shaping of waveforms by the electronics readout chain. Like detector parameters, these are common to all waveforms from the same detector.

A c software package developed within the GRETINA and MAJORANA collaborations [34] forms the framework of this model. The weighting potential and electric field for a given set of detector parameters are calculated using a program called `fieldgen`. Both are calculated via numerical relaxation on a grid, set for this work to 0.1 mm, using the detector geometry described in Section 2.2.1. We assume azimuthal detector symmetry, and so the relaxation is performed over only the radial and axial dimensions. The weighting potential calculation solves the 2-D Laplace equation given boundary conditions of 1 on the n+ contact and 0 on the p+ point contact. The electric field calculation solves the 2-D Poisson equation, including the charge from ionized impurities using a model of impurity distribution described in 2.2.1. The n+ contact boundary condition is set to the detector operating voltage, and the point contact set to 0 V.

Using the output of `fieldgen`, `siggen` calculates charge trajectories for a given set of waveform and detector parameters. Given an initial position of energy deposition, `siggen` calculates a drift velocity for the local electric field as described in Section 2.2.1. This velocity is used to determine the position after some time step Δt, set for this work to 1 ns. The signal path is determined by iterating over a number of time steps until the charge reaches an electrode. The induced signal at the point contact is calculated for each time step by using the Shockley-Ramo theorem (Equation 2.1), where $\Delta\varphi_0$ is given by the difference in weighting potential at the final and initial position of the step.

It should be noted that `fieldgen` and `siggen` only attempt to model the detector bulk. Edge effects which are not modeled include reduced velocity for charges near the passivated surface and diffusion of charge across the n+-p boundary. These effects are only significant in waveforms with energy deposition $\sim$<1 mm from the detector surface.

As part of the work of this dissertation, the `siggen` code has been updated and incor-

porated into a Python package called `pysiggen`[35]. Included in `pysiggen` are updates to the model for drift carrier velocities and electronics shaping. The python implementation provides several advantages, including rapid development and ease of interface with numerous external packages. By writing computationally intensive sections in `c` and wrapping into Python using `cython` [36], these advantages can be realized without sacrificing much in the way of computation time.

2.2.1:   Detector Parameters

Dimensions

The Majorana collaboration contracted AMETEK/ORTEC [37] to produce PPC detectors for the Demonstrator using enriched ^{76}Ge material. In order to maximize total enriched detector mass cut from each crystal boule, ORTEC allowed the exact dimensions of each cylindrical detector to vary slightly. The typical mass is on the order of 1 kg, with diameters ranging from 60-70 mm and lengths from 40-50 mm. The point contact, located at bottom center of the detector, is recessed in a "dimple" with a radius and depth which vary from detector to detector by up to a millimeter. This dimple is modeled in `fieldgen` as a hemispheroidal indentation. The outer diameter of each detector has a "taper," a 45° chamfer on the bottom outside edges, which is nominally 4.5 mm in height. Finally, the top outside edges have a "bulletization" rounding radius, which is around 1.2 mm. Upon acceptance of the detectors, Majorana collaborators measured the detector dimensions using a high precision optical measuring device.

The $n+$ contact on these detectors is created by diffusing lithium into the crystal. The thickness of this layer is on the order of hundreds of microns, and can reach a millimeter. Because the depleted volume of the detector is on the p side of the p-$n+$ junction, the $n+$ layer thickness reduces the active detector volume. In order to simulate only the active volume, the detector dimensions are therefore reduced from the measured values by the thickness of the $n+$ layer.

Impurities

Through the zone refinement and crystal pulling process described in Section 2.1.2, OR-TEC is able to create detectors with impurity concentrations on the order of 10^9 atoms per cubic centimeter. In general, the seed end of the crystal has higher impurity levels, and pulling the crystal introduces a gradient along the crystal length. The detector is manufactured in line with the direction of pulling, such that the impurity gradient aligns with the axial direction of the detector. The charge introduced by the impurities affects the electric field within the detector during operation. The gradient varies greatly between crystals, but is generally between $\sim 0.01 - 0.2 \times 10^{10} \text{cm}^{-4}$.

The `siggen` model used in this work allows for a linear impurity gradient, and assumes no radial impurity gradient. It is possible that a radial gradient is also present [38], but for successful crystal pulls at ORTEC, it is assumed this effect is negligible.

Impurities can be measured using the Hall effect, but the technique has relatively high uncertainties [21]. However, the impurity profile, together with the detector dimensions, determine the bias voltage at which the entire crystal volume depletes. The signals of a depleted and undepleted detector are significantly different, with the latter producing distinctly "rounded", slower pulses due to diffusion in the undepleted region. Performing a careful characterization of the depletion voltage, and comparing to predicted depletion voltage for a given impurity profile, provides an independent check to the Hall measurements.

Operating Voltage

In general, germanium detectors are operated at some margin above the depletion voltage to ensure depletion, and to increase field strength and therefore drift velocity. Beyond the depletion voltage, field strength scales linearly with increased bias. Suggested operating voltages are supplied by ORTEC, which are applied to DEMONSTRATOR detectors with a high precision WIINER MPOD crate [39] with an iseg EHS 8260P high voltage supply [40]. Typical bias voltages range from 1500-5000 V.

Drift Velocity

Accurately modeling signals in germanium detectors depends critically on understanding carrier drift velocities. Because the charges are propagating in a crystal, carrier mobilities are governed by the solid state physics of the system. This introduces nontrivial dependence of the drift velocity on both the magnitude of the electric field and the orientation of field with respect to the axes of crystal symmetry.

At low electric field and high temperature, the carrier drift velocity in germanium is Ohmic, meaning it increases linearly with electric field. In this regime, the velocity for each carrier is characterized by a mobility μ_0, where

$$\vec{v} = \mu_0 \vec{E} \tag{2.2}$$

The drift velocity, $\vec{v}$, will necessarily align with the electric field, $\vec{E}$. As the energy of the carriers relative to the lattice temperature increases, scattering with the crystal lattice causes a deviation from Ohmic behavior [41]. This scattering causes the drift velocity to reach a point of saturation, such that the drift velocity becomes constant with increasing field. Carriers with energies high enough relative to the thermal energy of the crystal to deviate from Ohmic behavior are referred to as "hot" electrons or holes. Empirically, the high field drift velocity follows the form [42]:

$$v(E) = \frac{\mu_0 E}{(1 + (E/E_0)^\beta)^{1/\beta}} \tag{2.3}$$

At low fields ($E \ll E_0$), the denominator term goes to one, and the Ohmic behavior $v \propto \mu_0 E$ is recovered. At high field ($E \gg E_0$), the velocity approaches asymptotically to a saturation velocity, $v = \mu_0 E_0$. Here, E_0 is a parameter with units of electric field which sets the saturation velocity. The general shape is visible in Figure 2.7. The β parameter controls the curvature between the linear and saturated regions, and typically falls between 0.2-2. This

29

equation can describe germanium carrier mobilities at a range of temperatures [43].

In addition to scattering off the lattice, it is possible for carriers to scatter with charged crystal impurities [44, 45]. It is therefore possible for the observed parameters of Equation 2.3 to vary between germanium detector crystals even when operated at the same temperature.

Germanium has a diamond cubic lattice, with crystallographic basis vectors in the $\langle 100 \rangle$, $\langle 110 \rangle$ and $\langle 111 \rangle$ directions. The band structure (Figure 2.8) varies in energy depending upon the orientation of the wave vector $\vec{k}$ with respect to the crystal axis. At high fields, this generates an anisotropy in the induced carrier drift velocity, where the velocity magnitude varies depending on the orientation of the field. In addition, the induced current at high fields is no longer necessarily parallel to the electric field [48] [1]. The details on how this arises are somewhat different for electrons, which propagate in the conduction band, and holes, which are affected by the valence band structure.

The lowest minima in the germanium conduction band are four ellipsoidal "valleys." The long axis of each ellipsoid is aligned with an $\langle 111 \rangle$ crystallographic axis [50, 51], as shown in Figures 2.9. Electrons populating this minimum with a wave vector $\vec{k}$ parallel to the long axis have larger effective masses than those with a k value perpendicular to the axis, and therefore have reduced conductivity and contribute less to the current.

To understand how this generates misalignment between the field and current vector, consider first an electric field applied at some arbitrary angle with respect to a single ellipsoid. Because of the difference in mass, an electric field which is aligned with the small-mass ellipsoid axis will generate a larger current than one aligned in the large-mass $\langle 111 \rangle$ direction. In addition, the higher conductivity in the small-mass direction causes the current vector to rotate somewhat toward the low-mass direction. Each of these effects is present regardless of the electric field strength.

For a field applied in a true germanium crystal, the net current will be a vector sum of

---

[1] This effect is also observed in other semiconductors, including silicon.
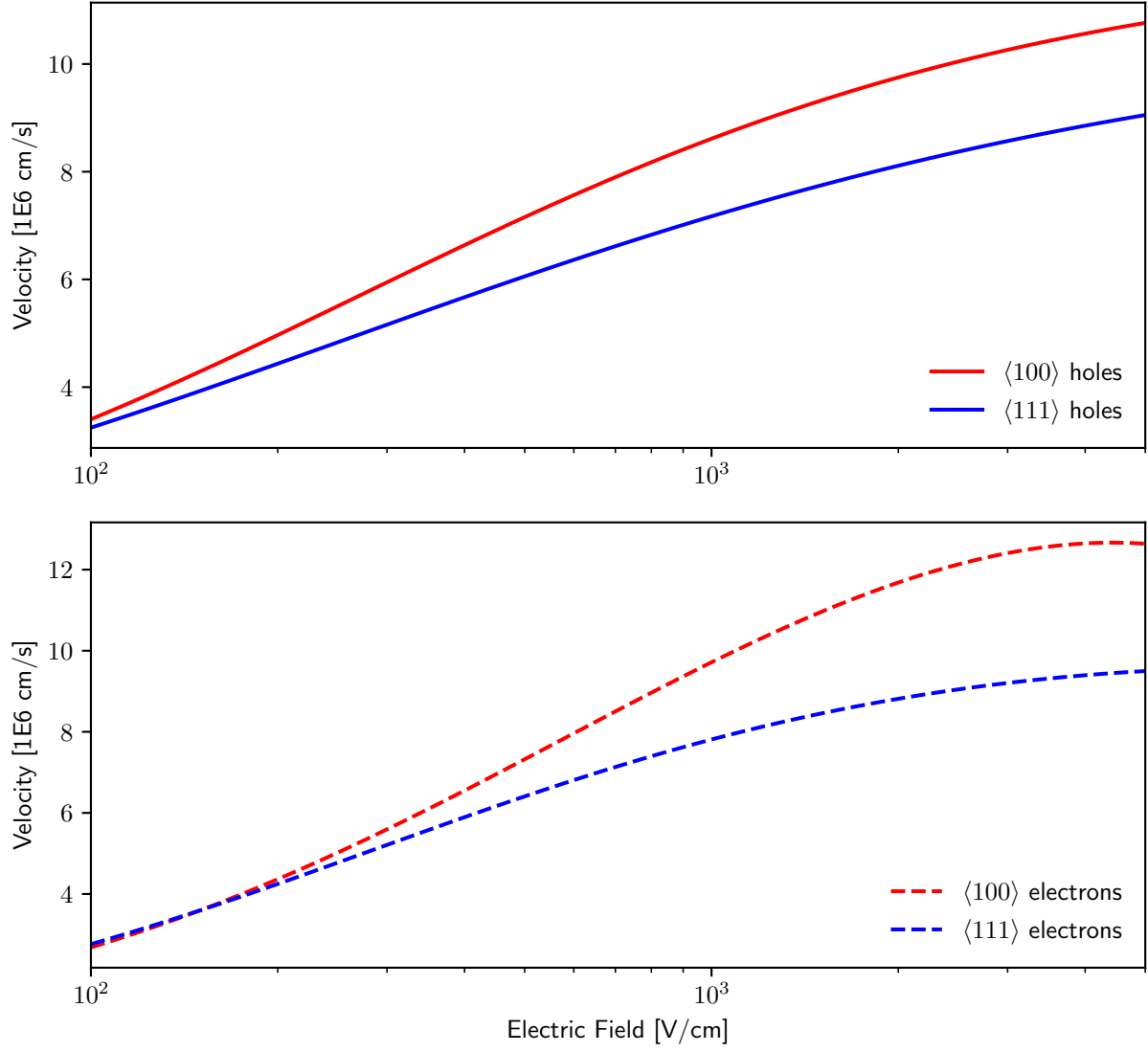
Figure 2.7: Drift velocity of holes and electrons at 78 K for the ⟨100⟩ and ⟨111⟩ axes. Hole values are from [46], electrons from [47]. The curves are produced using Equation 2.3 fit to experimental data, with parameters obtained by fitting to experimental data.
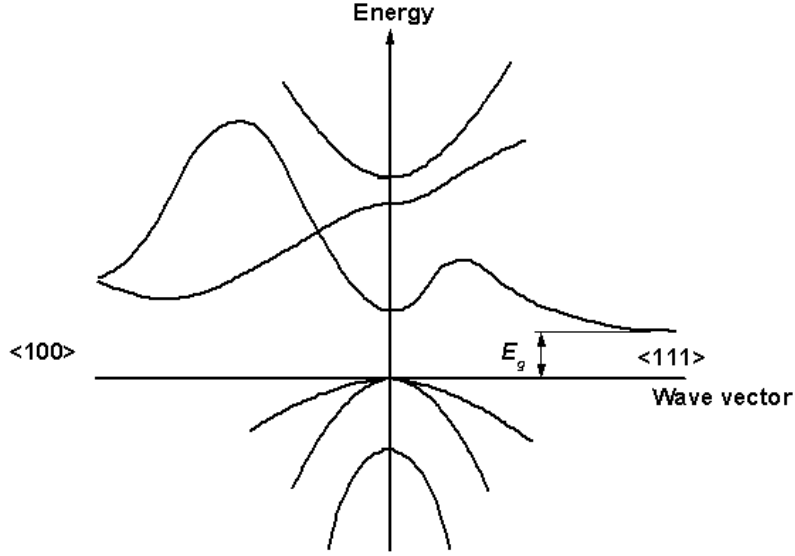
Figure 2.8: Band structure in germanium. Curves indicate the energy associated with various bands as a function of wavevector. The conduction band minimum at $\langle 111 \rangle$, labelled E_g, is associated with the ellipsoidal surfaces shown in Figure 2.9. At the valence band maximum enery, two degenerate bands meet, giving rise to hole populations with two distinct effective masses. The wider of the two is the so-called "heavy hole" band. Figure from [49].
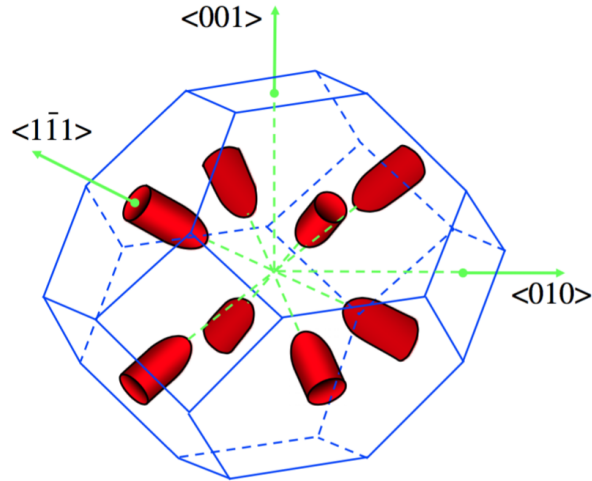
Figure 2.9: Surfaces of equal energy around the minima in the germanium conduction band. Each of the minima are ellipsoidal in shape, with long axes aligned along the $\langle 111 \rangle$ axis. Minima on parallel surfaces represent opposite ends of one complete ellipsoid, so there are four distinct symmetry-related minima. Figure from [52].

32

the current vectors produced from each of the four minima. At low fields, cubic symmetry necessitates that the components misaligned with the field cancel, and the net current vector is parallel to the field. However, at high fields, the velocities of the electrons are high enough that behavior is no longer Ohmic. Electrons in ellipsoids for which the field is most aligned with the small-mass axis become hotter, and therefore have reduced mobilities compared to large-mass dominated valleys. This breaks the symmetry and results in a current which no longer aligns with the electric field [41]. The anisotropy is enhanced by "inter-valley" electron transfer with higher local minima in the band structure [53].

The situation for holes is somewhat more complicated due to the degeneracy in the valence band maximum, seen at $\vec{k} = 0$ in Figure 2.8. Two bands have the same ground state energy, but different shape, creating a situation in which ground state holes can have two different masses. The wider of the two bands has a higher effective mass and is labelled the "heavy" hole band. The shape of the heavy hole band, shown in Figure 2.10, is warped–spherical, with some dependence on crystal axis. The anisotropy produces effective masses which are largest in the ⟨111⟩ direction and smallest in the ⟨100⟩. By the same argument as electrons above, this condition causes misalignment of the current and electric field vectors.

The light hole band is almost perfectly spherical and does not significantly introduce anisotropic drift. In addition, at thermal equilibrium, the heavy holes makes up 96% of the hole population [41]. For these reasons, it is possible to model the hole transport using only the heavy hole band without significant loss of accuracy [54].

For fields aligned with crystallographic axes, rotational symmetry requires that the drift velocity aligns with field. However, the effect of the band structure described above causes decreased mobility for the ⟨111⟩ compared to the ⟨100⟩ direction, for both holes and electrons. Each axis can be then independently described by Equation 2.3. Fits to experimental data for holes and electrons along the ⟨111⟩ and ⟨100⟩ axes are shown in Figure 2.7.

Due to the localized weighting potential in a PPC detector, the hole contribution domi-nates signals for signals originating from the vast majority of the detector (see Section 2.1.3).
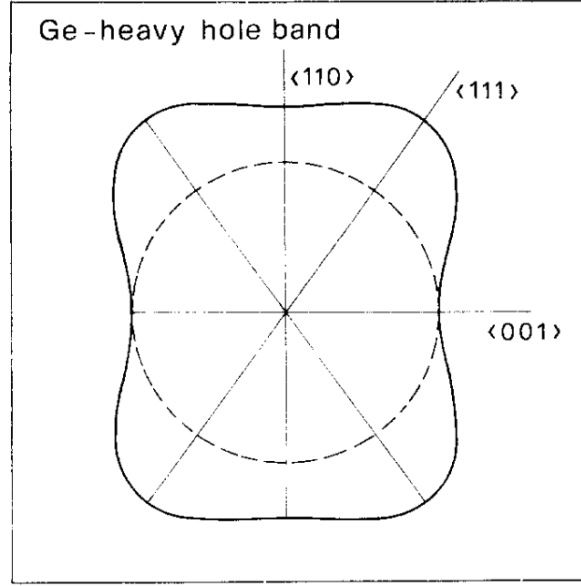
Figure 2.10: Energy in the $\langle 110 \rangle$ plane of the heavy hole germanium valence band, shown with the solid line. The radial direction indicates valence band energy for the wavector at matching the polar angle. The dashed circle is included to emphasize deviations from a spherical structure. Figure from [54].

The hole drift velocity is therefore more important to model correctly. However, because of the complex valence band structure discussed above, precise calculation of the velocity for arbitrary field direction is difficult, and usually performed with computationally expensive Monte Carlo calculations [46]. To decrease calculation time in `pysiggen`, we have implemented an approximate model of hole velocity for fields with arbitrary crystal orientation first developed for gamma ray tracking experiments like AGATA [52]. Given the field orientation and velocity in the $\langle 111 \rangle$ and $\langle 100 \rangle$ directions, this model provides an approximate magnitude and direction for the velocity. The approximation is able to predict the $\langle 110 \rangle$ drift velocity to high accuracy.

We therefore require as input to `pysiggen` six parameters to describe the hole velocities: the three parameters μ_0, E_0 and β of Equation 2.3 for both the $\langle 111 \rangle$ and $\langle 100 \rangle$ crystal axes. This allows the flexibility to account for differences in drift velocity between detectors due to differences, for example, in temperature or impurity concentration.

The currently implemented model of electron velocity in `siggen` is somewhat simpler.

With input of velocities in all three crystallographic directions, it performs polynomial interpolation to calculate field in a given orientation. The velocity values are taken from data collected in [43]. More precise models exist [47, 52] and could be implemented in the future. While it is possible this introduces percent-level error in waveform shape for events originating near the point contact, in the crystal bulk we expect accuracy to the order of parts per ten thousand.

Charge Trapping and release

The presence of impurity sites and crystal defects introduces potential wells for charge carriers. It is possible for some carriers to become trapped at these sites as they drift through the detector. Charge that becomes trapped will no longer induce signal at the electrode, so trapping attenuates the observed pulse amplitude. This effect can be modeled in terms of a mean free drift time of charge carriers before encountering trapping [55].

In the assumed model, the magnitude of drifting charge is reduced by a constant fraction per unit time, resulting in an exponential attenuation in charge as a function of drift time:

$$q(t) = q_0 e^{-t/\tau_T} \tag{2.4}$$

where $q(t)$ is the time-dependent charge contained in the carrier cloud, q_0 is the charge created in the initial interaction, and τ_T is the mean free drift time of trapping. We have implemented this model in `siggen` by reducing the effective charge of the charge cloud as a function of time. An example of waveforms generated with varying charge trapping constants is shown in Figure 2.11.

The mean hole trapping time will vary between detectors due to differences in trapping site concentration between crystals. Because the magnitude of charge trapped depends on drift time, pulses further from the point contact will experience greater charge trapping and therefore lower amplitude than closer interactions of the same energy. The energy of each
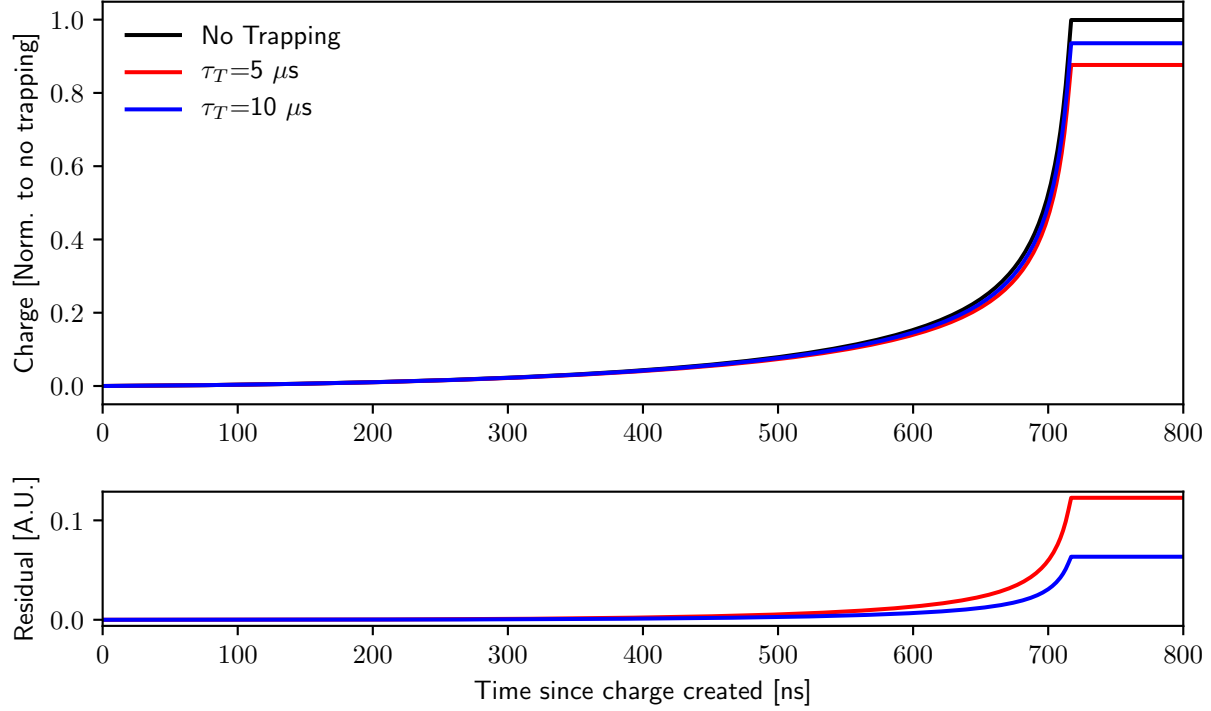
Figure 2.11: Simulated effect of charge trapping on hole compontent of PPC waveforms using the model described in Section 2.2.1. The black waveform shows the induced charge from hole carriers with no charge trapping. The red and blue show charge trapped waveforms mean charge trapping constants of 5μs and 10μs. These are much shorter than the hundreds of μs constants observed in the DEMONSTRATOR, but are chosen to more clearly display the effect. The residuals at bottom, which compare each trapped waveform to the no–trapping case, measure the the total charge trapped as a function of time.

waveform is reconstructed from the amplitude of the pulse, so charge trapping results in a degradation of energy resolution. This effect has been observed in the Demonstrator detectors with trapping constants on the order of hundreds of microseconds, corresponding to an energy reduction on the order of parts per thousand. The collaboration has developed a technique to correct the energy reconstruction which improves the resolution at 2614 keV by a factor as high as two.

Charges will eventually thermally escape from trapping sites. Release of trapped charge contributes a delayed current, resulting in a slow component of the signal. Like trapping, the release can be modeled in terms of an exponential time constant, τ_R. This effect is modeled very simply in pysiggen by summing the total current trapped charge and exponentially returning the charge to the cloud. During the drift time, the charge is added to the current location of the charge cloud, and not released at the point it was trapped. Once the charge cloud has reached the contact, amplitude is added to the signal exponentially until it reaches the amplitude expected for a signal with no trapping. The model is shown in Figure 2.12

This model for charge release is an imperfect approximation. A more complete model should re-inject charge at the point it was trapped. Doing so, however, would greatly increase the computational complexity of a waveform simulation.

The current implementation of `pysiggen` includes only one set of charge trapping and release constants which is applied to both the hole and electron signal. Physically, the two carriers are trapping at different types of impurities, and therefore there is no reason to expect their trapping and release times should be identical. However, since the hole contribution dominates the signal in a PPC detector, this approximation should not introduce significant error.
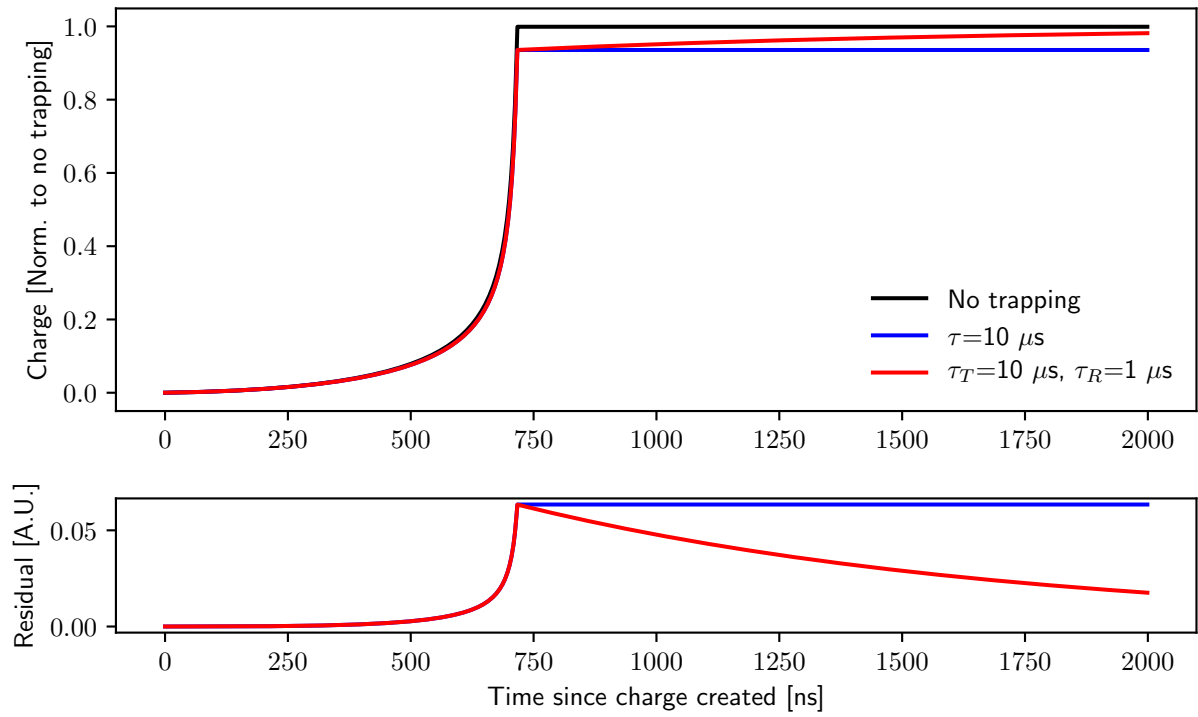
Figure 2.12: Effect of charge release from trapping sites as modeled in `pysiggen`. The addition of charge release causes the red waveform to return exponentially in amplitude from its trapped value, in blue, to the no-trapping value, in black.

2.2.2: Waveform Parameters

Interaction Position

Given a weighting potential, electric field, and drift velocity model, the generated signal for interactions for any a given position can be calculated. Both the drift path and total drift time of the charge cloud are obviously sensitive to the radial and axial position. To first order, the shape of the waveform is dominated by the hole drift time, and therefore the isochrone of origin. However, Figure 2.13 shows that there are percent-level differences in waveforms even along an isochrone.
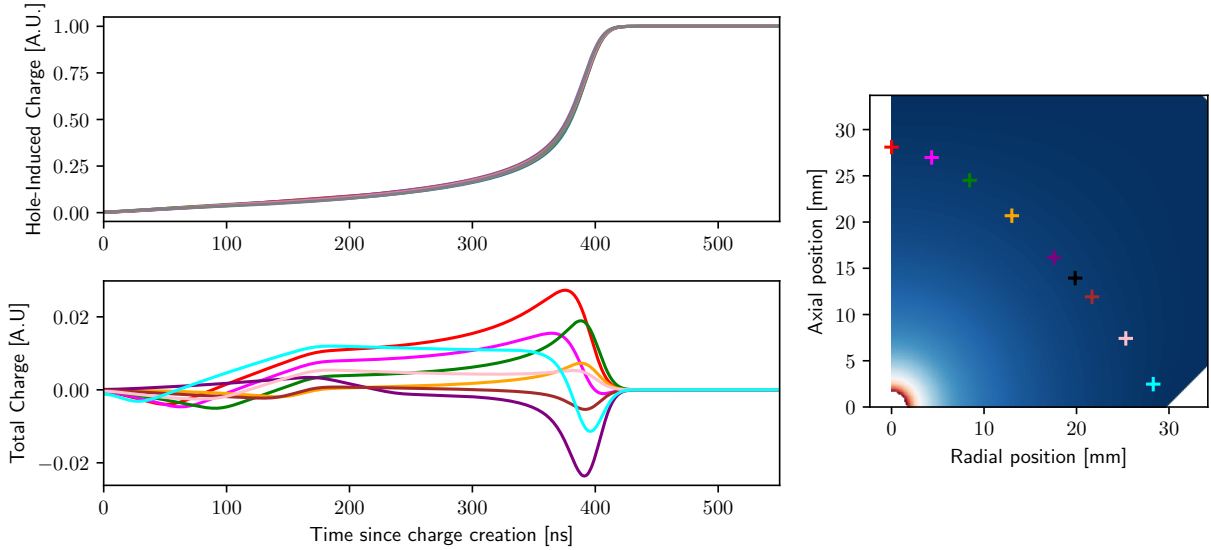
Figure 2.13: A comparison of shape between simulated waveforms with the same drift time (400 ns). Each event originates at azimuthal position $\phi = 0$, which corresponds to the $\langle 100 \rangle$ crystal axis. The position of each event, shown at right, corresponds with the waveform color at left. Residuals are drawn compared to a waveform from the position at right marked in black. Differences between the waveforms are at the part per hundred level, which originate because of differences in weighting potential and electric field along the drift path.

