Quasi-Background-Free Neutrinoless Double-Beta Decay Searches with LEGEND: Statistical Methods and Cryogenic SiPM Characterization

Samuel Borden

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington
2025

Reading Committee:
Jason Detwiler, Chair
Alvaro Chavarria
Vincenzo Cirigliano

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

Quasi-Background-Free Neutrinoless Double-Beta Decay Searches with LEGEND: Statistical Methods and Cryogenic SiPM Characterization

Samuel Borden

Chair of the Supervisory Committee:

Jason Detwiler
Department of Physics

As an electrically neutral and massive fermion, the neutrino is the only Standard Model particle whose Lagrangian mass term could include a Majorana term. The nature of the neutrino mass is exciting because it is inherently Beyond the Standard Model: neutrino oscillations show that the neutrino has a mass and that it violates conservation of individual lepton numbers, in direct contradiction to predictions of the Standard Model. The most sensitive probe of the Majorana nature of the neutrino is the search for a hypothetical second-order weak decay called neutrinoless double-beta decay. Observation of this decay would prove that conservation of total baryon minus lepton number is violated and that the neutrino has a Majorana mass term.

The Large Enriched Germanium Experiment for Neutrinoless-double beta Decay (LEGEND) collaboration is searching for the neutrinoless double-beta decay of ^{76}Ge by deploying an array of highly enriched germanium detectors inside of a liquid argon cryostat. In order to fully cover the allowed parameter space for inverted ordered light Majorana neutrinos, the LEGEND collaboration is pursuing a phased approach. The already-built LEGEND-200 experiment aims for a half-life discovery sensitivity of 10^{27} years using 200 kg of enriched detectors, and the proposed ton-scale follow-on experiment LEGEND-1000 aims for a discovery sensitivity of 10^{28} years. The success of these experiments rests in their great ability to reduce external background through techniques such as pulse shape discrimination and liquid argon scintillation

anti-coincidence vetoing.

The LEGEND-200 experiment has completed its first year of searching for neutrinoless double-beta decay with an accumulated exposure of 61 kg·yr. This thesis reports the first frequentist statistical analysis performed on data from the LEGEND experiment. A new Python-based framework, `freqfit`, is introduced to robustly and rapidly produce frequentist statistical inference on unbinned data. Using this framework, no evidence of neutrinoless double-beta decay is found ($p = 0.1$), and a lower limit on the half-life is placed at $T^{0\nu}_{1/2} > 5 \times 10^{25}$ years (90% confidence level). The background index for the group of detectors mainly comprised of the geometry to be used in LEGEND-1000 is computed to be $5^{+3}_{-2} \times 10^{-4}$ counts/keV/kg/yr. A frequentist joint analysis incorporating LEGEND-200 data and the data from two recent ^{76}Ge experiments, the MAJORANA DEMONSTRATOR (MJD) and Germanium Detector Array (GERDA), is also presented. The combined analysis also observes no evidence for a signal ($p = 0.29$) and results in the strongest lower limit, and highest half-life sensitivity, on the half-life in ^{76}Ge to-date: $T^{0\nu}_{1/2} > 1.9 \times 10^{26}$ years (90% confidence level). This half-life limit can be converted into a limit of the effective Majorana mass of the neutrino, assuming mediation by a light Majorana neutrino and a range of phenomenological nuclear matrix elements, and yields $m_{\beta\beta} < 75 - 200$ meV (90% confidence level).

With fewer than one expected count due to accidental radioactive background near the Q-value of neutrinoless double-beta decay, LEGEND-1000 will be a quasi-background-free experiment. The behavior of the frequentist profile likelihood ratio treatment for statistical inference on unbinned data from quasi-background-free experiments is investigated. It is found that test statistic distributions in this regime deviate from the asymptotic form predicted by Wilks' theorem — it is required to generate pseudoexperiments for these statistical analyses. The coverage is computed for ensembles of pseudoexperiments generated with a known Poisson-distributed background index. We show that the profile likelihood ratio treatment guarantees coverage very close to the nominal value for these ensembles.

In order for LEGEND-1000 to reach its nominal background index goal of 1×10^{-5} counts/keV/kg/yr, it is necessary that its liquid argon detector collects as much scintillation light in its silicon photomultiplier (SiPM) readout as possible. The detection efficiency of the liquid argon system is directly proportional to the photon detection efficiency (PDE) of the SiPMs. The PDE is a well-known quantity reported by the manufacturer, *at room temperature*; however, LEGEND operates its SiPMs at cryogenic temperature.

This work describes the operation and results of a cryogenic SiPM characterization test stand built at the University of Washington. We report that two SiPMs exhibit a nearly 20% drop in their PDE at liquid nitrogen temperature relative to their room temperature values. This drop was measured for the green wavelengths (562 nm) of light that match the optical emission spectra of light collection technology for LEGEND-1000. This drop in the PDE represents an important input for forecasting the background index for LEGEND-1000 in order to guarantee that it reaches its discovery sensitivity goal.

Acknowledgements

They say it takes a village, and what a village I have. I would first like to thank the members of my reading committee, Alvaro, Jason, and Vincenzo, for their insightful feedback and comments on my dissertation.

I am deeply indebted to the mentorship and support of Jason. I could not have asked for a better PI during graduate school. Your kindness, patience, and overwhelming enthusiasm (and aptitude!) for physics are everything one could hope for in a mentor. You are the kind of teacher that everyone should aspire to be. I am especially thankful that you gave me the freedom to pursue the topics that interested me, along with some gentle guidance, during my time as a student. I will cherish the time I spent in your group, and deeply miss it.

I would like to acknowledge the support from my colleagues in the LEGEND collaboration: Julieta, Matteo, Ian, Erin, Brady, George, and Jita, to name a few. Thank you for being such friendly and brilliant collaborators — you make LEGEND an exceptional experiment to work on. Special thanks to Danielle Schaper for being an amazing scientist and the perfect co-worker for a snowy January on-site, and to Florian Henkes for numerous helpful analysis discussions and teaching me how to play Schafkopf.

The mentorship I have received under the two postdocs I worked with in the LEGEND lab at UW has been an invaluable experience that I would not trade for anything. Clint, thank you for teaching me (among a lot of other things) Wiseman's law, which I repeat here in full: if you think it's plugged in, and have check that it is plugged in, then no, it is not plugged in. Louis, your depth of expertise and the clarity with which you share it are qualities I aspire to emulate. Thank you for guiding me through the lab with such patience, and for teaching me all the statistics that I know! Your genuine kindness, steady leadership, and thoughtful mentorship pushed me to grow into the best researcher I could be. I'm deeply grateful for everything you've taught me and for the care you put into helping me succeed.

Teachers rock. Thank you especially to April Elmore and Brian Brewer for kindling my passion for science at an early age. And to Benjamin, Brett, and Trent — thank you for expanding my Weltanschauung. Witold Skiba and Steve Sharpe both deserve kudos for excellently teaching some of the best courses I have ever taken as a student. I am also extremely grateful to Dave Moore: thank you for guiding me through both the graduate application process and my first foray into physics research.

Having entered my PhD program at the height of the global coronavirus pandemic in 2020, I am also eternally thankful for the friendships I formed throughout my time at UW. Kent Wilson, Murali Saravanan, Eli Lilleskov, Maddy Olson, Gillian Shen, and Nikita Zemlevkiy, thanks for listening to records with me. I would specifically like to thank Eric Lester, who I first met at the 2019 DNP conference, for being my oldest friend in Seattle and for helping me get through hard times.

The Center for Nuclear Physics and Astrophysics (CENPA) is the best place to work, hands down. I am deeply grateful for the institution and all that it represents. Alejandro, thank you for always having your door open to me and for answering my questions. Claudio, our time working together was too short, I could have learned so much more from you; you are an excellent and compassionate teacher, and your SiPM wizardry is impressive — you are a delightful part of CENPA and the LEGEND group at UW. Thanks to Elise and Dave for giving me excellent career advice. And thanks to Jens for always keeping the lunch table discussion lively. Thank you to Tim, David, Ryan, Nate, Brittany, Mason, Clint, Gary, and the rest of CENPA engineering and staff for being the beating heart of CENPA. The resources and community CENPA has make it unparalleled; I am lucky to have worked here.

The bulk of my time as a graduate student has been spent at the blue picnic table in front of CENPA. So, thank you to the lunch crew for passing the time with me: CJ Nave, Grace Song, Jimmy Sinnis, and Harry Ni (and honorarily Winston DeGraw). Thank you to all the senior graduate students who welcomed our cohort into CENPA: Nick, Alex, Heather, Brynn, Drew, Josh, and many others. A big shout out to some of the wonderful and brilliant undergraduates I have met at CENPA: Ian (good luck with the SiPMs at UNC!), Forrest, and Nicholas. CJ and Grace, I hope that I can find coworkers who are as good as you. I'll really miss you; you were the only people who could make three-hour group meetings worth it.

Finally, thank you to my family. To my mom and dad, thank you for reading so much to me as a child, and for your endless love and support. To my brother Luke and sister Sarah, thank you for putting up with

me. And last, but not least, thank you to my partner Ellis, for always picking me up when I'm down and for inspiring me to be a better scientist: I aspire to your level of dedication, intelligence, and compassion. And thanks to my cat Minnie.

Thank you to anyone who I have forgot to mention here!

DEDICATION

To my family and friends

"But to the likes of employees such as Roger Mexico it is music, not without its majesty, this power series $Ne^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\cdots+\frac{m^{n-1}}{(n-1)!}\right)$, terms numbered according to rocketfalls per square, the Poisson dispensation ruling not only these annihilations no man can run from, but also calvary accidents, blood counts, radioactive decay, number of wars per year . . . "

— Thomas Pynchon, *Gravity's Rainbow*

Contents

List of Figures

List of Tables

List of Acronyms

$0\nu\beta\beta$ neutrinoless double-beta. 20, 26, 45, 46, 50–52, 54, 55, 57, 59, 61, 66, 71, 100, 113, 115, 125, 169, 178

$2\nu\beta\beta$ two-neutrino double-beta. 44, 46, 52–54, 59, 71, 118, 120, 169

AC anti-coincidence. 94

ADC analog-to-digital converter. 95, 103–105, 170, 205

AP afterpulsing. 226

APD avalanche photodiode. 234

BEGe broad energy germanium. 75, 83, 106, 115–117, 120

BSM Beyond the Standard Model. 25–27, 41, 43, 45, 125

CDF cumulative distribution function. 144

ckky counts/keV/kg/yr. 80

CL confidence level. 143, 150, 152, 156, 164–167, 181

CSA charge sensitive amplifier. 68

CUPID CUORE with Particle Identification. 178

DAQ data acquisition. 83, 94–97, 112, 211

DCR dark count rate. 197, 200, 201, 208, 211, 219, 226

DCR delayed charge recovery. 79, 197

DeCT delayed crosstalk. 226

DEP double escape peak. 100, 103, 104, 107–111, 115

DiCT direct crosstalk. 226

DSP digital signal processing. 98

FBK Fondazione Bruno Kessler. 236, 238

FEP full energy peak. 96, 100, 105–107

FWHM full width at half maximum. 84

GERDA Germanium Detector Array. 2, 16, 17, 59, 75, 80, 84, 86, 103, 107, 110, 114, 120, 150, 151, 169–171, 183, 237, 239

HPGe high purity germanium. 14, 15, 59, 79–85, 88, 90–92, 95, 97, 113, 116

IC inverted coaxial. 76, 78, 89, 90, 120

ICPC inverted coaxial p-type point-contact. 76, 83, 102, 104, 106, 107, 114, 116, 117, 120, 160

IR infrared. 186, 238

irreps irreducible representations. 30

IV current-vs-voltage. 189, 229

LAr liquid argon. 80, 83, 84, 91, 92, 112, 113, 137, 184, 185, 197, 219, 236

LED light emitting diode. 208, 220–222, 224, 225, 230–233

LEGEND Large Enriched Germanium Experiment for Neutrinoless-double beta Decay. 1–3, 14–17, 20, 59, 60, 66, 68, 69, 74, 76, 78–83, 85–95, 102, 114, 115, 118–122, 127, 132, 135–138, 140, 143, 145–150, 152, 154–163, 165–185, 190, 197, 216, 220, 235–240

LN liquid nitrogen. 221, 229

LNGS Laboratori Nazionali del Gran Sasso. 80, 84, 90, 92

LQ late charge. 79, 106, 109

MJD MAJORANA DEMONSTRATOR. 2, 59, 75, 79, 80, 102, 103, 110, 111, 114, 150, 170, 237

ML machine learning. 105, 111

MLE maximum likelihood estimator. 123

MSE multi-site event. 73, 75–77, 89, 100, 107–109, 111

NIR near-infrared. 202

NME nuclear matrix element. 47–50, 54, 58, 89, 159, 171, 172

PDE photon detection efficiency. 2, 185, 193–195, 202, 203, 206, 207, 219–225, 227, 229–236, 238, 239

PDF probability density function. 17, 123, 127, 128, 131, 135, 139, 140, 148–150, 153, 154, 160, 163, 164, 168, 175–178

PEN polyethylene naphthalate. 86

PHS pulse height spectrum. 190, 191

PMT photomultiplier tube. 84, 95, 97, 100, 112

PPC p-type point contact. 75, 83, 106, 117, 120, 152, 168

PSD pulse shape discrimination. 62, 89, 92, 94, 97, 98, 106, 109–111, 115, 117, 118, 120, 137, 162, 169, 172, 236

QDC (SiPM) charge spectrum. 191

RMS root-mean square. 98, 100, 105, 212

SEP single escape peak. 100, 103, 104, 108

SF survival fraction. 108

SiPM silicon photomultiplier. 2, 3, 18, 19, 85, 86, 94, 95, 97, 100, 112, 113, 184–186, 188, 189, 191–197, 200–204, 206, 207, 211–225, 228–236, 238

SM Standard Model. 35

SNR signal-to-noise ratio. 211

SPAD single-photon avalanche diode. 186–190, 202, 203, 226, 227

SRH Shockley-Read-Hall. 198–200

SSE single-site event. 73, 75–77, 100, 107

TAUP Topics in Astrophysics and Underground Physics. 119

TPB tetraphenyl butadiene. 85, 86, 184

TPC Time Projection Chamber. 53, 55

TTL transistor-transistor logic. 95

UV ultraviolet. 186, 187, 196, 234, 235, 238

VUV vacuum ultraviolet. 84–86, 184–186, 233

WLS wavelength-shifting. 184, 220

WSLR wavelength-shifting light reflector. 85, 86

Chapter 1

Introduction

1.1 Motivation

In the Standard Model the neutrino is massless: this model input was based on decades of research that showed that this was a well-justified assumption. However, neutrino oscillation experiments have shown that the neutrino does in fact have a finite mass — or, at least two of the three mass eigenstates have a non-zero mass. This means that there is physics beyond the Standard Model responsible for the generation of the neutrino mass.

The simplest way to incorporate neutrino mass is through the same Higgs mechanism that gives the charged leptons and quarks their mass. However, due to the radically different size of the Yukawa coupling compared to other Standard Model couplings, this mechanism is suspicious due to the fine-tuning required to generate such a small neutrino mass. In this scenario, with only the Higgs generating the mass, the neutrino is called a **Dirac neutrino** and the total lepton number L is conserved: the neutrino is a distinct particle from the antineutrino.

Another possibility for the neutrino exists. If the total lepton number L is not a symmetry of nature, and the neutrino is indistinguishable from the antineutrino, then the neutrino is known as a **Majorana neutrino**. There exist many different Beyond the Standard Model (BSM) mechanisms that can generate a Majorana neutrino mass. The simplest of these mass-generation mechanisms is the only dimension-5 $SU(2)_L \times U(1)_Y$ invariant operator, and this operator can naturally generate a small neutrino mass through what is known as

the see-saw mechanism. Due to its lack of electromagnetic charge, the neutrino is the only known fermion that conceivably could have a Majorana mass component. The Majorana model of the neutrino requires that only two field components are present instead of the four required by Dirac neutrinos; the economy of this description is buttressed by the experimental fact that only neutrino field components have been observed in Nature so far. Furthermore, the neutrino mass already breaks conservation of individual lepton number L_i — as seen in neutrino oscillations — so it would be interesting if it also broke conservation of the total lepton number L, as well as baryon minus lepton number $B - L$, which is the last remaining accidental symmetry of the Standard Model. It is very plausible that the neutrino is a Majorana particle.

It is of fundamental importance then to determine if the neutrino is a Dirac or Majorana particle: doing so can give hints to the Beyond the Standard Model mechanism that generates the neutrino mass. The most sensitive experimental probe of the Majorana nature of the neutrino is the search for the rare nuclear decay called neutrinoless double-beta ($0\nu\beta\beta$) decay. The Standard Model predicts that, for nuclei in which double-beta decay is allowed, two anti-neutrinos are emitted during the process; however if the neutrino is its own antiparticle, there is a possibility for Beyond the Standard Model physics to allow the decay to occur without the emission of any neutrinos — hence the name "neutrinoless" double-beta decay. However, it should be noted that there exist many BSM processes that can give rise to $0\nu\beta\beta$ decay (without generating a Majorana mass for the neutrino). Thus, if the decay is observed, multiple future experiments would need to be performed to determine the mechanism behind the decay before the Majorana nature of the neutrino is confirmed or refuted.

Even so, observation of neutrinoless double-beta decay would be evidence that Nature places matter and antimatter on different footing: leptons would be created without the creation of antileptons. It would confirm that not only total lepton number L, but also total baryon minus lepton number, $B - L$, are not conserved quantities — the last remaining global symmetries of the Standard Model would be broken by observation of this decay. Because $B - L$ violation is a necessary requirement for leptogenesis models, many models incorporate or predict $0\nu\beta\beta$ decay. Leptogenesis is a leading model that can explain the observed matter-antimatter asymmetry in our current universe. Furthermore, if $0\nu\beta\beta$ decay is mediated by the so-called light Majorana neutrino exchange, then observations of the decay can help give information on the the absolute mass of the neutrino and the neutrino mass ordering.

The search for neutrinoless double-beta decay is an exciting experimental search that can open doors into new BSM physics and help us understand the origin of our matter-dominated universe. It is also devilishly tricky: lower bounds on the half-life are greater than 10^{26} years (a quadrillion times longer than the age of the universe) [1, 2]. The search for $0\nu\beta\beta$ decay has driven progress in novel detector technologies and external radioactive background reduction techniques. With the landscape of experimental progress opening up to ton-scale experiments probing even longer half-lives, the field is burgeoning with optimism and desire for discovery. Now is the age of the neutrino.