As described in Section 2.2.1, the velocity of charge carriers in high electric fields are anisotropic with respect to crystal axis. The crystal growth process ensures that the detector is oriented axially along the $\langle 001 \rangle$ axis. The crystal structure of germanium is diamond cubic, so the plane perpendicular to the axial will contain a axis and, separated by $\pi/4$ radians, a
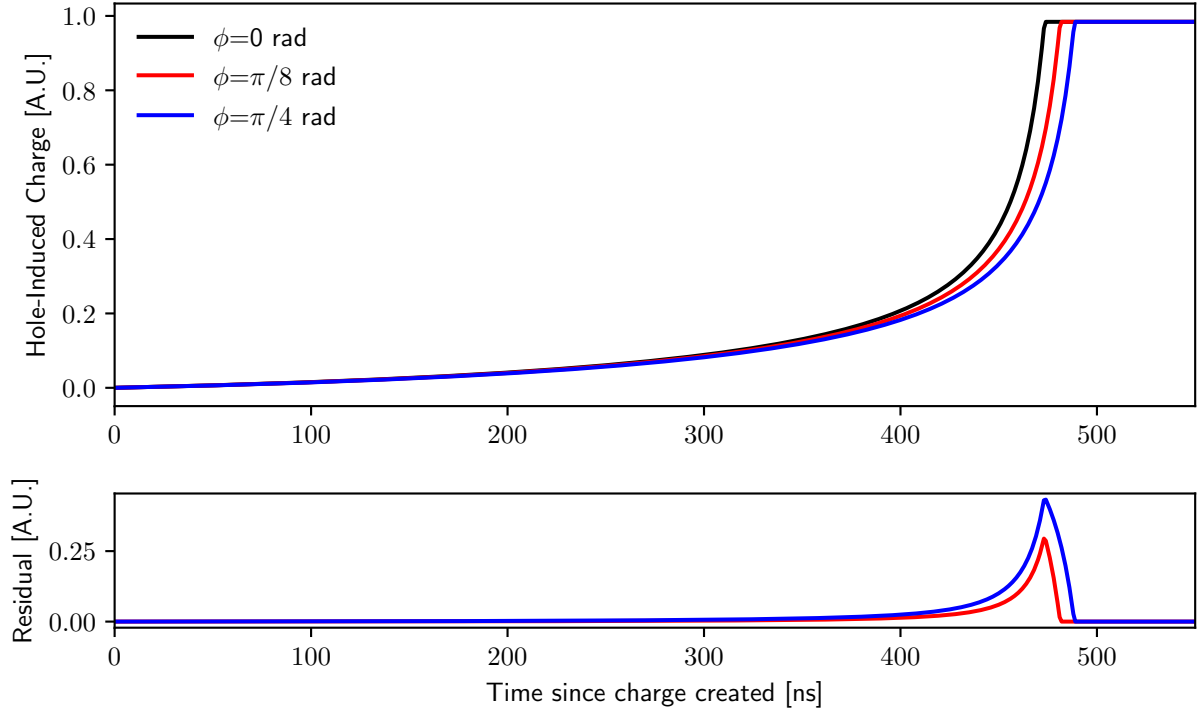
Figure 2.14: Comparison between waveforms simulated at the same radial and axial position $(r = 20, z = 20)$, but varying the azimuthal position. The crystal axis anisotropy in drift velocity at high electric field is responsible for the change in drift time. The fastest velocities occur along the $\langle 100 \rangle$ axis, corresponding to $\phi = 0$, while the $\langle 110 \rangle$ axis corresponds with $\phi = \pi/4$. Residuals are shown relative to the $\phi = 0$ waveform.

$\langle 110 \rangle$ axis. The waveform is therefore sensitive to the azimuthal position of the event, but with an eight-fold degeneracy. An example of waveforms from the same position generated at different azimuthal positions is shown in Figure 2.14. The fractional difference in risetime, and therefore the residual difference, between waveforms from the $\langle 100 \rangle$ and $\langle 110 \rangle$ axis will vary as a function of position (Figure 2.15). In general, maximum residual differences are on the order of parts per hundred.

The position of the waveform is therefore modeled in three dimensional cylindrical coordinates. The parameters r and z, which express the radial and axial position, can vary anywhere within the detector volume. The azimuthal ϕ position is modeled only from $[0, \pi/4]$, corresponding to the crystal axis symmetry.
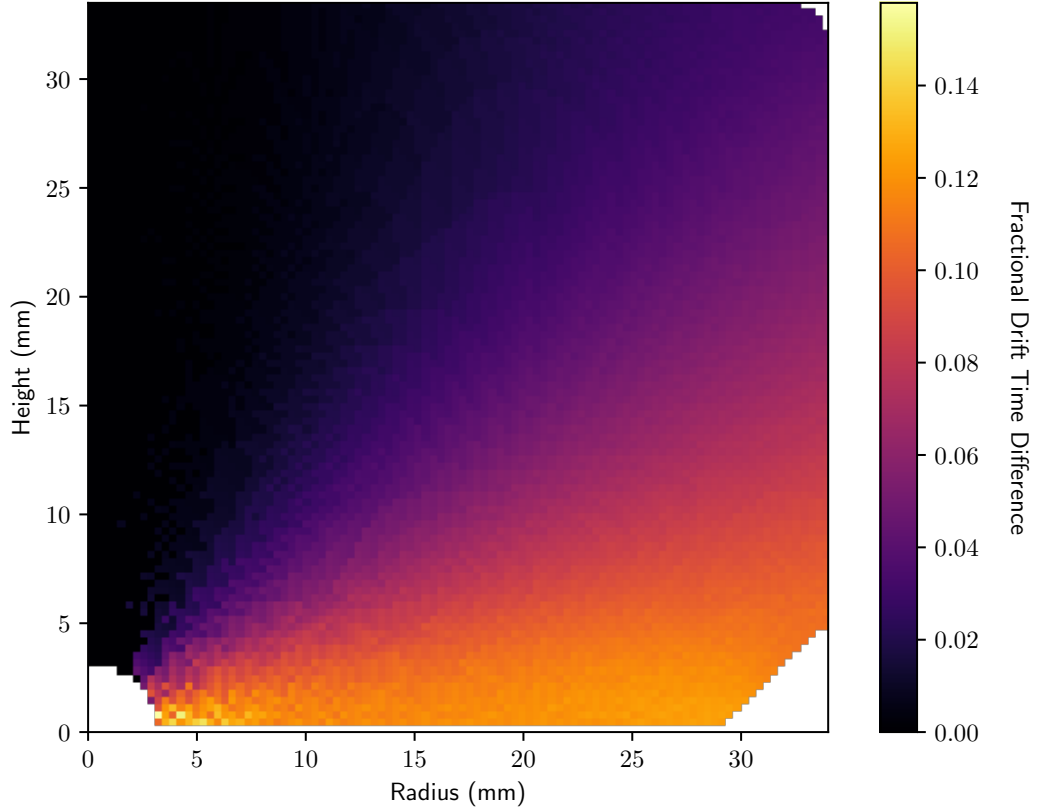
Figure 2.15: Fractional difference in rise time calculated between events which occur at the ⟨100⟩ and ⟨110⟩ axes as a function of position. The drift velocity is lower along the ⟨110⟩. The rise time difference increases with increasing radius due to the integrated effect of the lower drift velocity. The difference is highest at lower axial position because here the the calculated cloud drifts purely along the ⟨100⟩ or ⟨110⟩ axis to the point contact, whereas for higher positions both clouds must also drift along the ⟨001⟩ axis (aligned with the axial direction), which reduces the percent of the drift time the two clouds experience different drift velocities.

Energy

Assuming that all the charge generated in an event is collected, the signal amplitude depends linearly on energy. In `pysiggen`, the simulated signal is adjusted for energy before the charge trapping correction is applied. When incorporating the shaping of the electronics (see Section 2.2.3), all filters applied are corrected to preserve unity gain of the signal. By this procedure, the energy parameter remains a true measurement of energy, independent of any other parameter's effect on amplitude.

Charge cloud size & shape

The size and shape of the ionization charge cloud created in the crystal varies depending on the type of particle interacting and the total energy deposited. After creation, the cloud size can further grow due to diffusion and self-repulsion. We assume this cloud is created spherical gaussian in nature, reflecting a point-like energy deposition[2]. As charges accelerate and decelerate, the cloud stretches in space, but remains roughly gaussian in time. Since the majority of signal is generated very near the point contact, we make the further approximation that the cloud size correction can be performed with a single gaussian convolution in the time domain. The gaussian sigma is parameterized in time, which corresponds to the spatial width via the final velocity of the charge.

This approximation was made to reduce computational complexity, but does not fully encapsulate the shape of the signal. To quantify the degree of error introduced, the true signal produced by spherical gaussian signal was created using Monte Carlo. A million random points were generated around the interaction point, with each of the r, ϕ and z components and normally distributed with $\sigma = 0.5$ mm. A waveform was simulated for each of these points, and averaged to create a final signal. This was compared to the simplified gaussian convolution, choosing the gaussian sigma which minimized the least square difference from

---

[2]For radiation with longer interaction track lengths, such as MeV-scale electrons, the charge will be elongated from a simple spherical gaussian cloud. This effect is not currently modeled.

the Monte Carlo waveform. Figure 2.16 shows the result of this study. For realistic charge cloud sizes at any point in the detector, the calculated difference between the full Monte Carlo model and the simplified gaussian convolution is on the order of parts per thousand.
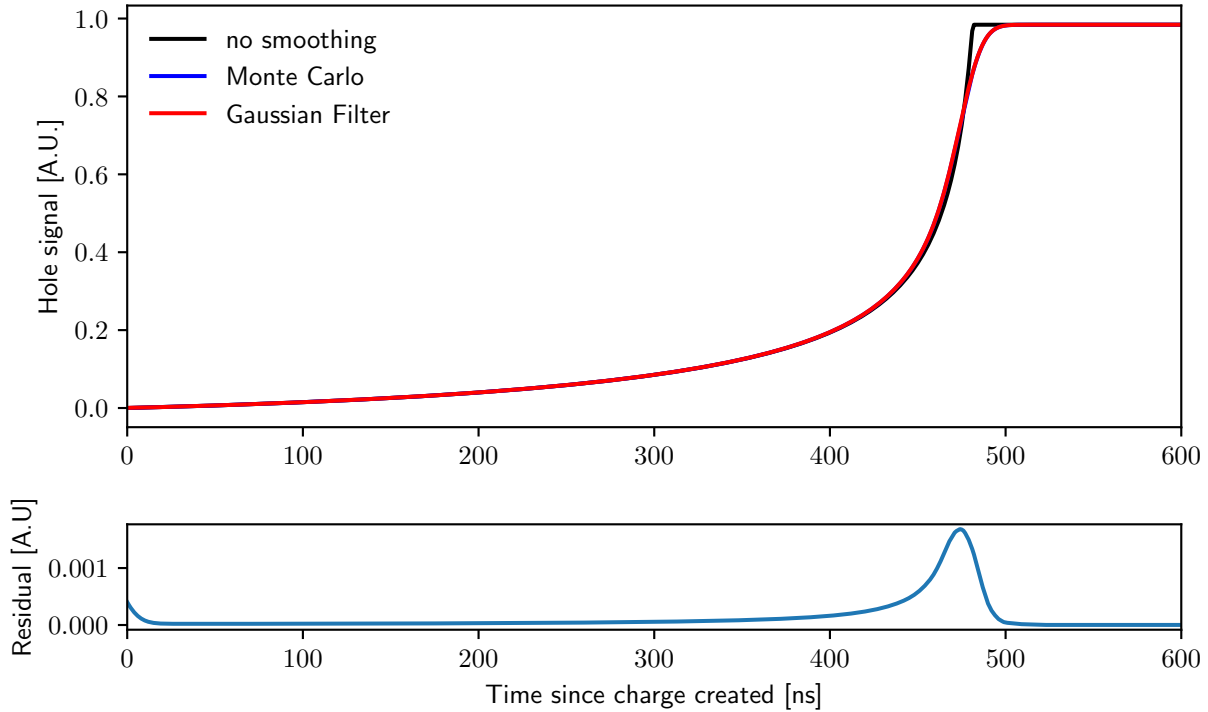
Figure 2.16: Measure of the error introduced by approximating the charge cloud shape as a gaussian convolution in the time domain of a single-point waveform. The red waveform shows the approximated waveform, while the blue waveform is the average of a set of many waveforms drawn from a spherical gaussian distribution around the same point. The residual below shows the difference between the Monte Carlo and Gaussian Filter waveforms.

In addition to the charge cloud size, `pysiggen` includes an optional parameter to modify the shape of the gaussian cloud. The spherical gaussian charge cloud assumes a point-like energy deposition. For events in which the energy despotion has some finite width, the charge cloud distribution should initially be more like a uniform sphere. However, diffusion as the cloud travels still spreads the charge in a gaussian manner. To combine the effects, `pysiggen` utilized a "generalized Gaussian" distribution, which introduces a parameter p to

the cloud shape distribution such that the convolution weighting distribution $w(t)$ becomes

$$w(t) \propto \exp\left(-\frac{1}{2}\left|\frac{t}{\sigma}\right|^{2p}\right),$$ (2.5)

where σ still controls the width, and p controls the shape. The usual gaussian distribution is recovered for $p = 1$ and, in the limit $p \to \infty$, $w(t)$ becomes a uniform distribution over $(-\sigma, \sigma)$. See Figure 2.17 for examples of this distribution.
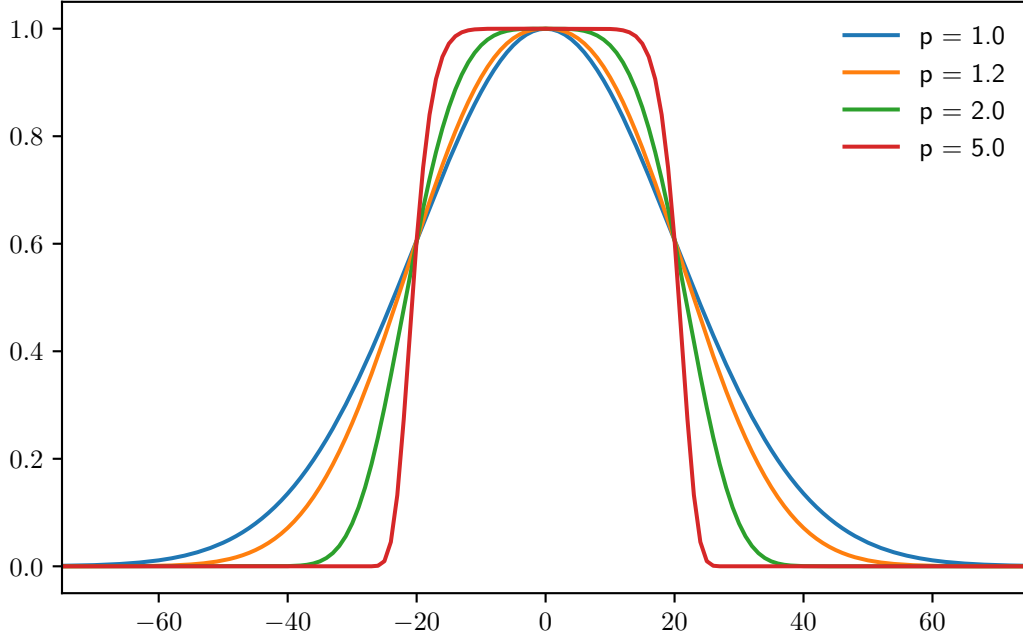
Figure 2.17: Generalized gaussian charge cloud shape described in Section 2.2.2. The distribution of Equation 2.5 is drawn with $\sigma = 20$ for four different values of p.

2.2.3: Electronics Parameters

The MAJORANA collaboration designed an optimized detector electronics readout and amplification system in an effort to minimize noise while achieving low radioactive backgrounds, which is shown schematically in Figure 2.18. The readout is performed with a resistive feedback circuit, containing an input field effect transistor (FET) and amorphous
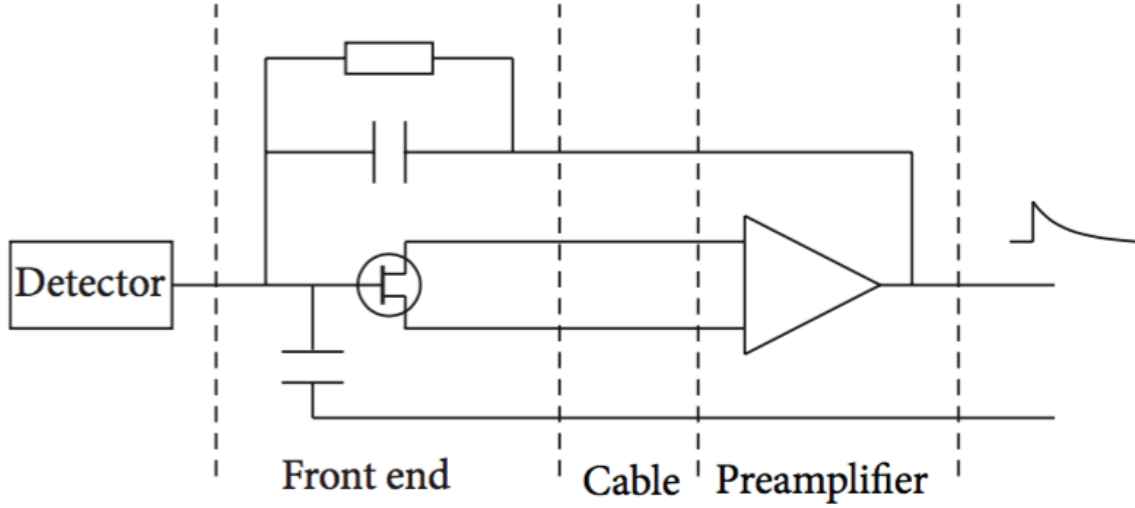
Figure 2.18: Schematic of the DEMONSTRATOR signal readout chain. The low mass front end (LMFE) board is located within the cryostat, very near to the detector point contact, as seen in Figure 1.4. The LMFE contains a resistive feedback element (RC element at top), field effect transistor with drain and source lines, and capacitively coupled pulser input (bottom). The four lines are bundles into one cable, which runs ~2 m to additional amplification elements outside the shield. Figure from [33].

germanium resistor, called the "low mass front end" (LMFE) board [19]. The LMFE is located within a few millimeters of the detector, reducing stray input capacitance and therefore series noise. However, given this positioning, the LMFE hardware must be must be minimal and constructed only with very radiopure materials, while the rest of the preamplifier system is located outside of the cryostat. Approximately 2 m of 0.4 mm diameter cable runs between the LMFE and first stage of the preamplifier, thereby creating a feedback loop 4 m long. A second stage preamplifier, also outside the cryostat, is capacitively coupled to a second stage of amplification. The signal is then digitized with 14 bit precision at 100 MHz with GRETINA digitizer boards [56], which were originally developed for the GRETINA experiment.

Given the extremely long feedback loop, and because the FET response depends sensitively on temperature, it is difficult to accurately model the response of the electronics readout with a circuit simulator. A capacitively-coupled pulser line runs to the input of the

FET, but the pulser input is also shaped by 2 meters of cable running into the cryostat. Without a direct measurement of the transfer function, we have developed an empirical parameterization.

The LMFE and inductance from the cable attenuate high frequency components of the signal, and can be modeled as a low-pass filter. A transfer function of order n can be generically modeled as a digital filter in the z domain

$$H(z) = \frac{c_n z^n + c_{n-1} z^{n-1} + \ldots + c_0}{d_n z^n + d_{n-1} z^{n-1} + \ldots + d_0} \tag{2.6}$$

To determine which order was appropriate for the DEMONSTRATOR electronics, we attempted to fit waveforms with a generic filter of first, second, and third order. Qualitatively, the second order filter was found to match the shape of the observed waveforms significantly better than a first order filter. The third order filter showed no dramatic improvement over the second order. For this reason, we chose to use a second order transfer function.

The fit to the second order preferred values of c_0 near to zero, giving us a total digital filter function with four free parameters

$$H_{low}(z) = \frac{az^2 + bz}{z^2 + 2cz + d^2}. \tag{2.7}$$

In the time domain, this corresponds to a convolution with a decaying oscillatory kernel,

$$H(t) \propto e^{-t} \cos(\omega t + \phi), \tag{2.8}$$

with the correspondence between the time and z domain is given by:

$$\omega = \cos^{-1}\left(\frac{c}{d}\right) \quad \text{and} \quad \phi = \tan^{-1}\left(\frac{ac - b}{a\sqrt{d^2 - c^2}}\right) \tag{2.9}$$

46
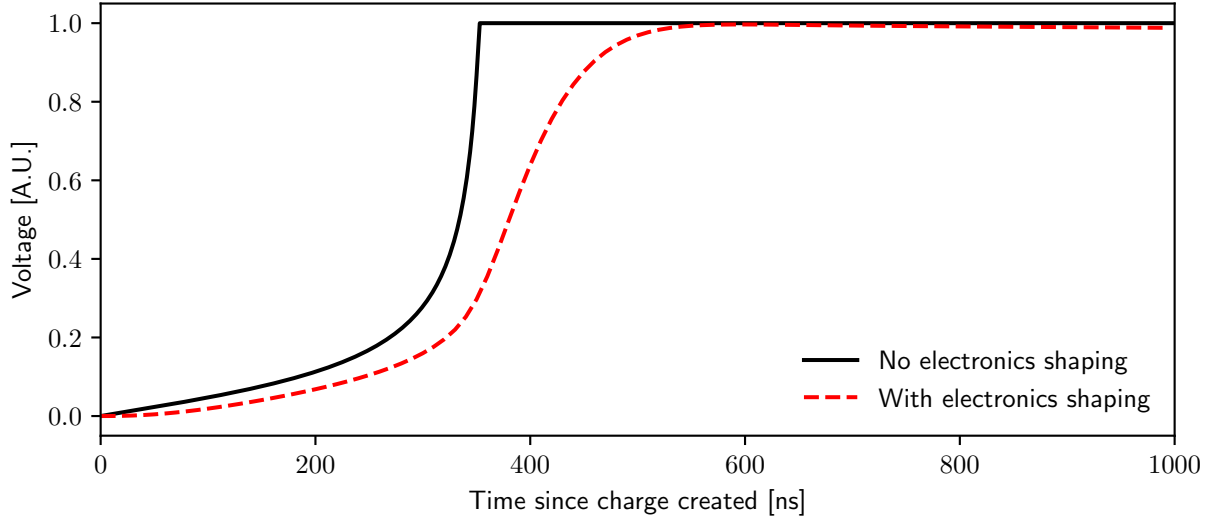
Figure 2.19: Simulated effect of the electronics chain on waveform shape. Note that the risetime is increased by $\sim$100 ns.

The steady-state, or DC, gain for the filter is given by

$$\lim_{z \to 1} \frac{az^2 + bz}{z^2 + 2cz + d^2} = \frac{a + b}{1 + 2c + d^2} \tag{2.10}$$

and therefore increasing the value of $a+b$ only linearly scales the amplitude without otherwise affecting the shape.

Capacitative couplings, most notably between the first and second preamp stages, cause the waveform to exponentially decay. Empirically, we observe two strong coupling constants with different strength. We model this with three additional parameters describing a linear combination of two exponential decay functions:

$$H_{hi}(z) = c \frac{z - 1}{z - \exp(-\frac{T}{\tau_1})} + (1 - c) \frac{z - 1}{z - \exp(-\frac{T}{\tau_2})}. \tag{2.11}$$

Here, τ_1 is set by the RC constant of the coupling between the first and second stage amplifiers, usually around 72 μs, and τ_2 is a $\sim 2\mu$s coupling we believe is intrinsic to the digitizer. The constant c expresses the fractional mixing between the decay constants.

The total transfer function originating from the electronics is the product of H_{low} and

47
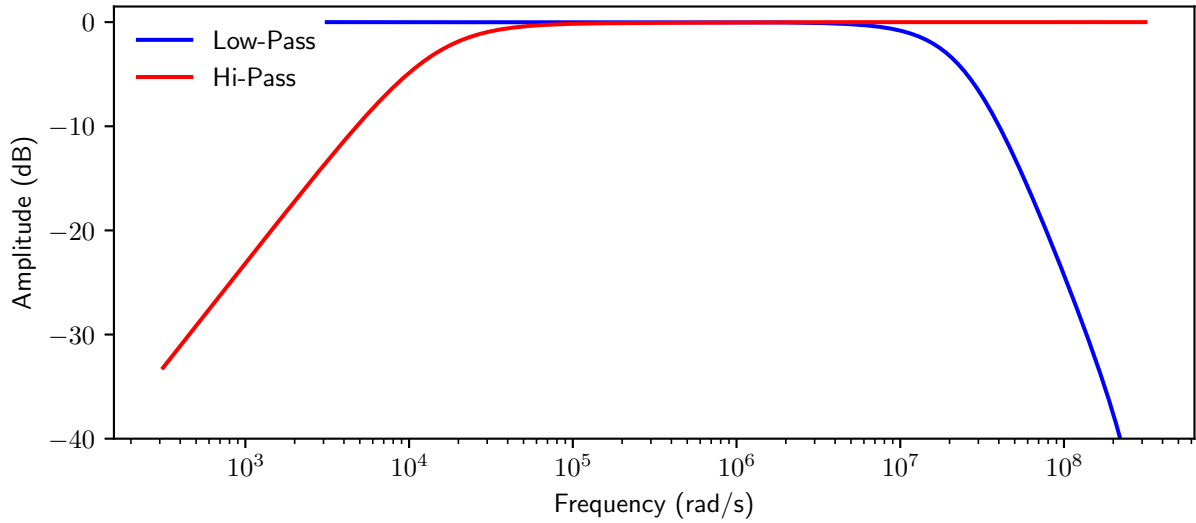
Figure 2.20: Bode diagram of preamp gain using the model described in Section 2.2.3. The total gain (on this logarithmic plot) is the sum of the two curves.

H_{hi}. The shaping introduced by the electronics chain using this model, based on parameters observed in the DEMONSTRATOR, is shown in Figure 2.19. The length of cable in the feedback loop is responsible for the dramatic increase in rise time [33]. A Bode diagram showing the frequency response is shown in Figure 2.20.

The process of digitizing can also change the shape of the recorded waveform. There is some inherent nonlinearity in the relationship between analog amplitude and digitally converted amplitude due to limitations of the analog-to-digital converters on board the digitizer. For the DEMONSTRATOR, this nonlinearity of the GRETINA digitizer has been studied and shown to introduce no more than ≤ 1 ADC unit error for any given sample. Because this is less than the noise amplitude, which is several ADC units, we ignore nonlinearity in our model.

Section 2.3: Conclusion

Germanium detectors are reverse-biased semiconductor diodes which measure the holes and electrons produced by ionizing radiation. Internal electric fields drift the carrier charges

48

to a readout electrode, where they are read by a data acquisition system. A model, based fundamentally on the Shockley-Ramo theorem, has been developed to calculate the signal induced at the electrode during charge collection. This model is parameterized to account for differences between PPC detectors. As an extension specific to the data acquisition system of the MAJORANA DEMONSTRATOR, an empirical representation of an electronics transfer function has been added.

Accurately simulating the waveforms from any specific detector with this model requires the input of over a dozen parameters. Some are well-understood, but others are measured poorly or not at all. Chapter 3 will focus on a technique to optimally estimate these parameters.

CHAPTER 3:   Machine learning of germanium detector parameters

The parameters outlined in the previous chapter sufficiently encompass the physics of waveform generation to enable high precision modeling of PPC detector signals. However, successfully modeling any specific detector requires precise tuning of over a dozen parameters, many of which are either unknown or highly uncertain. In the absence of independent measurements, it is possible to gain understanding of parameters by studying observed waveforms.

The shape of each waveform carries information about each of the detector parameters. In principle, if the model described above is sufficiently accurate, this information can be extracted by fitting the free model parameters to the waveform. This approach is limited, however, due to the presence of noise, the high dimensionality and high degree of correlation in the fit. When such a fit is performed on a single waveform, it tends to "overfit," or select parameters which fit the current waveform well but can't be used to fit additional waveforms.

To address this issue, we have developed an algorithm which pools the information from many waveforms in order to maximize the extracted knowledge about the parameters. Because it can infer information from a large and unspecialized data set, we label this a machine learning algorithm. Specifically, we use the framework of Bayesian modeling [57], which allows us to separately model each waveform's individual parameters while sharing common parameters related to the detector and electronics. The hierarchical model is trained to the data using Markov chain Monte Carlo (MCMC) [58].

This chapter contains a detailed description of the model formulation, the MCMC sampling and the training procedure.

Section 3.1: Bayesian modeling

A Bayesian model attempts to infer the probability of the value of a set of parameters, θ, given a data set x. We can write this as $p(\theta|x)$, which represents the probability density for θ conditional on the observation of x. The foundation of any Bayesian method is Bayes' law [57]:

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)}, \tag{3.1}$$

which expresses $p(\theta|x)$ (known as the posterior distribution) in terms of $p(\theta)$ (the prior distribution), and $p(x|\theta)$ (the likelihood function). The prior is a distribution which represents the knowledge of θ before the measurement. The likelihood represents the probability of measuring x if the parameter set θ is true, and is computed using a model. The denominator, $p(x)$, is a constant normalization factor obtained by integrating over all possible values of θ. Since it is a constant, $p(x)$ is only necessary to compare posteriors between models.

3.1.1: Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a method for numerically estimating a distribution by sampling from it via a random walk [57]. Given a starting point, θ_0, the MCMC algorithm steps randomly through a sequence of values for the parameter $\theta_1, \theta_2, \ldots$ such that the sampling density of the parameter values converges to a "target" distribution. When used for Bayesian analysis, the posterior is chosen as the target distribution.

The algorithm is based on a "Markov chain," meaning that the value selected for the n th sample, θ_n, depends only on the previous step in the sequence, θ_{n-1}. The steps along the Markov chain are governed by a "transition distribution" $T(\theta_n|\theta_{n-1})$, which describes the probability distribution of selecting θ_n given the value of the previous step in the chain. The random walk, or Monte Carlo, aspect of the algorithm is that the value chosen for the given step is probabilistic according to the transition distribution. The crux of designing of an MCMC algorithm is choosing a transition distribution such that it samples from the

target distribution and converges quickly enough to be computationally achievable.

A simple and commonly used sampling method for MCMC is the Metropolis-Hastings algorithm [59]. In Metropolis-Hastings, a "proposal" value θ' is chosen from a proposal distribution $q(\theta'|\theta^{n-1})$. The proposal is then accepted with a probability, α, proportional to the ratio of the target distribution at the current and proposed distribution. Assuming the target distribution is the posterior $p(\theta|x)$, the acceptance probability is

$$\alpha = \min\left[1, \frac{p(\theta'|x)/q(\theta'|\theta^{n-1})}{p(\theta^{n-1}|x)/q(\theta^{n-1}|\theta')}\right] \tag{3.2}$$

If the proposal is accepted, then $\theta^n = \theta'$. Otherwise, the value remains the same, $\theta^n = \theta^{n-1}$.

3.1.2: Diffusive nested sampling

The Metropolis-Hastings algorithm is prone to failure for strongly multimodal target distributions, because it is difficult to define a proposal distribution which can suitably sample from well-separated maxima in the target distribution. Similarly, multidimensional models with highly correlated variables are difficult to properly sample since the proposal must correctly adjust multiple parameters simultaneously to correctly explore the target distribution. These issues were observed when a Metropolis-Hastings sampler was applied to the model of Chapter 2. The estimation of position, especially, appears to be strongly multimodal.

Diffusive nested sampling [60] is an algorithm which extends upon Metropolis-Hastings in a way which can more robustly sample difficult distributions. It is an extension of "classic" nested sampling [61], which constrains sampling to progressively constrained regions of the prior space, called levels. The first level is simply the full prior space. After a set number of iterations, which can be sampled by with Metropolis-Hastings MCMC, a new level is created which encompasses e^{-1} percent of the likelihoods. As the algorithm runs, new levels are successively created, each compressed (in terms of enclosed probability) by a factor of

e^{-1}. Given a likelihood threshold l for a level, and a likelihood function $L(\theta)$, the target distribution for the MCMC algorithm is

$$p(\theta; l) = \frac{p(\theta)\mathbb{1}[L(\theta) > l]}{X(l)} \tag{3.3}$$

where

$$X(l) = \int p(\theta)\mathbb{1}[L(\theta) > l]d\theta \tag{3.4}$$

and $\mathbb{1}$ is a function that is unity when the conditional argument is true and 0 otherwise. $X(l)$ is a measure of the cumulative probability above the threshold value for a given level, and so should for level N should be e^{-N} as a result of the level creation scheme. Figure 3.1 shows the progressive compression of the prior space. By only sampling the progressively higher regions of likelihood space, nested sampling is able to accurately converge around the highest region of the posterior distribution.

Diffusive nested sampling extends the classic nested sampling algorithm by sampling instead from a mixture of levels, weighted toward the highest likelihood levels. This allows the algorithm to "backtrack" and explore different regions of the prior space. This capacity improves the ability to sample highly correlated and multimodal target distributions, also shown in Figure 3.1. The sampling is now over all levels, so the sampling distribution becomes

$$p(\theta) = \sum_j \frac{w_j p(\theta)\mathbb{1}[L(\theta) > l_j]}{X(l_j)} \tag{3.5}$$

where l_j is the likelihood threshold for level j, and w_j is the weight assigned to that level, which are exponentially weighted toward the highest likelihood levels with some "backtracking length" scale λ:

$$w_i \sim \exp(j/\lambda). \tag{3.6}$$

As the diffusive nested sampler runs, MCMC steps from the target distribution (Equation 3.5) are saved. An approximate value of X for each sample based on its likelihood

53

and the likelihood of the adjacent levels. Given the sample likelihood, this can be used to compute the weight of a set of parameters in terms of the posterior distribution.

To improve the ability to map a difficult posterior, it is possible to evolve simultaneously multiple "particles," each of which independently estimates the parameters of the model. Each particle evolves along a Markov Chain independently from a starting position chosen randomly from the prior. This is especially useful for complicated posterior spaces, which might be strongly multimodal. If the likelihood for the parameters of given particle falls too far behind the rest of the particle population, it can be eliminated and replaced with a duplicate of a higher likelihood particle, thereby ensuring no particles become "stuck" in a relatively low likelihood local maximum.

We use a software implementation of diffusive nested sampling named `DNest4` [62]. The individual particles used in diffusive nested sampling can be run in parallel, with their results pooled at regular intervals. In `DNest4`, the parallelization is implemented via shared memory on a single machine. In order to improve the number of particles that can be run simultaneously, we have extended the `DNest4` code to run in parallel on scientific computing clusters with the Message Passing Interface (MPI) standard using the OpenMPI implementation [63]. When parallelized to utilize a core for each particle used in the DNS algorithm, the corresponding reduction in computation time scales with the number of particles, $\mathcal{O}(n_{\mathrm{particles}})$.

In addition, the likelihood function can be parallelized by computing each waveform's `pysiggen` model in parallel. We can maximize efficiency by parallelizing both the diffusive nested sampling algorithm and evaluation of the likelihood function. For evaluation of the likelihood of a single particle, `pysiggen` must be run once for each waveform in the training set. By parallelizing the call to `pysiggen`, we are able to improve computation time an additional factor of $n_{\mathrm{waveforms}}$. By combining this with the parallelization of particles, utilizing $\sim n_{\mathrm{particles}} \cdot n_{\mathrm{waveforms}}$ cores, we are able to improve the computation time by $\mathcal{O}(n_{\mathrm{particles}} \cdot n_{\mathrm{waveforms}})$.
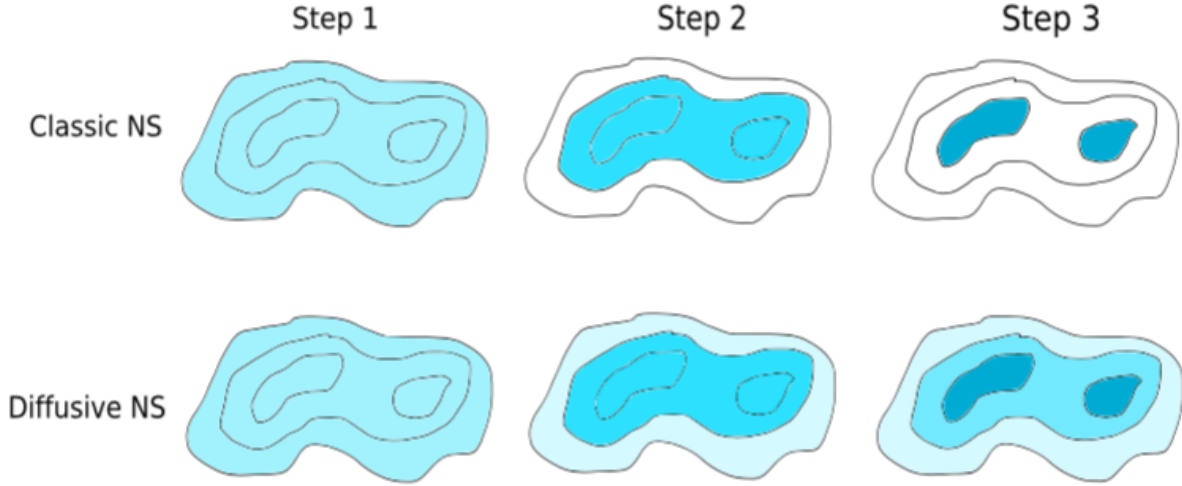
Figure 3.1: Illustration of the nested sampling algorithms. As the algorithm proceeds in step number, new levels are created, shown as progressively darker shades of blue. In diffusive nested sampling, the algorithm has some probability to "backtrack" and sample from lower likelihood regions. Figure from [60].

Section 3.2:   Learning model implementation

The goal of the machine learning model is to use a set of training waveforms to make an accurate prediction of the detector and electronics parameters. The set of waveforms is chosen from calibration data of the highest available energy (2614 keV ^{208}Tl photopeak) to maximize the signal to noise ratio, and drawn with a variety of rise-times to sample events from across the detector volume. The model is constructed such that the set of parameters which describe the detector or electronics, ϕ, are shared among all waveforms. An individual set of parameters θ_i is assigned to each waveform to describe attributes which are unique to that event, such as position, and compared to the observed data for that waveform, x_i. The total unnormalized posterior for a set of n waveforms is then constructed as a sum of each individual posterior:

$$p(\boldsymbol{\theta}, \phi | \mathbf{x}) \propto \sum_{i=0}^{n} p(x_i | \theta_i, \phi) p(\theta_i) p(\phi). \tag{3.7}$$

55

3.2.1: Likelihood function

We wish to compare simulated waveforms from `pysiggen` with observed waveforms, which have been measured as part of the Demonstrator experiment. The digitizer boards records 2038 samples in 10 ns increments, so in principle, the dataset x can consist of as many as 2038 elements. In reality, most of this 20 μs window is either flat "baseline" preceding the signal, or simple RC decay as the preamplifier returns to baseline (see Figure 3.2). We therefore window the signal around the $\sim \mu$s–long rising edge where the signal is of most interest.

Under the assumption that `pysiggen` produces a deterministic and correctly–modeled waveform, the likelihood of measuring a given value x_j for an individual digitized sample is a function of the difference between x_j and the `pysiggen` prediction y_j and the noise present on the waveform. We model the noise on each waveform as Gaussian and white. While there is some frequency structure in the Demonstrator noise, the waveforms under consideration are high enough in signal to noise ratio that any effect should be very minor. The noise present on the baseline, and its distribution, is shown in Figure 3.3. The noise standard deviation, σ is measured from the digitized baseline prior to the waveform and used to find a value. The likelihood for each individual sample j, then, follows a normal distribution. Using the standard notation $N(\mu, \sigma^2)$ for a normal distribution with mean μ and variance σ^2, we have

$$p(x_j|\theta, \phi) \sim N(x_j - y_j, \sigma^2), \tag{3.8}$$

and the total likelihood is the sum of the likelihood for each sample within the data window, such that for waveform i with m samples,

$$p(x_i|\theta_i, \phi) \sim \sum_{j=0}^{m} N(x_j - y_j, \sigma^2). \tag{3.9}$$
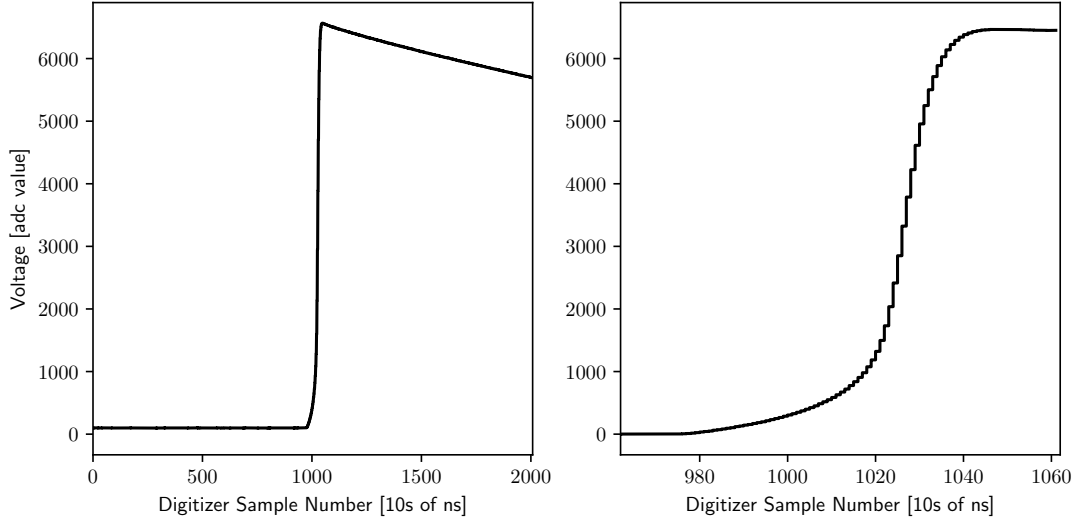
Figure 3.2: A digitized DEMONSTRATOR waveform at 2614 keV, shown in full at left, and windowed around the rising edge at right. Each sample is plotted as a horizontal step to emphasize the digitization. Note that the baseline value is 100 ADC counts above zero. There is noise present on the waveforms, but because it is at the level of ~ 5 ADC counts, it is not visible in these plots.
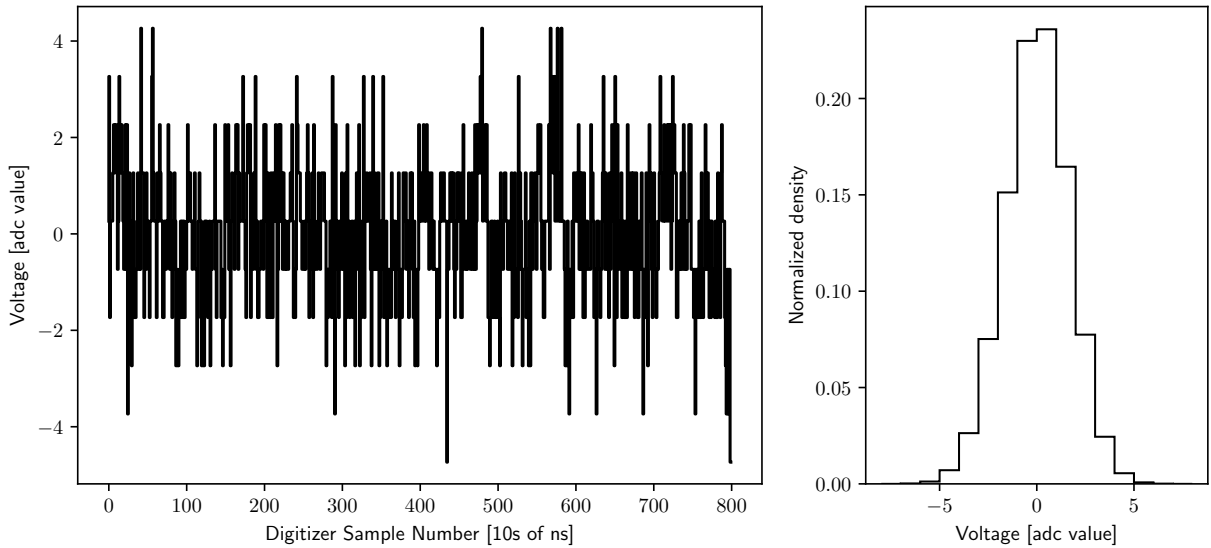
Figure 3.3: Noise characteristics of DEMONSTRATOR waveforms. The plot at left shows the noise present on flat baseline from the first 800 samples of the waveform in Figure 3.2. The mean baseline value has been subtracted such that the mean value is zero. The plot at right is a histogram of the subtracted baseline value for 32 waveforms, normalized such that the sum of all bins is unity.

57

3.2.2:   Waveform parameterization and priors

The `pysiggen` parameters specific to each waveform, as discussed in Section 2.2.2, are position, energy, and charge cloud shape. However, for the purposes of the learning model, we use training waveforms from a monoenergetic photopeak, which allows us to model them with a common energy parameter[1]. This eases a strong correlation in shape that improves convergence of the charge trapping parameter, and will be described in greater detail in Section 3.2.3.

In order to compare the `pysiggen` waveform with digitized data, an additional parameter is required to align the waveform in time with the digitized signal. The parameterization and prior for the time alignment, as well position and charge shape, are discussed below.

Position

Although it would be natural to model the position in cylindrical coordinates, it is important to recognize that the hole drift time is a principal driver of the overall waveform shape. From the isochrones seen in Figure 2.4, it is apparent that there is a strong anticorrelation in drift time between the radial and axial directions. To reduce the impact of this correlation, we perform the fit using polar coordinates, with the origin located at the point contact.

At present, it is assumed that there is no *a priori* knowledge of a waveform's position, so the prior for each of the three positions is uniform over the appropriate range of values. For the radial and polar coordinates, this is anywhere within the detector active volume. The crystal axis dependence on drift velocity, discussed in Section 2.2.2, is periodic with a unique angular span of $\pi/4$ radians. We therefore set the azimuthal coordinate prior as uniform in the $[0, \pi/4]$ quadrant, and recognize that any estimation of this parameter is eight-fold degenerate with respect to the true azimuthal position in the detector.

---

[1]Within a photopeak, there is a spread in energy caused by the statistical variance of the creation of charge carriers, which is a Poisson process. The training waveforms are taken from a narrow energy band around the photopeak maximum to account for this.

Charge cloud

The model uses the generalized gaussian model of Equation 2.5. During fitting, we have seen that waveforms at 2614 keV tend to cluster around $\sigma = 20$ ns and $p = 2$. We therefore set each prior with a wide normal distribution around these values: a standard deviation of 20 ns for σ and 10 for p. The sigma prior is limited by the range (1,40) ns, where the lower bound is chosen to prevent σ of less than one sample. The p parameter range bounds are (1,20), which corresponds to a gaussian distribution, and 20 is a value high enough to be indistinguishable from a uniform distribution.

Time alignment

The DEMONSTRATOR digitizer continuously digitizes the signal on each input channel at 100 MHz, storing the values in a temporary buffer. When the signal exceeds a threshold over the flat baseline value, defined by the output of an on-board trapezoidal filter, it triggers the digitizer to save the event to disk. To avoid triggering on noise, it is necessary to set the trigger threshold several standard deviations above the noise amplitude. A waveform with a short drift time will therefore cross the trigger threshold closer in time to its true start time than a longer drifting waveform, as shown in Figure 3.4. There is some additional uncertainty on the exact trigger point due to noise and the inherent 10 ns width of the digitization sample.

To enable high precision time alignment, it is therefore necessary to add a parameter which aligns the `pysiggen` waveform with the digitized waveform. Aligning by the start point of the waveform would be most convenient. However, because drift time is highly correlated with position, this would tightly correlate the alignment time and position parameters. It advantageous to instead choose an alignment time point at a high fraction of the waveform rising edge, where all waveforms have relatively similar shapes. In addition to reducing correlation with other parameters, it is easier to make a robust prediction of a time point on the rising edge than the start time, especially for long drift-time waveforms which have very
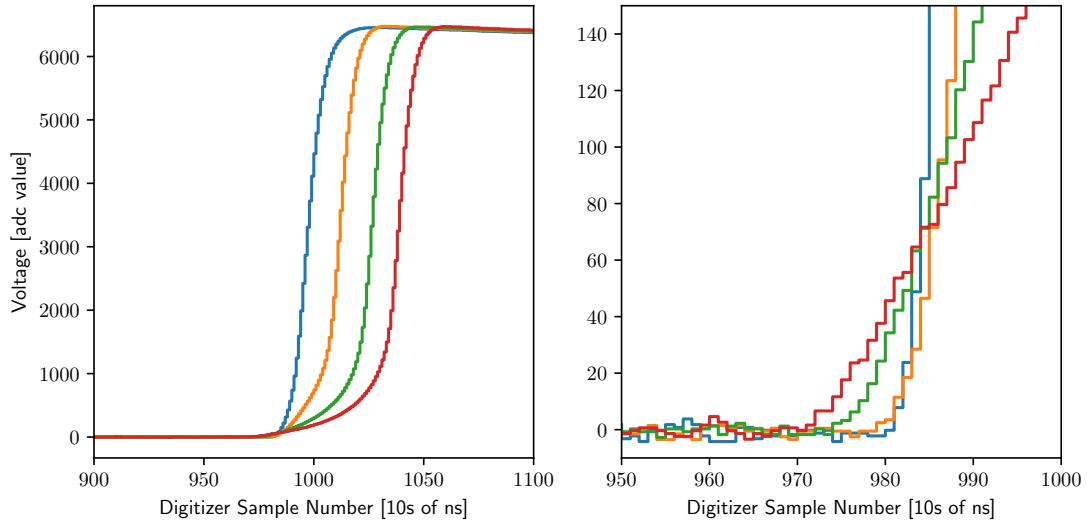
Figure 3.4: Four digitized DEMONSTRATOR waveforms from the 2614 keV Tl^{208} photopeak. The plot at left shows the two microseconds of the signal roughly centered around the rising edge. The same waveforms are shown at right, more tightly windowed around the trigger point. The baseline value is subtracted from each waveform. The point at sample ~ 985 where the waveforms roughly intersect is where the digitizer is triggered. The waveform with the shortest drift time, in blue, has less elapsed time between signal start and the trigger point than the longer drifting waveforms.

little signal amplitude at their start.

Because the `pysiggen` waveform is calculated in discrete time steps, aligning at an arbitrary, continuous-valued time point requires interpolation. The `siggen` calculation produces an output in 1 ns time steps. To compare with digitized data, `pysiggen` downsamples to 10 ns increments and performs a piecewise-linear interpolation between the calculated points.

We have chosen to align based on the 95% rise-time of the waveform amplitude as the alignment point, defined as the point where the waveform reaches 95% of its maximum value. The prior is normally distributed around an estimated value from the parameter, which is calculated simply as the sample number where the waveform exceeds 95% of its maximum value. The prior has a standard deviation of one sample (10 ns).

3.2.3: Detector and electronics parameterization and priors

The parameters which are shared between all waveforms in the training set affect characteristics of either the detector or the electronics shaping.

Electronics transfer function

The electronics shaping, as described in Section 2.2.3, is divided into a high-pass and low-pass contribution:

$$H_{hi}(z) = c\frac{z-1}{z - \exp(-\frac{T}{\tau_1})} + (1-c)\frac{z-1}{z - \exp(-\frac{T}{\tau_2})}, \qquad H_{low}(z) = \frac{az^2 + bz}{z^2 + 2cz + d^2}. \qquad (3.10)$$

The high-pass filter, characterized by three parameters, describes the decay of the signal to baseline after energy is collected. Because these parameters affect the signal shape after the rising edge, they are largely uncorrelated with the parameters which describe the signal formation process. It is therefore possible to perform a simple maximum likelihood fit on the decaying tail of individual waveforms to gain an understanding of these parameters. The fit is performed on the tail only, which we define as starting 20 samples (200 ns) after the

61
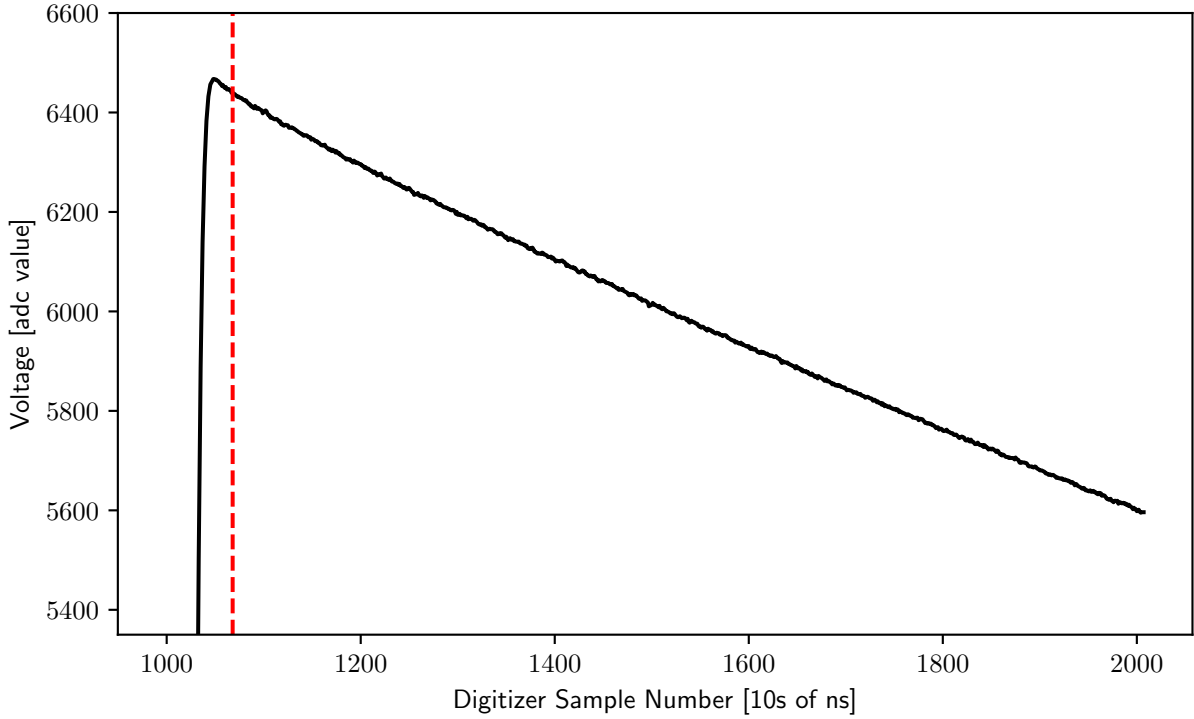
Figure 3.5: Decaying tail of the waveform shown in Figure 3.2. The vertical dashed red line is 20 samples past the waveform maximum. The fit to the tail shown in Figure 3.6 uses the samples after this point.

waveform maximum, and extending to the end of the digitization window (see Figure 3.5). The results of this fit over many waveforms, shown for one detector in Figure 3.6, is used to set a normal prior distribution for each parameter.

The parameters associated with the low-pass filter affect the rising edge and therefore can't be fit separately in the same way as those for the high-pass filter. The transfer function is applied in `pysiggen` as a digital filter in the z domain in the form of Equation 2.7, which has four free parameters. However, since the DC gain of the filter does not effect the shape, and the signal amplitude is fit as an energy parameter, we can reduce this expression to a unity gain filter with three parameters:

$$H_{low}(z) = \frac{1 + 2c + d^2}{1 + b} \cdot \frac{z^2 + bz}{z^2 + 2cz + d^2} \tag{3.11}$$
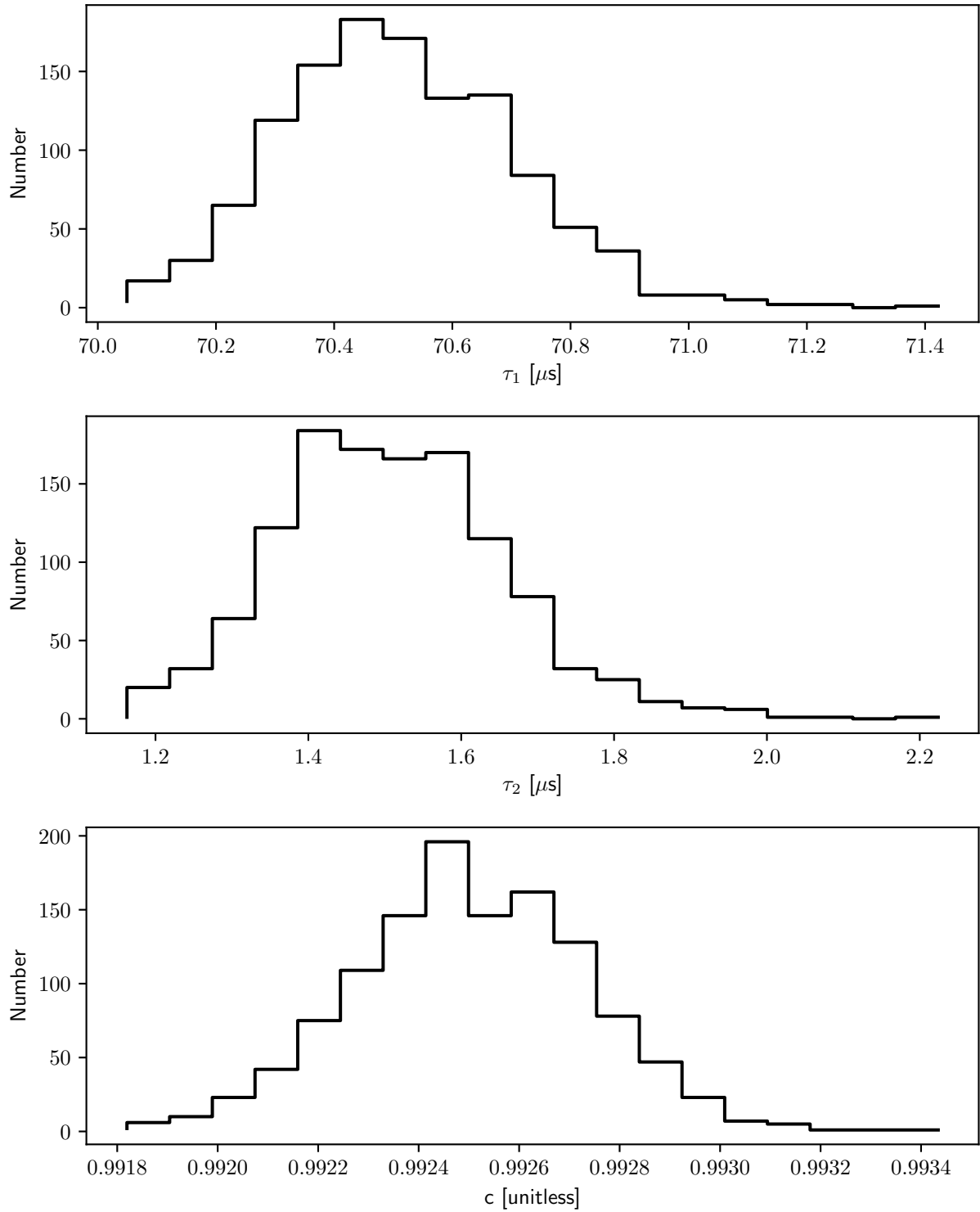
Figure 3.6: Results of a three-parameter maximum likelihood fit to high pass filter shaping (Equation 2.11) with the decaying tail of 1208 waveforms. All waveforms are from the 2614 keV photopeak and are from the same DEMONSTRATOR detector (P42661A). The mean and variance for each parameter are used to set priors for the full machine learning fit.