1.2 History of the neutrino

1.2.1 Beta decay: from postulation to the discovery of the neutrino

Neutrinos were originally postulated by Pauli in 1930 [3] in order to solve spin-statistics discrepancies in certain nuclei and to explain the continuous nature of beta-decay spectra. At the time, it was believed that beta decay was a two-body decay, in which the nucleus emits an electron and leaves a final state nucleus. However, experiments had showed that the beta decay energy spectrum was continuous, and that the total energy of the emitted electron was less than the available energy of the decay [4]. This caused Bohr to hypothesize that energy was not conserved during beta-decay (!) and that the energy loss of the electron followed a statistical description [1] [6]. Pauli's particle restored energy conservation by making beta decay a three-body decay; the neutrino carries away some of the available energy, and this makes the beta energy spectrum continuous.

Fermi successfully incorporated this new neutral particle into his theory of beta decay and named it the *neutrino*. Because of its spin-1/2 nature, the neutrino is thus a fermion. Fermi's theory was extremely successful at predicting the shape of beta decay curves; however, skepticism about the existence of the neutrino, a required part of Fermi's model, persisted. In 1946 Pontecorvo suggested a direct detection mechanism through inverse beta-decay [7]. In this process an incoming antineutrino is absorbed by a proton, converting into a neutron and positron: $\bar{\nu} + p \rightarrow n + e^{+}$. In 1956 Cowan and Reines first detected reactor antineutrinos through inverse beta decay, thus confirming the existence of Pauli's ghost particle [8]. Immediately,

[1] It is interesting to note that Ellis and Wooster also floated this idea in one of their papers, but concluded that "an explanation of this type would only be justified when everything else has failed." [5]

scientists set out to measure the mass of this new particle. Limits on the neutrino mass were set through experiments that measured the beta decay endpoint. The limits on the neutrino mass, m_ν, were so much smaller than other elementary particles that m_ν was often taken to be zero [6][9].

After subsequent decades of research — and numerous missteps (including the "discovery" of a 17 keV sterile neutrino! [10]) — the story was mostly set in stone. Three flavors of left-handed neutrinos and right-handed antineutrinos had been observed, and the neutrino was believed to be massless. This was all codified into the Standard Model. However, at the turn of the millennium, the belief that neutrinos are massless was about to be broken.

1.2.2 Neutrino oscillation

After the discovery of strangeness oscillations in neutral kaons, Pontecorvo began considering if similar oscillations were also possible in neutrinos in 1957 [11] and in 1958 [12]. The possibility of neutrino oscillation lies in the assumption that the neutrinos we observe, ν_e, ν_μ, ν_τ, are the eigenstates of the weak interaction, and can be expressed as a linear combination of different mass eigenstates. Analogous to mixing in the quark sector, a mixing matrix can be used to describe this phenomenon, and for neutrinos it is called the PNMS matrix and is denoted by U [13]. It is typical for three mass eigenstates to be included in the definition of the mixing matrix, but there is no good reason to *a priori* assume that there are not more mass eigenstates. The term "sterile neutrino" is used to refer to weak interaction eigenstates that are beyond the three current generations. In this three-weak and three-mass-eigenstate model the weak eigenstates can be expressed by

$$|\nu_l\rangle = \sum_i^3 U_{il} \, |\nu_l\rangle \, . \tag{1.1}$$

For the three known neutrino weak eigenstates, the PMNS matrix U is a 3×3 unitary matrix (the determinant must be equal to 1 to conserve probability); as such, it can be parameterized by 9 degrees of freedom. It is typical to write these in terms of 3 Euler angles called $\theta_{12}, \theta_{13}, \theta_{23}$ because these are the parameters that experiments can measure. The remaining 6 degrees of freedom are phases, and three of them can be absorbed into definitions of the fields [13]. The 3 phases left over are δ_{CP}, the measure of CP-violation in the mixing, and α and β are called the Majorana phases. Using the notation $s_{ij} = \sin\theta_{ij}$,

$c_{ij} = \cos\theta_{ij}$ the PMNS matrix can be written as:

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{CP}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{CP}} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} e^{i\alpha} & 0 & 0 \\ 0 & e^{i\beta} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (1.2)$$

The mixing angles θ_{ij} are all measured from various experiments. Future experiments are needed to determine δ_{CP}, α, and β.

Neutrino oscillations were first observed in both the Super-Kamiokande and SNO experiments that used solar neutrinos and atmospheric neutrinos, respectively, as sources. Due to matter effects in the sun, it turns out that solar neutrino experiments are sensitive to the mixing angle θ_{12}, as well as the mass difference Δm^2_{21}, where $\Delta m_{ij} \equiv m_i^2 - m_j^2$. The discovery of solar neutrino oscillation showed that the neutrino has a mass. This fact will be important later. It was also determined that $\Delta m^2_{21} > 0$. Atmospheric neutrino oscillation experiments have measured a different mass splitting $|\Delta m^2_{23}| > \Delta m^2_{12}$. Unfortunately, current neutrino oscillation experiments are not sensitive enough to determine the sign of Δm^2_{23}. This uncertainty in sign leads to what is known as the neutrino mass ordering problem.

If $\Delta m^2_{23} > 0$, then the splittings between the mass eigenstates give what is known as the **inverted ordering** where $m_2 > m_3$, and the lightest neutrino is ν_3, the neutrino mass eigenstate with the smallest admixture of electron neutrino flavor. If $\Delta m^2_{23} < 0$, then the splittings between the mass eigenstates give what is known as the **normal ordering** where $m_2 < m_3$ and the lightest neutrino is ν_1, the neutrino mass eigenstate with the largest admixture of electron flavor. Figure 1.1 shows the two allowed mass orderings. It is unknown which ordering describes the neutrino.

Future measurements of both the sign of Δm^2_{23} and the absolute value of the lightest neutrino mass will complete our knowledge of the neutrino and its flavor oscillations. The neutrino mass can be measured directly through beta decay: knowledge of the decay's kinematics can be used to solve for the mass [15]. But what is important is that neutrino oscillations show that **neutrinos have mass**. Because neutrinos have mass, and are also electrically neutral, this means that they can be either a Majorana or a Dirac particle. Let us now explore what that means.

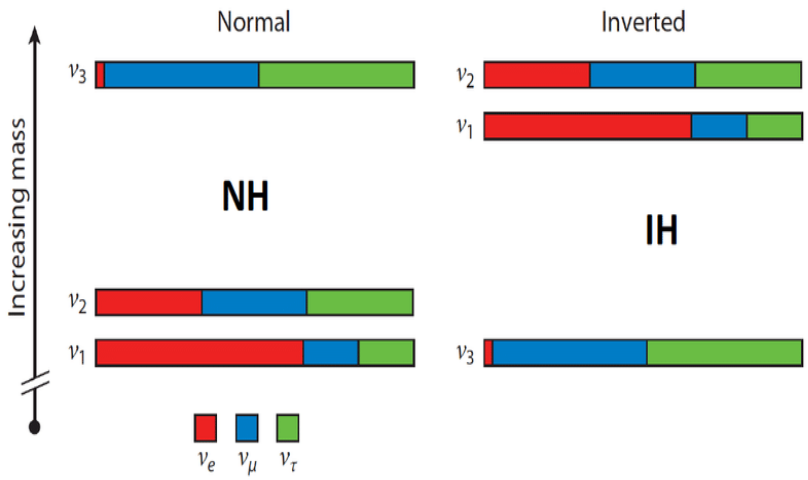

Figure 1.1: Possible orderings of neutrino mass eigenstates based on current experimental understanding. Picture from Ref. [14].

1.3 Majorana vs. Dirac

1.3.1 The Dirac equation

The Dirac equation is the equation of motion for spin-1/2 particles, called fermions, with mass [16]. Neutrinos, as just described in section 1.2.2, are massive fermions. In terms of the γ matrices, the Dirac equation can be expressed as

$$(i\gamma^\mu \partial_\mu - m)\psi = 0. \tag{1.3}$$

Because ψ must represent an object that is Lorentz invariant in four space-time dimensions, the wave function ψ must necessarily be a 4-component object and is called a **Dirac spinor**. It is common to express γ^μ in terms of the Pauli-Dirac basis. However, there are numerous irreducible representations (irreps) of the Lorentz group $SO(3,1)$, so the γ matrices can take on different forms. A unitary change of basis can transform the Pauli-Dirac gamma matrices into the so-called **Weyl basis**. In the Weyl representation of the γ matrices, the Dirac spinor is easily broken down into two 2-component objects called bispinors:

$$\psi = \begin{pmatrix} \xi \\ \chi \end{pmatrix}. \tag{1.4}$$

The decomposition of the Dirac spinor into bispinors is done because these two bispinors transform

differently under a Lorentz boost. To see this, consider the operator γ^5 as represented in the Weyl basis:

$$\gamma^5 = \begin{pmatrix} -\mathbb{1}_{2x2} & 0 \\ 0 & \mathbb{1}_{2x2} \end{pmatrix}. \tag{1.5}$$

The eigenstates of γ^5 with eigenvalue -1 are called **left-chiral** states, while those with eigenvalue $+1$ are called **right-chiral** states. We immediately see that the choice of the Weyl basis means that the Dirac spinor is written in terms of two definite-chirality bispinors:

$$\gamma^5 \begin{pmatrix} \xi \\ 0 \end{pmatrix} = - \begin{pmatrix} \xi \\ 0 \end{pmatrix} \implies \psi_L = \begin{pmatrix} \xi \\ 0 \end{pmatrix}, \tag{1.6}$$

$$\gamma^5 \begin{pmatrix} 0 \\ \chi \end{pmatrix} = + \begin{pmatrix} 0 \\ \chi \end{pmatrix} \implies \psi_R = \begin{pmatrix} 0 \\ \chi \end{pmatrix}. \tag{1.7}$$

It is important to note that for Standard Model particles, any particles' field can be written in terms of eigenstates of γ^5 and thus decomposed into fields of definite chirality. Chiral states transform uniquely under charge conjugation. If we introduce the charge conjugation operator $C = i\gamma^2\gamma^0$, then the charge conjugate of a Dirac spinor is given by

$$\psi^C = i\gamma^2\psi^*. \tag{1.8}$$

Therefore we can see that the charge conjugate of left-handed spinors are right-handed objects $(\psi_L)^C = (\psi^C)_R$, and vice-versa for right-handed spinors $(\psi_R)^C = (\psi^C)_L$.

1.3.2 Majorana fermions

In 1939 Majorana searched for solutions to the Dirac equation which required fewer than the 4 components that other solutions required. He found that this condition could be achieved if he required the field to be its own charge conjugate [17]. This is known as the Majorana condition, that a particle given by the 4-component Dirac spinor θ is its own anti-particle (up to a phase):

$$\theta^C = \pm\theta \tag{1.9}$$

To see how this reduces the number of degrees of freedom for the spinor from 4 to 2, introduce the following two **Majorana fields** with **Majorana spinors**:

$$\omega = \psi_L + (\psi_L)^C = \begin{pmatrix} \xi \\ i\sigma_2\xi^* \end{pmatrix}, \tag{1.10}$$

$$\rho = \psi_R - (\psi_R)^C = \begin{pmatrix} i\sigma_2\chi^* \\ \chi \end{pmatrix}. \tag{1.11}$$

These fields are their own charge conjugates, and their spinors show that only one bispinor with two degrees of freedom is necessary to describe the field. Contrast this with how a run-of-the-mill Dirac spinor transforms under charge conjugation:

$$\psi^C = i\gamma^2 \begin{pmatrix} \xi \\ \chi \end{pmatrix} = \begin{pmatrix} -i\sigma_2\chi \\ i\sigma_2\xi \end{pmatrix} \neq \psi. \tag{1.12}$$

Majorana spinors are therefore fundamentally different than Dirac spinors. If a field satisfies Eqn. 1.9 it is a **Majorana field**, and if it doesn't, it is a **Dirac field**. Majorana particles are their own antiparticles, while Dirac antiparticles are distinct from their particles.

In a quantum field theory, the Lagrangian density is a more convenient quantity than the equation of motion. Let us now consider the Lagrangian density that gives rise to the Dirac equation:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \tag{1.13}$$

where the term $\overline{\psi}$ is defined as $\overline{\psi} = \psi^\dagger\gamma^0$, so that $\overline{\psi}\psi$ is a Lorentz-invariant object. The first term is the kinetic term, and the second is the mass term. If we decompose this mass Lagrangian density in terms of the left and right chiral fields of the Dirac spinor then we get:

Dirac mass term

$$\mathcal{L}_D = -m\overline{\psi}\psi = -m(\overline{\psi_L + \psi_R})(\psi_L + \psi_R) = -m(\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L), \tag{1.14}$$

where we have used the fact that $\overline{\psi}_L\psi_L = \overline{\psi}_R\psi_R = 0$ because γ^0 is off-diagonal in the Weyl basis [16][18].

Now, what if ψ was a Majorana spinor, without loss of generality say ω? Plugging the field ω into the Dirac equation Lagrangian:

$$\mathcal{L}_m = -m\overline{\omega}\omega = -m\overline{(\psi_L + \psi_L^C)}(\psi_L + \psi_L^C) = -m(\overline{(\psi_L)^C}\psi_L + \overline{\psi}_L(\psi_L)^C), \tag{1.15}$$

which follows from the fact that $(\psi_L)^C$ behaves like a right-handed object, as we have seen, and thus any term coupling right-to-right or left-to-left drops out. At first glance, this looks like a great and valid mass term. However, because this is quadratic in ψ_L, a factor of $1/2$ must be introduced to ensure that the Dirac equation of motion can be recovered from the Lagrangian, as the Lagrangian is quadratic in terms of ψ_L [13, 19]. This overall term is called a **Majorana mass term**:

Majorana mass term

$$\mathcal{L}_{Majorana} = -\frac{m}{2}(\overline{(\psi_L)^C}\psi_L + \overline{\psi}_L(\psi_L)^C). \tag{1.16}$$

Because ω and ρ are in principle independent Majorana spinors, both of them can be present in the Lagrangian, so the most general Majorana mass term can be written in terms of Majorana spinors, or equivalently, in terms of the chiral components of the underlying field ψ:

$$\mathcal{L}_{Majorana} = -\frac{m}{2}(\overline{(\psi_L)^C}\psi_L + \overline{\psi}_L(\psi_L)^C) - \frac{m}{2}(\overline{(\psi_R)^C}\psi_R + \overline{\psi}_R(\psi_R)^C). \tag{1.17}$$

This is called a Majorana mass term because any field with a term like this in its Lagrangian will satisfy the Majorana condition (Eqn. 1.9) once the Lagrangian is diagonalized into the mass eigenstate basis, as we will see eventually.

Because the Majorana mass term couples particles to antiparticles, if the field ψ described a charged particle, then the Lagrangian would allow for transitions like $l^+ \leftrightarrow l^-$ which would break the $U(1)_Y$ symmetry responsible for the conservation of electric charge. Therefore only electrically neutral fermions could be Majorana in nature, and the only neutral fermion in the Standard Model is the neutrino. As a mass term, Eqn. 1.17 also requires, well, a non-zero mass. The question is: are neutrinos Dirac or Majorana fermions?

1.4 Neutrinos in the Standard Model

Within the Standard Model, there are three flavors of neutrino that are massless and that do not mix [20][13]. The neutrinos are placed in a left-handed isospin doublet with members of the same flavor generation; right-handed neutrino singlets are absent from the theory. To understand how the neutrinos are massless in the Standard Model, it is instructional to see how other lepton masses are generated through the Higgs mechanism.

The Higgs mechanism presupposes an $SU(2)$ doublet ϕ with two complex scalar fields that can be expressed as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.18}$$

before spontaneous symmetry breaking. The Higgs doublet can then be expressed in the unitary gauge after spontaneous symmetry breaking as

$$\phi = \begin{pmatrix} 0 \\ \frac{h+v}{\sqrt{2}} \end{pmatrix}, \tag{1.19}$$

where h is the Higgs field and v is its vacuum expectation value [21]. The $SU(2)_L$ isospin doublets into which leptons are placed can be expressed as $L = \begin{pmatrix} (\nu_e)_L \\ e_L \end{pmatrix}$, and the other leptons that are not neutrinos are placed in $SU(2)_L$ singlets, $R = e_R$.

In order to form a Lagrangian from these terms, the Lagrangian needs to respect the electroweak $SU(2)_L \times U(1)_Y$ local gauge symmetry before spontaneous symmetry breaking. The combination $\overline{L}\phi R$ is

$SU(2)_L$ invariant. Thus an $SU(2)_L \times U(1)_Y$ invariant Lagrangian before spontaneous symmetry breaking can be written as Yukawa interaction terms given by

$$\mathcal{L} = -g_e[\overline{L}\phi R + \overline{R}\phi^\dagger L], \tag{1.20}$$

where g_e is the coupling strength. After spontaneous symmetry breaking when the Higgs doublet acquires a vacuum expectation value and can be expressed by Eqn. 1.19, if we scale the coupling as $g_e = \sqrt{2}\frac{m_e}{v}$ where m_e is the invariant mass of the electron, then we can rewrite the Lagrangian as

$$\mathcal{L} = -m_e\overline{e}e - \frac{m_e}{v}\overline{e}eh. \tag{1.21}$$

The first term can be identified as the term that generates the mass of the electron, and the second term is the electron's direct coupling to the Higgs boson.

Importantly, we see that neutrinos drop out in this model; they obtain no mass because ν_R is not present in a singlet. The choice of omitting ν_R from the Standard Model was predicated on the fact that neutrinos did not have an observable mass, and that only left-handed neutrinos and right-handed anti-neutrinos were observed in experiments.