We have found the MCMC converges fastest when fitting the time-domain representation of the oscillation parameters, $\omega = \cos^{-1}(c/d)$ and $\phi = \tan^{-1}\left(\frac{c-b}{\sqrt{d^2-c^2}}\right)$. A benefit of this parameterization is that, since the convolution kernel goes as $\cos(\omega t + \phi)$, each is naturally cyclical with 2π and can therefore be bounded and given a uniform prior.

There is freedom in how to parameterize the third free parameter. We have found that the shape is quite sensitive to the combination $g = \frac{1}{1+2c+d^2}$, which is the contribution to the gain from the denominator. The numerator parameters can be scaled without changing the gain, but because of the z^2 term in the denominator, c and d cannot be scaled without changing the shape. The combination g then represents sort of an intrinsic gain when $a + b = 1$. The choice is somewhat arbitrary, but appears to converge fairly well. There is no natural bound to this parameter, so a very wide gaussian distribution is used around a reasonable value.

The effect of each of the three parameters is shown in Figure 3.7.

Drift velocity

Recalling Section 2.2.1, we must fit the three parameters of Equation 2.3 for both the $\langle 100 \rangle$ and $\langle 111 \rangle$ crystal axes. Again, the equation, together with its physical limits, is:

$$v(E) = \frac{\mu_0 E}{(1 + (E/E_0)^\beta)^{1/\beta}}, \qquad \lim_{E \to 0} v(E) = \mu_0 E, \qquad \lim_{E \to \infty} v(E) = \mu_0 E_0 \qquad (3.12)$$

We therefore have two independent handles on velocity at very low and very high field, and the parameter β controls how "sharply" they connect. In the middle field region, the three parameters combine nonlinearly in a way that would be difficult to fit without the input of the high and low field behavior. Figure 3.8 shows that, for holes in germanium, the linear low field behavior falls off at $\sim$10V, and the saturation velocity is reached $\sim$> 10,000V.

Figures 3.9 and 3.10 show the electric field inside a DEMONSTRATOR enriched germanium detector. The span of electric field values are well within the range of moderate field where
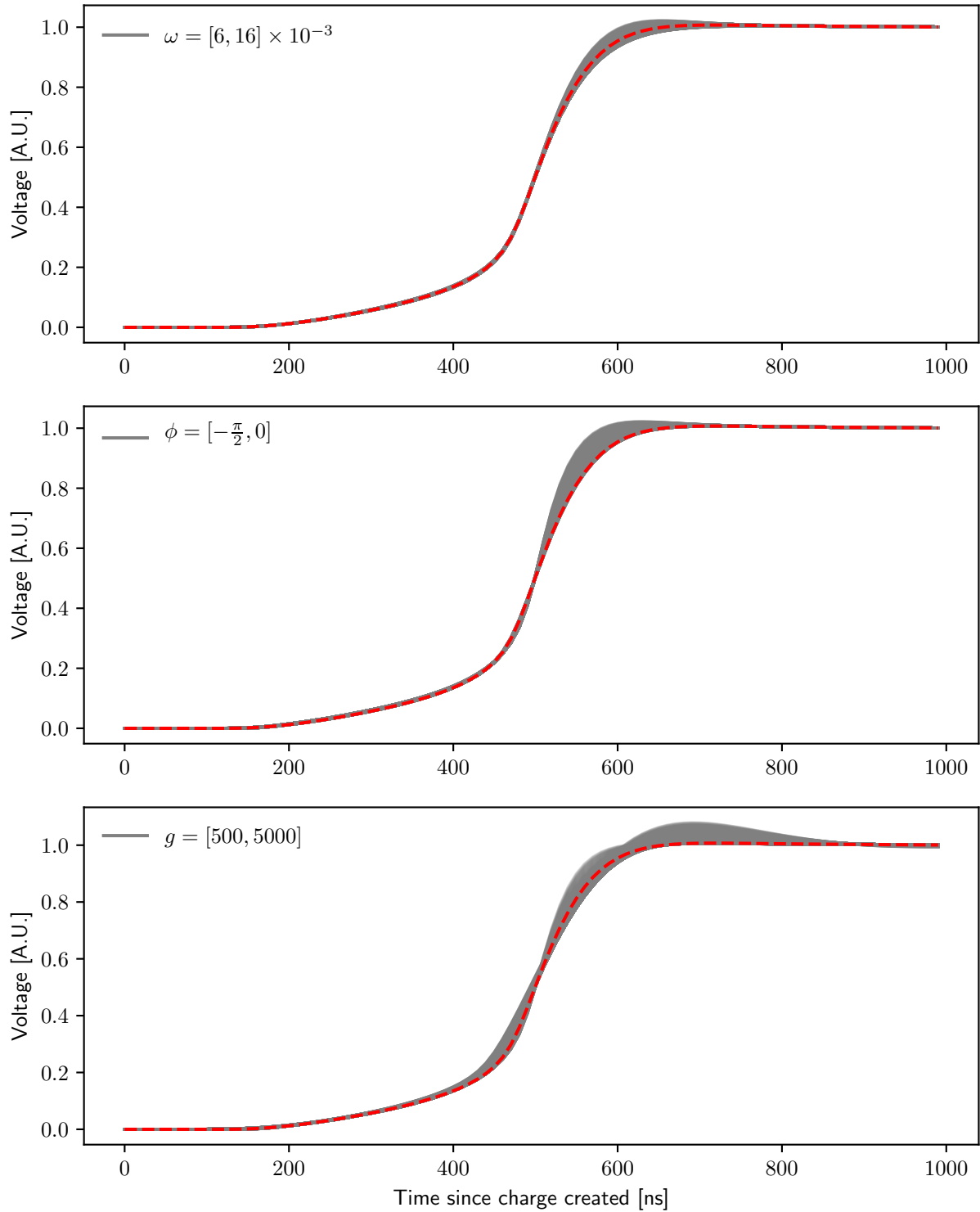
Figure 3.7: Effect of the three low-pass electronics transfer function parameters on the waveform shape. From top to bottom, the ω, ϕ and g parameters are plotted holding the other two constant. Each parameter has a grey band, showing the effect of varying the value over the range shown in the label. The dashed red line, which is the same in each plot, shows the curve for a waveform with values trained by DEMONSTRATOR data. The range for each parameter is chosen such that the waveform stays near in shape to the red curve.
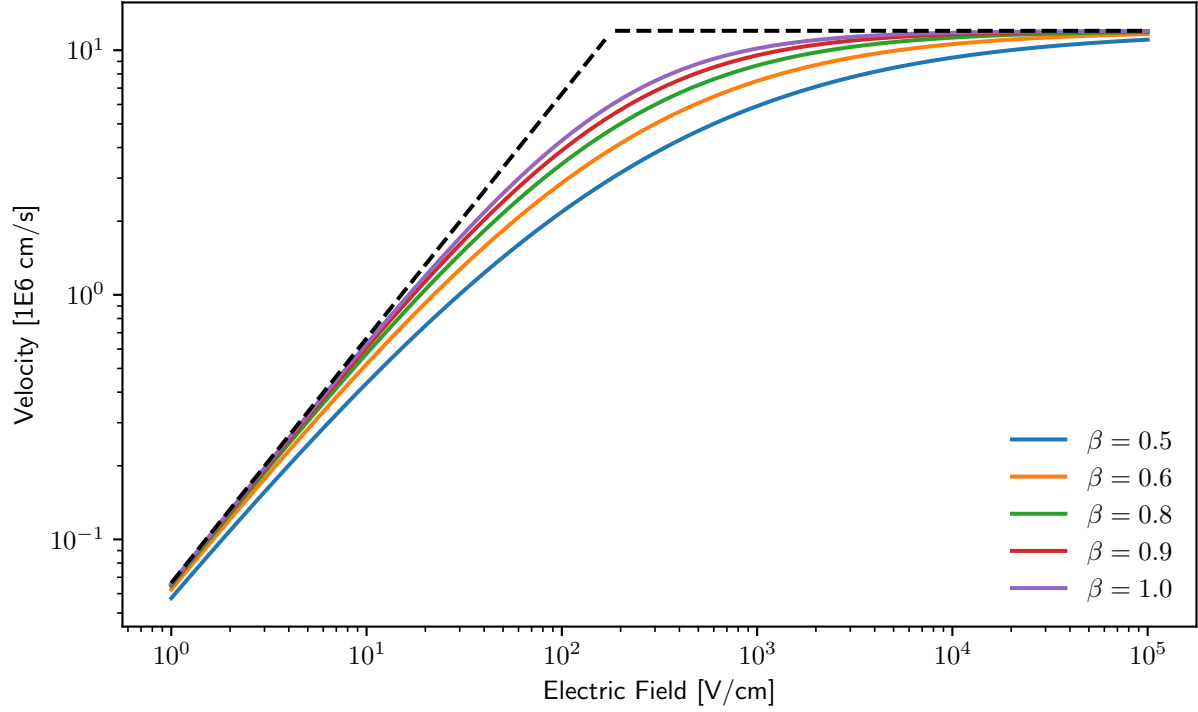
Figure 3.8: Drift velocity curves using the Equation 2.3 parameterization, varying the value of β. The black dashed lines show the linear $v \sim \mu_0 E$ behavior at low field and constant $v \sim \mu_0 E_0$ velocity at high field. The values $E_0 = 181$ and $\mu_0 = 66333$ are from $\langle 100 \rangle$ direction observed by [46].
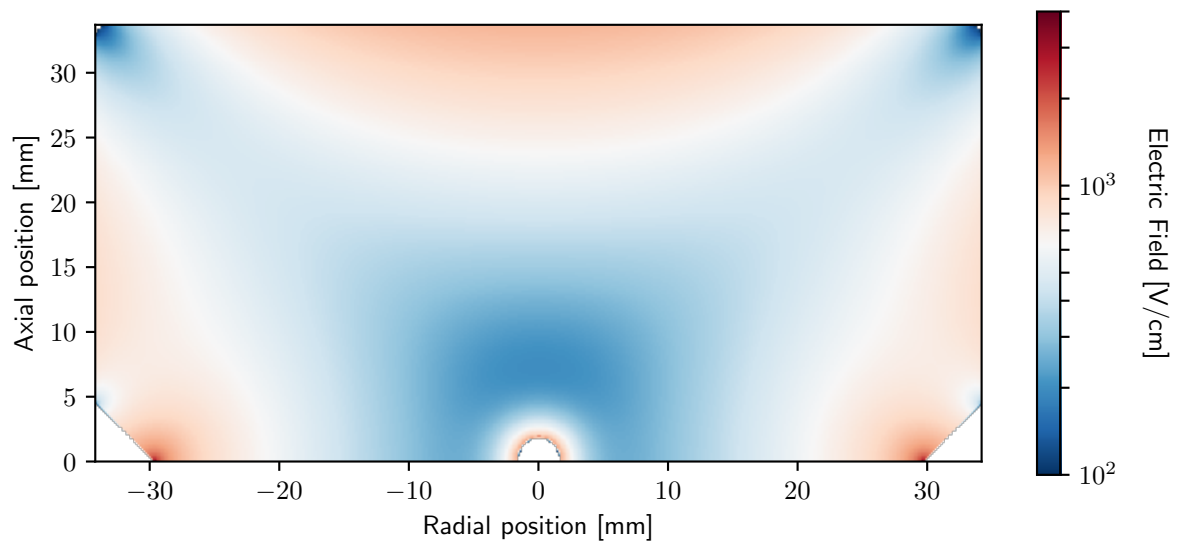
Figure 3.9: Electric field in a representative DEMONSTRATOR enriched detector.
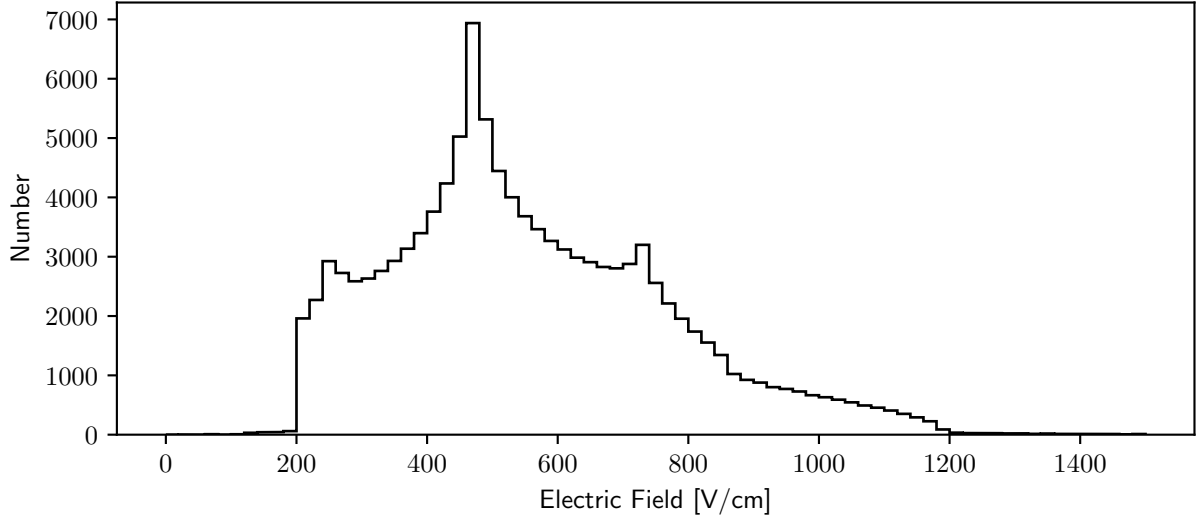
Figure 3.10: A one dimensional histogram of the electric field values in Figure 3.9. Fields in the detector mainly fall between 200-1200 V/cm.

the velocity parameters are more correlated and therefore more difficult to fit. We therefore wish to reparameterize the equation in terms of parameters which are more directly influenced by the electric fields present in our detectors. Specifically, it would be helpful to express the curve in terms of velocities at the bottom and top end of the observed range of fields.

If we begin with the velocity at two fields E_A and E_B,

$$v_A = \frac{\mu_0 E_A}{(1 + (E_A/E_0)^\beta)^{1/\beta}}, \qquad v_B = \frac{\mu_0 E_B}{(1 + (E_B/E_0)^\beta)^{1/\beta}}, \tag{3.13}$$

we can express μ_0 and E_0 in terms of v_A, E_A, v_B and E_B,

$$E_0 = \left(\frac{\psi^\beta E_B^\beta - E_A^\beta}{1 - \psi^\beta} \right)^{1/\beta}, \quad \text{where} \quad \psi = \frac{E_A v_B}{E_B v_A} \tag{3.14}$$

and

$$\mu_0 = \frac{v_A}{E_A} \left(1 + \left(\frac{E_A}{E_0} \right)^\beta \right)^{1/\beta}. \tag{3.15}$$

The three parameters v_A, v_B and β are used to perform the fit. Figure 3.11 shows the reparameterization using $E_A = 250$V and $E_B = 1000$V. The velocity is well described in the

67
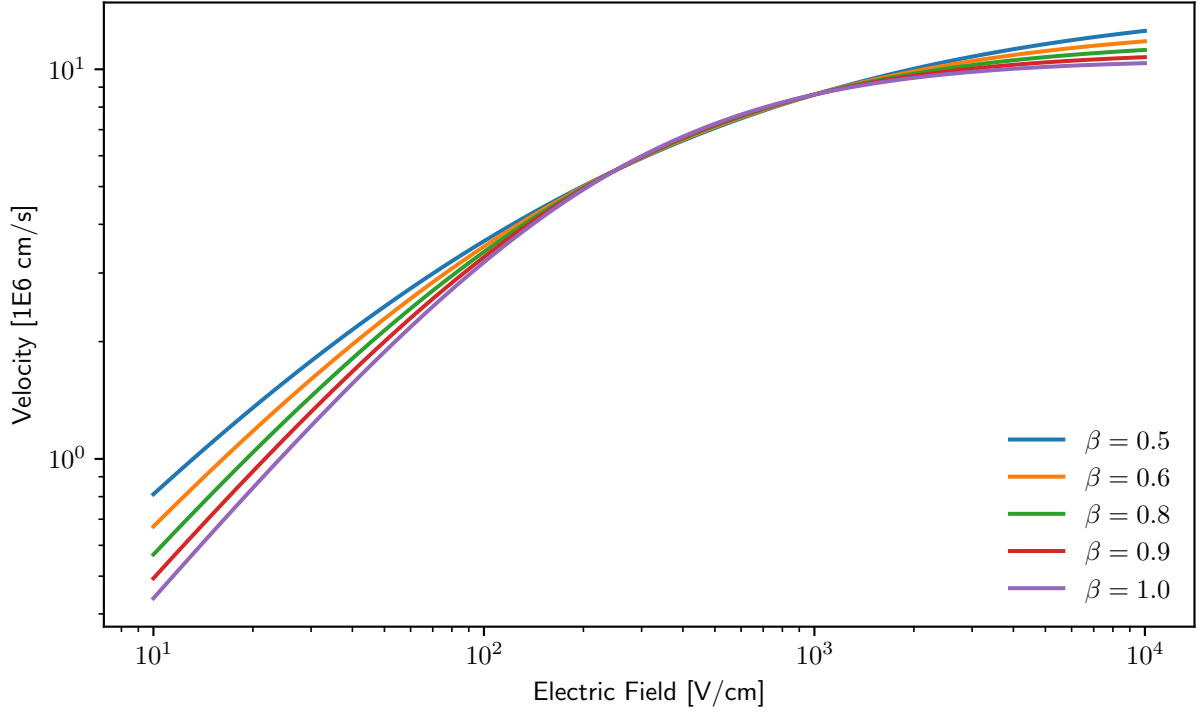
Figure 3.11: Velocity curves using the parameterization of Equation 3.14. The velocity reference points used in this plot are $E_A = 250V$ and $E_B = 1000V$, using the same $\langle 100 \rangle$ direction velocity parameters from [46] as Figure 3.8.

range of fields observed in the detector by the parameters v_A and v_B, but the β parameter no longer has much leverage over the curve shape in this region. We have seen significant improvement in overall convergence using (v_A, v_B, β) model, so the this is implemented in the machine learning fit.

The priors for v_A and v_B are normally distributed, with means set using the values given by [46] and 20% standard deviation. Measured values of β can vary greatly, and our fit value of β does not tend to converge reliably in any case. For this reason, we conservatively set a uniform prior on β from 0 to 1.

Point contact dimensions

Because the weighting potential in a PPC detector is very strongly weighted near the point contact, the generated signal depends very sensitively on the dimensions of the point contact.

68

The point contact dimple is modeled as a hemispheroidal indentation. Precision optical measurements of its diameter and depth are available for some, but not all, DEMONSTRATOR detectors. A separate measurement was made at ORTEC during manufacturing but with uncertain accuracy. For detectors measured at both ORTEC and with the DEMONSTRATOR optical device, values can vary by as much as a few tenths of a millimeter.

For this reason, we allow the radius and depth of the point contact dimple to float in the model. For detectors with optical device measurements, the measured values used to set a gaussian prior with 0.1mm standard deviation. Detectors without the optical measurement are given a flat prior in a range of 0.5mm around the ORTEC provided value.

Because point contact parameters affect the electric field and weighting potential, changing them requires calculating fields with `fieldgen`, which is an expensive process. To improve the speed of the machine learning process, we precompute the fields for a grid of point contact values. The field for arbitrary point contact dimensions is then calculated via bilinear interpolation between the precalculated values. To facilitate interpolation, the range of allowed point contact values must be bound between the minimum and maximum values precomputed for each parameter. We use a 0.5mm window calculated in 0.1mm increments around the measured value.

Impurity profile

As discussed in Section 2.2.1, there is considerable uncertainty in the measurement of impurity. Since the impurity is measured at each end of the detector, the impurity gradient is even less certain.

Like the point contact dimensions above, the impurity values affect the field values and therefore must be precomputed. The range for precalculation is chosen to encompass a range of reasonable values based roughly on 50% range around the measured values, but taking into account the range over which the detector fully depletes. The prior is simply set as flat over this range.

Charge trapping, charge releasing and energy

Charge trapping, as described in Section 2.2.1 , reduces the amount of charge in the carrier cloud as a function of time. When trapping effects are nonnegligable, the amplitude of waveforms with the same interaction energy will depend on the drift time of the charge cloud, with longer drifting waveforms more strongly attenuated. This effect is clearly visible in DEMONSTRATOR data, as shown in Figure 3.12.

To first order, the difference in waveform shape due to charge trapping is simply the reduction in amplitude. When each waveform in a training set is given an individual energy parameter, the fit therefore has a strong local maximum to simply reduce the amplitude with the energy parameter and choose a very long mean charge trapping time constant. To avoid this, we select a training set with all waveforms from the same energy peak, and share a common energy parameter. This forces the fit to use charge trapping, and not an individual energy value, to match the observed amplitude.

The energy prior is normally distributed around the maximum the training set waveform with the shortest observed drift time, with a 10 ADC standard deviation.

There is little prior knowledge of the charge trapping constant. The reduction in amplitude in Figure 3.12 can be reproduced with charge trapping on the order of a few hundred μs. However, since charge can subsequently be released from the traps, the observed reduction is actually a combination of both the trapping and untrapping rates.

To be cautious, we use a uniform prior over a wide range of possible charge trapping and untrapping values. The trapping constant is allowed to vary between 10μs and 1000μs, while the untrapping constant varies from 1μs to 1000μs. We have observed good convergence in the trapping parameters with this approach.

Summary of free parameters

In total, there are 19 parameters shared between all waveforms, and 6 parameters which describe each individual waveform. Table 3.1 summarizes the parameters.
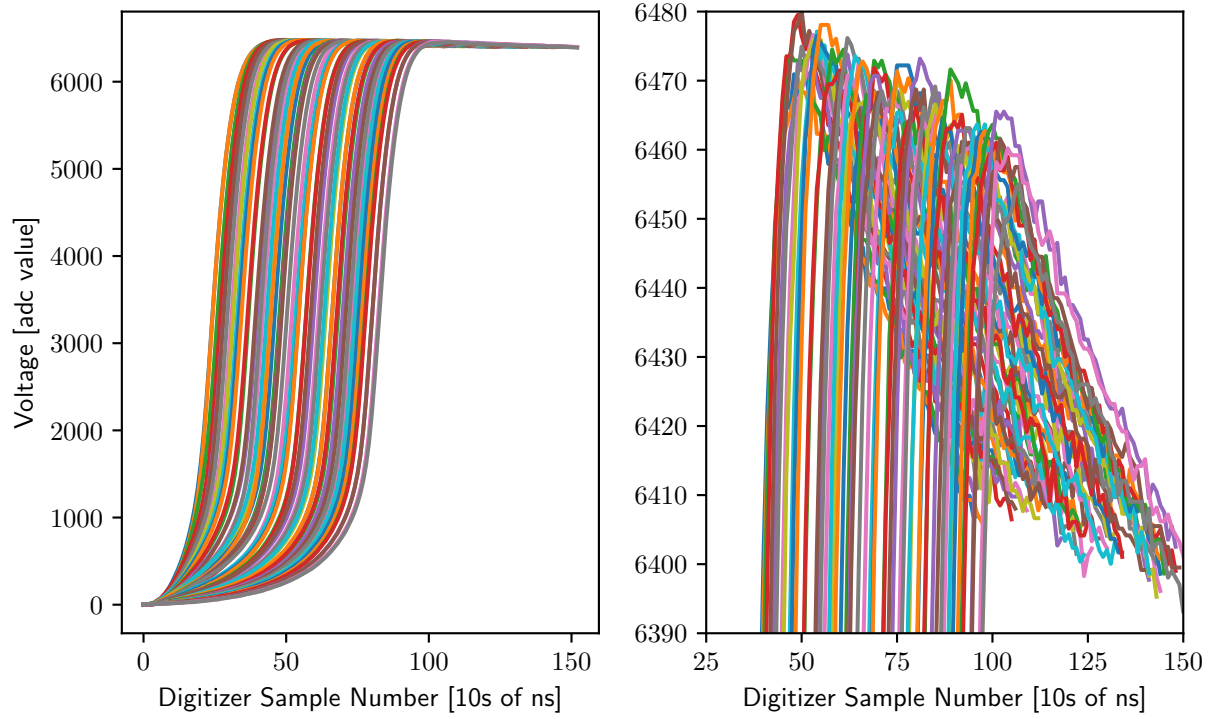
Figure 3.12: Sixty waveforms from the 2614 keV photopeak for a DEMONSTRATOR detector. Each color is a separate waveform. The plot at right shows the full waveforms, which are time aligned by their estimated start time. The plot at right is windowed to emphasize the amplitude of the waveform. Longer drift time waveforms have lower amplitudes, which is caused by charge trapping.

Type:	Detector	Waveform	Electronics
	Point contact length	Interaction position (x3)	Hi-pass (x3)
	Point contact depth	Alignment time	Low-pass (x3)
	Average impurity concentration	Charge cloud (x2)	
	Impurity gradient		
	Charge trapping constant		
	Charge release constant		
	Training set energy		
	$\langle 100 \rangle : \beta, \mu_0, E_0$		
	$\langle 111 \rangle : \beta, \mu_0, E_0$		
Total:	13	6	6

Table 3.1: Summary of the free parameters in the model.

71

3.2.4:   Sampling strategy

The proposal distribution for each parameter is set according to the prior. Following the recommendation of [62], we use a "heavy tailed" proposal distribution rather than tune proposal step sizes to improve sampling. The distribution is a student-t distribution multiplied by the width of the prior distribution. For normally distributed parameters, this is the distribution standard deviation, and for uniformly distributed parameters, it is the width of the allowed region.

For any given step in the chain, we update one or more parameters from either the global parameter set (ϕ) or an individual waveform's parameter set (θ_i). By updating more than one parameter simultaneously, we can better sample correlated parameters. However, parameters from different waveform sets should not be correlated, so it is desirable to update only one set of parameters at a time. To satisfy both of these conditions, we block update either detector or waveform parameters, each with 50% probability. If updating waveforms, we update one or more waveforms simultaneously, with the number drawn randomly from an exponential t-tailed distribution. This step chooses which parameter sets to update.

Once the parameter sets are chosen, we must update the parameters within the set. Following the example of [62] for heavily sampling from correlated distributions, we draw a random number n from $100^{N(0,1)}$. Then, for n iterations, we randomly update one of the parameters in the set.

Section 3.3:   Choosing a training set

The machine learning algorithm trains the parameters of the model described above by fitting a set of training waveforms. In order for the model to be accurate, it is necessary that the training waveforms arise from representative, single site interactions within the detector bulk. Additionally, it is advantageous that the waveforms have as high a signal to noise ratio as possible.

Figure 3.13: The DEMONSTRATOR calibration system. A line source is inserted via the clear plastic tubing track through the shield, at left. The track source wraps helically around the copper cryostat.

Training waveforms are selected from DEMONSTRATOR calibration data. The calibration is performed by inserting a ^{228}Th line source into the shield, which wraps around the module cryostat (see Figure 3.13) [64].

The spectrum from a ^{228}Th calibration is shown in Figure 3.14. The highest energy feature in the spectrum is the 2614 keV photopeak from the decay of ^{208}Tl, a daughter nuclide in the ^{228}Th chain. To maximize the signal to noise ratio of our training waveforms, we choose waveforms from this peak. However, a large number of interactions in this peak originate either from the incoming gamma either Compton scattering or pair producing in the detector. In both cases, for the full energy of the interaction to be recorded, it is necessary that the event be multisite in nature. We can apply a cut based on the current pulse amplitude to energy ratio (A/E) to eliminate the majority of the multisite events.