1.4.1 Higgs generation of Dirac neutrino Mass

We now know that neutrinos have a finite rest mass from oscillation experiments, so the Standard Model no longer accurately describes the true nature of the neutrino. In principle, the neutrino mass can also be generated by the exact same Higgs mechanism. For simplicity, we shall only consider a world where one flavor eigenstate is present. All that is required as a further input to the Standard Model is the presence of the ν_R singlet. This additional particle content would keep the gauge invariance, and thus symmetry group, of the SM [22]. If we add ν_R then we get terms in the Lagrangian like Eqn. 1.20, but with ϕ_C (the charge conjugate of ϕ) replacing ϕ. Expressed in the unitary gauge, this equation becomes

$$\phi_C = -i\sigma_2\phi = -\frac{1}{\sqrt{2}}\begin{pmatrix} h+v \\ 0 \end{pmatrix}, \tag{1.22}$$

$$\implies \mathcal{L} = -\frac{g_\nu v}{\sqrt{2}}(\overline{\nu}_L\nu_R + \overline{\nu}_R\nu_L) - \frac{g_\nu h}{\sqrt{2}}(\overline{\nu}_L\nu_R + \overline{\nu}_R\nu_L). \tag{1.23}$$

The relationship between the left- and right-handed fields interacting with the Higgs to give a mass can be seen diagrammatically in the second row in Fig. 1.2. Now, if we set the coupling to the bare neutrino mass $g_\nu = \sqrt{2}\frac{m_\nu}{v}$ we get:

$$\mathcal{L} = -m_\nu\overline{\nu}\nu - \frac{m_\nu}{v}\overline{\nu}\nu h. \tag{1.24}$$

Now we can identify the term in the Lagrangian that generates the neutrino mass, and in turn call this the **Dirac mass term** in the Lagrangian:

$$\mathcal{L}_D = -m_D\overline{\nu}\nu. \tag{1.25}$$

We can see that there is a conserved quantum number by considering the action of an arbitrary $U(1)$ transformation in which case the fields transform like $\nu_L \to e^{i\alpha}\nu_L, \nu_R \to e^{i\alpha}\nu_R$,

$$\implies \mathcal{L}' = -m_\nu(e^{i\alpha}e^{-i\alpha}\overline{\nu}_L\nu_R + e^{i\alpha}e^{-i\alpha}\overline{\nu}_R\nu_L) = -m_\nu\overline{\nu}\nu = \mathcal{L}. \tag{1.26}$$

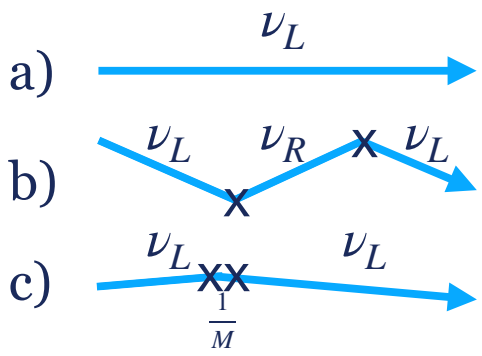

Figure 1.2: Pictorial depiction of different mass generation mechanisms for the neutrino, based on a figure from Ref. [23]. Subplot a) shows the Standard Model neutrino which has no mass and does not interact with the Higgs. Subplot b) shows a totally Dirac neutrino that inherits its mass from the Higgs mechanism, which is shown by the kinks in the line. Subplot c) shows a Majorana neutrino with both Dirac and Majorana mass terms; the mass is generated from the Weinberg operator, where the left-handed neutrino is given its small mass from the presence of a massive right-handed mostly sterile neutrino.

Therefore there is a $U(1)$ gauge symmetry, and by Noether's theorem there is a conserved current: in this case, the conserved current is called the total lepton number. Because the lepton number is conserved, the neutrino is distinct from the antineutrino, and thus the neutrino is called a **Dirac** particle. This argument was only presented for one flavor, but it can easily be extended to all three flavors.

However, something is afoot. The magnitude of the Yukawa coupling deserves a closer inspection. Although the mass of the neutrino is not known, there are upper bounds on the neutrino mass from cosmological studies as well as direct mass measurements. Based on a recent KATRIN result showing that $m_\nu < 0.9$ eV for the electron neutrino, the estimated Yukawa coupling becomes at most [15]

$$g_\nu \leq \sqrt{2}\frac{0.9 \text{ eV}}{246 \text{ GeV}} = 10^{-13}. \tag{1.27}$$

Compared to the other Yukawa couplings in the standard model which are all between $\mathcal{O}(1e^{-6}) - \mathcal{O}(1)$, this is extremely small[13]! The smallness of this parameter is somewhat unpalatable. In addition, the generation of the Dirac mass through the Higgs mechanism requires the addition of two whole new bispinor fields: both ν_R and its charge conjugate, the latter of which is a left-handed antineutrino. This would put the total number of neutrino fields at 4 in the Standard Model; however, as discussed before, only two fields have been observed to date in Nature. Moreover, both the ν_R and its charge conjugate would be sterile fields because they would not couple to the W and Z bosons. This has given enough reason to search for other Beyond the Standard Model processes that could generate the neutrino mass in a perhaps more "natural" way.

1.5 Generation of neutrino masses through Majorana mass terms

Because the neutrino carries no electric charge, it can potentially be endowed with a Majorana mass component. By enforcing the Majorana condition that $\nu^C = \nu$ we can show how Majorana mass components can perhaps more "naturally" generate the mass of the neutrino.

1.5.1 Left-handed Majorana mass term: Type-II see-saw

Recall that the charge-conjugation of a left-handed field is a right-handed field. Because the right-handed neutrino is absent from the Standard Model, it is more natural to first introduce a Majorana mass term with the left-handed neutrino. Writing out the Majorana mass term Lagrangian:

$$\mathcal{L}_{m_L} = -\frac{1}{2} m_L (\overline{\nu}_L (\nu_L)^C + h.c.). \tag{1.28}$$

Expressing this in terms of a mass matrix:

$$\mathcal{L}_{m_L} = \begin{pmatrix} \overline{\nu}_L & \overline{(\nu_L)^C} \end{pmatrix} \begin{pmatrix} 0 & m_L \\ m_L & 0 \end{pmatrix} \begin{pmatrix} \nu_L \\ (\nu_L)^C \end{pmatrix}. \tag{1.29}$$

Diagonalization of this mass matrix leads to the mass eigenstate $\nu = \nu_L + (\nu_L)^C$, which satisfies the Majorana condition $\nu^C = \nu_L + (\nu_L)^C = \nu$. From use of the chiral projection operators P_L, P_R, we conclude that $\nu_R = (\nu_L)^C = (\nu^C)_R$, and therefore the right-handed and left-handed parts of Majorana fermions are connected. Taking a closer look at the mass Lagrangian we can determine more interesting features of this left-handed chiral Lagrangian:

$$\mathcal{L}_{m_L} = -m_L (\overline{\nu}_L (\nu_L)^C + h.c.). \tag{1.30}$$

This operator annihilates a left-handed antineutrino and creates a left-handed neutrino, and is called the *negative chiral Majorana mass term*. But, a problem can be seen because ν_L is in a weak hypercharge isospin doublet: mass terms of this type do not conserve hypercharge and violate the $SU(2)_L \times U(1)_Y$ symmetry of the Standard Model [19]. In order for this term to conserve hypercharge, an $SU(2)$ triplet scalar Higgs would need to be present in the Standard Model — in this case, this Lagrangian can still generate a neutrino mass through what is known as the type-II see-saw mechanism if a Dirac mass term is also present in the Lagrangian [13].

Also note that under the $U(1)$ transformation the Lagrangian is no longer invariant:

$$\mathcal{L} \to \mathcal{L}' = -m_L (e^{-2i\alpha} \overline{\nu}_L (\nu_L)^C + h.c.). \tag{1.31}$$

We see that the lepton number that was conserved in the Dirac mass Lagrangian is violated by two units.

1.5.2 Right-handed Majorana mass term: Type-I see-saw

Instead of adding a left-handed Majorana mass term and a scalar triplet, it can more economical to instead enforce the Majorana condition using the right-handed neutrino:

$$\nu = \nu_R + (\nu_R)^C. \tag{1.32}$$

And form a mass term:

$$\mathcal{L}_{m_R} = -m_R(\overline{\nu}_R(\nu_R)^C + h.c.). \tag{1.33}$$

This term is a-ok to add to the Standard Model because ν_R is in an $SU(2)_L$ singlet and thus has no associated gauge charges; in fact, because it is in a singlet, we are free to choose its mass [18, 24]. This is known as the *positive chiral Majorana mass term*.

In general, if the left and right chiral fields are independent, then the most general mass Lagrangian that could be built is the sum of Dirac, positive-chiral Majorana, and negative-chiral Majorana mass terms. Here we are abstracting what it means to be a Majorana mass term. Because the charge conjugation operator flips the chirality of a particle, it is allowable to build a mass term out of the contraction of the field with its charge conjugate: this is called a Majorana mass term. In this scenario, we abandon the easier-to-understand Majorana notion that $\nu_X = (\nu_X)^C$ and simply create these mathematical structures called the Majorana mass terms that are formed from the chiral field and its conjugate. We will eventually show that the presence of terms like this in the Lagrangian makes it so that the physical fields — the mass eigenstates — satisfy the broader Majorana condition that $\nu^C = \nu$. Back to the most general Lagrangian:

$$\mathcal{L}_{D+M} = -m_D(\overline{\nu}_R\nu_L + \overline{\nu}_L\nu_R) - m_R(\overline{\nu_R}(\nu_R)^C + h.c.) - m_L(\overline{\nu}_L(\nu_L)^C + h.c.), \tag{1.34}$$

where $h.c.$ is the Hermitian conjugate. Which can be rewritten as

$$\mathcal{L} = \begin{pmatrix} \overline{\nu}_L & \overline{(\nu_R)^C} \end{pmatrix} \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix} \begin{pmatrix} (\nu_L)^C \\ \nu_R \end{pmatrix} + h.c. \tag{1.35}$$

In order to respect the gauge symmetry of the Standard Model we let $m_L \to 0$. This is analogous to what happens for $K^0 - \overline{K}^0$ oscillations: because the mass matrix is not diagonal, the fields ν_L, ν_R are not physical mass eigenstates, but are instead linear combinations of actual mass eigenstates. The goal is then to diagonalize the matrix. Using the assumption that $\frac{m_D}{m_R} \ll 1$ we can use the matrix

$$\Lambda = \begin{pmatrix} 1 & -\frac{m_D}{m_R} \\ \frac{m_D}{m_R} & 1 \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}, \tag{1.36}$$

to diagonalize the mass matrix. The second matrix comes from the requirement that we want our diagonal matrix to have positive eigenvalues [25][13]. In this case the diagonal mass matrix becomes

$$M \approx \begin{pmatrix} \frac{m_D^2}{m_R} & 0 \\ 0 & m_R \end{pmatrix}. \tag{1.37}$$

The eigenstates of the original mass matrix, ν, N , can be expressed by using the diagonalization matrix:

$$\begin{pmatrix} \nu \\ N \end{pmatrix} = \begin{pmatrix} 1 & -\frac{m_D}{m_R} \\ \frac{m_D}{m_R} & 1 \end{pmatrix} \begin{pmatrix} \nu_L \\ (\nu_R)^C \end{pmatrix}, \tag{1.38}$$

where the phases have been reabsorbed into the definitions of the fields. Performing the multiplication gives:

$$\nu = \nu_L + \frac{m_D}{m_R}(\nu_R)^C, \tag{1.39}$$

$$N = (\nu_R)^C - \frac{m_D}{m_R}\nu_L. \tag{1.40}$$

But so far we have ignored the Hermitian conjugate part of the initial Lagrangian. Adding its terms back in gives:

$$\nu = \left(\nu_L + (\nu_L)^C\right) + \frac{m_D}{m_R}\left(\nu_R + (\nu)_R^C\right), \tag{1.41}$$

$$N = \left(\nu_R + (\nu)_R^C\right) - \frac{m_D}{m_R}\left(\nu_L + (\nu_L)^C\right). \tag{1.42}$$

Because $\frac{m_D}{m_R} \ll 1$, we see that the field ν in Eqn. 1.41 can be considered as being mostly left-handed, while N in Eqn. 1.42 is mostly right-handed. The overall Lagrangian can then be rewritten as:

$$\mathcal{L}_{D+M} = -\frac{m_D^2}{m_R}\overline{\nu}\nu - m_R\overline{N}N, \tag{1.43}$$

where N is a massive fermion. We see that if the Majorana mass m_R is extremely large, then the mass of the ν field is lowered; this is known as the type-I see-saw mechanism [18]. The type-I see-saw gives the interacting neutrinos a small mass by introducing a mostly sterile heavy neutrino N with mass m_R. The type-I see-saw is perhaps a more natural mass generation mechanism because the Dirac masses of the neutrinos m_D can then be set on the same scale as other charge leptons $m_D \approx 1$ MeV, and then when the mostly right-handed neutrino N is added to the theory, its mass can be set arbitrarily large so that $\frac{m_D^2}{m_R}$ is approximately the same order as the observed neutrino masses. In particular, setting $m_R \sim E_{\mathrm{GUT}}$ — where E_{GUT} is the energy scale where we expect to see new physics — gives that $m_\nu \sim \mathcal{O}(10\text{ meV})$ [26]. The lighter mass eigenstate ν is mostly comprised of the left-handed weak eigenstates, and the heavy state N is comprised mostly of the right-handed singlet state ν_R. A pictorial description of this see-saw mechanism is shown in the third row of Fig. 1.2.

Even more importantly, these mass eigenstates are their own charge conjugates $N^C = N, \nu^C = \nu$. This would even be true in the case where $m_L \neq 0$; as long as a Majorana term of the form $\nu^C\nu$ is present in the Lagrangian, then the physical mass states satisfy the Majorana condition. See [25] for the derivation of this. The original Dirac neutrino is split into two Majorana neutrinos. It is important to note that even if the Lagrangian has a Dirac mass term, as long as it has a Majorana mass term the neutrino is a Majorana fermion. This is an important consideration when there exist possibilities for the two Majorana phases to almost completely destructively interfere with each other so that the neutrino would technically be Majorana in nature, but it would behave almost entirely like a Dirac neutrino.

1.6 Why else look for a Majorana neutrino?

As discussed above, the Majorana nature of the neutrino could hint at a BSM mechanism behind the neutrino mass. Neutrinos are the only charge-free Standard Model particle, and as such, are the only candidate Majorana particles in the Standard Model. It is thus important to determine their fundamental nature as an input to an extended Standard Model. In addition, if the neutrino is a Majorana particle, then lepton number

could be violated by two units, and $B-L$ symmetry would no longer hold because the baryon number is not affected by neutrinos [22]. Violation of $B-L$ would be an important discovery because it would remove the last global symmetry of the Standard Model.

1.6.1 Leptogenesis

The violation of the global $B-L$ symmetry could also play an important part in understanding how our universe became to be matter dominated. Cosmological models of the Big Bang predict that matter and antimatter should be present in equal ratios [2]; the universe we inhabit, however, is mostly made up of matter. The asymmetry between matter and antimatter can be parameterized by η, the baryon asymmetry [20]:

$$\eta = \frac{n_B - n_{\overline{B}}}{n_\gamma} \approx 6 \times 10^{-10}. \tag{1.44}$$

The question is: how did the universe come to be so staggeringly matter dominated? One possible way is for a lepton asymmetry to exist in the early Universe, and then Standard Model processes converted the lepton asymmetry into a baryon asymmetry; this process is known as leptogenesis. Sakharov postulated three conditions necessary for baryon asymmetry through *baryogenesis* [27]:

Sakharov's conditions for baryogenesis

1. **The presence of a baryon number violating process**: this is necessary in order to create baryons without antibaryons
2. **CP violation**: so that the baryon number violating process favors the production of baryons over antibaryons
3. **Departure from thermal equilibrium**: so that creation of baryons occurs more often than the process that annihilates baryons.

Due to sphaleron processes that exist within predictions of the Standard Model, an excess of antileptons can be converted into an excess of baryons [18]. As such, these three conditions can be re-cast in terms of

[2] If baryon asymmetry were an initial condition, then inflation would have diluted it away [20]. Furthermore, even if such an initial condition *could* work, it would need to be extremely fine tuned to explain the current level of asymmetry.

processes involving leptons; however, the lepton number violating process and CP violation must favor the creation of antileptons over the creation of leptons. The first condition for leptogenesis requires the presence of a lepton number violating process. There are many different mechanisms that could generate a lepton number violation; luckily, we have seen that the presence of a Majorana mass term fundamentally violates lepton number conservation. In models of leptogenesis through the type-I see-saw, such as the one Fukugita and Yanagida proposed [18], the heavy Majorana neutrino preferentially decays into anti-leptons and creates an excess. In general, there are other BSM processes that can provide a lepton number violating process, and many of them involve Majorana neutrinos [18][28]. But here we have seen a simple way that incorporating just heavy Majorana neutrinos alone could give rise to such a process.

1.6.2 Why the neutrino may Be Majorana

In short, here is a summary of reasons described in prior sections as to why the neutrino may be a Majorana particle:

Why look for Majorana neutrinos?

1. Neutrinos are massive, and a Dirac mass term alone requires the Yukawa coupling to the Higgs to be extremely small and is potentially unpalatable. And even if right-handed neutrino singlets are added to the Standard Model, there is no symmetry that prevents a Majorana mass term.

2. The generation of a Majorana mass is possible through a minimal extension of the Standard Model and incorporates the only lowest dimensional non-renormalizable operator feasible. And because this operator is the lowest order, it is naively the first non-renormalizable BSM physics we would have access to. The Majorana mass term can naturally give rise to the small neutrino mass, as seen through the effective Weinberg operator.

3. We have only seen left-handed neutrinos and right-handed anti-neutrinos, it would be economical if the neutrino was a Majorana fermion to explain why we have only seen these two components. The see-saw mechanism for the generation of the Majorana neutrino mass can naturally explain this, see Eqn. 1.41, as well as why the neutrino mass is so small.

4. Majorana neutrinos allow for $\Delta L = \pm 2$ processes. Similar processes in the early universe can give rise to baryogenesis and help explain the antimatter-matter asymmetry observable in the universe.

1.7 Neutrinoless double-beta decay

The most sensitive probe of the Majorana nature of the neutrino is the search for neutrinoless double-beta decay [22][13][25][3]. The ability for such a decay to probe the Majorana neutrino was pointed out by Furry in 1939 [29], which provided the first calculations of the decay rate (Furry also pointed out that searching for a Majorana neutrino would be difficult in 1938 [30]), building on the work of Goeppert-Mayer's proposal of two-neutrino double-beta ($2\nu\beta\beta$) decay in 1935 [31], which is an allowed Standard Model nuclear decay, and is only observable for only a handful of isotopes. Double-beta decay is only observable for certain even-even nuclei where the nuclear pairing force forbids a transition through a single beta decay (the new proton produced has no other neutron in the same shell with which to pair up, so the ground state energy increases) but allows a transition through two simultaneous beta decays (so that the two protons produced can pair up in the same shell) to a nucleon where the ground-state energy is lower than the original nucleus. Within the Standard Model, these two beta decays emit two electrons and two electron antineutrinos:

Two neutrino double-beta decay

$$X(A, Z) \rightarrow X(A, Z+2) + 2e^{-} + 2\nu^{C} \tag{1.45}$$

Because this process involves two beta decays, it is second order in the weak interaction coupling and therefore has an extremely long half-life. $2\nu\beta\beta$ decay was first observed (in a lab) by Elliott *et al.* in Selenium-82 in 1987, with a half-life of 10^{20} years [32]. In general, most $2\nu\beta\beta$ decay half-lives observed to-date fall in the range of $10^{18} - 10^{21}$ years.

If the neutrino is a Majorana particle, then it is possible for double-beta decay to occur without detection of any neutrinos from the nucleus: the neutrino emitted by one of the nucleons can interact with the other

[3] For a good reference explaining why it is difficult to look for Majorana neutrinos using accelerator-produced beams of neutrinos, see Ref. [24]

neutrino before leaving the nucleus [4], leading to the process:

Neutrinoless double-beta decay

$$X(A, Z) \rightarrow X(A, Z + 2) + 2e^{-}. \tag{1.46}$$

This process is called neutrinoless double-beta ($0\nu\beta\beta$) decay. A small caveat: light Majorana neutrinos might not be the main mechanism behind $0\nu\beta\beta$ decay, there could be other BSM physics that causes $0\nu\beta\beta$ decay. In general, any BSM lepton number violating process that gives rise to operators of odd dimension ≥ 5 will generate $0\nu\beta\beta$ decay [22, 24]. But, luckily any BSM physics that drives the $\Delta L = 2$ process in $0\nu\beta\beta$ decay will generate a Majorana mass term for the neutrino in what is known as the Schechter-Valle [33] or black-box theorem [13]. Observation of neutrinoless double beta decay directly implies that the neutrino is a Majorana fermion. However, it could be that certain BSM physics drives $0\nu\beta\beta$ decay but leaves the Majorana mass of the neutrino extremely small — small enough to the point that the neutrino behaves almost entirely like a Dirac fermion. Thus, if $0\nu\beta\beta$ decay is observed, future experiments would need to be performed to determine the actual mechanism driving the reaction. But, regardless if the neutrino is a Majorana fermion or not, the observation of $0\nu\beta\beta$ decay would demonstrate the violation of $B - L$. Such an observation would lend credence to leptogenesis and thus is important to search for $0\nu\beta\beta$ decay regardless of the underlying mechanism.

1.7.1 Light neutrino exchange and the half-life: why is it so long?

The most common and well-motivated BSM physics behind $0\nu\beta\beta$ decay is the case where the process is mediated by the exchange of light, Majorana neutrinos. In this scenario, a dimension-5 operator is present and dominant to endow the neutrinos with a Majorana mass term (as in the see-saw mechanism!) so that they are their own antiparticle, and to drive down the mass of the Standard Model neutrino [22][25][20]. The rest of the vertices in the interaction remain Standard Model vertices. The Feynman diagram for this process is shown in Fig. 1.3. In this picture, the neutrino emitted from one neutron interacts with the W^{-} boson from the other neutron. In order for this decay to proceed through the weak interaction vertex, the neutrino

[4] It is perhaps more correct to say that the neutrino emitted at one vertex interacts as an antineutrino with the other weak vertex and that the whole process is virtual and that no "real" (detectable) neutrinos are ever emitted at all.

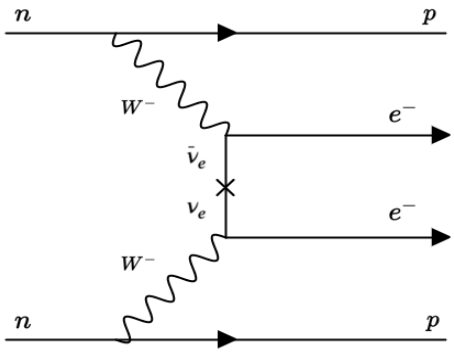

Figure 1.3: Feynman diagram for light neutrino mediated neutrinoless double-beta decay. The neutrinos remain virtual, and all the vertices are Standard Model vertices.

needs to have mass in order to flip its chirality (and oscillation experiments show that it does have mass), and it must be its own antiparticle. So we see that, again, the $0\nu\beta\beta$ decay rate is suppressed also by $\frac{m}{E}$ in light neutrino exchange due to the chirality flip. However, the reason that $0\nu\beta\beta$ decay is the most sensitive probe of the Majorana nature of the neutrino is that its dominant background is from $2\nu\beta\beta$ decays — which is also a rare decay — so that there is a much lower background than in an accelerator experiment [24]. The question is: how often does this $0\nu\beta\beta$ decay occur?

Answer: on at least very, very, *very* long time scales. Neutrinoless double-beta decay is a second-order weak interaction, so its matrix element carries a factor of G_F^4 which suppresses the decay rate. As derived (very cleanly) in Ref. [24] , the decay rate can be expressed as a product of the phase space factor $G_{0\nu}$, the nuclear matrix element M, and the effective Majorana mass $m_{\beta\beta}$ [18][25][22]:

Neutrinoless double-beta decay half-rate

$$\Gamma^{0\nu}_{1/2} = G_{0\nu}|M_{0\nu}|^2 m^2_{\beta\beta}. \tag{1.47}$$

Let's break all of these pieces down and rhapsodize why the decay rate ends up so small.

Phase space factor

The phase space factor $G_{0\nu}$ accounts for the weak interaction couplings and the kinematics of the decay. Intuitively we may expect that the larger the Q value is, the more likely the decay will occur due to entropy

(more final states available). This is encoded in the phase space factor, which can be calculated to show that

$$G_{0\nu} \propto Q_{\beta\beta}^5$$

although, this ignores any Coulomb interactions between the daughter nucleus and the outgoing electrons. So the actual phase space calculations depend on both $Q_{\beta\beta}$ and Z.

Nuclear matrix elements

Because of the complexity of the nucleus (the decaying nucleons interact strongly with many different non-participating nucleons), the nuclear matrix element (NME) is not as well known as the phase space factor. To make matter worse, the virtual neutrino that is emitted requires one to integrate over all possible momenta in the loop. Thus calculations require computation of the many-body wavefunctions of all the accessible intermediate nuclear states — a very computationally intensive task. Calculation of the NME depends strongly on what model of the nucleus is being used and how the calculations are being performed. Because the nuclear matrix element is, well, a matrix element, it involves an overlap between the final and initial nuclear states; the complex interaction of the nucleons makes it necessary to use a model to calculate these initial and final wavefunctions. Comparison of current values of $M_{0\nu}$ for different isotopes and different calculation methods show a spread of a factor of 3 in some cases. Hence, there is a large uncertainty in the actual value of the NME. A short description of four phenomenological many-body methods to calculate the NMEs follow [28][34][20]:

Summary of common phenomenological many-body approaches to calculating the NME

- **Interacting Shell Model:** In this model, the nucleus is modeled by a harmonic oscillator with strong spin-orbit couplings. Nuclear levels follow a shell structure, much like the placement of electrons in the atomic shell model. In these calculations, only the nucleon-nucleon interactions in the shell near the Fermi surface are considered. This yields a Hamiltonian that can be numerically diagonalized for the wavefunctions.

- **Quasiparticle Random Phase Approximation:** The nuclear pairing force is added to this model using Cooper-like pairs of neutron-neutrons and proton-protons via BCS theory. The quasiparticle part of the name comes from the transformation into a basis where nucleons are partially particles and partially holes. The transition amplitude is then computed in this basis where a large number of single-particle orbits but few correlations are included.

- **Energy Density Functional Theory:** This approach uses multiple mean fields, and this approximation reduces that of many interacting particles to non-interacting particles in a mean-field. Minimization of the energy density functional calculated from these mean fields yields the nuclear ground state.

- **Interacting Boson Model:** Intermediate between QRPA and ISM, this model treats the lowest energy shells as inert and treats nucleon pairs as bosons. One- and two-body interactions between these bosons are placed into a Hamiltonian that is then diagonalized to obtain the wavefunctions.

These phenomenological methods rely on use of measurements from experiments instead of deriving the NME directly from first principles using QCD. Common empirical data include single beta decay rates, transition strengths and other nuclear data. Oftentimes these models are tuned to reproduce observed experimental results — one way of doing this is to use effective operators and effective interactions. This is most commonly done for some computational methods such as the interacting shell model by introducing "quenching" for the axial-vector coupling constant g_A, where its value is reduced in order to reproduce two neutrino double-beta decay half-rates [35]. The quenching of the axial vector is often justified by appealing to complex interactions that are not well-modeled occurring in the nuclear medium.

Regardless of what method is used to calculate the matrix element, its uncertainty is large.

New methods and recent research: recent approaches using chiral effective field theory have made several advances in calculating the nuclear matrix element. These approaches treat the QCD Lagrangian as an effective field theory, so that the nuclear potential can be calculated from this Lagrangian. QCD is typically non-perturbative, so calculations are extremely difficult; in the chiral effective field theory approach, at low energy the Lagrangian is treated as having a chiral $SU(2)_L \times SU(2)_R$ symmetry, and this allows

for perturbative calculations in the chiral energy scale of the nuclear potential [35]. These methods that use the QCD Lagrangian from first principles are known as *ab-initio* methods for computing the nuclear matrix element. In the development of this procedure, it was realized in Ref. [36] that an important contribution to the nuclear matrix element was missing in most calculations: calculations neglected interactions from high-momentum (on the order of 1 GeV), short-range neutrinos. This short range interaction, also called the contact term, has a potentially large impact on the computed nuclear matrix element because it is also a leading-order term in the calculation.

Because *ab-initio* methods are computationally expensive, and the complexity scales with the number of nucleons, it has only been recently that these methods have produced nuclear matrix elements for neutrinoless double-beta decay isotopes. The NME for germanium-76 was computed using *ab-initio* methods in Ref. [37] — this results represents the first nuclear matrix element calculation that has fully-quantified theoretical uncertainty.

Effective Majorana mass

Finally, we come to the rate killer. Because weak-eigenstate neutrinos are known to be a superposition of mass eigenstates, the amplitude must then be proportional to a sum of the contributions of these mass eigenstates weighted by their mixing into the flavor eigenstate. This leads to the **effective Majorana mass** of the neutrino:

$$m_{\beta\beta} = \left| \sum_{i=1}^{3} m_i U_{ei}^2 \right|, \tag{1.48}$$

$$m_{\beta\beta} = \left| m_1 c_{12}^2 c_{13}^2 e^{2i\alpha} + m_2 s_{12}^2 c_{13}^2 e^{2i\beta} + m_3 s_{13}^2 e \right|. \tag{1.49}$$

where U_{ei} are the electron flavor entries in the PMNS matrix. The reason it enters as squared is that there are two vertices for the virtual neutrino [5] [13][24]. The terms m_i are the masses of the 3 mass eigenstates. The reason this is called the effective Majorana mass is that it also contains a dependence on the two Majorana phases α, β as well as on the values of the neutrino mass eigenstates. A non-zero measurement of the effective Majorana mass would give some information on the mass of the electron neutrino.

[5] Or, another way to think about it is that this equation is just re-describing $\nu_e = c_1\nu_1 + c_2\nu_2 + c_3\nu_3$, and the constants that we put in the mixing matrix are just the squares of these constants c_i.

The reason that the rate of neutrinoless double-beta decay would be slow is partially because of the smallness of $m_{\beta\beta}$. The mass of the neutrino is just so small. To see the dependence of $m_{\beta\beta}$ on the lightest neutrino mass, m_1, also called $m_{lightest}$, the expression can be rewritten

$$m_{\beta\beta} = m_1 c_{12}^2 c_{13}^2 e^{2i\alpha} + \sqrt{m_1^2 + \Delta m_{12}^2} s_{12}^2 c_{13}^2 e^{2i\beta} + \sqrt{m_1^2 \pm |\Delta m_{23}^2|} s_{13}^2 \tag{1.50}$$

Because the lightest neutrino mass is bounded by the current limits on the electron neutrino mass from direct measurements, $m_{\beta\beta}$ works out to be extremely tiny. Current neutrinoless double-beta decay experiments have shown that $m_{\beta\beta} < 28 - 122$ meV [1]. This corresponds to a limit on the half-life of neutrinoless double-beta decay of $T_{1/2}^{0\nu} > 3.8 \times 10^{26}$ years in ^{136}Xe.

1.7.2 What would we learn if we saw neutrinoless double-beta decay?

Regardless of the mechanism behind it, observation of $0\nu\beta\beta$ decay would be the first known lepton number violating process. This would demonstrate a violation of the last accidental global symmetry of the Standard Model ($B - L$) and could have important implications for the baryon asymmetry of the universe. If the mechanism behind it can be determined to come from light neutrino exchange, more information about the nature of the neutrino could be gleaned.

If neutrinoless double-beta decay is discovered, and if the uncertainties on the NMEs were reduced, then the decay rate could be used to calculate the effective Majorana mass $m_{\beta\beta}$. As seen from Eqn. 1.50, $m_{\beta\beta}$ is sensitive to the lightest neutrino mass as well as four other unknown parameters: the sign of the mass difference between m_2 and m_3, the Majorana phases, and the CP violating phase. The sign of the mass splitting determines the ordering of the neutrino masses, as discussed above in section 1.2.2. Based on the two choices for this sign, we can then vary the remaining parameters and report $m_{\beta\beta}$ as a function of m_1. This results in what is known as the "lobster claw plot" as seen in Fig. 1.4.

Next-generation experiments will be sensitive to half-lives that correspond to just beyond the lower limit for the inverted ordering. Non-observation of neutrinoless double-beta decay at the half-life level of $10^{27} - 10^{28}$ years, depending on the isotope, would thus rule out a Majorana neutrino in the inverted ordering scenario. This doesn't exclude the neutrino from being Dirac and in the inverted ordering though. Plotting the $m_{\beta\beta}$ value for a specific half-life limit in Fig. 1.4 shows this inverted ordering exclusion, as

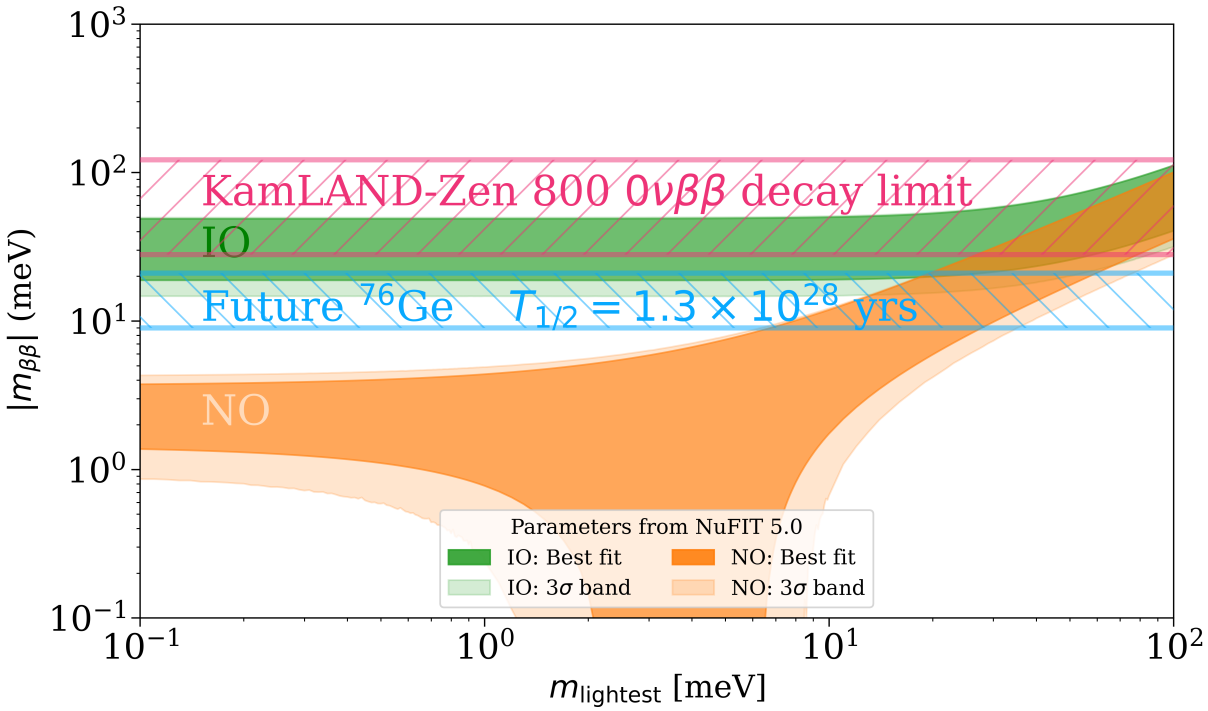

Figure 1.4: The famous "lobster claw plot" that shows the allowed phase space for $m_{\beta\beta}$ given different values of $m_{lightest}$, allowing the unknown parameters to vary maximally, plotted using software from Ref. [38]. Modern neutrinoless double-beta decay experiments have scratched the top of the flat band of the inverted ordering. Future ton scale experiments hope to place a bound on $m_{\beta\beta}$ below the inverted ordering, as shown by the blue line for a specific half-life limit in germanium-76.

well as the importance of the interplay between various types of neutrino experiments. Future neutrino oscillation experiments can constrain δ_{CP} and lower the width of the bands in the lobster plot [39]. Direct measurements of the neutrino mass m_β through tritium decays — such as searches performed with magnetic spectrometers in Ref. [15] or with cyclotron radiation emission spectroscopy in Ref. [40] — can also help constrain the parameter space. An interesting outcome of measurements of m_β might measure m_1 to a point where Eqn. 1.50 and Eqn. 1.47 would predict a half-life shorter than the upper bound placed by experiments. Thus the non-observation of $0\nu\beta\beta$ decay would imply that the neutrino is a (dominantly) Dirac particle[6]. Furthermore, limits on the sum of the neutrino masses from cosmological measurements can also help constrain the phase space for $m_{\beta\beta}$ [41].

It is currently unknown which of the two neutrino mass orderings describe our physical reality, because

[6]That is, assuming no competing mechanisms interfere destructively with light neutrino exchange

current-generation neutrino oscillation experiments are largely insensitive to the sign of the Δm^2_{32} mass splitting [42]. However, experiments in the near future will hopefully determine the mass ordering. One such experiment is the Jiangmen Underground Neutrino Observatory (JUNO). This experiment plans to resolve the mass ordering by measuring reactor antineutrino survival probabilities precisely — and across a wide spectrum — to observe an interference pattern caused by quasi-vacuum neutrino oscillations; this method is independent of the parameters δ_{CP} and θ_{23}, which are fit parameters that are degenerate with the mass ordering for current generation neutrino oscillation experiments [43]. A confirmation of the inverted ordering would be the most exciting prospect, as next generation neutrinoless double-beta decay experiments will be sensitive to discoveries in this parameter space. A positive measurement of the normal ordering, however, would leave a larger parameter space open at much smaller effective Majorana masses. The situation is not hopeless though, because a cosmological measurement of the sum of the neutrino masses would set a lower bound on $m_{\beta\beta}$ and thus greatly reduce the available parameter space, even in the case of the normal ordering [22].