As described in Section 2.1.4, a single site waveform has only one current peak, while a multisite waveform has multiple. Since the energy of a current waveform scales with its integral, at a given energy a multi-peaked current waveform will be wider than a single-peaked waveform, and therefore have a lower maximum value. By removing events with
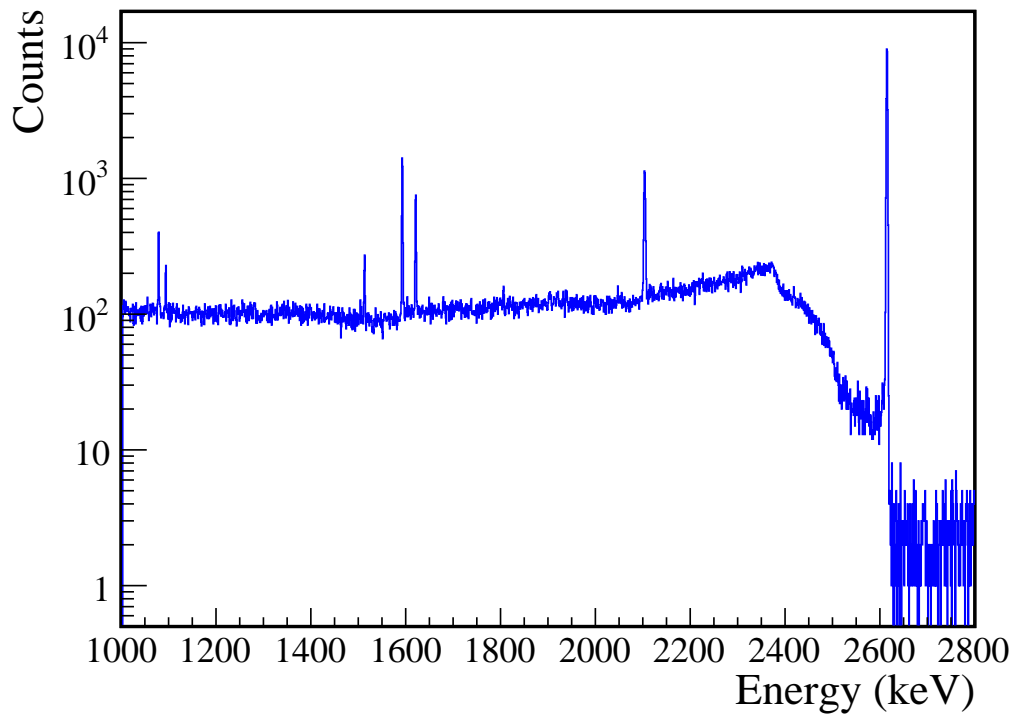
Figure 3.14: Calibration spectrum using a ^{228}Th source in the SMALL CAPS DEMONSTRATOR. The highest energy peak in the spectrum is the 2614 keV peak, a photopeak from the decay of ^{208}Tl, a ^{228}Th daughter nuclide. The next two highest intensity features are the ^{208}Tl double and single escape peaks at 1592 keV and 2103 keV, respectively. The remaining peaks are from various ^{228}Th chain decays.
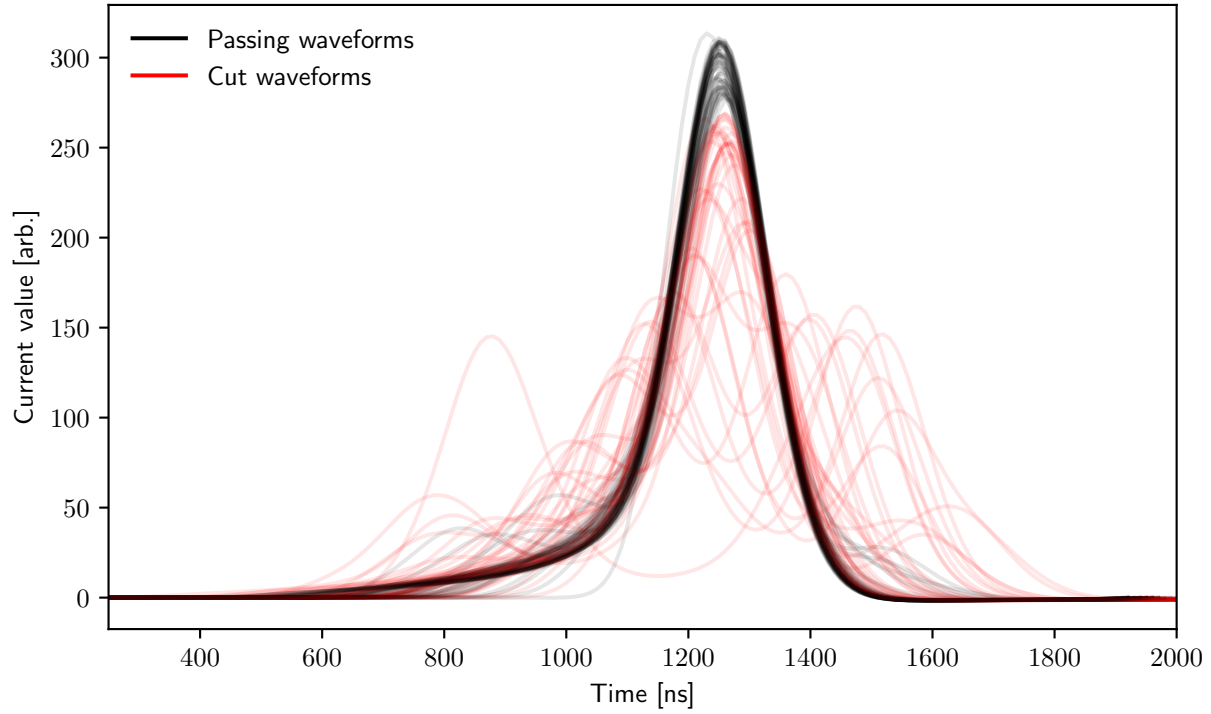
Figure 3.15: Current waveforms for 100 events from the 2614 keV ^{228}Tl photopeak. Each charge waveform is simultaneously differentiated and smoothed via convolution with a window which is the first derivative of a gaussian, with a 50 ns standard deviation. Waveforms in red fail the A/E cut due to the low amplitude of the current pulse. Many of the failing events have multiple current peaks, clearly marking them as multisite events.
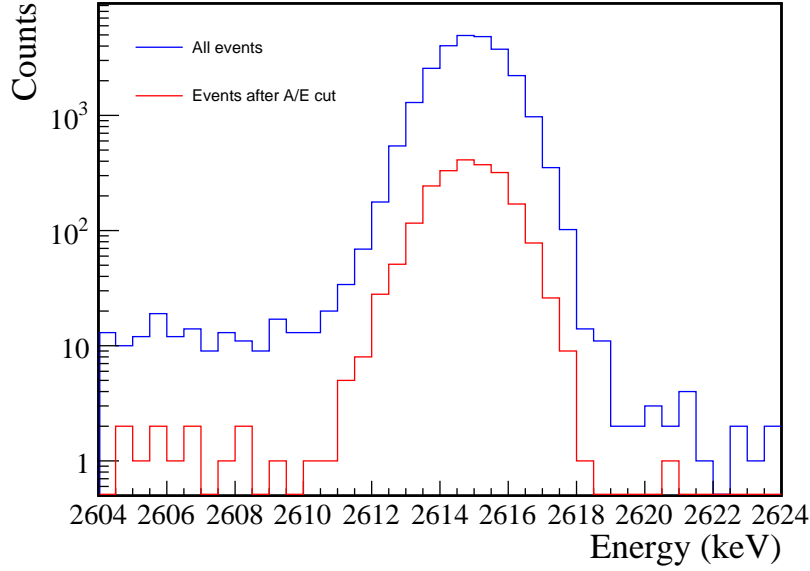
Figure 3.16: Effect of an A/E cut on the amplitude of the 2614 keV ^{228}Tl photopeak. The cut reduces the peak for this calibration run from 26,108 to 2,185 events, or by 92%.

low current peak amplitude, shown in Figure 3.15, the A/E discriminates against multisite events. The value for the A/E cut is chosen to accept 90% of the ^{208}Tl double escape peak, which is dominated by single site events. Figure 3.16 shows reduction in the amplitude to 8% of the 2614 keV peak after the A/E cut. Examining the waveforms remaining after the cut in Figure 3.17, the variance in shape before ~800 ns is due to varying drift times. However, the outlier waveforms near the maximum at ~1000 ns show there remain clearly multisite events in the set. A multisite event could create this shape if a relatively high energy charge cloud arrives at the point contact before a second lower-energy cloud. Evidently, the A/E cut is inefficient at removing this type of multisite event.

Figure 3.18 shows the current waveforms for the events which survive the A/E cut. In current space, it is easier to understand the failure mode of the A/E cut: waveforms which have a high energy cloud with a short drift time and a low energy cloud with a longer drift time create current peaks which are nearly mirrors of single site events with long drift times.
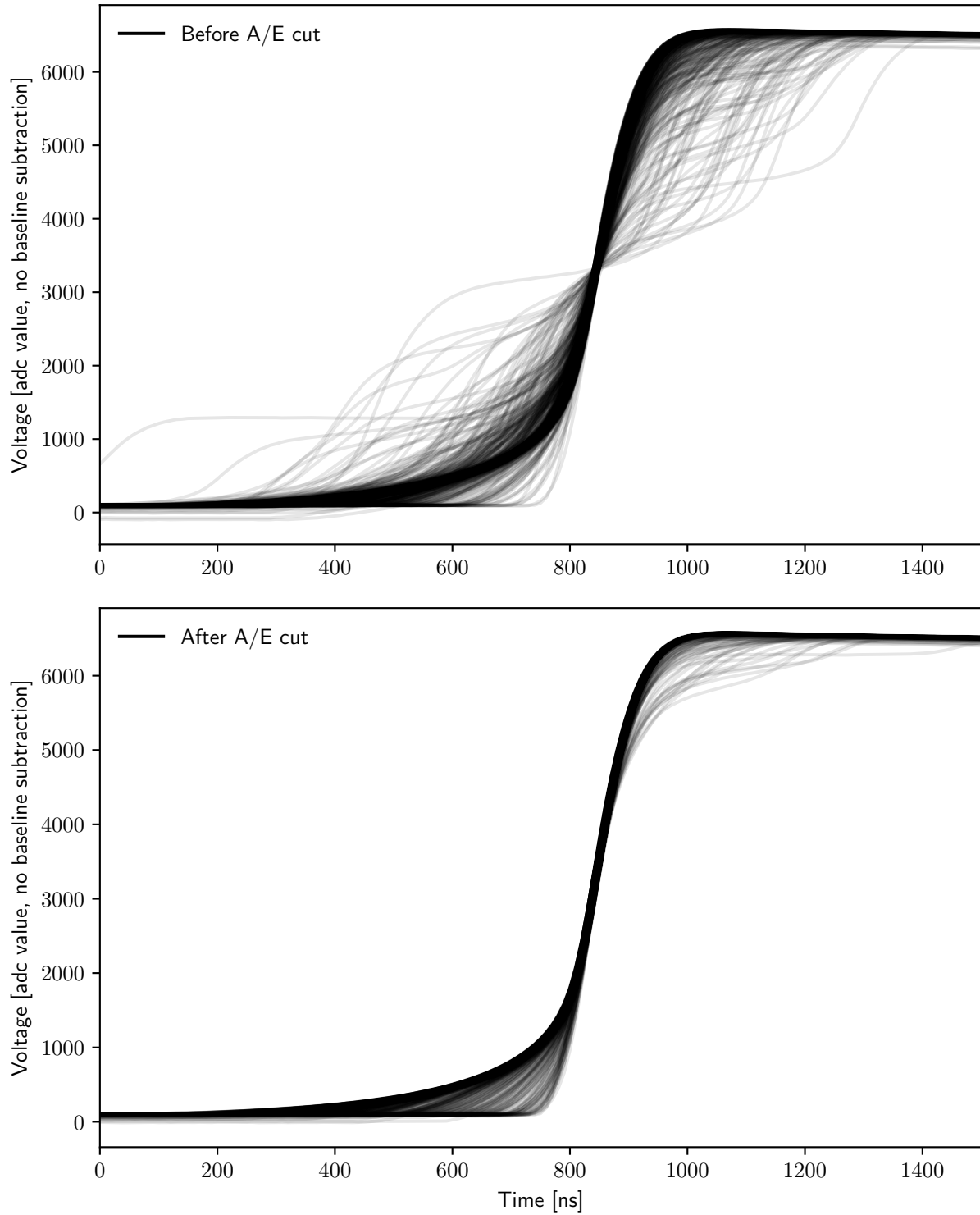
Figure 3.17: On top, 500 waveforms from the 2614 keV photopeak before an A/E cut is applied. At bottom, 500 waveforms that survive the A/E cut. Each waveform drawn is translucent, such that darker bands are more densely populated regions. Waveforms are aligned at their 50% risetime time point.
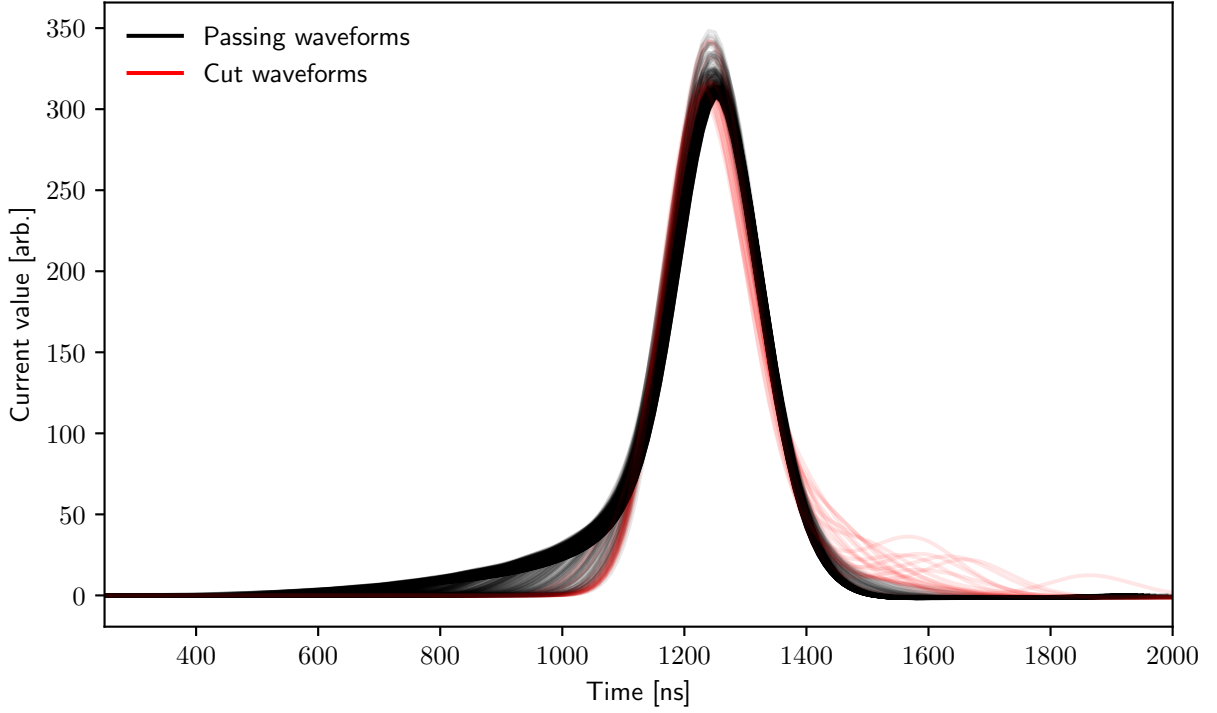
Figure 3.18: Current waveforms for events which survive the A/E cut. Each charge waveform is simultaneously differentiated and smoothed via convolution with a window which is the first derivative of a gaussian, with a 50 ns standard deviation. Waveforms in red are multisite, and tagged based on the extended tail at on the falling edge of the current peak.

To eliminate these events, we introduce a cut based on the fall-time from the current peak to the baseline. Since our goal is only a representative training set, the efficiency of this cut has not been studied in detail.

An additional cut is made based on the mean baseline value. For a given detector, the baseline value for all events should be the same. However, the electronics take several tens of μs to settle back to baseline after an event is measured. If a second event occurs before the baseline is fully recovered, the observed baseline in the first few microseconds of the digitization window will deviate from the equilibrium window. The shape of the waveform will be distorted as the electronics continue to settle, so these waveforms must be rejected

from the training set[2]. Figure 3.19 shows the waveforms this cut removes.

Finally, we wish to ensure that all waveforms in the training set originate from interactions in the "bulk" of the detector. If the interaction occurs in the small volume around the point contact where the initial weighting potential of the interaction is relatively high, the electron contribution to the waveform will be significant. Several approximations made in the model described in Section 2.2 assume that the hole contribution dominates, so these waveforms should be excluded from the training set. Since events closer to the point contact have shorter drift times, these events can be simply removed by a drift time cut. Figure 3.20 shows the distribution of rise times once all other cuts have been applied. Conservatively, we exclude events with drift times less than 500 ns.

Once all cuts have been applied, we must choose a subset of the remaining waveforms to use in the training set. As a proxy for sampling events originating from positions throughout the detector, we aim to select waveforms uniformly distributed in drift time. To this end, we group the remaining waveforms by drift time into a bins, choosing a number of bins corresponding to the size of the desired training set. One waveform is then chosen randomly from each bin to create a training set.

While this procedure is largely able to discriminate against multisite waveforms, it is still possible for "slightly" multisite waveforms to pass through the cut. We have been able to identify such events based on their failure to converge during the machine learning fit. In these cases, the event is manually excluded and replaced with a waveform of similar drift time. Without a pure source of single site events, having to iterate on training set to exclude multisite events is probably unavoidable.

Similarly, because the energy parameter is shared between all waveforms in the set, it is imperative that all waveforms truly have the same energy. However, there is some natural statistical fluctuation in the number of electron-hole pairs produced by an interaction. The

---

[2]This situation is much more common during calibration, when the count rate is quite high, than during $0\nu\beta\beta$ data collection.
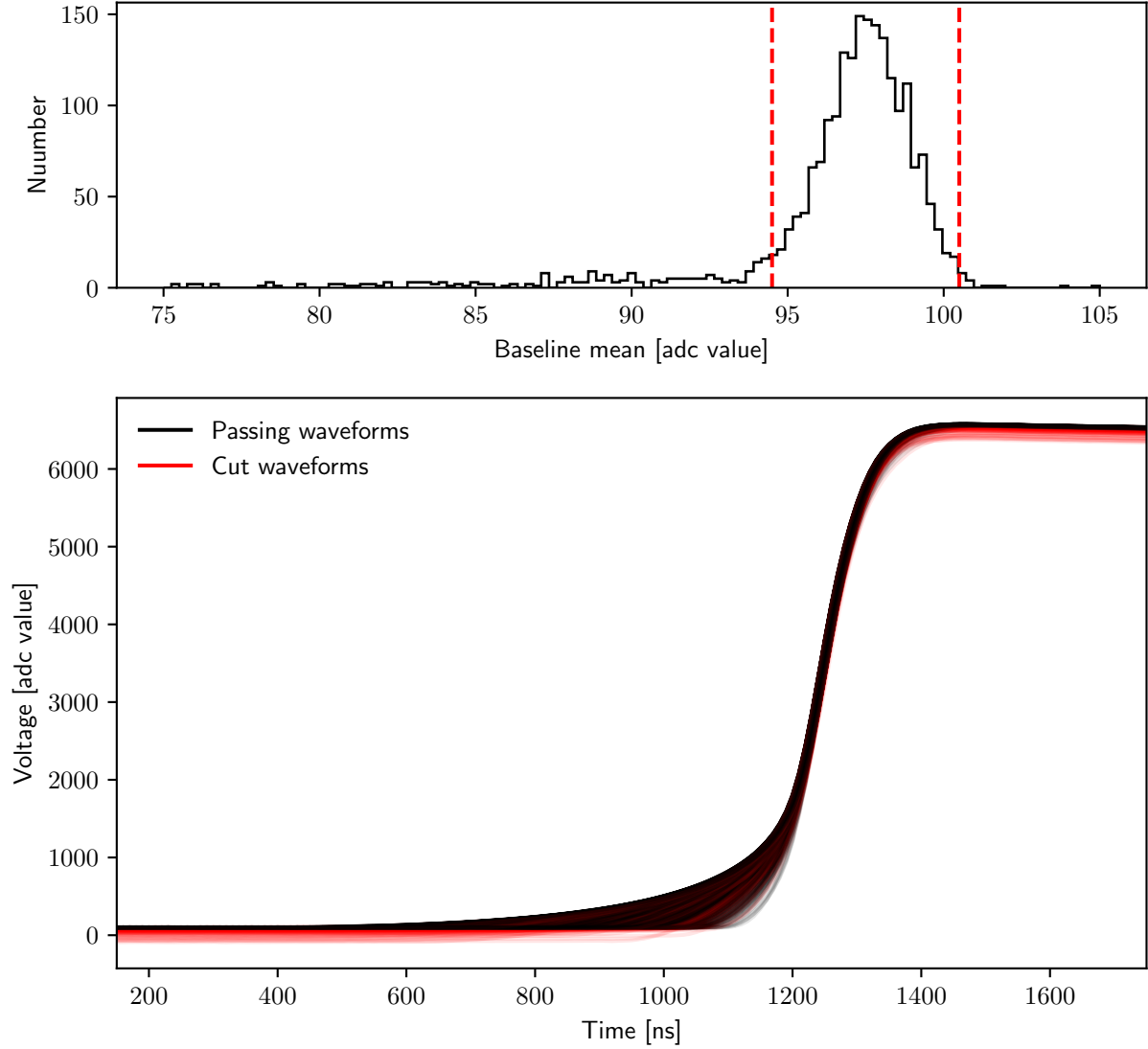
Figure 3.19: Cut based on baseline value. The top plot shows a histogram of the mean baseline value for the first 800 samples of 2614 keV waveforms remaining after the multisite event cuts. Only waveforms with baselines between the red dashed lines, drawn 3 ADC counts from the mode, are accepted. The bottom plot shows the waveforms accepted and rejected by this cut, in black and red respectively. Of 1,842 waveforms, 1,583 (85%) of waveforms pass this cut.
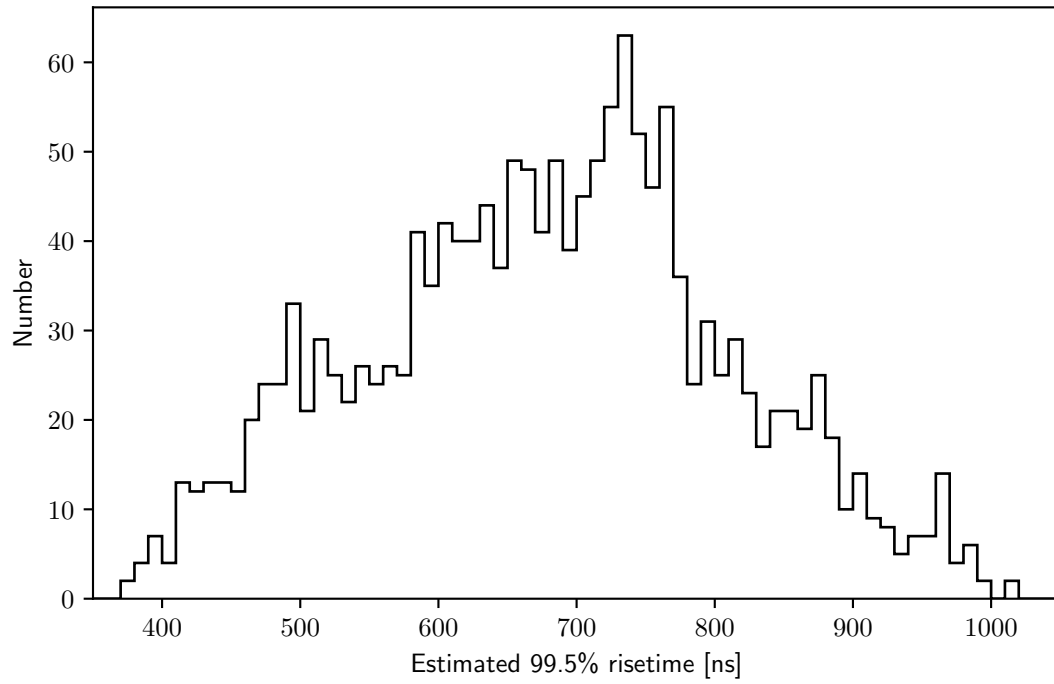
Figure 3.20: Distribution of estimated drift time for waveforms which pass the multisite discrimination and baseline cuts.

fit is sensitive to this fluctuation, and higher or lower than average energy waveforms will not have the correct amplitude. Again, these waveforms are excluded from the training set manually.

Each waveform in the set is windowed around the rising edge. Because the model before the waveform start is only a flat baseline, adding window before the start of the waveform adds no information. It is only necessary to ensure the window begins early enough to capture the beginning of each pulse. The tail contains information about the transfer function decay parameters, and it is useful to keep enough tail in the window to ensure those parameters converge. However, increasing the length in turn increases the required computation time. We have found that a 250 sample window, centered around the 99.5% risetime, works acceptably.

It is possible that the procedure of cuts described above accidentally eliminates some single site events which should be retained in the training set. Given the current accuracy of the model, we believe the set of training waveforms originate a sufficiently diverse set of locations to provide the algorithm with information from isochrones across the detector. However, should the accuracy improve in the future, more careful consideration should be made to ensure that the training waveforms do not discriminate against any specific regions of origin.

Section 3.4:   Conclusions

By constructing a Bayesian model that combines multiple waveforms with shared detector parameters, it is possible to learn about the detector and electronics model parameters by solving for the posterior distribution. Because the model is both high dimensional and strongly correlated, MCMC with diffusive nested sampling is a good choice of algorithm for estimating the posterior. Still, we are able to improve the convergence properties of the algorithm by choosing a parameterization of the model which attempts to minimize correlations, and choosing a varied set of single site training waveforms.

CHAPTER 4: Results & Applications

Section 4.1: Machine learning results

The learning algorithm described above was implemented by integrating `pysiggen` into the framework of `DNest4`. We then used diffusive nested sampling to estimate the parameters for an enriched germanium detector from the DEMONSTRATOR, which is identified by the serial number P42661A. This detector is the third from the top on a string opposite the crossarm opening, which is given the DEMONSTRATOR designation P5D3. The training set selected for detector P42661A is shown in Figure 4.1, chosen at sixteen waveforms to keep computation time to convergence reasonable. The event identification number corresponding to each waveform is listed in Table 4.1.

The diffusive nested sampling algorithm was run with 20 particles, which was the maximum number which could be run in parallel on the UNC killdevil computing cluster. A new level was created each time 10,000 samples exceeded the likelihood of the current threshold, with each level compressed by e^{-1} relative to the previous. Every 1,000 samples, one particle at random was saved to disk as a way to store the MCMC target distribution. Each particle was run for 680,000 iterations, when the Markov Chain was empirically found to be converged.

Sampling the posterior with high enough density to fully map the distribution would require roughly an order of magnitude more iterations, which was not computationally feasible. Instead, we can simply take the highest likelihood sample in the MCMC as the best fit.

The best fit to the training waveforms is shown in Figure 4.2. The residual shows that the model is accurate nearly to the level of the noise amplitude. There is, however, some
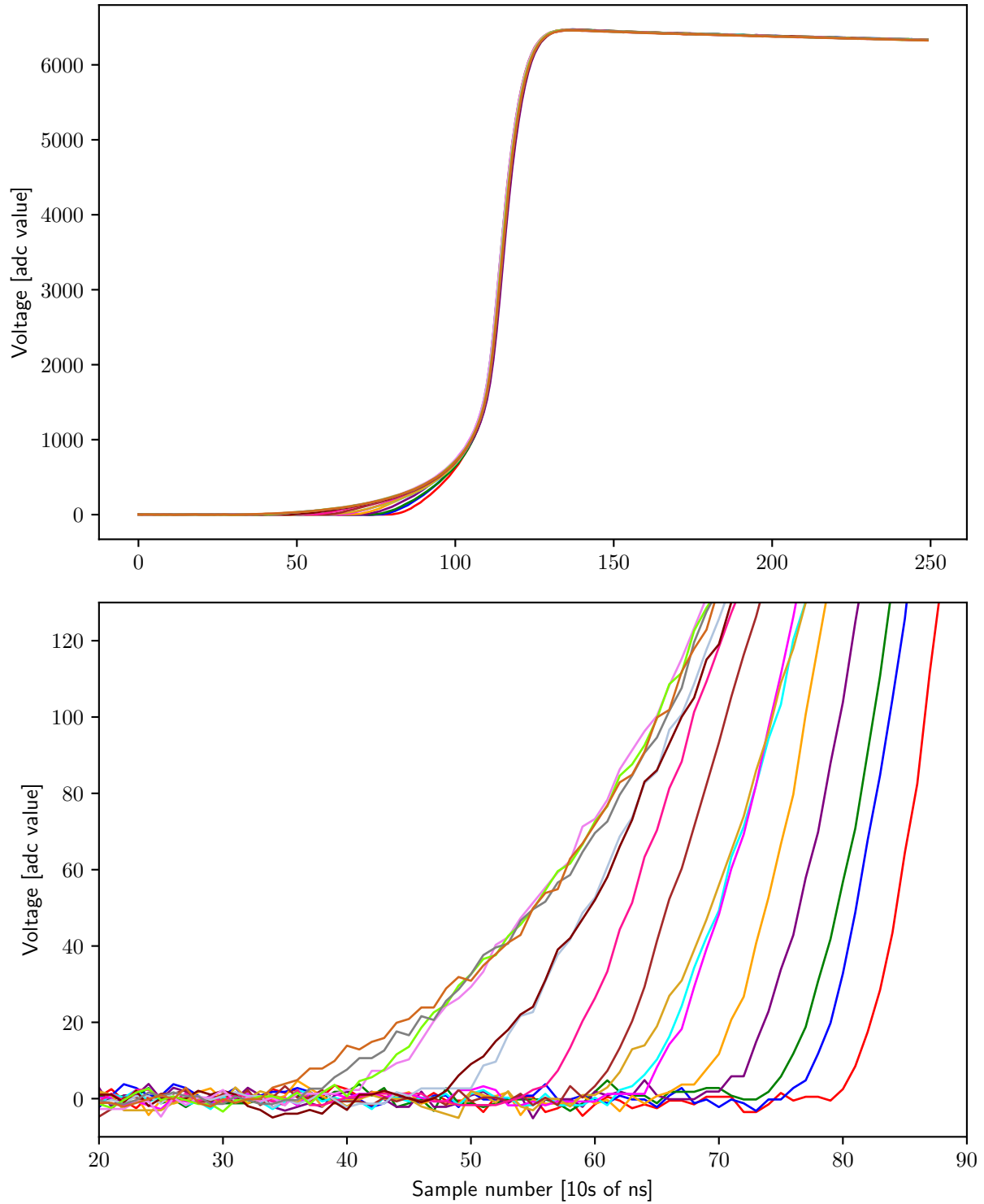
Figure 4.1: Training set used for machine learning fit for detector P42661A. There are sixteen waveforms in the set. The top plot shows the full 250-sample window used in the fit, centered around the estimated 95% rise time. The bottom plot shows the differences in the shape at the start of each waveform.

Number	Color	Rise Time [ns]	Run number	Entry number
1	red	528	11515	128965
2	blue	565	11511	96633
3	green	584	11510	317207
4	purple	621	11511	48362
5	orange	646	11510	56040
6	cyan	682	11511	202862
7	magenta	684	11515	62152
8	goldenrod	705	11512	71698
9	brown	724	11511	95195
10	deeppink	770	11514	208461
11	lightsteelblue	824	11511	13079
12	maroon	839	11516	92744
13	violet	904	11515	225892
14	lawngreen	915	11513	1225
15	grey	939	11513	6869
16	chocolate	969	11542	8401

Table 4.1: Training set of sixteen waveforms for detector P42661A. The color is used for all plots of these waveforms within this work. Rise time is estimated from the start time to the 99.5% rise time. Run number and entry number are unique identifiers from the DEMONSTRATOR dataset. Each waveform is taken from the high gain amplification channel.

structure evident in the residual. The structure indicates either that the model does not fully capture the physics of the signal, or that the fit was not able to properly converge on the global maximum.

To help determine the quality of convergence, we performed a set on a second set of sixteen training waveforms. The residual structure for the second set is very similar to the first training set, as seen in Figure 4.3.

4.1.1: Fit parameters

The parameters for fits from both training sets are compiled in Table 4.2.