1.8 How to build a neutrinoless double-beta decay experiment

1.8.1 Guiding principles

If it is important to look for neutrinoless double-beta decay, then how would we detect it? What is the experimental signature of neutrinoless double-beta decay? Because $0\nu\beta\beta$ decay is a three-body decay, the final state energy $Q_{\beta\beta}$ is shared between the daughter nuclei and the two outgoing electrons. However, the recoil energy of the nucleus is small compared to σ_E, the energy resolution of the experiment, and thus the outgoing electrons must carry away all of the energy $Q_{\beta\beta}$. This means that if the sum of the energies of the outgoing electrons were measured, the result would be $Q_{\beta\beta}$. This is in contrast to $2\nu\beta\beta$ decay which is a 5-body decay: the outgoing neutrinos carry away some of the available energy according to some probability distribution. Thus measurement of the summed electron energies in the $2\nu\beta\beta$ decay case will result in a continuum. Figure 1.5 shows an example of the monoenergetic peak at $Q_{\beta\beta}$ as an experimental signature of neutrinoless double-beta decay. In order to discriminate $2\nu\beta\beta$ decay from $0\nu\beta\beta$ decay the observable we need to measure is therefore the summed energies of the outgoing electrons with high resolution.

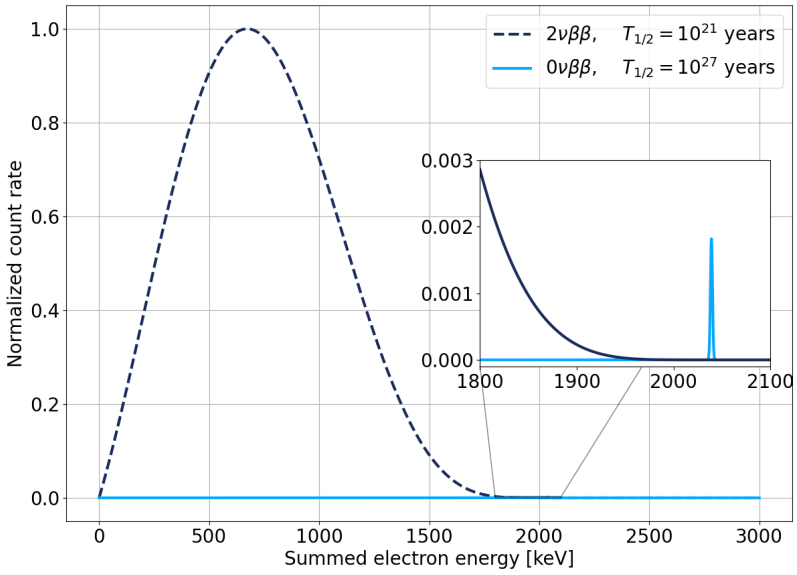

Figure 1.5: A hypothetical spectrum from a neutrinoless double-beta decay experiment showing the summed energy of the electrons versus the count rate. The experimental signature of $0\nu\beta\beta$ is the monoenergetic peak at the $Q_{\beta\beta}$ value of the decay. The $2\nu\beta\beta$ decay can be a substantial background to the $0\nu\beta\beta$ decay.

There are many different experimental techniques to measure the energy spectrum of the final-state electrons. A short table summarizes them and their detection principles below:

Experimental strategies to detect $0\nu\beta\beta$ decay

- Bolometric calorimeters: media containing the decaying isotope are cooled to cryogenic temperatures so that the phonons the escaping electrons produce in the lattice can be measured with sensitive thermometers [44].
- Xenon Time Projection Chambers (TPCs): the decay electrons generate scintillation light that is immediately detected, along with ionization that is drifted to charge collection areas. The drift time can be used to reconstruct the 3D location of where the decay occurred [45].
- Germanium detectors: decay electrons excite electron-hole pairs that are collected to the detector contacts, and the readout charge is proportional to the decay energy [2].

- Large liquid scintillators: scintillation light produced by ionizing radiation in the liquid scintillator is detected, and the amount of collected light is proportional to the event energy [1].
- Amorphous selenium charge-coupled detectors (like Ref. [46]), and so, so many more clever ideas. Too many to list, see Ref. [22] for more!

Regardless of what technique is used to detect the electrons, these experiments are all **counting experiments** and thus share a similar set of experimental challenges and design philosophies. The expected number of detected $0\nu\beta\beta$ decay events, $N_{0\nu\beta\beta}$, in a counting experiment can be expressed in terms of the signal detection efficiency ϵ of the experiment, the mass M of the decay isotope, the run-time t of the experiment, Avogadro's number N_A, the molar mass of the decay isotope $m_{\text{iso.}}$, and $T_{0\nu}^{1/2}$ the half-life of the decay:

$$N_{0\nu\beta\beta} = \ln 2 \frac{N_A M t \epsilon}{m_{\text{iso.}} T_{0\nu}^{1/2}}. \tag{1.51}$$

We would want to detect as many neutrinoless double-beta decay signals as we possibly could, in order to confirm a discovery. Therefore, the experimental quantities we want to optimize for the most ideal experiment are as follows:

Ideal $0\nu\beta\beta$ decay Experiment Shopping List

- A large active mass of the decay isotope. The more active mass, the higher total rate.
- A large $Q_{\beta\beta}$. This is something you have control over by choice of isotope. The larger the Q-value the more above natural radioactive backgrounds the event signal will occur.
- Favorable NME so that Eqn. 1.47 drives down the half-life.
- Long run time. This requires a stable detector.
- Excellent energy resolution. This is so the sharp peak from $0\nu\beta\beta$ decay can be distinguished. This is crucial to separate the mono-energetic $0\nu\beta\beta$ decay peak from the $2\nu\beta\beta$ decay background continuum.
- Excellent background rejection. This is so that background processes that might mimic a $0\nu\beta\beta$

decay signal can be removed from the analysis. This is also important, as we will later see, to allow for fewer observed $0\nu\beta\beta$ decay events to be classified as a discovery.

One way that many experiments have achieved a large active mass with a low background is by using the decay isotope as the detector itself; this reduces the need for extraneous detector technology that increases the radioactive background near the decay sample. Typically this process involves enriching already feasible detector isotopes with decay isotopes — like xenon-136 for liquid xenon TPCs and germanium-76 for germanium semiconductor detectors. This is often called the "source=detector" paradigm. Choice of $Q_{\beta\beta}$ depends on the choice of the isotope; however, isotope choice can also be driven by isotopic abundance and compatibility with detector technologies. This makes selecting an isotope challenging and explains why there are many different experiments using different isotopes — in addition to the fact that we would want to measure the decay rate in various isotopes to help understand the decay mechanism if a discovery was made.

1.8.2 Discovery and exclusion sensitivities

Because the detection mechanisms between different experiments vary so widely, it is useful to use a common metric to quantify these experiments' capabilities to detect neutrinoless double-beta decay and exclude possible half-rates ($\Gamma_{0\nu}^{1/2}$), which is the reciprocal of the half-life [7]. Two such metrics are the median discovery and exclusion sensitivities. The median discovery sensitivity is the number of signal counts (or, equivalently, the signal rate or half-life) for which an experiment has a 50% chance of measuring a signal 3σ above the background; this is a metric that quantifies an experiment's ability to discover $0\nu\beta\beta$ decay. If however, the $0\nu\beta\beta$ decay half-life is larger than the experiment's discovery sensitivity, then the experiment's exclusion sensitivity can be used to characterize the typical bound the experiment would place on the half-life: the exclusion sensitivity is the median 90% confidence level half-life upper limit that an experiment is expected to achieve in the absence of a signal [22]. For a given number of expected background counts n_b within the vicinity of $Q_{\beta\beta}$, these definitions can be used to determine the median number of signal counts n_s that set the discovery or exclusion sensitivity. In general, the sensitivities increase with increased expo-

[7] The decay constant $\Gamma_{0\nu}$ is then $\ln(2)$ times the half-rate.

sure and decrease with increased background. These signal counts can be converted into either the half-life discovery sensitivity or the half-life exclusion sensitivity by using Eqn. 1.51.

Because both the signal and background counts come from (different) radioactive decays, they are described by (independent) Poisson random processes, but an experiment only ever observes their sum $n_s + n_b$ without knowing the true values of either; however, the value of n_b can typically estimated from the spectrum further away from $Q_{\beta\beta}$. The required n_s for the discovery sensitivity depends on n_b: the larger the n_b, the larger n_s must be. An experiment can enter the **background free** regime when the background is so low that a single signal count can qualify as a 3σ discovery. Alternatively, experiments where the expected number of background counts is less than one are in the so-called **quasi-background free** regime; here, experiments have a nearly linear scaling of the discovery sensitivity with exposure and reap nearly all the benefits that a background free experiment would. Fig. 1.6 shows the relationship between half-life discovery sensitivity and the number of background counts: the larger the background, the longer an experiment would need to run to reach a similar half-life sensitivity. It is, therefore, crucial to lower the background as much as possible.

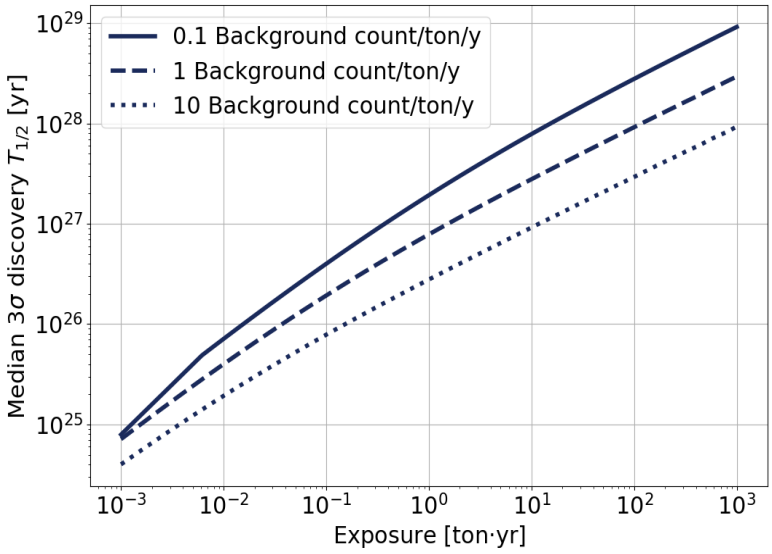

Figure 1.6: The larger the background rate, the longer an experiment needs to run to reach the same half-life discovery sensitivity as a lower background experiment, as this plot shows. The discovery sensitivity relationships as discussed in the above section are used to calculate this specific plot in the case of ^{76}Ge isotope loading.

Experiment	Isotope	$T_{1/2}$ [yr]	$m_{\beta\beta}$ [meV]
LEGEND-200 [2]	^{76}Ge	$> 1.9 \times 10^{26}$	$< 75 - 200$
CUPID-0 [47]	^{82}Se	$> 4.6 \times 10^{24}$	$< 263 - 545$
AMoRE [48]	^{100}Mo	$> 2.9 \times 10^{24}$	$< 210 - 610$
CUORE [44]	^{130}Te	$> 2.2 \times 10^{25}$	$< 90 - 305$
KamLAND-Zen [1]	^{136}Xe	$> 3.8 \times 10^{26}$	$< 28 - 122$

Table 1.1: Current $0\nu\beta\beta$ decay half-life limits for various isotopes.

1.8.3 Current results and future directions

To date, no one has observed neutrinoless double-beta decay [8]. A summary of the lower bounds on the half-life for neutrinoless double-beta decay, and its conversion into a bound on the effective Majorana mass, is shown in Table 1.1. The current best limits come from ^{136}Xe and have a lower bound on the half-life on the order of 10^{26} years [1].

Currently running experiments are using on the order of $\mathcal{O}(100\text{kg})$ of isotope to hit a $10^{26} - 10^{27}$ year sensitivity [22]. As seen from Fig. 1.4, experiments that can reach a half-life sensitivity of $10^{27} - 10^{28}$ years would be able to robustly probe the inverted mass ordering. This makes a $10^{27} - 10^{28}$ year half-life sensitivity a natural goal of next-generation neutrinoless double-beta decay experiments.

The items from List 1.8.1 that can most readily be improved to increase sensitivity are an increased active mass and improved background suppression. Next-generation experiments hope to use an isotopic mass of $\sim$ one to several tons in order to reach the necessary sensitivity to probe the inverted ordering. As the detectors scale up in size, it is crucial to improve background rejection techniques because more active mass will necessarily introduce more background radiation. Fig. 1.7 shows the "experiment racetrack," or a brief summary of planned experiments' sensitivities to half-life and the effective Majorana mass. The community has 10^{28} year half-life sensitivities within its reach; things are heating up in the search for neutrinoless double-beta decay.

[8] There are *claims* of discovery; however, the half-lives of these discoveries have been refuted by numerous subsequent experiments.

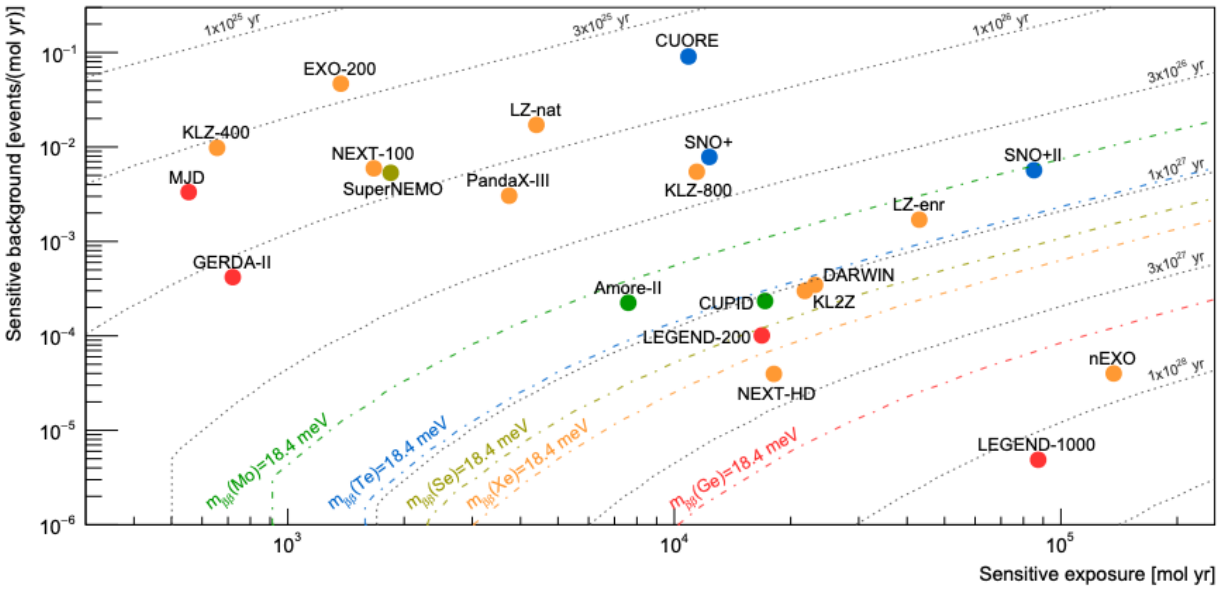

Figure 1.7: An overview of the sensitivity of past and future neutrinoless double-beta decay experiments. The half-life discovery sensitivities are calculated by using the above discussion. The effective Majorana mass for different isotopes is calculated using the corresponding NMEs. Figure from [22]

Chapter 2

The LEGEND experimental program

The Large Enriched Germanium Experiment for Neutrinoless-double beta Decay (LEGEND) collaboration is searching for $0\nu\beta\beta$ decay in germanium-76. Germanium-76 is an even-even nucleus and was identified as an isotope of interest for double-beta decay experiments (potentially) as early as 1952 when a sample of germanium was placed in an emulsion to detect its beta decay [49]; the first double-beta decay experiment to use a (natural abundance) germanium source as the detector — in the form of a Ge(Li) semiconductor detector — was performed in 1967 [50]. The $2\nu\beta\beta$ decay process was confirmed to occur in germanium-76 when it was measured directly in enriched germanium detectors in the 1990s [51–53]: in order to increase the active mass M these detectors were enriched to >80% in germanium-76, which is much larger than the natural abundance of 7.7% [54] of germanium-76. Enriched detectors fabricated with similar techniques are known as enriched high purity germanium (HPGe) detectors. Recent experiments have driven major progress in enriched HPGe detector design and background suppression: the MAJORANA DEMONSTRATOR (MJD) collaboration had the best energy resolution for a $0\nu\beta\beta$ decay search [55], and the Germanium Detector Array (GERDA) collaboration had the lowest background per full width at half max for a $0\nu\beta\beta$ search [56]. The LEGEND collaboration is formed from the union of the GERDA and MJD collaborations. The LEGEND experimental program continues the long history of searches for neutrinoless double-beta decay in germanium-76 [1].

The LEGEND collaboration aims to fully cover the allowed parameter space for inverted ordered Majo-

[1] A history that contains purported (and excluded) discoveries as well!

rana neutrinos in its search for neutrinoless double-beta decay. To that end, it has taken a phased approach with the first step towards the inverted ordering being the LEGEND-200 experiment, which has a target discovery sensitivity of 10^{27} years, and the proposed ton-scale experiment LEGEND-1000 being designed with a discovery sensitivity of 10^{28} years to fully cover the inverted ordering. The LEGEND-200 experiment represents a proving ground for detector technology and background rejection techniques that will enable the LEGEND-1000 experiment to reach its target discovery sensitivity. Besides an upgrade in raw detector mass from LEGEND-200's target of 200 kg to LEGEND-1000's 1000 kg, the LEGEND-1000 experiment will also feature numerous technological advances that are described in Sec. 2.3.

This chapter will outline the designs of the LEGEND apparatuses and how their designs optimize the key analysis parameters that influence the discovery sensitivity according to Eqn. 1.51. The operating principles of germanium detectors for a neutrinoless double-beta decay search will also be reviewed.

2.1 Germanium detectors

Because it is a semiconductor, germanium is naturally suited for the detection of ionizing radiation. When ionizing radiation passes through a germanium crystal, valence-orbit electrons are excited proportionally to unoccupied energy states and can freely traverse the crystal — it takes at least 0.67 eV of energy to promote one electron at 300 K (and about 2.6 eV on average) [57, 58]. Germanium semiconductor detectors take advantage of this proportional relationship by turning the crystal into a diode and adding an applied electric field: these free charge carriers can then be collected, and the collected charge is directly proportional to the amount of energy deposited in the crystal. Thus, if neutrinoless double-beta decay were to occur in a germanium diode, the two emitted electrons would deposit all $Q_{\beta\beta}$ of their energy in the crystal and the correspondingly liberated electrons would be collected and read out and the energy of the event determined. Germanium detectors offer many advantages for a neutrinoless double-beta decay search:

Properties that make germanium detectors great for searching for neutrinoless double-beta decay

- Highly enrichable. Germanium detectors can be made from >90% germanium-76 reliably, meaning that $\varepsilon_{\text{enrichment}}$ can be made high. And because the neutrinoless double-beta decay source is the detector this reduces extraneous detector readout material that can increase the radioactive background.