The drift velocity curves over the range of fields relevant to the detector for training set one are plotted in Figure 4.4, compared to the measured value from [46] which formed the basis of the prior. The magnitude of the anisotropy between the $\langle 100 \rangle$ and $\langle 111 \rangle$ axes is
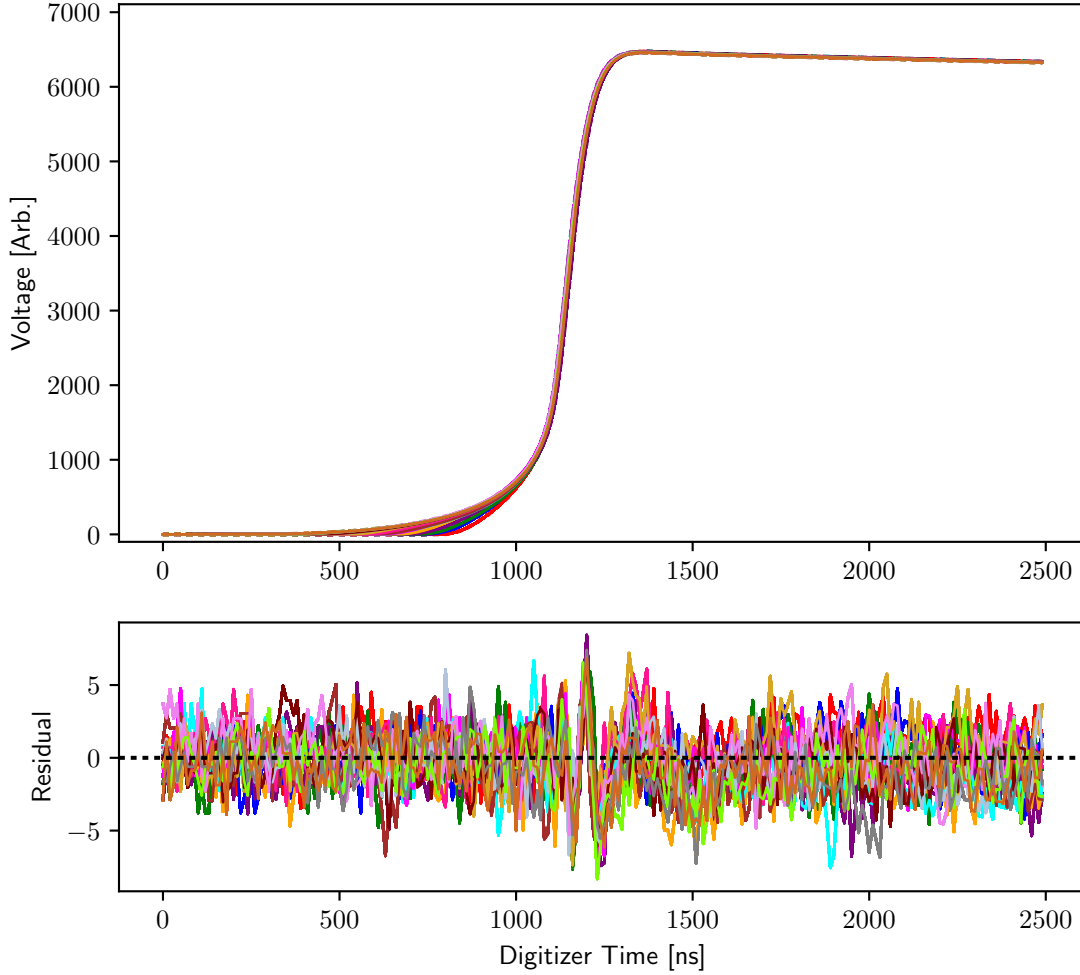
Figure 4.2: Best fit of the training waveforms to the model. At top, all waveforms are shown, with the best fit overlaid. The residual, on bottom, shows that the fit is accurate nearly to the level of noise. However, some structure remains, which is more clearly visible when averaged, as in Figure 4.3.

roughly the same as expected from previous measurements, but the overall drift velocity is lower by ∼10%. The curves are similar for training set two. The disagreement between the measured and fit values of the β parameters are not surprising, since the reparameterization in Section 3.2.3 leaves β with little influence over the velocity over the fields of interest.

The impurity and point contact dimensions estimated for the detector are shown in
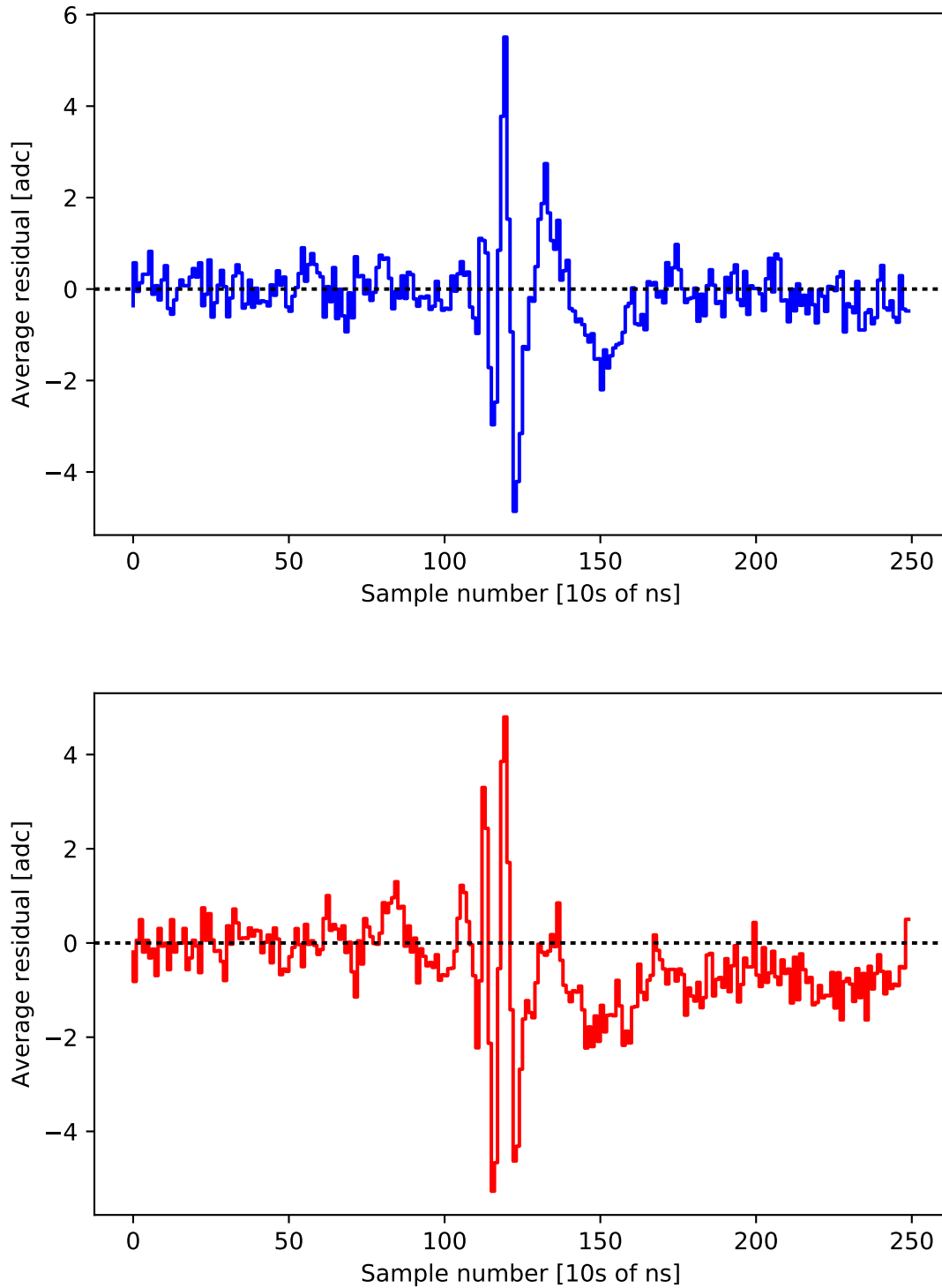
Figure 4.3: The residual between the training waveform and fit, averaged over the sixteen waveforms in the training set. The top and bottom plots use two different sets of sixteen waveforms as training samples. If the model were perfectly accurate, the average residual would be flat at zero, shown by the dashed black line. The structure after sample 100 suggests the model, though accurate at approximately the part per thousand level, could be further improved.
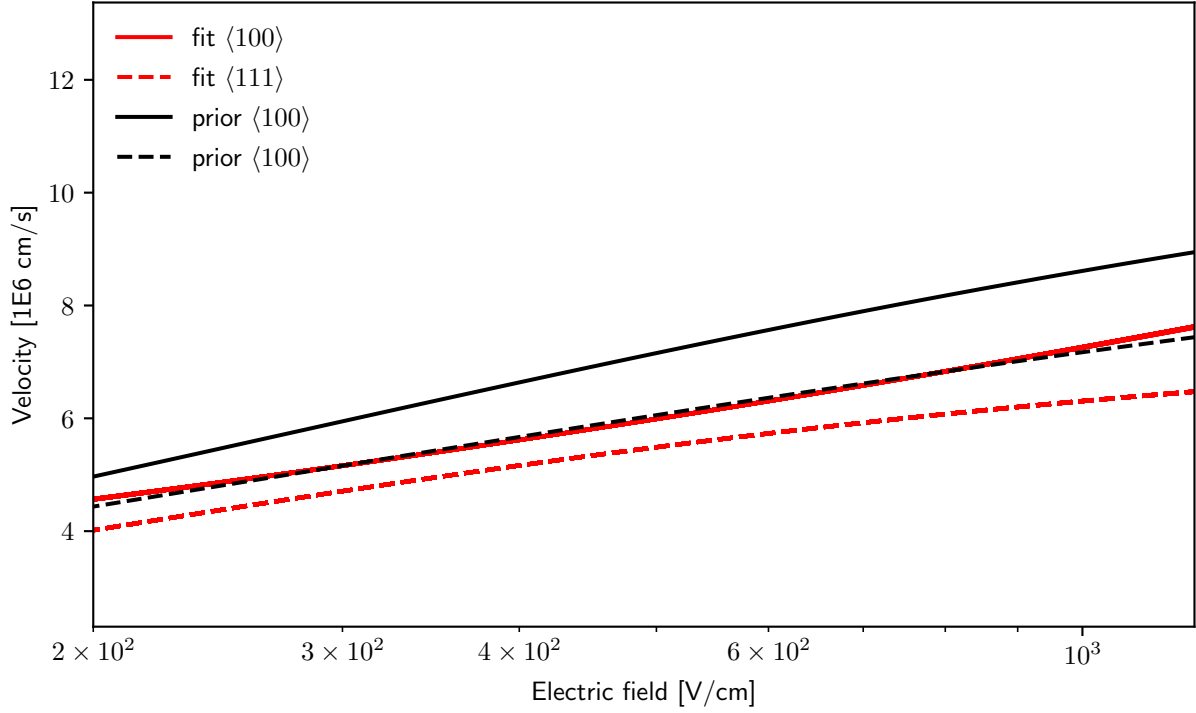
Figure 4.4: Drift velocity curves for the $\langle 100 \rangle$ and $\langle 111 \rangle$ axes. The values in red are from the fit to detector P42661A. The values in black are from the measurement by [46].

Table 4.2. Optical measurements of the point contact for this detector were not available, so we compare to the values measured at ORTEC. The measured depletion voltage for this detector is 1200 V. With the parameters taken from the fit, the calculated depletion voltage is 1400 ± 25 V. The difference is likely due to an incorrectly estimated impurity profile.

4.1.2: Fit to second detector

To ensure the process is applicable across different crystals, the same training process was applied a second DEMONSTRATOR detector. Detector P42574A hails from a different crystal boule than P42661A, and was pulled at an earlier date. It is located as the third detector from the top in a string located crosswise to the crossarm, designated P6D3. The training waveforms are shown in Figure 4.5, with more details in Table 4.3. The residual structure for the fit to sixteen P42574A waveforms is shown in Figure 4.6. The model is

Velocity parameters			
Parameter	Fit Value (Set 1)	Fit Value (Set 2)	Measured Value
$\langle 100 \rangle : v_{250}$	4.89×10^6 cm/s	5.05×10^6 cm/s	5.50×10^6 cm/s
$\langle 100 \rangle : v_{1000}$	7.26×10^6 cm/s	7.16×10^6 cm/s	8.61×10^6 cm/s
$\langle 100 \rangle : \beta$	0.1	0.41	0.744
$\langle 111 \rangle : v_{250}$	4.40×10^6 cm/s	4.41×10^6 cm/s	4.83×10^6 cm/s
$\langle 111 \rangle : v_{1000}$	6.31×10^6 cm/s	6.48×10^6 cm/s	7.17×10^6 cm/s
$\langle 111 \rangle : \beta$	0.9	0.47	0.580

Electronics parameters		
Parameter	Fit Value (Set 1)	Fit Value (Set 2)
ϕ	-1.44 mm	-1.52 mm
ω	0.01124 mm	0.01123
gain	1554.7	1521.9
τ_1	73.4 μs	71.0μs
τ_2	1.4 μs	1.1μs
c	0.90	0.99

Detector parameters			
Parameter	Fit Value (Set 1)	Fit Value (Set 2)	Measured Value
PC Radius	1.54 mm	1.6 mm	1.5 mm
PC Length	1.21 mm	1.64 mm	1.7 mm
Avg. Impurity Conc.	0.55 $\times 10^{10}$ cm^{-3}	0.44 $\times 10^{10}$ cm^{-3}	0.59 $\times 10^{10}$ cm^{-3}
Impurity Gradient	0.014 $\times 10^{10}$ cm^{-4}	0.071 $\times 10^{10}$ cm^{-4}	0.016 $\times 10^{10}$ cm^{-4}
τ_{trap}	294.3 μs	267.3 μs	
τ_{release}	5.1 μs	2.8 μs	
Waveform Amplitude	6432.4 ADC	6444.2 ADC	

Table 4.2: Fit values for the hole drift velocities for both training sets for P42661A. Measured values for velocity are from [46] and for detector parameters are from manufacturer measurements. There are no reported uncertainties associated with any of the measured values.

once again able to fit to residuals at the ~6 ADC level, and the structure in the residual looks very similar to that for detector P42661A.

Number	Color	Rise Time [ns]	Run number	Entry number
1	red	424	11514	113363
2	blue	448	11514	99694
3	green	483	11510	307803
4	purple	513	11511	134776
5	orange	578	11510	16756
6	cyan	604	11514	96625
7	magenta	657	11512	29496
8	goldenrod	675	11511	33826
9	brown	709	11510	308551
10	deeppink	758	11510	257709
11	lightsteelblue	778	11512	112523
12	maroon	838	11511	7198
13	violet	878	11517	28223
14	lawngreen	891	11510	200387
15	grey	946	11516	169409
16	chocolate	955	11588	90466

Table 4.3: Training set of sixteen waveforms for detector P2574A. The color corresponds to the waveforms in Figure 4.5. Rise time is estimated from the start time to the 99.5% rise time. Run number and entry number are unique identifiers from the DEMONSTRATOR dataset. Each waveform is taken from the high gain amplification channel.

Section 4.2:   Validation

To validate the parameters selected by the machine learning fit, we can apply the model to estimate the position of a large population of single site events. Events from calibration data should be distributed roughly evenly throughout the detector volume. If the model cannot accurately reconstruct waveforms from any given volume of the detector, it should be apparent by examining the reconstructed position distribution. We can also evaluate the ability of the model to discern between populations of single and multisite events, comparing its accuracy to the A/E algorithm.

It would be computationally intractable to run the full model described in Chapter 3 for

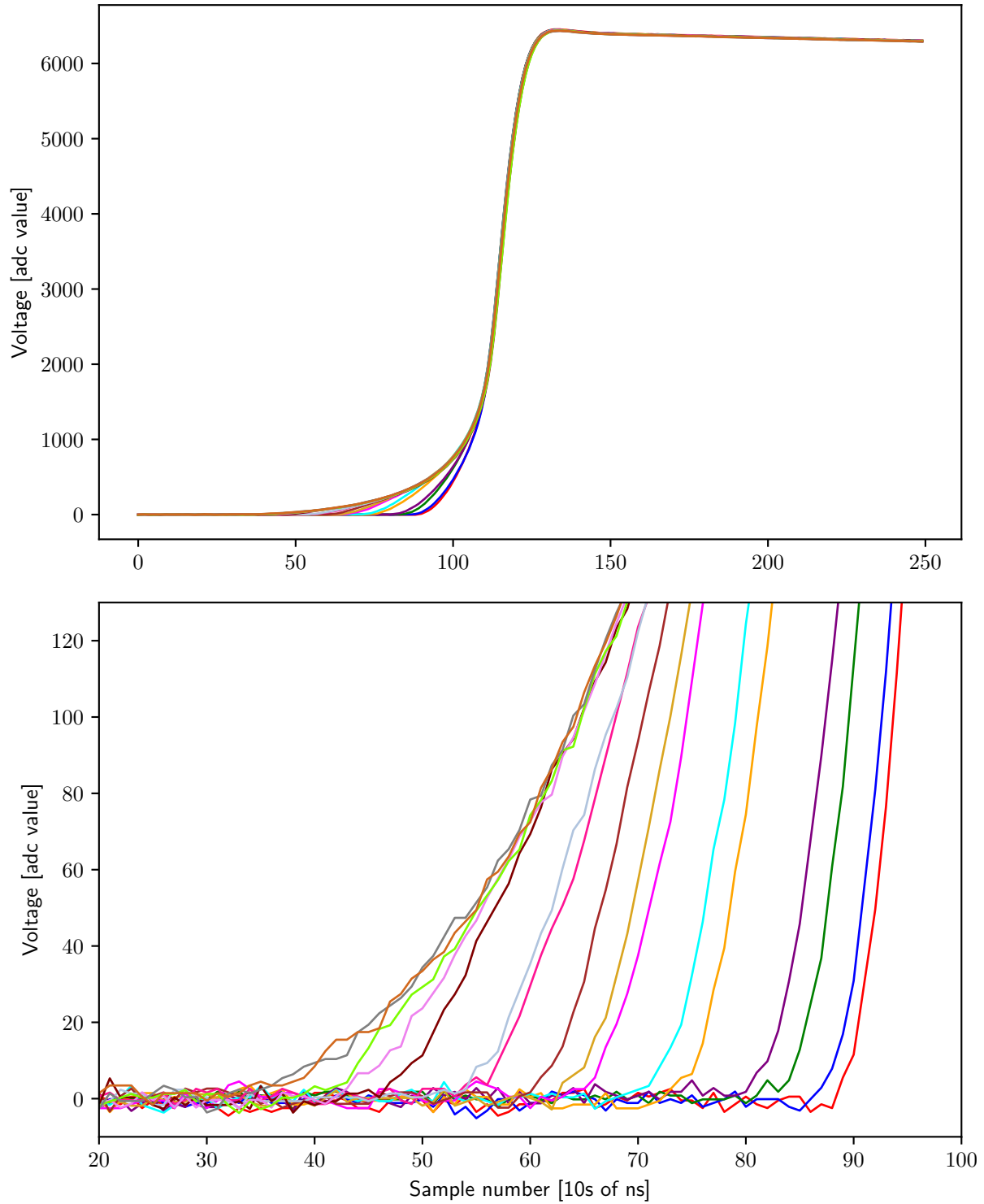Figure 4.5: Training set of sixteen waveforms used for machine learning fit for detector P2574A. The top plot shows the full 250-sample window used in the fit, centered around the estimated 95% rise time. The bottom plot shows the differences in the shape at the start of each waveform.
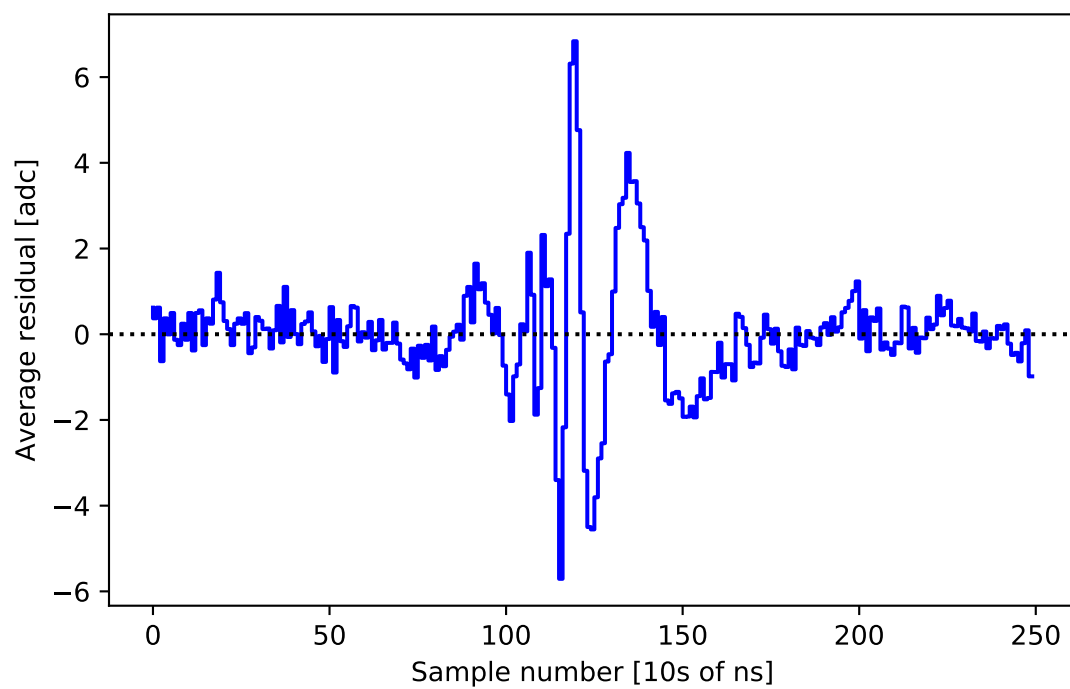
Figure 4.6: The residual between the training waveform and fit, averaged over the sixteen waveforms in the training set, for detector P42574A.

several thousand waveforms. Instead, we have developed a subset of the model which uses `DNest4` to fit only the parameters individual to each waveform. The shared detector and electronics parameters are fixed to the values computed in the full machine learning fit. This gives us a much simpler fit which can be performed on each waveform individually.

For a source of single and multi site events, we can use the Tl^{208} escape peaks. The escape peaks are features in an energy spectrum created from pair production events from a photopeak at higher energy. When the created positron annihilates, it creates a pair of 511 keV gammas inside the detector. One or both of these gammas can go on to "escape" the detector volume, thereby reducing the total measured energy of the event by 511 keV or 1022 keV, respectively creating a "single escape" peak (SEP) and "double escape" peak (DEP) at energies 511 keV and 1022 keV below the photopeak. In the case of double escape peak events, since both 511 keV gammas escaped, there is a high probability that the interaction was confined to the single site of the pair production. For the single escape peak, one of the 511 keV gammas must have interacted in the detector, very likely at a distance separated from the original interaction location. The single escape peak is therefore largely composed of multisite events. For the ^{208}Tl photopeak, the double and single escape peaks are at 1592 keV and 2103 keV, respectively. They are visible in the spectrum in Figure 3.14.

We used the detector and electronics parameters from the fit to P42661A to fit 11,466 calibration events, 6,378 of which are from the SEP and 5,088 from the DEP. The fit likelihood distribution for each event in shown in Figure 4.7, which shows that the DEP events are reconstructed with higher likelihood values than single escape peak events.

The A/E cut value is defined such that it retains 90% of events in the DEP. We can attempt to create a similar cut using a fit likelihood threshold, since our model should fit single site events better than multisite. The results of a cut which retains 90% of DEP events in shown in Figure 4.8.

The waveforms for all events in the DEP are shown in Figure 4.9. Many of the failing events are characterized by an extremely long time for the waveform to rise from $\sim 90\%$ of
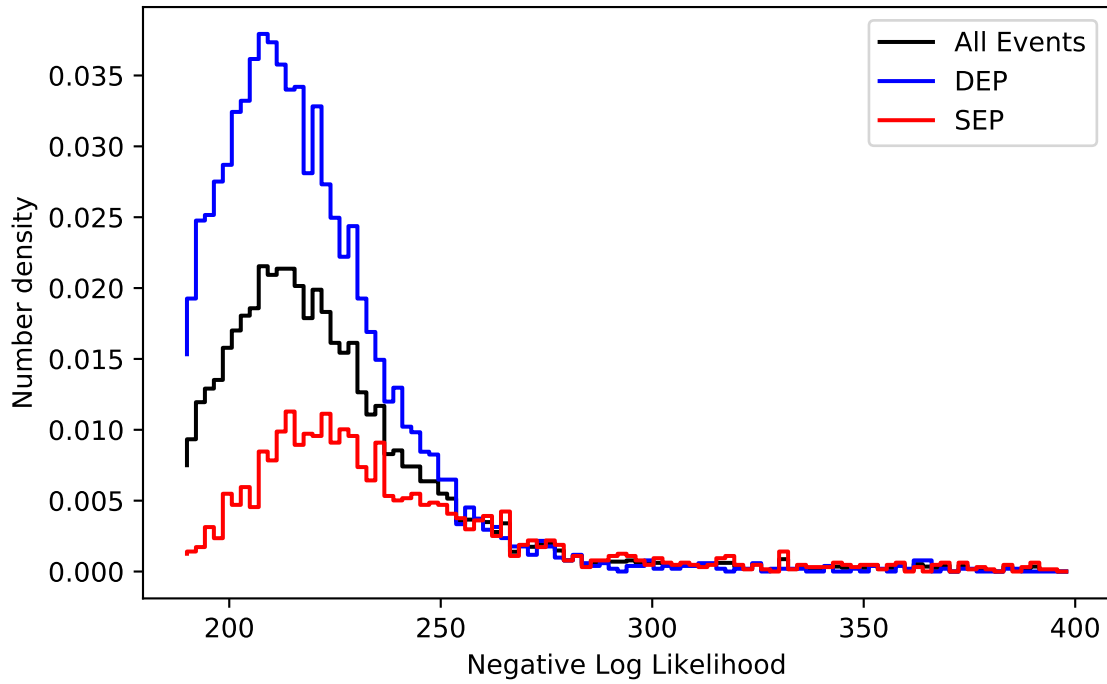
Figure 4.7: The plot shows the negative log likelihood (NLL) for fits to waveforms from the ^{208}Tl single and double escape peaks. Because it is negative likelihood, a lower value corresponds to a better fit. The distribution for the DEP and SEP are each normalized to the total number of events in that peak. Multisite waveforms should have higher NLL values than single site events.The SEP distribution is significantly lower in amplitude because many multisite events from the distribution have extremely high NLL values and are therefore beyond the scale of this plot.
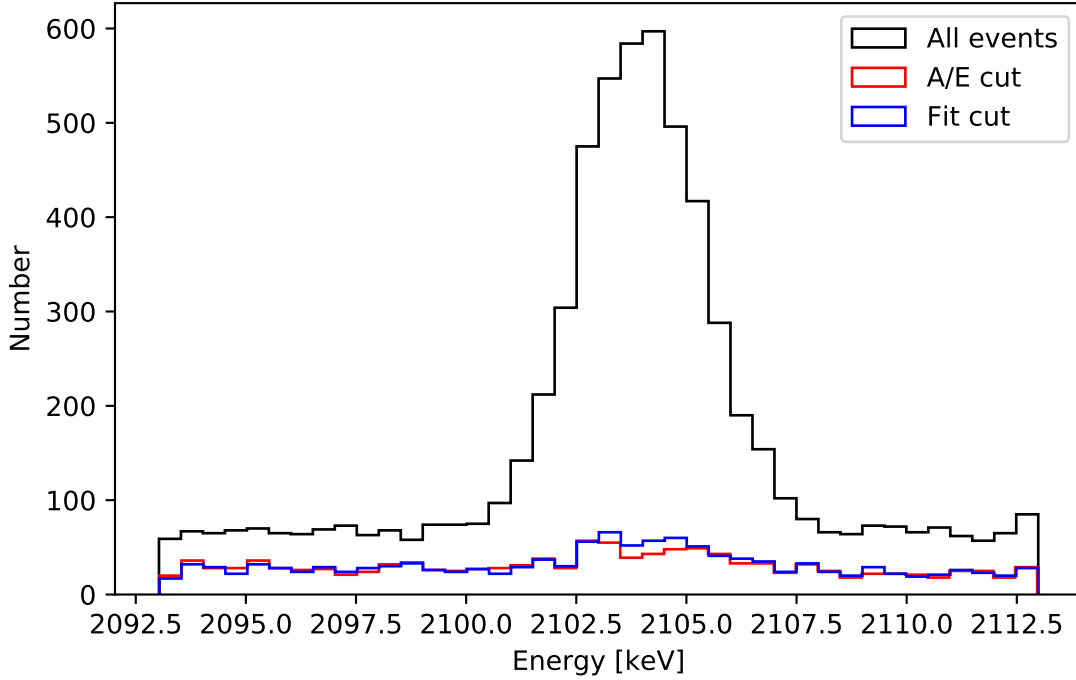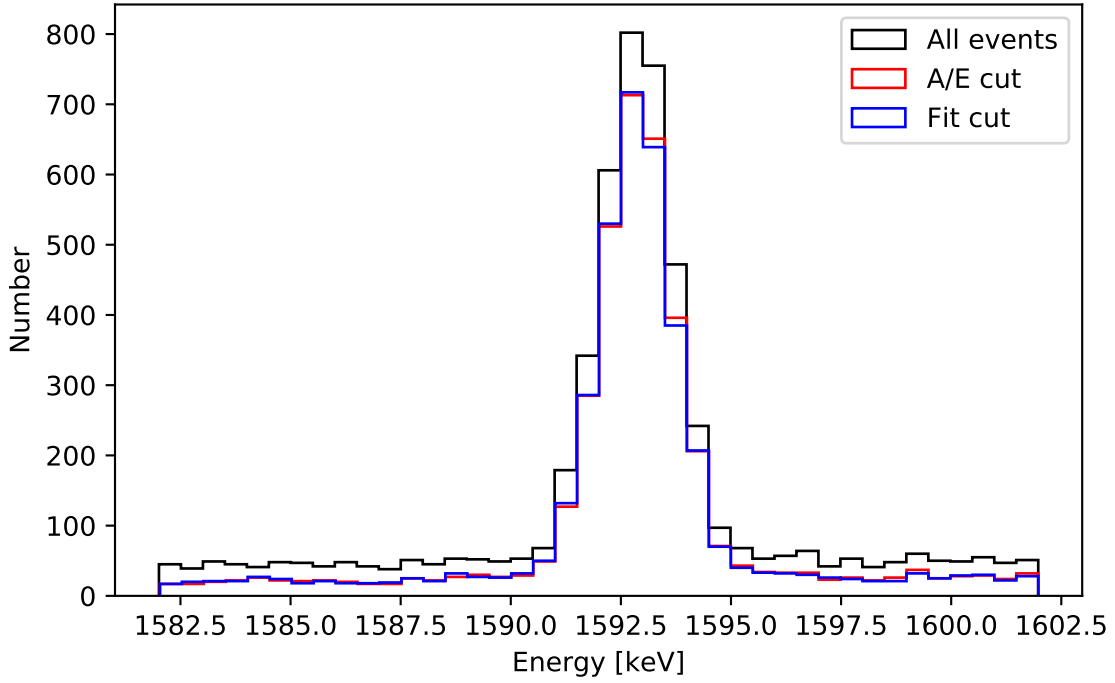
Figure 4.8: Comparison of A/E cut and likelihood fit cuts on the ^{208}Tl double (at top) and single (at bottom) escape peaks. Both cuts are defined to retain 90% of events in the DEP. Using an extremely simple cut, we are able to discriminate against multisite events in the SEP nearly as efficiently as the A/E cut.
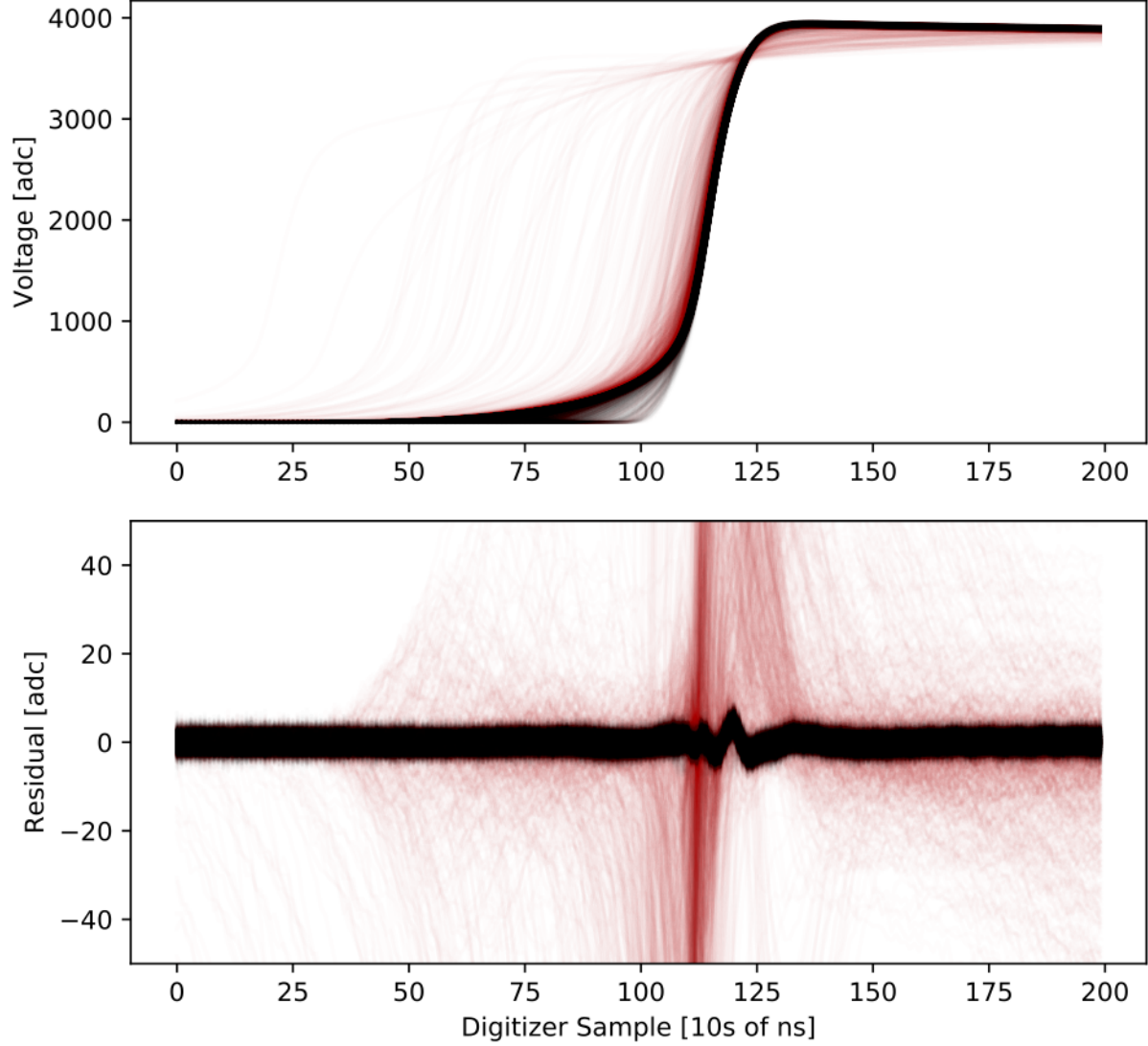
Figure 4.9: Waveforms and estimated residuals for DEP waveforms in detector P42661A. Red waveforms fail the likelihood cut, while black waveforms pass. Many of the cut waveforms show a very long time between the time points corresponding to ∼90% to ∼99% of the maximum amplitude, suggesting that these events may have a percentage of their holes slowed by diffusion in the $n+$ layer.

the amplitude to the maximum. It is likely that these events deposited energy very near to the n+-p boundary at the outside of the detector. Holes in low field regions around this boundary can diffuse slowly into the depleted region of the detector, creating a delayed signal component. Since this diffusion dominated region is not part of the current model, it is no surprise that they should fit poorly.

The estimated radial and azimuthal position of DEP events is shown in Figure 4.10. Given the clustering of poorly fitting events at low r and z, whose waveforms have the fast rise time corresponding to isochrones near the point contact, it is clear that events near the point contact are not fitting with high precision. Waveforms in this region will have non-negligable electron contributions to the waveform. Given the approximations made in the model that the hole contribution dominates, it is no surprise that point contact events should fit poorly. The residual for waveforms which do pass the cut have approximately the same structure as seen in the training waveforms.

The estimated positions can be compared to Figure 4.11, which shows the GEANT4 simulated position of double escape events in the detector. There is some preference for the events to occur at the edges of the detector, where there is less material for annihilation gammas to penetrate before escaping. However, it is clear that the simulated events do not cluster strongly in any specific location of the detector.

In the MCMC estimated positions, there is a strong cluster in the corner of the detector opposite the point contact. There is an additional cluster at $\sim (r = 500$ mm$^2, z = 20$ mm). Each are likely caused by some inaccuracy in the model. In addition, there is an area just above the taper at high r and low z which contains nearly no reconstructed events. We model this taper as a perfectly sharp angle, which is somewhat unphysical, and probably causes distortion in the modeled fields in the region, seen in Figure 3.9. In reality, there is some curvature radius on the taper edge which should be incorporated into the model.

The estimated azimuthal position is shown in Figure 4.12. The model appears to cluster waveforms near $\phi = 0, \pi/8$ and $\pi/4$. This likely points to some issue with the estimated

97
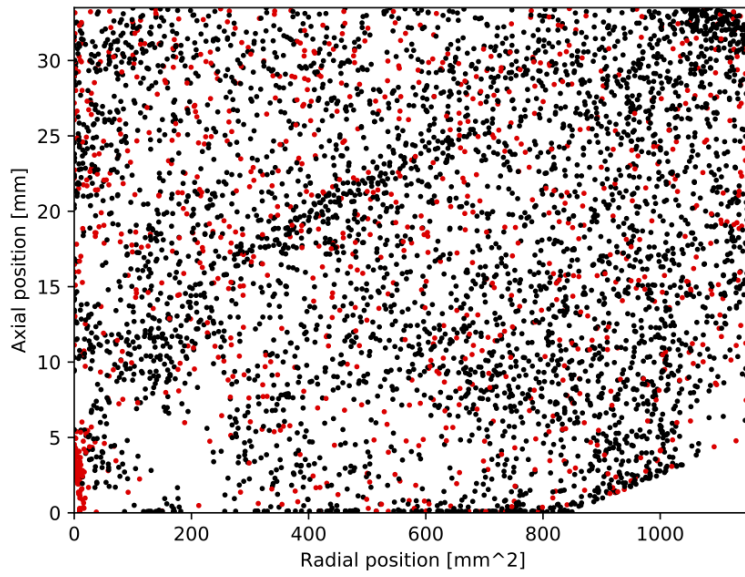
Figure 4.10: Estimated position for 5,088 DEP events in detector P42661A using values from training set one. Events which fail the 90% DEP cut are shown in red, events which pass are in black. The radial position is squared to normalize for the increased detector volume in a given voxel at higher radius. In these coordinates, the position distribution should be roughly uniform (see Figure 4.11). The taper is visible in the corner at low axial position and high radial position.
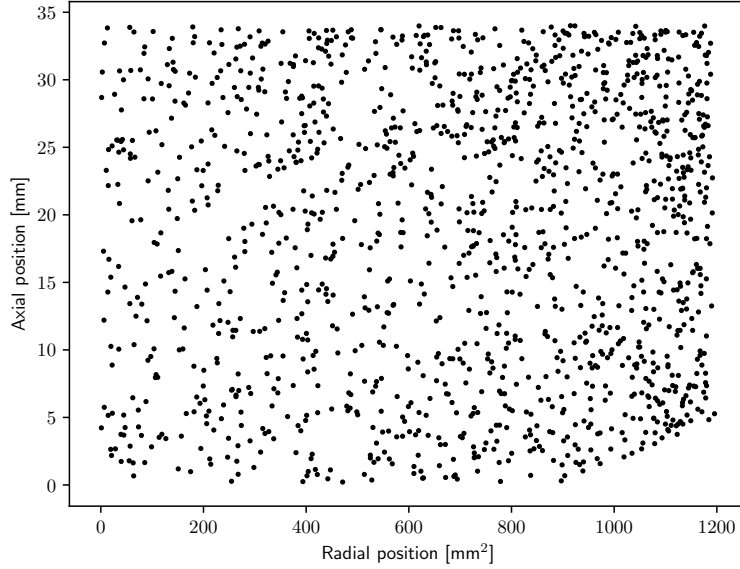
Figure 4.11: Position distribution for DEP events in P42661A using a GEANT4-based simulation of the DEMONSTRATOR geometry with the thorium source inserted. Double escape events are tagged by energy and the presence of a positron during the track. Some clustering is visible at the corners of the detector.

drift velocity curves in Figure 4.4, since the azimuthal position reconstruction depends on the drift velocity anisotropy between crystal axes.

The DEP analysis was also repeated for detector P42574A. Figure 4.13 shows the waveforms corresponding to this fit, while Figure 4.14 shows the positions. The behavior in this detector is broadly similar to P42661A. A cluster of failing events appears near the point contact, and there is some strong clustering of positions. Especially interesting is the linear cluster at $\sim (r = 500 \text{ mm}^2, z = 27 \text{ mm})$, which is similar to one which appears at a proportionally similar position within detector P42661A.

There are two interesting differences between the position maps. While P42661A shows strong clustering in the corner opposite the point contact, in P42574A there is a strong cluster just below the corner. It is possible this is related to poorly estimated impurity profile parameters. Additionally, detector P42574A has a large sparse region of high likelihood events reconstructed at low radial and high axial position.
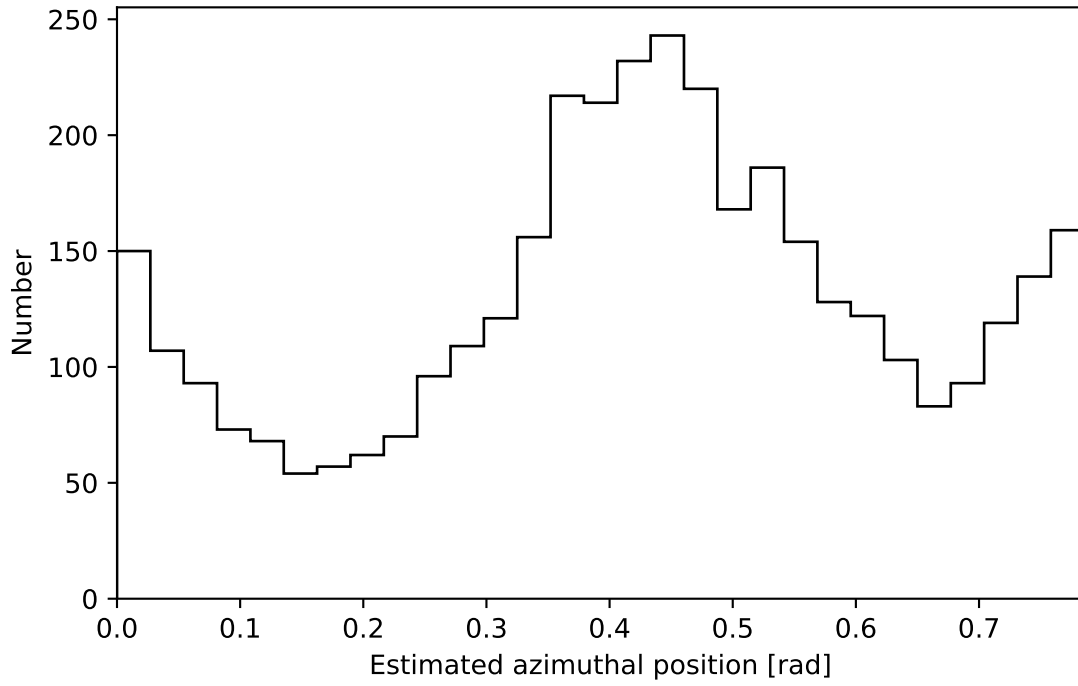
Figure 4.12: Estimated azimuthal position for DEP events which pass the 90% cut. If the model were perfect, we should expect this distribution to be flat between 0 and $\pi/4$.
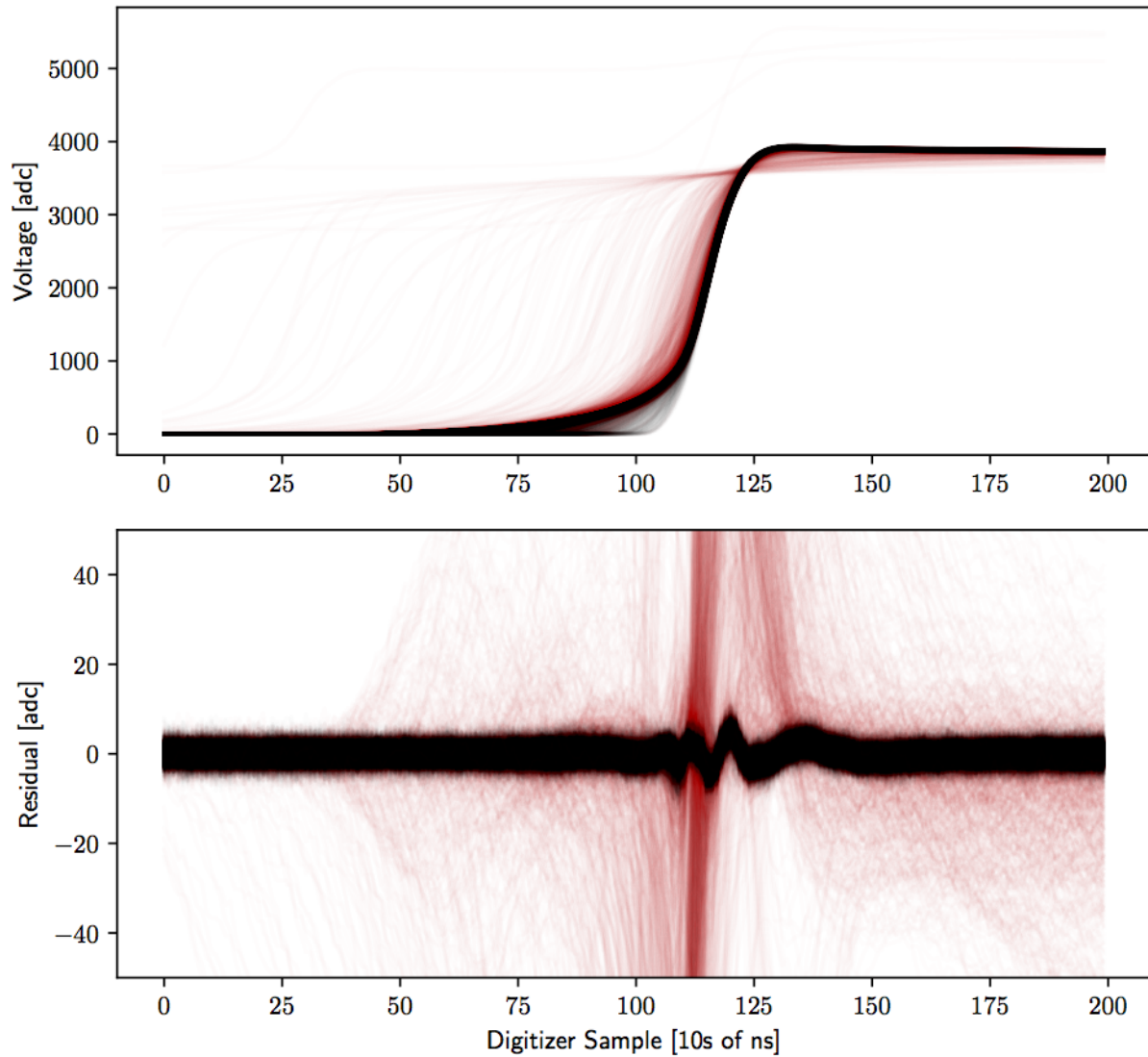
Figure 4.13: Waveforms and estimated residuals for DEP waveforms in detector P42574A. Red waveforms fail the likelihood cut, while black waveforms pass. Again, a large fraction of the failing events have the long drift time between ~90% and ~99% of the maximum amplitude.
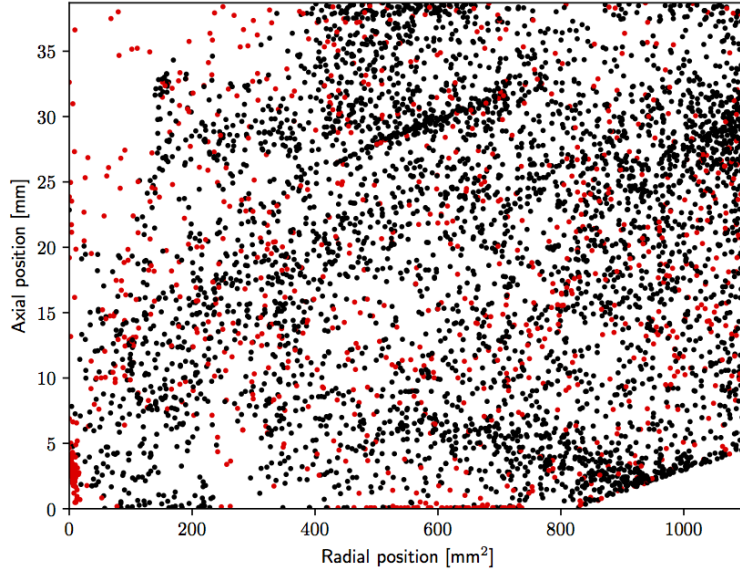
Figure 4.14: Estimated position for DEP events in detector P42574A. Red points fail a 90% efficiency likelihood cut, while black points pass the cut. Similar to detector P42661A, there is clustering in certain locations. Additionally, the area at high axial and low radial position is particularly sparse.

Section 4.3: Uncertainties

There are several possible sources of uncertainty in the estimation of values from the detector training algorithm. Because we are training the model on a finite set of waveforms, it is possible the estimated value is biased based on the waveforms selected for the training set. Additionally, if the Markov chain cannot smoothly sample the parameter space, it can become "stuck" on one local maximum likelihood value, and not find the true global maximum.

In general, choosing a training set with the largest possible number of waveforms should produce the most accurate estimate of the parameters. However, to make the algorithm computationally feasible, we chose to train with only sixteen waveforms. To attempt to understand the uncertainty associated with the selection of this training set, we ran the algorithm with two different sets of training waveforms for detector P42661A. The estimated parameters for both sets are shown in Table 4.2. For velocity and electronics filter

parameters, the difference in estimated values is roughly on order of 10%.

There is a larger discrepancy between sets for impurity concentration and point contact dimension parameters. It is difficult to definitively ascribe the difference in estimated value to physical differences between the training sets, because it is also possible that the Markov chain is not correctly converging on the optimum value for these parameters. Qualitatively, we can judge how well the a Markov chain is converging to a true global maximum likelihood based on the structure of a histogram of the values explored by the chain for a given parameter. Figure 4.15 shows a histogram of the final 1,000 values of the chain for electronics shaping parameters. The chain is exploring a continuous region of the possible space for each of these parameters. In contrast, Figure 4.16 and Figure 4.17 show the final 1,000 values for the velocity and detector parameters. Most of these histograms show sharp peaks clustered around disparate values (each of which corresponds to the location in parameter space of a distinct particle in the diffusive nested sampler). It is possible that the likelihood space is truly so strongly multimodal that this is a correct representation of the target distribution. However, it is more likely that the chain is unable to smoothly sample the parameter space due to the complexity of the model.

It should be expected that the velocity and detector parameters are more difficult to sample than the electronics parameters. Figure 2.19 shows that the electronics is a dominant effect on the waveform shape. Because the electronics shape the signal after it has formed in the detector, the extremely long rise time of the DEMONSTRATOR amp effectively washes out the imprint of the detector and velocity parameters. In addition, there are a number of very strong correlations in the model which describes signal formation. The model for hole drift velocity, for example, depends both on the parameters of the velocity curve as well as the modeled electric field. This creates a strong correlation between the impurity concentration and velocity parameters.

It is possible to approximate an uncertainty for parameters with poor chain performance by evaluating the region of parameter space the chain explores before it becomes fragmented.

Figure 4.15: Histogram of the final 10,000 saved steps in the Markov chain for electronics filter parameters in P42661A training set one. For each of these parameters, the chain is exploring a continuous region around a single value. This is the proper behavior for a well-converged chain.

Figure 4.16: Histogram of the final 10,000 saved steps in the Markov chain for velocity parameters in P42661A training set one. For these parameters, the chain is exploring different well-separated pockets of different values. This indicates either that the likelihood space is extremely multi-modal likelihood space or that the chain is having difficulty sampling over the allowed space for these parameters.
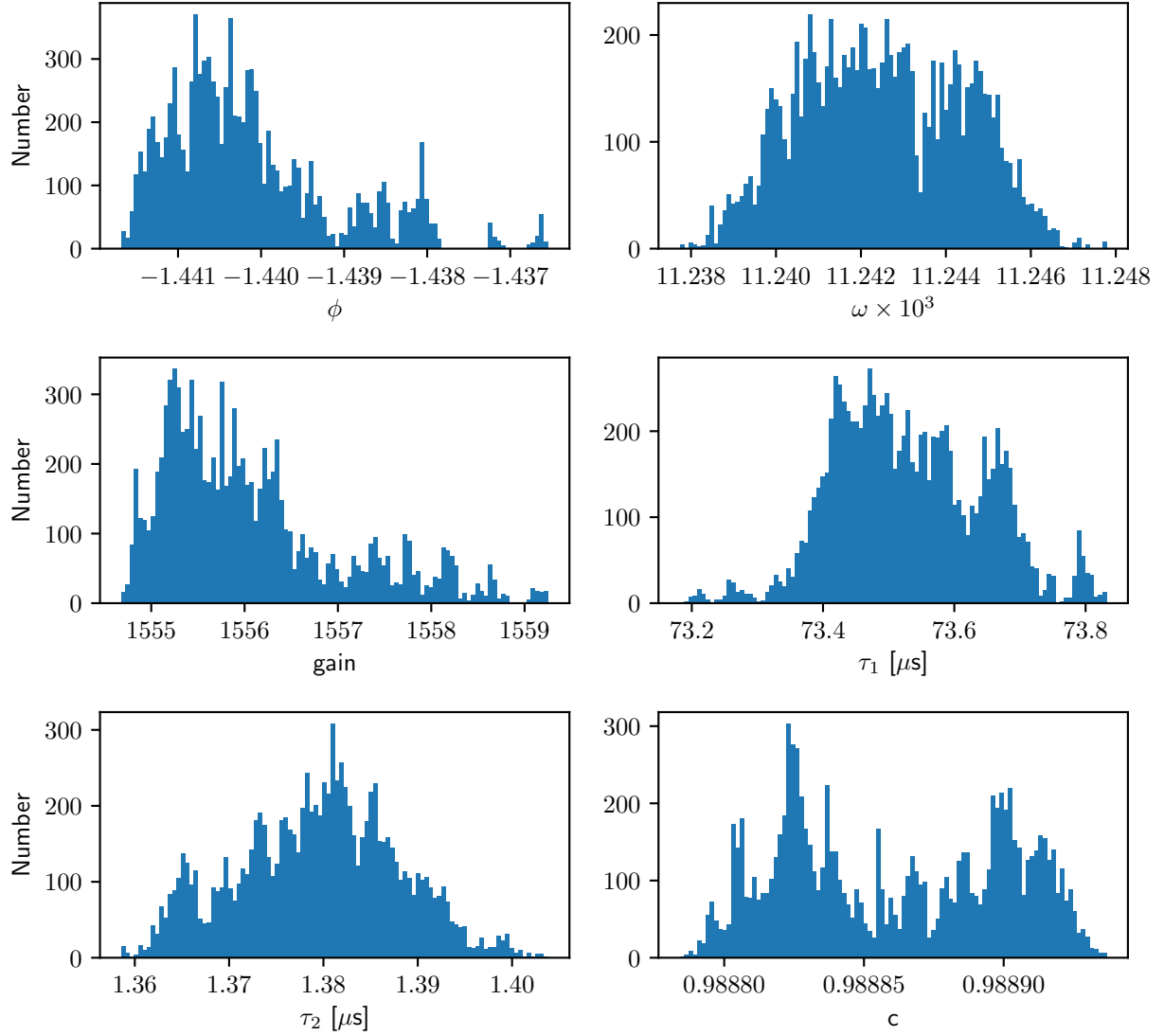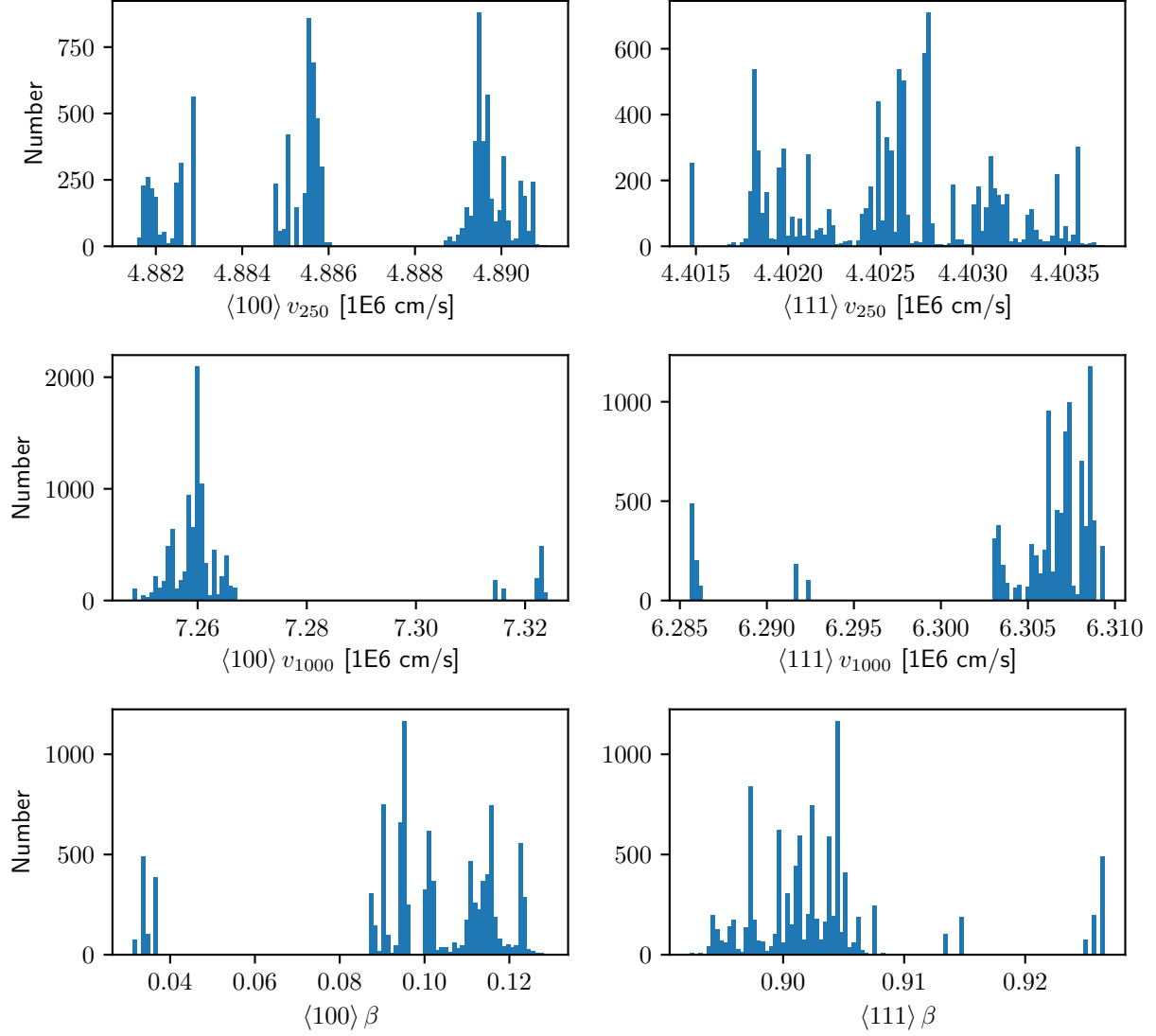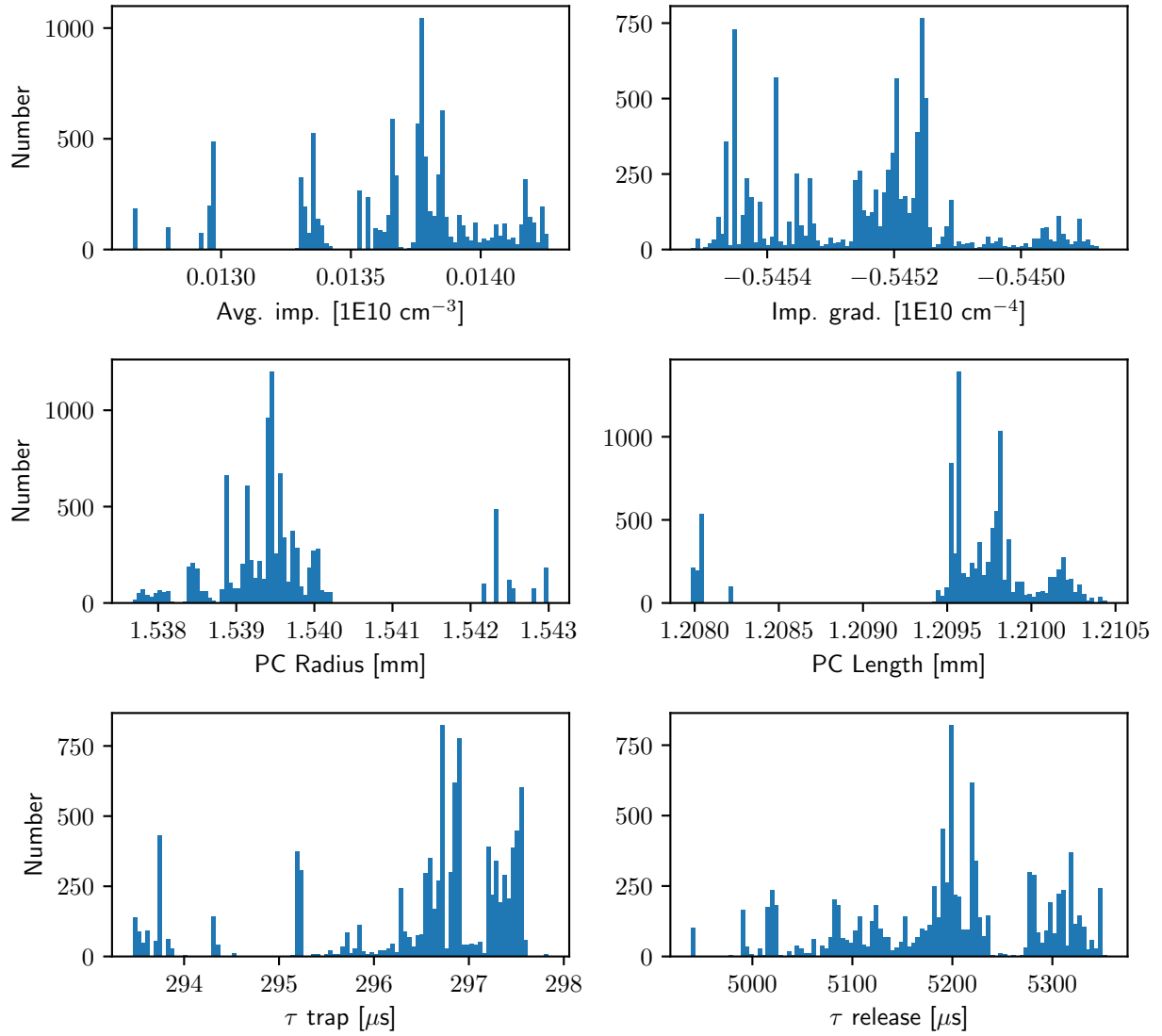
Figure 4.17: Histogram of the final 10,000 saved steps in the Markov chain for detector parameters in P42661A training set one. Like in Figure 4.16, the chain is not smoothly sampling over these parameter values.