- Highly pure. Germanium detectors can be made with almost no contamination from elements other than germanium (and dopants) which lowers accidental radioactive background and increases $\varepsilon_{\text{enrichment}}$

- Very large active fraction. With recent developments in crystal growth and detector manufacturing, germanium detectors can be made with masses >$\mathcal{O}(3)$ kg. Even at these very large sizes, these detectors have an active volume — around $\mathcal{O}(90\%)$ — that subtends most of the detector due to clever doping profiles to enhance charge collection. This leads to a large M for a germanium experiment.

- Easily manufactured. Germanium detectors are a well-established technology that can be manufactured by commercial companies.

- Great energy resolution. Because the partitioning of an incident particle's energy loss between charge carrier generation and other mechanisms are statistically correlated, fluctuations in the number of charge carriers created per unit energy deposit are strongly reduced compared to Poisson statistics. This leads to a better energy resolution σ: germanium detectors have achieved the best energy resolution — around $\mathcal{O}(0.1\%)$ at $Q_{\beta\beta}$ — out of all large scale $0\nu\beta\beta$ decay detector technologies to-date [59].

- Mature background rejection analysis techniques. The electric fields of germanium detectors can be constructed so that the charge collection process contains information that enables discrimination between signal-like and background-like energy depositions. This is known as

pulse shape discrimination (PSD). The electric field structure also leads to excellent fiducialization, where radioactive background events that occur on the surface of a detector can be reliably removed from analysis. This results in a lower background index BI.

This section covers the physics behind the operation of a germanium diode as an ionizing radiation detector for a neutrinoless-double beta decay search.

2.1.1 Semiconductors

Atomic germanium forms a crystal with a face-centered cubic (diamond-like) geometry [60]. The plot of allowed energy states versus crystal momentum is shown for germanium in Fig. 2.1; because germanium is a three-dimensional crystal, it is typical to plot this versus one-dimensional crystal momentum where the dimensions point along lines connecting high symmetry points of the Fourier transformed lattice (called the reciprocal lattice). The bands of a crystal will be populated by electrons within the crystal according to the Fermi-Dirac distribution:

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_f}{k_B T}\right)}, \tag{2.1}$$

where k_B is Boltzmann's constant, T is the temperature of the crystal, and E_f is the Fermi energy or the chemical potential of the electrons in the crystal. In (intrinsic) semiconductors and insulators at zero temperature the lowest energy band is completely filled and the Fermi energy located exactly at the middle of the band gap E_g [58, 62]. Germanium is an example of an indirect band gap, where the maximum of the filled valence band is at a different value of $\vec{k}$ than the minimum of the conduction band. Transitions between indirect gaps thus require changes in momentum. When the band gap is less than $\sim$4 eV a material is called a semiconductor, otherwise it is known as an insulator.

The ability to use germanium as a radiation detector rests on the doping of germanium in order to turn it into a diode. Doping is the process of adding known impurities (an impurity is any element other than what makes up the pure crystal) to a crystal to modify its electronic properties [57, 58, 62]. When there are no impurities present the crystal is known as intrinsic. For intrinsic semiconductors, when an electron is promoted to the conduction band, one hole is created — the number of holes is equal to the number of

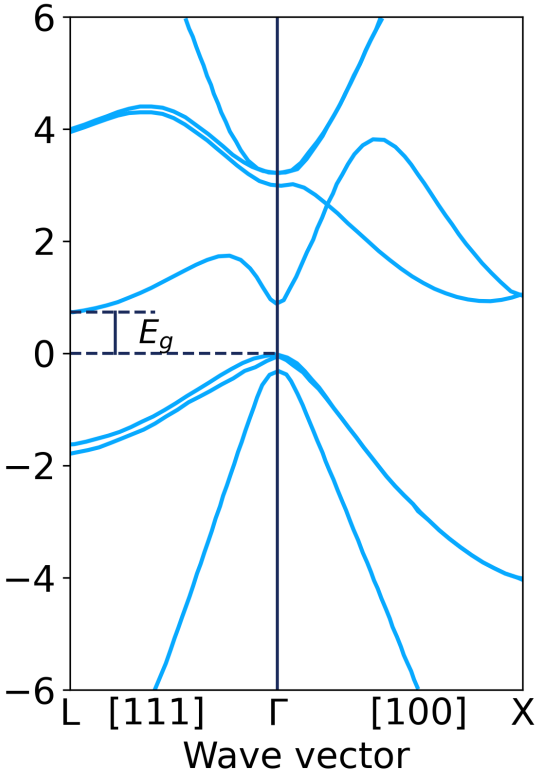

Figure 2.1: Band structure of germanium calculated at absolute zero. The band structure is taken from Ref. [61] and was computed using a $\vec{k} \cdot \vec{p}$ method that involves perturbative expansion of the $\vec{k} = 0$ Hamiltonian in terms of crystal momentum's inner product with the real momentum $\vec{p}$. The band gap, $E_g \approx 0.67$ eV, is identified as the smallest energy difference between the valence band and the lowest conduction band: notice that the minimal conduction band energy does not happen at $\vec{k} = 0$ (called the Γ point) where the valence band is maximum, and thus the band gap is deemed to be indirect.

conduction electrons. Semiconductors are useful for radiation detection because ionizing radiation creates a number of electron-hole pairs that is proportional to the energy deposited in the detector. If we dope a crystal with an element that wants to donate its valence electron (known as a donor element) this extra electron goes into the conduction band: this type of doping is called n-doping where n is the number density of electrons in the conduction band. The ionized positive atom is locked immobile into the lattice and does not contribute to the density of holes. Common donor elements include phosphorous and lithium. If we instead add an element that accepts extra electrons (called an acceptor element) this removes electrons from the valence band, or equivalently adds holes: this type of doping is known as p-doping where p is the number density of holes. Common acceptor elements are aluminum and boron. It is critical to remember that even though doping adds or removes electron states from the bands, the overall crystal is electrically neutral. Typical doping levels are on the order of 10^{15}/cm^3 [58].

2.1.2 p-n junctions

When a p-type semiconductor is brought into perfect mechanical contact with an n-type semiconductor (usually through implantation or diffusion), a $p-n$ junction forms [57, 58, 62]. At the interface between these two regions, free electron charge carriers from the n-type region will annihilate free hole charge carriers from the p-type region; this process leads to a small amount of net charge due to the immobile, and now positively charged, donors on the n-type side and negatively charged immobile acceptors on the p-type side. This small amount of net charge sets up an internal electric field that prohibits the free charge carriers from further annihilating. This spatial region that has no more free charge carriers is called the depletion region. A picture of this is shown in Fig. 2.2. The depletion region is ideal for radiation detection because the resistivity of the region is much higher than intrinsic or doped material: there are no free charge carriers left to carry a current and thus the conductivity is reduced [57].

The $p-n$ junction behaves as a diode when a voltage is applied. When a diode is reverse biased, current does not flow freely though the device. If a positive voltage is applied to the n side and a negative voltage to the p side, then a net electric field is established that causes the depletion region to enlarge and establishes a large voltage barrier. When the reverse bias voltage is large enough such that the depletion region subtends a majority of the volume of the detector, the detector is then termed as "fully depleted". If the reverse

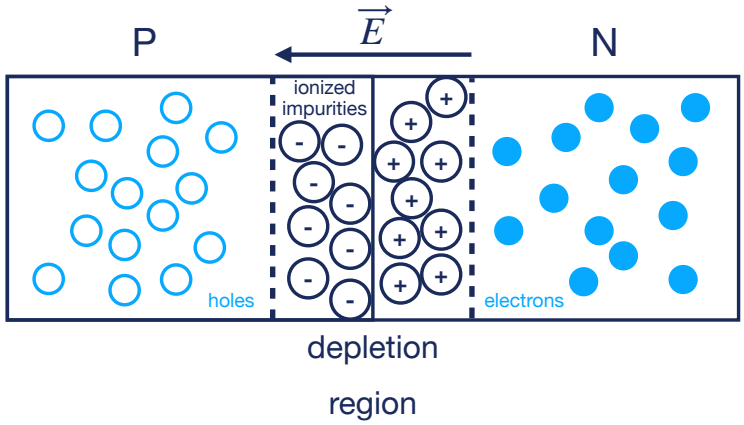

Figure 2.2: Depiction of a $p-n$ semiconductor junction and the depletion region.

bias is set too large, however, the electric field inside the detector can be large enough to sustain avalanche multiplication of thermal carriers and the detector "breaks down" with a large sustained current. A detector in this state is said to be above its breakdown voltage — this is a useful property for some radiation detectors to count total rates, see section 5.2 for its application for photon detection, but not for reconstructing energy of ionizing events.

The reverse biased $p-n$ junction forms the basis of a low-noise radiation detector. Any radiation that passes through the depletion region creates electron-hole pairs that are swept to charge collecting contacts via the applied electric field. Importantly, the number of electron-hole pairs — and thus the collected charge — is proportional to the energy deposited in the detector. Ionizing radiation that deposits energy outside of the depletion region creates charge carriers that feel no electric field; the area outside of the depletion region is subdivided into a dead layer and a transition layer. Charge carriers in the transition layer have some chance of diffusing into the depletion region and being collected appropriately, however a large fraction of the initial charge carriers recombine and thus only partial charge collection is possible. Charge carriers in the dead layer have an infinitesimal chance of diffusing into the depletion region and thus have no charge collection.

Germanium detectors for neutrinoless double-beta decay searches are formed from $p-n$ junctions. These detectors are made out of a majority of intrinsic material that is slightly p-type, with an n-type con-

tact and a p-type contact used to apply high voltage and establish a $p-n$ junction. The intrinsic region has an impurity concentration of $\mathcal{O}(10^{10})$ cm^{-3} that is several orders of magnitude smaller than the concentration at the heavily-doped p-type and n-type contacts [63]. Typically, only the contact of the same type as the intrinsic region is used for signal readout. The large intrinsic region enables a large depletion region to be established; this increases the sensitive exposure of a germanium $0\nu\beta\beta$ decay experiment. The intrinsic region is made slightly p-type because detectors with n-type intrinsic regions have worse energy resolution: electrons, which generate most of the signal (see the end of Sec. 2.1.3 for discussion) for an n-type detector, have a larger charge trapping in lattice dislocations [64]. In addition, holes have a larger effective mass and thus the resistivity is higher for p-type material [65]. Another advantageous feature is that p-type detectors have a very thick dead layer on the n-doped contact and this forms passive shielding from surface contamination, as illustrated in Fig. 2.3 [64]. For LEGEND detectors, the p-type contact is doped with boron via ion implantation, and the n-type contact is doped with lithium via diffusion.

2.1.3 Signal formation

We are interested in the time profile of the charge collection in our detectors because it gives information on where the signal occurred in the detector. The instantaneous current from the induced charge on an electrode is given by the Shockley-Ramo theorem [66]:

$$i = -q(\vec{v}_e + \vec{v}_h) \cdot \vec{\nabla}\Phi_0, \tag{2.2}$$

where q is the charge of the carrier, $\vec{v}$ is the carrier's instantaneous velocity, and Φ_0 is the weighting potential [57, 65]. The weighting potential for an electrode is obtained by solving Laplace's equation with the boundary conditions that the electric potential is set to 1 on the electrode of interest and to 0 on all other surfaces. The weighting potential is essentially the fractional contribution to the total field of the one contact, and gives the coupling of a charge to the electrode.

For simple detector geometries with sufficient symmetries, such as planar or coaxial detectors, the solution to Laplace's equation are analytical. However for most detector designs used in LEGEND, the weighting potential must be solved using numerical methods. In addition, to use the Shockley-Ramo equation to simulate pulse shapes, it is also necessary to know the electric field inside the detector because the electric

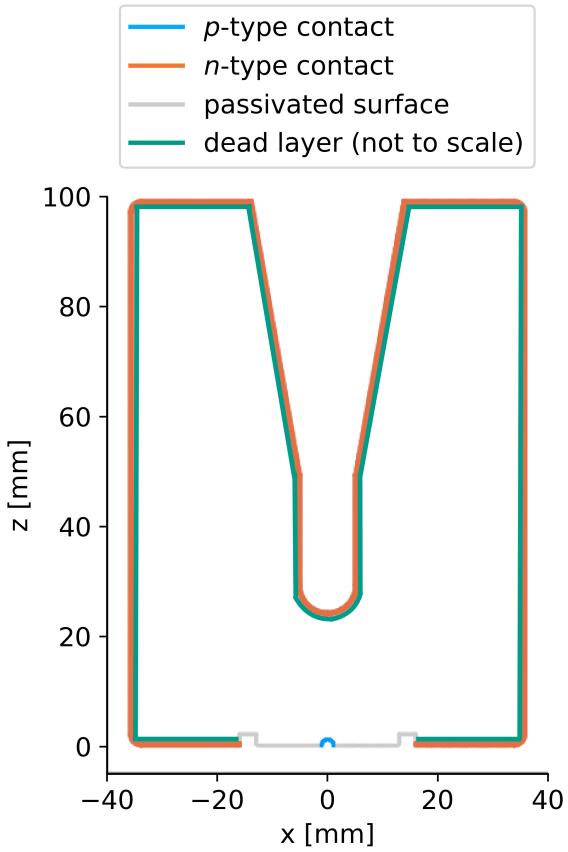

Figure 2.3: Illustration of the p-type and n-type surfaces, as well as the dead layer, for a germanium detector. This plot also shows the passivated surface — made of amorphous germanium — that holds off the voltage difference between the p and n contacts.

field gives the charge carrier's instantaneous velocity via the Drude transport model $\vec{v} = \mu\vec{E}$, where μ is a tensor describing the mobility of the charge carriers [62]. Thus it is also necessary to solve Poisson's equation,

$$\nabla^2\Phi = -\frac{\rho}{\varepsilon}, \tag{2.3}$$

with sufficient boundary conditions from the applied high voltage, where ρ is the space charge density. The space charge density inside the depleted detector is given by the left-over ionized impurities; for totally depleted p-type detectors, the space charge is thus given by $-q\rho(\vec{x})$ where $\rho(\vec{x})$ is the doping profile of the detector. Computational tools exist to solve the Laplace and Poisson equations, typically through methods such as finite element methods to solve the partial differential equation. Two pieces of software are used by the LEGEND collaboration to calculate the electric field and weighting potential for germanium detectors: `siggen` [67] and `ssd.jl` [68].

All LEGEND detectors use a slightly p-type intrinsic crystal for the bulk of the detector, and only have charge readout on the heavily-doped p-type contact. This means that the majority of the readout signal is generated by the drift of holes: electrons do not experience as large of a change in the weighting potential as the holes do near the p-type contact, and thus holes contribute more to the signal via Eqn. 2.2.

Because the important information in a germanium signal is the total amount of charge collected — because it is proportional to the energy deposited in the detector — and the time profile of the charge collection, a charge sensitive preamplifier is used instead of a current-sensitive amplifier. A simplified circuit diagram for a charge sensitive amplifier (CSA) consists of a feedback capacitor and resistor in parallel with an operation amplifier that is in series with the detector, as shown in Fig. 2.4.

The current from Eqn. 2.2 is integrated on to a feedback capacitor C_F, and the voltage V_0 across C_F is used to measure the induced charge during signal generation as

$$V_0 = \frac{Q}{C_F}. \tag{2.4}$$

Both the electron and hole components of Eqn. 2.2 are necessary so that the charge in Eqn. 2.4 is proportional (up to charge trapping losses) to the number of electron-hole pairs generated during an energy deposition. The number of electron-hole pairs is proportional to the energy of the event, and thus ionizing

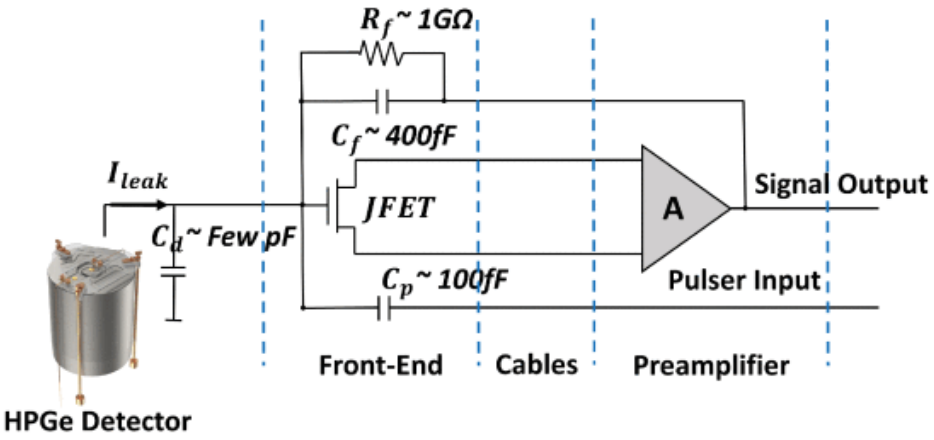

Figure 2.4: Charge sensitive preamplifier simplified circuit diagram for germanium detector signal readout in LEGEND-200, from Ref. [69].

event energies can be reconstructed.

Charge sensitive amplifiers have the added benefit that the feedback capacitor and open-loop gain render any changes in the detector capacitance virtually unnoticeable [65]. The feedback resistor is present in the circuit to drain the charges off the feedback capacitor to help improve noise performance and eliminate the DC noise of the op-amp [65]. The addition of this resistor however changes the time profile of the charge collection: a decaying exponential with time constant $\tau = R_F C_F$ is convolved with the voltage signal. This leads to characteristic decay tails on germanium waveforms.

An example of a simulated induced charge on a p-type contact as a function of time — called a waveform — for a LEGEND detector is shown in Fig. 2.5. The charge profile is usually broken up into three distinct regions:

Germanium signal structure

1. The **baseline** is the flat region before charge carriers are created in the detector
2. The **rising edge** is the region of time that charge is induced on the contact. This region of the waveform contains valuable information on the location of the event within the detector, as the time profile of the induced charge depends on the specific path taken through the weighting potential.