Figure 4.18 shows a series histogram of the estimated radial and axial position as the chain iterates. We can estimate the uncertainty in the position estimate as the rough area the chain explores in subplot (b).

In principle, the issue of poor chain performance could be addressed by increasing the number of particles run in the diffusive nested sampling algorithm. However, the availability of computational resources has limited us to 20 particles for the sixteen waveform set considered above.

Section 4.4:   Conclusions

Diffusive nested sampling was utilized to fit a training set of sixteen waveforms from two different detectors from the DEMONSTRATOR. The model is able to fit the waveforms with residuals on the order of parts per thousand of the overall signal amplitude. There is, however, some structure which remains in the residuals, indicating either that the model is incorrect or that it has converged on a suboptimal set of parameters.

The performance of the fit was evaluated by performing a simplified fit to individual waveforms from the ^{208}Tl escape peaks. Using the majority single site events of the DEP, we estimated the ability of the model to estimate position. In general, there are no large volumes it totally fails to sample correctly, with the exception of the region near the point contact, and the region near the taper. The nonuniform clustering seen in position, especially in ϕ, again suggests some deficiency or degeneracy in the model.

Given the complexity and high degree of correlation in the model, it is difficult to speculate as to the exact parameters which are causing these issues. It is clear that the residual affects all waveforms equally, which could point to the shaping due to electronics. The electronics response function chosen in Section 2.2.3 was arrived at empirically but without high-precision measurements using signals of known input. Future measurements should be dedicated to ensuring that this function is adequate at the level of accuracy now required of the model.
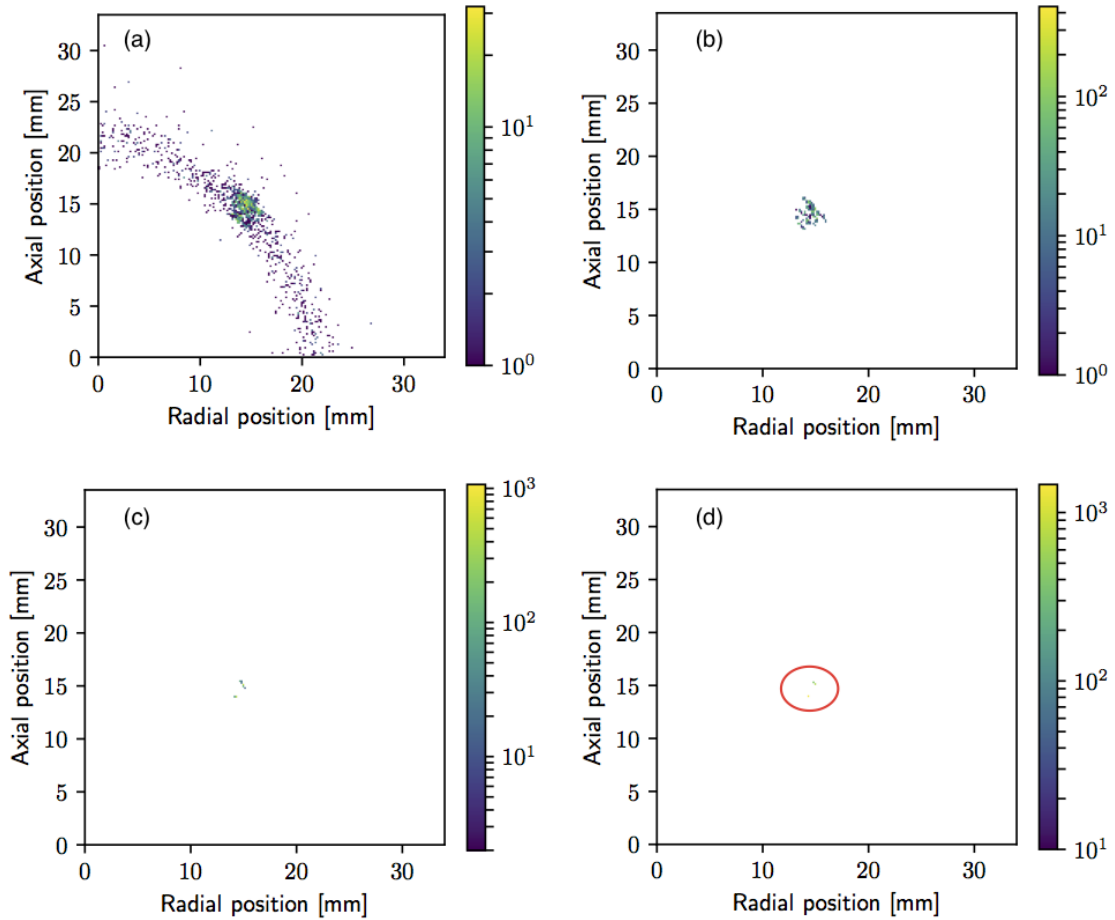
Figure 4.18: Histogram of position parameters for different slices of the Markov chain for P42661A training set one. The chain was run for 680,000 saved steps. Each plot shows a histogram of chain values for position over a different slice of the iteration number: (a) steps 1,000–7,700; (b) 7,700 – 14,400; (c) 14,400-21,100; (d)21,100-27,800. In (a), the chain is exploring the region generally along a drift time isochrone, and has shown preference for the region around ($r = 15$mm, $z = 15$mm). In (b) and (c), the chain splinters into on disconnected subspaces of position. By (d), the chain samples only a few of the 0.1mm × 0.1mm pixels. (The red circle is included to emphasize where on the plot the pixels are located).

Even with an imperfect model, a cut based on fit likelihood was shown to discriminate against multisite events with nearly the same efficiency as an A/E cut. Given that the cut is defined at 90% efficiency in the DEP, and we are sacrificing a large number of events to the point contact region, it is clear that the current model can at least discriminate between single and multisite events with accuracy.

CHAPTER 5:   Conclusions

Section 5.1:   Summary of results

We have developed a model which is able to simulate waveforms formed in MAJORANA DEMONSTRATOR PPC germanium detectors with high precision. This model incorporates semiconductor physics with an empirical representation of the shaping introduced by readout electronics. Using the framework of Bayesian statistics, we have utilized the model to build an algorithm to learn the correct model parameters for a given detector. The model is trained using only a small number of calibration waveforms which have been carefully selected to provide information about single site event shapes from across the detector volume.

When the algorithm is applied to a training set of sixteen waveforms from a DEMON-STRATOR detector, the model is able to reconstruct waveforms to the level of parts per thousand. However, the structure present in the residual suggests either the model is missing key physics, or the model is not correctly converging on a global maximum. Even with this limitation, the model is clearly able to distinguish between normal and abnormal events with high efficiency.

The distribution of estimated position for single site events is less uniform than would be expected. Again, this points to either the model or parameters being incorrect. In addition to clustering, both analyzed detectors showed some regions which are with fewer estimated events compared to simulation. Again, this points either to a deficiency in the model or a failure of the MCMC algorithm.

Section 5.2:   Additional studies

There are several studies which would improve understanding of the results described above. In Section 4.3, we showed that the MCMC is likely failing to properly explore the full parameter space. As a future study of the MCMC convergence, the algorithm can be run several times on the same training set with different random number seeds. Changing the seed will change the initial state of the Markov chain, which is drawn randomly from the prior distribution. If the parameter estimates differ substantially between seeds, it is proof that the chain is converging poorly.

It is also possible that increasing the number of waveforms in the training set could help improve the ability of the chain to converge. If the chain is failing because there is insufficient information in the sixteen waveform training set to create a strong preference in likelihood space, additional waveforms would provide more information. On the other hand, adding waveforms also increases the dimensionality of the fit, which makes convergence more difficult. Still, it is worth attempting to train on larger sets, keeping in mind that increasing the size also increases the required computational resources.

In addition, several studies are necessary to investigate possible dependence of the estimated parameters on the choice of training set waveforms. All the data discussed in this work originate from a single calibration set, taking place over the course of several hours. It is possible that some parameters could change in time due to environmental or other factors. For example, it is conceivable that ambient temperature could affect the electronics filter parameters. The learning algorithm should be applied to training waveforms from calibrations taken weeks and months apart to evaluate the stability of all parameters.

Similarly, all training waveforms in the results above were taken from the 2614 keV ^{208}Tl photopeak, which was chosen to maximize the signal to noise ratio. It is feasible that the estimated parameters might changes based on the energy of the training waveforms used. This can be examined by training with waveforms originating from different energies, such as the double escape peak or Compton continuum edge.

Finally, it would be extremely useful to begin dedicated measurements to validate the parameters estimated by the model. As a first step, collimated scans could be performed with a relatively high energy source. The estimated position distribution could then be compared to the location of the collimated beam as a more precise form of validation than simply looking at the distribution of DEP events.

Section 5.3:   Improvements

5.3.1:   Model improvements

We have put great effort into trying to resolve the origin of the structure in the residual by adjusting the model and incorporating additional physics. One possibility is that the model is correct, but the MCMC algorithm is unable to converge on the global likelihood maximum. Given the state of the Markov chain shown in Figures 4.15 and 4.16, it seems worth exploring improvements to the MCMC before attempting to introduce new changes to the model.

Should it prove necessary to adjust the model, there are a few areas which stand out as obvious candidates for improvement to the observed residual. Because the residual is of equal strength for waveforms of all drift times, its origin cannot be a localized effect. This eliminates the failure to float electron velocity parameters as a possible culprit. The electronics filter parameters stand out as clearly affecting all waveforms equally. It might be possible to increase the order of the low pass filter to fully encapsulate the DEMONSTRATOR electronics frequency response. A second possibility is that the approximations about charge cloud shape described in Section 2.2.2 are no longer accurate at this level of precision. Finally, since the holes from all waveforms are collected at the point contact, error in the field or weighting potential in this region could give rise to the residual. At present, the point contact is modeled as a perfect hemispherical indentation into the detector surface, which creates a sharp corner where the two join. In reality, this corner has some curvature, which could be

112

included in the model in the future.

In the long term, modeling waveforms which originate in the volume immediately surrounding the detector will require including an electron drift velocity parameterization in the model. Given that this region is a relatively small percentage of the overall detector volume, and would require increasing the dimensionality of the fit, this is probably a low priority.

As a parallel project, it would be interesting to attempt to incorporate into the model the effects which distort waveform shape at the edges of the detectors. For instance, we believe diffusion in low field regions near the $n+$ contact is responsible for the waveforms with slowed rise times from ${\sim}90\%$ to maximum in Figures 4.10 and 4.14. Modeling this effect might allow us to fit the slowed waveforms which originate near the n+ contact in This could, in turn, allow for improved understanding of the depth of the n+ contact and the true active volume of each detector.

5.3.2: Training improvements

Improving the uncertainties on parameters before introducing them to the model could improve convergence of the chain and reduce required computation time. Perhaps the simplest to implement is a dedicated measurement of the electronics filter parameters. By using a signal generator to inject a signal of known shape at the detector, the transfer function parameters could be fit completely independently of the detector and waveform parameters.

The work described in this document was started after construction on the DEMONSTRATOR was well underway, which prevented us from taking any dedicated detector characterization data. In the future, dedicated collimated scans could be used to create training sets of waveforms with well-understood location. Understanding the location would considerably shrink the size of the allowed parameter space from the prior and simplify the training algorithm.

5.3.3: Computational improvements

An impediment to broadly applying this technique is the heavy computational load it requires. The training algorithm as implemented in this work, using sixteen waveforms, takes approximately a day of wall time on the UNC killdevil cluster using 341 cores. Assuming the detector parameters are stable in time over the course of data taking for the experiment, this should only need to be run once per detector. Once detector parameters are frozen, fitting each individual waveform takes on the order of a minute (and is trivially parallelizable).

The simplest way to improve the computation time required for the training step is to improve the information input to the training step, as described above. With better information on electronics parameters and waveform position, the training step would take considerably less time to complete. Additionally, improving the model to account for the residual should smoothen the convergence process and improve computation time. Model improvement should similarly improve the convergence time for individual waveforms.

Finally, more work can focus on improving the computation itself. There is remaining optimization which should be done within `pysiggen`, especially in the application of convolutions for charge cloud size and electronics filtering. Some speedup should be possible by porting the python components to the c language.

Section 5.4:   Outlook for the DEMONSTRATOR and beyond

As we continue to improve the model, it is important to keep in mind that the ultimate measure of this technique is its applications for the DEMONSTRATOR and future tonne-scale experiments. In Section 4.2, we showed that a simple likelihood-based cut can effectively discriminate against multi-site events in the single escape peak, as well as reject unusual waveforms in the double escape peak. An obvious next step is to apply the cut to waveforms from DEMONSTRATOR background data. Given that a large number of the DEMONSTRATOR background events around the ROI have been identified as alpha events with a distorted

114

waveform shape, it would be interesting to evaluate the ability of the likelihood cut to reject alphas. Additionally, the $2\nu\beta\beta$ events in the spectrum will provide a check on the ability of the technique to fit $\beta\beta$ waveforms.

Although the position estimation shown in Section 4.2 is imperfect, it represents the first time event origin has been estimated inside a PPC detector. The availability of this new information can be used for several purposes in a $0\nu\beta\beta$ experiment. During the commissioning phase, should any unexpected background appear within the experiment, the location could help localize the position of the source. Additionally, since $\beta\beta$ events should be distributed uniformly through the detector volume, position information can serve as a useful check on systematics from analysis cuts.

As we move toward a tonne-scale germanium $0\nu\beta\beta$ experiment, understanding the origin and characteristics of every waveform will become paramount. This can be achieved by designing the tonne-scale experiment with the capabilities of this technique in mind. For example, the narrow bandwidth of the DEMONSTRATOR preamp electronics causes a dramatic increase in waveform rise time, and effectively blurs out information about the detector and waveform properties. A wider-bandwidth preampfification chain should increase the amount of this information which is available in the measured waveform, which could enable improved background rejection. It would be possible to quantify this improvement by studying the algorithm performance on waveforms simulated with a higher bandwidth preamp.

BIBLIOGRAPHY

[1] L. Camilleri, E. Lisi, and J. F. Wilkerson, "Neutrino Masses and Mixings: Status and Prospects," *Annu. Rev. Nucl. Part. Sci.* **58** (2008) 343–369.

[2] M. Goeppert-Mayer, "Double Beta-Disintegration," *Phys. Rev.* **48** (Sep, 1935) 512–516. `http://link.aps.org/doi/10.1103/PhysRev.48.512`.

[3] S. R. Elliott, A. A. Hahn, and M. K. Moe, "Direct evidence for two-neutrino double-beta decay in ^{82}Se," *Phys. Rev. Lett.* **59** (Nov, 1987) 2020–2023. `http://link.aps.org/doi/10.1103/PhysRevLett.59.2020`.

[4] E. Majorana, "Teoria simmetrica dell'elettrone e del positrone," *Il Nuovo Cimento (1924-1942)* **14** no. 4, (Sep, 2008) 171. `https://doi.org/10.1007/BF02961314`.

[5] W. H. Furry, "On Transition Probabilities in Double Beta-Disintegration," *Phys. Rev.* **56** (Dec, 1939) 1184–1193. `http://link.aps.org/doi/10.1103/PhysRev.56.1184`.

[6] I. Avignone, Frank T., S. R. Elliott, and J. Engel, "Double Beta Decay, Majorana Neutrinos, and Neutrino Mass," *Rev.Mod.Phys.* **80** (2008) 481–516, `arXiv:0708.1033 [nucl-ex]`.

[7] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the Unified Model of Elementary Particles," *Progress of Theoretical Physics* **28** no. 5, (1962) 870–880.

[8] S. Dell'Oro, S. Marcocci, and F. Vissani, "New expectations and uncertainties on neutrinoless double beta decay," *Phys. Rev. D* **90** (Aug, 2014) 033005. `https://link.aps.org/doi/10.1103/PhysRevD.90.033005`.

[9] J. Barea, J. Kotila, and F. Iachello, "Nuclear matrix elements for double-β decay," *Phys. Rev. C* **87** (Jan, 2013) 014315. `https://link.aps.org/doi/10.1103/PhysRevC.87.014315`.

[10] J. Engel and J. Menndez, "Status and future of nuclear matrix elements for neutrinoless double-beta decay: a review," *Reports on Progress in Physics* **80** no. 4, (2017) 046301. `http://stacks.iop.org/0034-4885/80/i=4/a=046301`.

[11] **Particle Data Group** Collaboration, C. Patrignani *et al.*, "Review of Particle Physics," *Chin. Phys.* **C40** no. 10, (2016) 100001.

[12] **GERDA** Collaboration, M. Agostini *et al.*, "Background-free search for neutrinoless double-β decay of ^{76}Ge with GERDA," *Nature* **544** no. 7648, (04, 2017) 47–52.

[13] **CUORE** Collaboration, K. Alfonso *et al.*, "Search for Neutrinoless Double-Beta Decay of ^{130}Te with CUORE-0," *Phys. Rev. Lett.* **115** (Sep, 2015) 102502. `https://link.aps.org/doi/10.1103/PhysRevLett.115.102502`.

[14] **EXO** Collaboration, J. B. Albert *et al.*, "Search for Majorana neutrinos with the first two years of EXO-200 data," *Nature* **510** no. 7504, (06, 2014) 229–234. http://dx.doi.org/10.1038/nature13432.

[15] **KamLAND-Zen** Collaboration, A. Gando *et al.*, "Search for Majorana Neutrinos Near the Inverted Mass Hierarchy Region with KamLAND-Zen," *Phys. Rev. Lett.* **117** (Aug, 2016) 082503. https://link.aps.org/doi/10.1103/PhysRevLett.117.082503.

[16] **Majorana** Collaboration, N. Abgrall *et al.*, "The Majorana Demonstrator Neutrinoless Double-Beta Decay Experiment," *Adv.High Energy Phys.* **2014** (2014) 365432, arXiv:1308.1633 [physics.ins-det].

[17] F. Šimkovic, A. Faessler, H. Müther, V. Rodin, and M. Stauf, "$0\nu\beta\beta$-decay nuclear matrix elements with self-consistent short-range correlations," *Phys. Rev. C* **79** (May, 2009) 055501. http://link.aps.org/doi/10.1103/PhysRevC.79.055501.

[18] **Majorana** Collaboration, N. Abgrall *et al.*, "Muon flux measurements at the davis campus of the sanford underground research facility with the majorana demonstrator veto system," *Astroparticle Physics* **93** (2017) 70 – 75. http://www.sciencedirect.com/science/article/pii/S0927650517300038.

[19] P. Barton, P. Luke, M. Amman, Y. D. Chan, J. Detwiler, J. Loach, R. Martin, A. Poon, C. Tindall, and K. Vetter, "Low-noise low-mass front end electronics for low-background physics experiments using germanium detectors," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, pp. 1976–1979. Oct, 2011.

[20] ORTEC, "Review of the Physics of Semiconductor Detectors." https://www.ortec-online.com/-/media/ametekortec/other/review_of_the_physics_of_semiconductor_detectors.pdf. Accessed: 2017-07-07.

[21] E. E. Haller, W. L. Hansen, and F. S. Goulding, "Physics of ultra-pure germanium," *Advances in Physics* **30** no. 1, (1981) 93–138.

[22] G. Knoll, *Radiation Detection and Measurement*. John Wiley & Sons, 2010. https://books.google.com/books?id=4vTJ7UDel5IC.

[23] P. N. Luke, F. S. Goulding, N. W. Madden, and R. H. Pehl, "Low capacitance large volume shaped-field germanium detector," *IEEE Transactions on Nuclear Science* **36** no. 1, (1989) 926–30.

[24] **CoGeNT Collaboration** Collaboration, C. E. Aalseth *et al.*, "Results from a Search for Light-Mass Dark Matter with a *p*-Type Point Contact Germanium Detector," *Phys. Rev. Lett.* **106** (Mar, 2011) 131301. https://link.aps.org/doi/10.1103/PhysRevLett.106.131301.

[25] **Majorana** Collaboration, N. Abgrall *et al.*, "New Limits on Bosonic Dark Matter, Solar Axions, Pauli Exclusion Principle Violation, and Electron Decay from the Majorana Demonstrator," *Phys. Rev. Lett.* **118** (Apr, 2017) 161801. https://link.aps.org/doi/10.1103/PhysRevLett.118.161801.

[26] P. S. Barbeau, J. I. Collar, and O. Tench, "Large-mass ultralow noise germanium detectors: performance and applications in neutrino and astroparticle physics," *Journal of Cosmology and Astroparticle Physics* **2007** no. 9, (2007) 9.

[27] N. Abgrall *et al.*, "Search for Pauli exclusion principle violating atomic transitions and electron decay with a p-type point contact germanium detector," *The European Physical Journal C* **76** no. 11, (2016) 619. http://dx.doi.org/10.1140/epjc/s10052-016-4467-0.

[28] ORTEC, "Overview of Semiconductor Photon Detectors." https://www.ortec-online.com/-/media/ametekortec/other/overview-of-semiconductor-photon-detectors.pdf. Accessed: 2017-05-09.

[29] W. Shockley, "Currents to Conductors Induced by a Moving Point Charge," *Journal of Applied Physics* **9** no. 10, (1938) 635–636. http://scitation.aip.org/content/aip/journal/jap/9/10/10.1063/1.1710367.

[30] S. Ramo, "Currents Induced by Electron Motion," *Proceedings of the IRE* **27** no. 9, (Sept, 1939) 584–585.

[31] M. Agostini *et al.*, "Pulse shape discrimination for GERDA Phase I data," *Eur. Phys. J.* **C73** no. 10, (2013) 2583, arXiv:1307.2610 [physics.ins-det].

[32] R. Cooper, D. Radford, K. Lagergren, *et al.*, "A pulse shape analysis technique for the Majorana experiment," *Nucl. Instrum. Methods Section A* **629** (2011) 303–310.

[33] N. Abgrall *et al.*, "The Majorana Demonstrator Neutrinoless Double-Beta Decay Experiment," *Advances in High Energy Physics* **2014** (2014) 18.

[34] "mjd_fieldgen and mjd_siggen software." http://radware.phy.ornl.gov/MJ/mjd_siggen/.

[35] https://github.com/benshanks/pysiggen. pysiggen software.

[36] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. Seljebotn, and K. Smith, "Cython: The Best of Both Worlds," *Computing in Science Engineering* **13** no. 2, (March-april, 2011) 31 –39.

[37] ORTEC, "Oak Ridge, TN, USA.".

[38] W. L. Hansen and E. Haller, "High-Purity Germanium Crystal Growing," *MRS Proceedings* **16** (1982) .

[39] W. P. Electronics, "MPOD Power Supplies." `http://www.wiener-d.com/sc/power-supplies/mpod--lvhv/mpod-hv-module.html`. Accessed: 2017-07-18.

[40] iseg High Voltage, "EHS High Voltage Supplies." `http://iseg-hv.com/en/products/detail/EHS`. Accessed: 2017-07-18.

[41] E. Conwell, *High Field Transport in Semiconductors*. Solid State Physics: Advances in Research and Applications. Supplement. Academic Press, 1967. `https://books.google.com/books?id=vdgWAQAAMAAJ`.

[42] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proceedings of the IEEE* **55** no. 12, (Dec, 1967) 2192–2193.

[43] M. Omar and L. Reggiani, "Drift velocity and diffusivity of hot carriers in germanium: Model calculations," *Solid-State Electronics* **30** no. 12, (1987) 1351 – 1354. `http://www.sciencedirect.com/science/article/pii/0038110187900633`.

[44] J. Gunn, "Effect of electron and impurity density on the field-dependence of mobility in germanium," *Journal of Physics and Chemistry of Solids* **8** (1959) 239 – 241. `http://www.sciencedirect.com/science/article/pii/0022369759903257`.

[45] D. M. Brown and R. Bray, "Analysis of Lattice and Ionized Impurity Scattering in *p*-Type Germanium," *Phys. Rev.* **127** (Sep, 1962) 1593–1602. `https://link.aps.org/doi/10.1103/PhysRev.127.1593`.

[46] L. Reggiani, C. Canali, F. Nava, and G. Ottaviani, "Hole drift velocity in germanium," *Phys. Rev. B* **16** (Sep, 1977) 2781–2791. `https://link.aps.org/doi/10.1103/PhysRevB.16.2781`.

[47] L. Mihailescu, W. Gast, R. Lieder, H. Brands, and H. Jger, "The influence of anisotropic electron drift velocity on the signal shapes of closed-end {HPGe} detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **447** no. 3, (2000) 350 – 360. `http://www.sciencedirect.com/science/article/pii/S0168900299012863`.

[48] W. Sasaki, M. Shibuya, and K. Mizuguchi, "Anisotropy of Hot Electrons in n-type Germanium," *Journal of the Physical Society of Japan* **13** no. 5, (1958) 456–460. `http://dx.doi.org/10.1143/JPSJ.13.456`.

[49] "Germanium band structure and carrier concentration." `http://www.ioffe.ru/SVA/NSM/Semicond/Ge/bandstr.html`. Accessed: 2017-07-01.

[50] G. Dresselhaus, A. F. Kip, and C. Kittel, "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals," *Phys. Rev.* **98** (Apr, 1955) 368–384.

https://link.aps.org/doi/10.1103/PhysRev.98.368.

[51] M. Cardona and F. H. Pollak, "Energy-Band Structure of Germanium and Silicon: The k·p Method," *Phys. Rev.* **142** (Feb, 1966) 530–543.
https://link.aps.org/doi/10.1103/PhysRev.142.530.

[52] B. Bruyneel, P. Reiter, and G. Pascovici, "Characterization of large volume {HPGe} detectors. Part I: Electron and hole mobility parameterization," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **569** no. 3, (2006) 764 – 773.
http://www.sciencedirect.com/science/article/pii/S0168900206015166.

[53] M. I. Nathan, "Anisotropy of the Conductivity of *n*-Type Germanium at High Electric Fields," *Phys. Rev.* **130** (Jun, 1963) 2201–2204.
https://link.aps.org/doi/10.1103/PhysRev.130.2201.

[54] L. Reggiani, "The anisotropy of the hot-hole drift velocity in Ge," *Journal of Physics and Chemistry of Solids* **37** no. 3, (1976) 293 – 300.
http://www.sciencedirect.com/science/article/pii/0022369776900901.

[55] M. Martini and T. McMath, "Trapping and detrapping effects in lithium-drifted germanium and silicon detectors," *Nuclear Instruments and Methods* **79** no. 2, (1970) 259 – 276.
http://www.sciencedirect.com/science/article/pii/0029554X70901497.

[56] S. Zimmermann, J. T. Anderson, D. Doering, J. Joseph, C. Lionberger, T. Stezelberger, and H. Yaver, "Implementation and Performance of the Electronics and Computing System of the Gamma Ray Energy Tracking In-Beam Nuclear Array (GRETINA)," *IEEE Transactions on Nuclear Science* **59** no. 5, (Oct, 2012) 2494–2500.

[57] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. https://books.google.com/books?id=ZXL6AQAAQBAJ.

[58] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995.
https://books.google.com/books?id=TRXrMWY_i2IC.

[59] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika* **57** no. 1, (1970) 97–109.
http://www.jstor.org/stable/2334940.

[60] B. J. Brewer, L. B. Pártay, and G. Csányi, "Diffusive nested sampling," *Statistics and Computing* **21** no. 4, (2011) 649–656.
http://dx.doi.org/10.1007/s11222-010-9198-8.

[61] J. Skilling, "Nested sampling for general Bayesian computation," *Bayesian Anal.* **1** no. 4, (12, 2006) 833–859. `http://dx.doi.org/10.1214/06-BA127`.

[62] D. F.-M. Brendon J. Brewer, "DNest4: Diffusive Nested Sampling in C++ and Python," `arXiv:1606.03757 [stat.CO]`.

[63] E. Gabriel, G. E. Fagg, G. Bosilca, T. Angskun, J. J. Dongarra, J. M. Squyres, V. Sahay, P. Kambadur, B. Barrett, A. Lumsdaine, R. H. Castain, D. J. Daniel, R. L. Graham, and T. S. Woodall, "Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation," in *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pp. 97–104. Budapest, Hungary, September, 2004.

[64] N. Abgrall *et al.*, "The Majorana Demonstrator calibration system," `arXiv:1702.02466 [physics.ins-det]`.