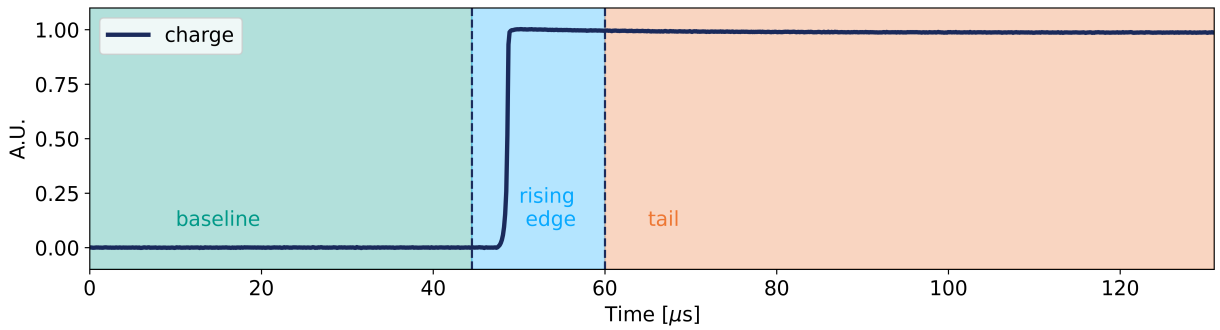

Figure 2.5: Exemplar waveform showing the different portions of the signal

3. The **tail** is the region of full charge collection when the signal ends. It is usually called a tail because read-out electronics add a decaying exponential to this region.

2.1.4 Energy calibration and energy resolution

Because the gain of a detector — the constant of proportionality between the collected charge and deposited energy — is not known with sufficient precision and varies in time, it is common to calibrate a germanium detector. A calibration uses the charge collected from depositions at a known energy to compute the correction factor to convert collected charge to actual energy. This process builds a calibration curve $f(E_{uncal.})$ which is a recipe to convert uncalibrated charge collection to calibrated energy. Typically external radioactive sources with known gamma energies are used, because these gamma energies can be monoenergetic and must reconstruct to a singular value. Calibrations often use energy depositions across a wide range of energy to create a calibration curve. A common calibration curve is similar to that for the energy resolution:

Common energy calibration

$$E_{cal.} = a + bE_{uncal.} + cE_{uncal.}^2, \tag{2.5}$$

where a comes from filtering effects [2], b from energy dependence of the charge carrier creation process, and c comes from amplification nonlinearities [71]. Depending on the stability of the readout electronics,

[2] the optimal energy estimator is obtained by taking the maximum of the filtered waveform, and this introduces a slight bias in the energy estimator for events with 0 deposited energy [70].

the gain of a germanium detector can shift frequently and thus calibrations must be performed on a regular basis.

The energy resolution of a germanium detector is a key analysis parameter; a smaller energy resolution means a larger sensitivity for neutrinoless-double beta decay as there is less chance for a uniformly distributed radioactive background to fall within the likely energy range for a true neutrinoless-double beta decay event. A smaller energy resolution also helps prevent smearing out the irreducible $2\nu\beta\beta$ decay spectrum which could obscure the monoenergetic $0\nu\beta\beta$ decay peak at $Q_{\beta\beta}$. The energy resolution of a detector represents how statistical fluctuations tend to smear out the reconstructed energy of a monoenergetic energy deposition [57]. For all ionizing radiation detectors, the optimal energy resolution is sadly not a Dirac delta, but is instead a complicated function due to the inherent statistical nature of the charge carrier creation process. The creation of individual charge carriers is not independent. The fluctuations in the number of charge carriers was first described in Ref. [72] as

$$\frac{\sigma_E}{E} = \sqrt{F\frac{E_{e-h}}{E}}, \tag{2.6}$$

where E_{e-h} is the average energy needed to create an electron-hole pair and F is the Fano factor that describes these statistical fluctuations appropriately. If $F < 1$ then the fluctuations are called sub-Poissonian because the energy resolution is smaller than what Poissonian statistics predicts with $F = 1$. The Fano factor is a quantity that is experimentally measurable, and a smaller Fano factor means a smaller and better energy resolution. The Fano factor for germanium at 77 K is roughly 0.1 [57].

To understand the origin of Eqn. 2.6, it helps to recall that energy deposited by ionizing radiation can convert into other channels besides charge carrier generation. One model conceptualizes this as an incident particle shedding its energy into only electron-hole pairs until it reaches a state where it can only thermalize by exciting phonon modes [65, 73]. Conservation of energy and momentum in this model then derives that the average pair creation energy is described by $E_{e-h} = 2.8E_g$, which agrees well with an empirical fit to data given by Ref. [73]:

$$E_{e-h} \approx 2.8E_g + 0.6 \text{ eV}. \tag{2.7}$$

Thus the average energy to create an electron-hole pair exceeds the band gap, but is proportional to its value. For germanium with $E_g = 0.67$ eV we recover $E_{e-h} \approx 2.53$ which is close to the standard 2.8 eV measured [57, 65]. We can then see that germanium partitions roughly $\frac{0.67}{2.8} \approx 20\%$ of its energy into creating electron-hole pairs and $\sim 80\%$ into phonon modes [65].

The sum of the energy lost creating electron-hole pairs and the energy lost creating phonons *must* sum to the initial energy of the ionizing radiation. If there is any statistical fluctuation in the amount of energy partitioned into creating electron-hole pairs, there must be an equal and opposite fluctuation in the energy going into creating phonons. This anti-correlation and constraint on the total energy are responsible for the origin of the Fano factor [57, 65]. Direct computation of the Fano factor is a complicated task that requires utilizing the phonon and electron densities of states [65].

In practice, the energy resolution is made worse by the fact that there are many different sources of statistical fluctuation of the collected charge carriers. One dominant irreducible source is the statistical fluctuation from the electronic read-out: this noise is constant even in the absence of drifting charge carriers due to fluctuations of average charge on the amplifier's C_F (due to leakage current) and fluctuations in the voltage measurement. As such, it is treated as being independent of energy [57]. One other dominant noise source is from incomplete or delayed charge collection due to impurities and charge traps in the crystal. This source of noise is directly proportional to the reconstructed energy of the event, as larger energy depositions have a larger interaction radius and thus an increased statistical spread in trapped carriers [59, 74]. All sources of noise are assumed to be symmetrically distributed and independent, so the overall energy resolution is the sum in quadrature of the independent contributions by the central limit theorem [57]:

$$\sigma^2 = \sigma_N^2 + \sigma_F^2 + \sigma_\tau^2, \tag{2.8}$$

so that the energy resolution as a function of energy is given by

Energy resolution as a function of energy

$$\sigma = \sqrt{a + bE + cE^2}, \tag{2.9}$$

where a is the contribution from electronics noise, b is the Fano/statistical noise, and c is the incomplete

charge collection noise [75]. The apparent widths of gamma lines at different energies observed during a calibration can then be used to compute the energy resolution as a function of energy.

2.1.5 Detector geometry

The geometry of a germanium detector impacts the shape of the electric field and thus the charge collection properties and waveform profiles of the detector. Precisely engineering the geometry allows for improved detector performance in energy resolution and pulse shape discrimination. Early germanium detector geometries included planar detectors and coaxial (also called semi-coaxial) detectors, as these geometries were easy to grow as crystals and turn into detectors. Most germanium detectors have axial symmetry from the method of crystal growth used [57, 63]. Coaxial detectors in particular were of interest to early germanium neutrinoless double-beta decay experiments as they could be made quite massive. They consist of a large p-type contact in the central borehole, and then the large n-type electrode on the exterior of the detector, allowing for easier handling [3]. An example of the coaxial geometry is shown in Fig. 2.6c.

It was realized quite early on in the 1980s that the shapes of waveforms generated in germanium detectors could be used to discriminate between neutrinoless double-beta decay-like events and other accidental radioactive background [76–78]. Experiments began to use these techniques in the 1990s and 2000s [79–81]. In particular, neutrinoless double-beta decay events should deposit all their energy within a $\mathcal{O}(1\ \mathrm{mm})$ volume of the detector [82]; this is in contrast to some gamma events that can Compton scatter and deposit energy in multiple locations across great distances in a detector. Neutrinoless double-beta decay events are thus an example of what is called a single-site event (SSE), while a Compton scatter is called a multi-site event (MSE). Because MSEs leave charge in different locations of the detector, the induced signal will behave differently than if all the charge were deposited at a single location. The characteristic feature of MSE is that their induced current will look like the sum of SSE. For detectors with a large change in weighting potential near the point contact, MSE have an induced current that has multiple peaks due to the different deposition sites having different drift times through the detector. For coaxial detectors however, the p-type contact is so large that the change in weighting potential around it is less steep and therefore rise times are slower and the differences in the drift times of components of the signal are not easily distinguishable, and

[3] Coaxial detectors are also widely available for gamma spectroscopy, where it was common for the borehole to be n-type so that the thinner p-type surface contact would allow for higher sensitivity to lower energy gammas.

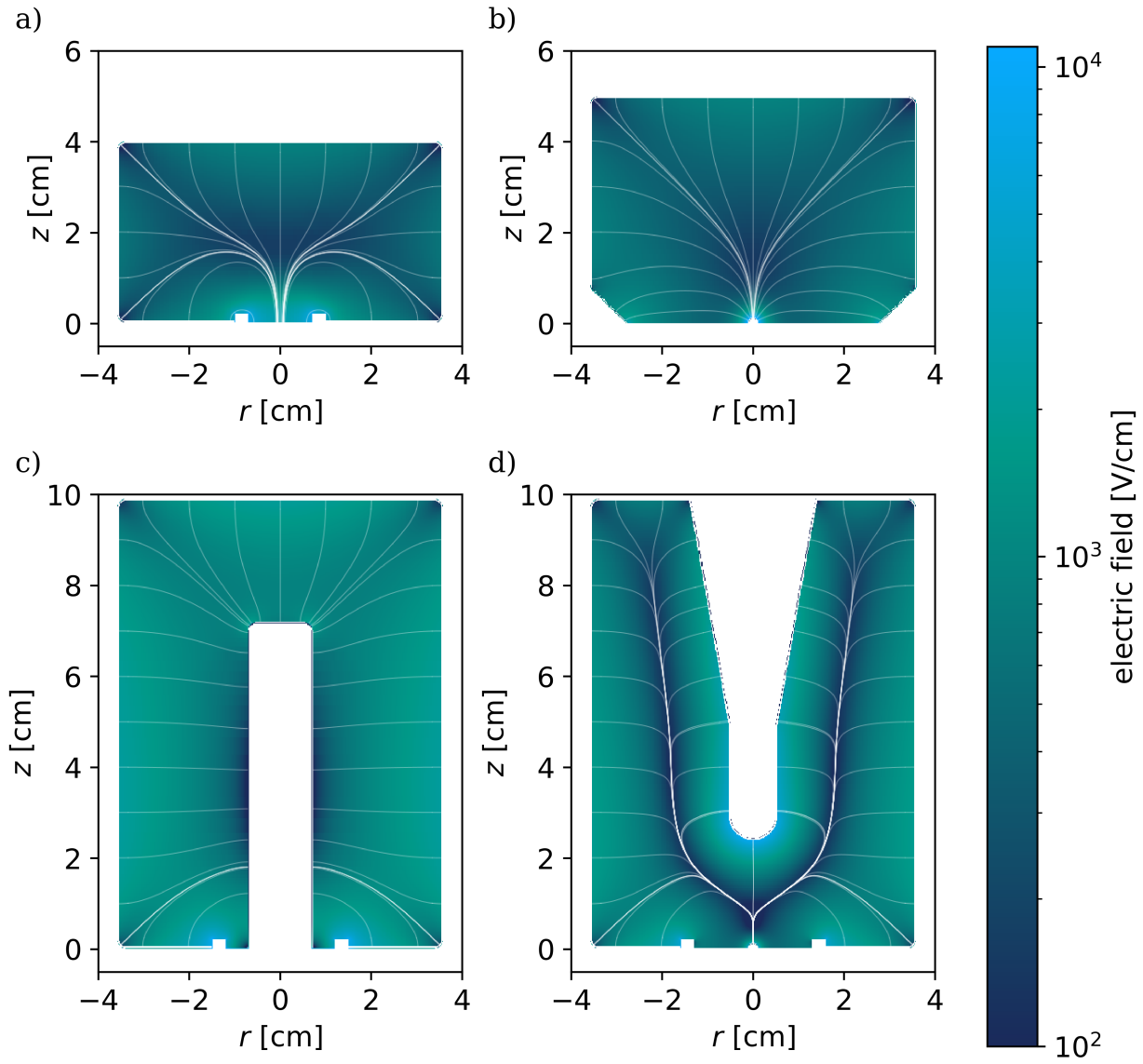

Figure 2.6: Electric fields of four germanium detector geometries used in the LEGEND-200 experiment, along with the drift paths for charge carriers superimposed in white lines. Subplot 'a)' shows the field from a broad energy germanium detector, 'b)' shows the field from a p-type point contact detector, 'c)' shows the field from a coaxial detector, and 'd)' shows the field for an inverted coaxial p-type point contact detector.

thus complex techniques — like machine learning — are necessary to perform this rejection [80, 81].

The noise performance of coaxial detectors was also not ideal: their capacitance is quite large and limits their energy resolution (recall that a smaller detector capacitance means better noise performance). It was realized that manufacturing detectors with a point-contact — which is just an extremely small surface area contact — would ensure a lower capacitance and that leads to better noise performance [64, 83]. The added benefit of this design is that point-contacts distort the weighting field such that the change in weighting field is extremely high in the vicinity of the point-contact and much lower elsewhere in the detector. Thus, drifting charges will induce a majority of their signal when they are close to the point-contact. The rise time of the signal is therefore much shorter than the difference in drift times for the different energy depositions in a MSE, with drift times differing on the order of ~ 1 μs [84]. This makes it much easier to discriminate between the time of arrival of induced current from different elements of a MSE energy deposition. Furthermore, signals generated anywhere in the bulk experience nearly the same gradient traversal and thus all SSE have approximately the same signal shape (except for events in the vicinity of the point contact).

Two initial approaches were taken to point-contact detectors for neutrinoless double-beta decay searches, building on the point-contact technology developed in Refs. [64, 83]: the broad energy germanium (BEGe) detector and the p-type point contact (PPC) detectors pursued by the GERDA and MAJORANA DEMONSTRATOR (MJD) collaborations respectively. Broad energy germanium (BEGe) detectors were a commercially available geometry from Canberra that featured a small point-like p-type contact (of the order $\sim$70 mm for those used in GERDA [85]). They also have a small ditch made of non-conducting amorphous germanium around the point-contact in order to stand-off the high voltage that is applied to the remaining n-type surface. They are roughly of order 0.6 kg in mass [85]. The p-type point contact (PPC) detectors engineered by the MAJORANA DEMONSTRATOR (MJD) collaboration following Ref. [64] and in collaboration with ORTEC had roughly the same dimensions and average mass of 0.8 kg [55]; however, the point contact is much smaller, of order of $\sim$1 mm [86], and there is no ditch surrounding the p-type contact. Instead, the surface containing the p-type contact is passivated with amorphous germanium all the way to the n-contact on the other edge. The amorphous germanium is non-conducting so it stands-off the high voltage. The geometries from these two types of detectors are shown in Fig. 2.6a,b. For point-contact geometry detectors, the n-type region extends roughly on the order of $\sim$1 mm into the detector, while the heavily doped p-type

point-contact region extends $\sim 1\mu$ m [86].

The doping profile sets the absolute length of these detectors, and thus their mass. It was pointed out in Ref. [87] that detectors of these geometries could not be made any longer within the technological limits of creating a doping profile, as the detectors would no longer be fully depleted at reasonable operating voltages. If a borehole was installed in the n-type contact, then the overall length of the detector could increase while ensuring the detector could fully deplete [87]. This geometry is known as inverted coaxial (IC); the geometry is an "inverted" coaxial because the doping profile of the contacts, and thus electrode geometry, is the opposite of a traditional coaxial detector. Figure 2.6d shows this geometry and its electric field. IC detectors can be larger than 10 cm long and are of order $\sim$3 kg in mass. These detector types form the baseline design of the LEGEND experimental program [88, 89]; they are as massive as possible, reducing background; they have low capacitance, making the energy resolution as good as possible; and they have a precisely engineered weighting field and excellent pulse shape discrimination, thus reducing background even further. IC detectors manufactured by ORTEC have a small p-type point contact and are therefore called inverted coaxial p-type point-contact (ICPC) detectors. This is in contrast to IC detectors manufactured by Mirion which have larger p-type (more disc-like in geometry than point-like) contacts.

2.1.6 Background rejection capabilities

As mentioned previously, the differences in arrival times of depositions of charge in multiple locations in a detector was identified as a way to discriminate between multi-site background events and single-site neutrinoless double-beta decay-like events [79–81]. This drift time cut was difficult to apply in coaxial detectors where the weighting potential varied constantly throughout the detector. With the advancement of point-contact like detectors where the weighting potential decreases very rapidly outside the vicinity of the point-contact, the MSE cut was improved and its analysis was simplified. The MSE cut can be implemented with one analysis parameter, the value of A/E, or the amplitude of the current pulse divided by the energy of the event. A SSE has a current waveform that looks like a bump; while a MSE of the same energy has multiple bumps due to the different charge carrier distributions passing through the high weighting potential at different time. This A/E parameter was realized to reflect these differences numerically in Ref. [90]: MSEs effectively deposit their energy more slowly and have a lower A/E than SSEs. Figure 2.7 shows a

plot of a SSE and MSE waveform as well as their A/E parameters.

Event topologies can also be determined using information from the signal shapes. It was realized that the weighting field from a point-contact also easily enabled discrimination between events originating on or near the point-contact, events originating on the n-type contact, and single-site events occurring in the bulk of the detector [86, 91–93]. Different types of external radiation leave different signatures in the germanium detector due to differences in their stopping power and interactions with the lattice. Due to their large stopping power, α and β types of radiation deposit a significant amount of energy in the surface of the detector and are thus called *surface events*; while γ type radiation typically does not deposit significant fractions of energy at the surface of the detector.

Point-contact background rejection: The only surfaces that α particles can penetrate are the p-type point contact and passivated surface; the n-type contact is a $\sim$1 mm thick lithium layer (comprising a thin dead layer and then a transition region) that the α particles cannot make it through [86]. The p-type point contact however is on the order of microns thick, so α particles can enter the detector (β particles can also enter through this contact). When a charged particle enters the detector at the point contact, almost all of the ionizing radiation is absorbed in the large-gradient weighting potential region around the point contact. These events then induce a signal that has a very steep rise time as the charges are collected very quickly and with a very high change in weighting potential (see the Shockley-Ramo theorem Eqn. 2.2). This gives these events a very large A/E parameter value. The lower-middle panel of Fig. 2.7 shows an example waveform of an event originating near the p-type point contact. Cutting events with a high A/E cut removes events originating and depositing energy near the point contact and thus it is very effective at removing α backgrounds. This cut represents a fiducilization of the detector, as it removes all events originating in a small vicinity around the point-contact.

Dead-layer background rejection: the remaining surfaces of the detector, the n-type contact and any passivated surfaces, have a $\sim$ 1 mm thick dead layer due to the impurity profile of the crystals; only β particles and γ rays can penetrate these surfaces; all α particles are stopped. This layer is called the dead layer because it is incapable of charge collection: carriers generated in the n-type region feel no electric field and must diffuse to the depletion region to be collected. The holes, the dominant contribution to the detector signal, that are generated in n-type region are minority carriers and thus very quickly recombine.

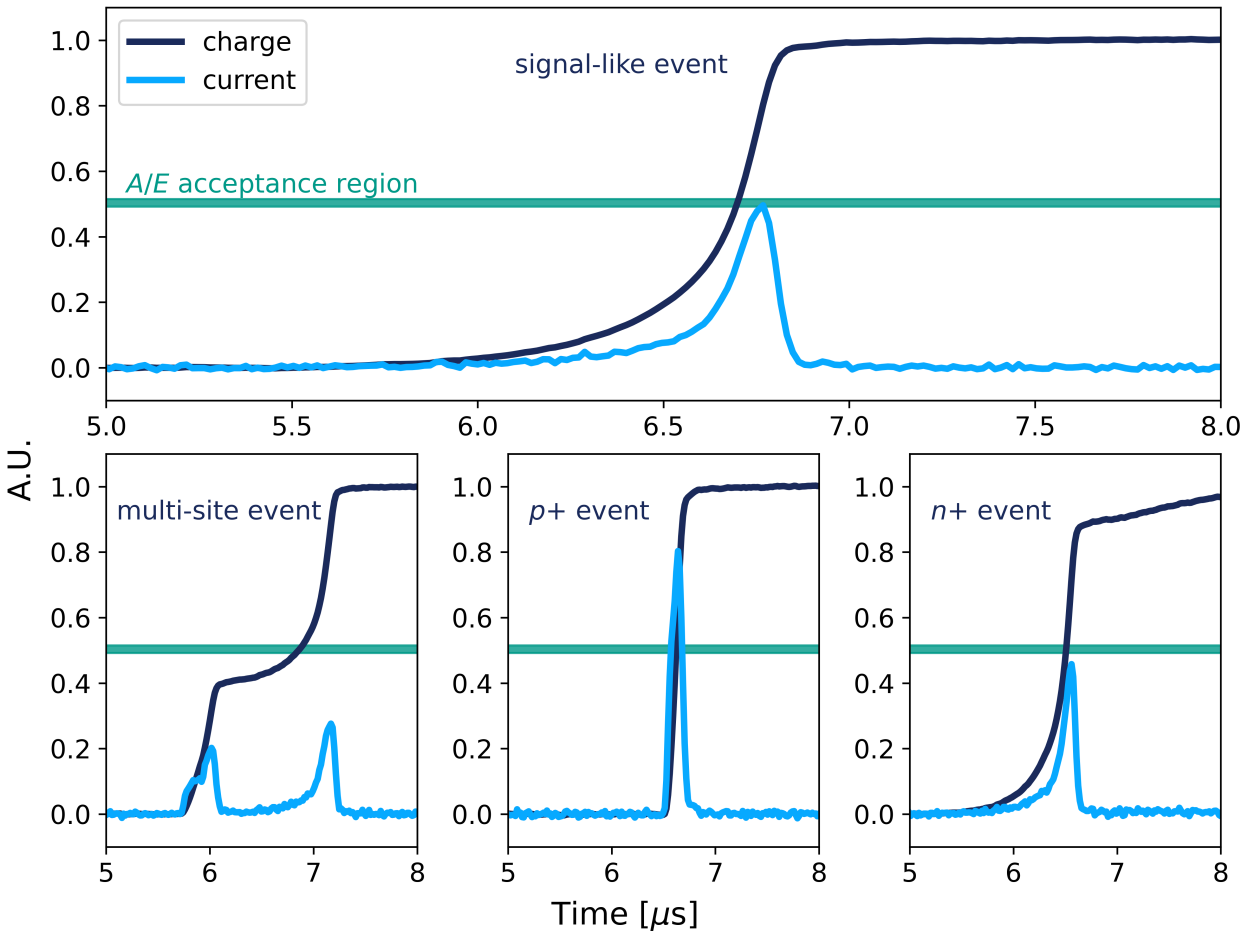

Figure 2.7: Pulse shapes from different background event topologies compared to a signal-like waveform shape. Data and calibrated A/E acceptance region are from the IC-type detector V07302A during a calibration run 0001 in period 03 of LEGEND-200 data taking. Charge waveforms are normalized to their maximum, and the current waveforms all have equal area. Top) a signal-like event whose A/E parameter falls within the acceptance band, shown in green, for this period. Bottom left) an example multi-site event whose A/E parameter falls below the acceptance band and characteristically has two peaks in the current waveform. Bottom middle) an event likely originating from near the p-type contact, as characterized by its extremely spiked current waveform and high A/E value. Bottom right) an event likely originating from the n-type contact where the delayed charge component from transition layer charge diffusion causes the A/E parameter to be low, and the waveform to have a slow rise after a majority of the charge has been collected.

The β particles that pass through the dead layer also deposit a significant fraction of their energy in the transition layer before reaching the depleted region where normal charge collection occurs. The transition layer is part of the n-type region close enough to the depletion region for partial charge collection due to diffusion of carriers into the depletion region. Thus β events originating on n-type and passivated surfaces are characterized by delayed charge collection: the energy the β does deposit into the depleted region is collected normally, in addition to charge carriers from deposition in the transition layer slowly diffusing into the depleted region. The charge carriers in the transition layer feel no electric field, so they are only collected if they diffuse (via a random walk) into the depleted region where the electric field is non-zero — provided that the charge carriers don't recombine in the transition layer. The lower-right panel of Fig. 2.7 shows a plot of a waveform originating from the n-type surface. These events can be classified by several different parameters. First, these events have a lower A/E as the charge takes a long time to collect. The delayed charge collection leads to a very rounded top of the waveform: this feature is picked up by both the late charge (LQ) cut and the delayed charge recovery (DCR) parameter. The DCR cut was developed in the MAJORANA DEMONSTRATOR (MJD) to remove events from the surface having a very small slope to the tail of the pole-zero corrected waveform [93]. The late charge (LQ) cut was developed and implemented in MAJORANA DEMONSTRATOR (MJD) and also implemented in the LEGEND-200 experiment and is defined by the integral above the rounded part of the waveform after t_80, the time at which 80% of the charge has been collected; delayed charge collection will have larger values of the late charge (LQ) classifier [55, 94]. This cut also represents a fiducial cut on the volume of the detector, removing events and detector mass near the transition region.

In summary, the electric field geometry of point-contact germanium detectors enables advanced pulse shape discrimination techniques that remove many different sources of radioactive background.

2.2 LEGEND-200

The LEGEND-200 detector aims to capitalize off the numerous advantages of HPGe detectors outlined above — high mass, excellent energy resolution, high enrichment, and advanced pulse shape discrimination — to build an experiment with a half-life discovery sensitivity of 10^{27} years within roughly ten years of operation with 200 kg of detector material [88]. This requires a background index goal of $2 \times$

10^{-4} counts/keV/kg/yr (ckky), a factor of three lower than what was observed in GERDA [56]. In order to reach this goal, the LEGEND collaboration builds on the technical achievements made by the MJD and GERDA experiments. Low-noise readout electronics developed by the MJD experiment enables world-leading energy resolution per full-width-at-half-maximum. Significant background reduction is achieved through the use of passive and active shielding, including a liquid argon detector used for anti-coincident background rejection, as developed by GERDA. The LEGEND-200 detector consists of an HPGe array that is suspended in a liquid argon (LAr) environment that acts as both coolant and active shield. Liquid argon scintillates when ionizing radiation passes through it; therefore, tagging this scintillation light can remove radioactive background events because **neutrinoless double-beta decay should almost always deposit all its energy entirely within one germanium detector**. The liquid argon cryostat is placed within a water tank that is used as a muon veto. In addition, materials used in the construction of the detector were assayed to ensure excellent radiopurity. The whole detector is installed in the underground Laboratori Nazionali del Gran Sasso (LNGS) in Italy with roughly 1400 m of rock overburden to shield against cosmic rays. Figure 2.8 illustrates the experimental design of the LEGEND-200 detector.

2.2.1 Germanium array

A closely packed array of HPGe detectors forms the basis of the LEGEND-200 experiment. Detectors are suspended vertically on twelve strings, with on average, ten detectors per string; although, in the first deployment, only ten strings were utilized. A computer aided design rendering of the germanium array is shown in Fig. 2.9. The dense spacing of the detectors enables further background rejection through a multiplicity cut: if substantial energy is deposited into multiple detectors, the event is most likely not signal-like. This array is assembled in an ultra-pure nitrogen purged glove box and is then lowered in its entirety into the liquid argon. The strings are formed out of underground electroformed copper rods — the electroforming ensures the copper is ultrapure, and growing it underground diminishes cosmogenic activation of the copper. Radioactive sources can be lowered into and removed from four points near the array in order to take calibration data. Each string of germanium detectors is surrounded by a nylon mini-shroud to prevent radioactive ions in the surrounding liquid argon medium from being swept to the surface of the detectors by convection and weak electric field drift.

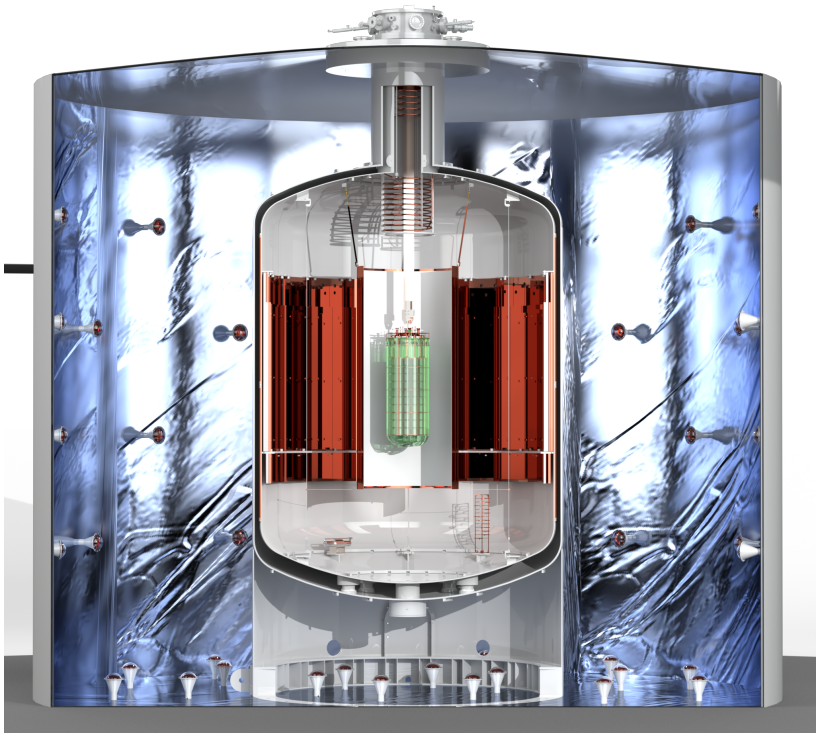

Figure 2.8: Artistic rendering of the LEGEND-200 experiment. Shown in green at the center of the figure is the twelve-string HPGe detector array surrounded by optical fibers for liquid argon scintillation light readout. The white backdrop behind the HPGe detector array is the wavelength shifting light reflector to help increase light collection efficiency. The detector array is suspended inside of the liquid argon cryostat. The liquid argon cryostat is surrounded by a water tank, shown in blue, instrumented with photomultiplier tubes for vetoing muon events.

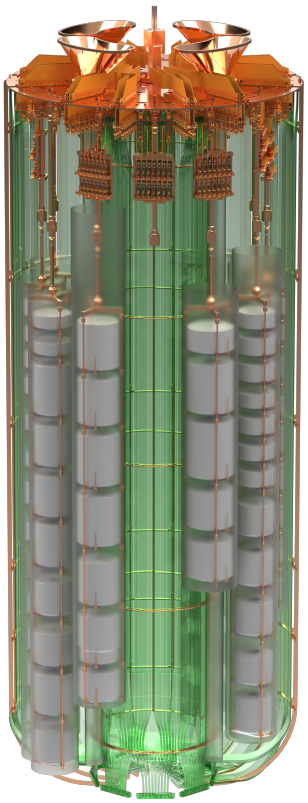

Figure 2.9: Computer aided design rendering of the HPGe array for the LEGEND-200 experiment. The germanium detectors, shown in gray, were deployed on ten strings in the first year of physics data taking. The low mass copper support structure for the detectors is shown. Surrounding each string of detectors is a nylon mini-shroud to reduce ^{42}K backgrounds. The inner optical fiber barrel and half of the outer fiber barrel are shown in green; these help collect and guide scintillation light to silicon photomultipliers positioned in strips, visible at the bottom of the figure, but also present at the top of the array. The copper funnels at the top of the array help guide calibration sources from outside the liquid argon environment into the position necessary for weekly calibrations.

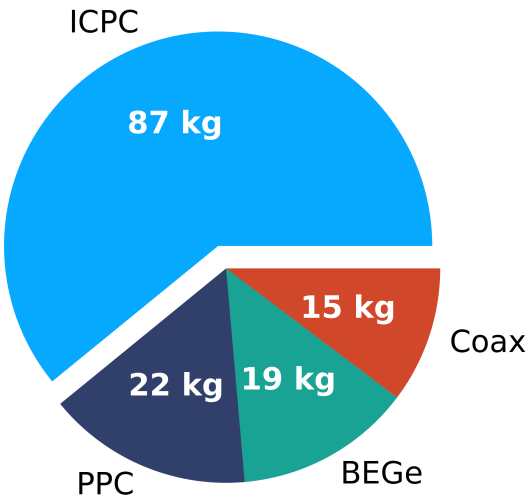

Figure 2.10: Fraction of total LEGEND-200 HPGe detector array by different detector types

For the first year of physics data taking, the array consisted of 101 different detectors — totaling 142 kg of mass — with four types of detector present. Figure 2.10 shows the breakdown by mass of the four detector types: coaxial, BEGe, PPC, and ICPC. Of these 101 detectors during the year of operation, 10 were inoperable due to disconnection (totaling 10 kg) [2]. A further 13.2 kg·yr of exposure was removed from the final dataset due to detectors that were unstable during operation or were not able to be fully depleted — these detectors were however used during the multiplicity cut.

Each detector has a high voltage cable, a charge sensitive amplifier, and a signal cable. The cables are unshielded pico coaxial cables from Axon' made from ultra-pure copper. The first stage of the charge sensitive amplifier is a custom-designed board called a low-mass front end amplifier: its design minimizes the mass of electronics near the sensitive p-type point contact of the detector. It consists of a 400 fF feedback capacitor and an $\sim$ 1 GΩ sputtered amorphous germanium resistor, where the capacitor is made from the sputtered Ti/Al traces on the board. The signal cables are routed to a platform $\sim$ 2 m above the detector array called the bird's nest; this platform contains a second-stage amplifier for the germanium detectors as well as high-voltage filtering. It additionally contains amplifiers for the detectors used for LAr scintillation light read-out. Signal cables pass out of the bird's nest to a feed-through flange that is then connected with the data acquisition (DAQ) system.

2.2.2 Active shields

While the use of radiopure materials and pulse shape discrimination in germanium is effective at cutting radioactive backgrounds, further background rejection capabilities can be provided through the use of active shields.

Water tank

A 590 cubic meter ultra-pure deionized water tank contains both the liquid argon cryostat and the HPGe array [2]. The water provides excellent neutron moderation from thermal neutrons generated by cosmic rays impinging on the steel cryostat walls or rock walls of LNGS. In addition, the water tank is instrumented with 63 photomultiplier tubes (PMTs) to detect Cherenkov light from cosmic muons that pass through the volume. Figure 2.8 shows the water tank and its instrumentation. GERDA found that without the muon veto the muons would contribute 3×10^{-3} counts/keV/kg/yr to the background index at $Q_{\beta\beta}$ [95].

Liquid argon detector

Argon scintillation: When ionizing radiation passes through argon, some argon atoms become ionized and can form excimer pairs. An excimer state is a quasi-bound state consisting of pair of argon atoms that share one electron — called a molecular Rydberg state — and is denoted by Ar_2^+ [96]. When these excimer states decay, they release some of their energy through photoemissions. Because of the spin of the electron, there are two possible states that have very different lifetimes. The lower energy state is the triplet state and has a lifetime of the order 1500 ns; the higher energy state is the singlet state and has a lifetime of 6 ns [96]. Both excimer states have the same scintillation emission spectrum, shown in Figure. 2.11. The emission is strongly peaked at 128 nm with a full width at half maximum (FWHM) of about 10 nm. This wavelength of light is shorter than ultraviolet light and thus termed vacuum ultraviolet (VUV).

On average, around 40,000 VUV photons are created per MeV of energy deposited in liquid argon [96]. Liquid argon is transparent to its own photoemissions, thus it is characterized as a scintillator.; the attenuation length for a VUV photon is therefore quite long in the LAr, on the order of 40 cm [98]; however, the VUV photon is ill-suited for current photodetectors. Most photodetectors rely on detection of light via amplifying electrons generated through the photoelectric effect [57]. But VUV photons typically have a

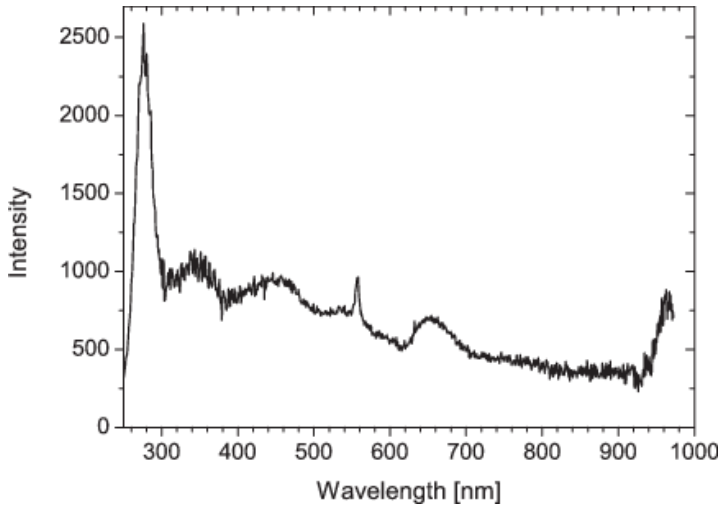

Figure 2.11: Emission spectrum for liquid argon scintillation light, from Ref. [97].

short absorption length in the optical materials used in photodetectors, and are therefore typically absorbed before reaching an active volume of a photodetector. It turns out, due to material properties, that the optimal quantum efficiency for most photodetectors occurs at blue-green wavelengths. It is therefore highly advantageous for LEGEND to wavelength shift the light before collection.

The liquid argon detector: the LEGEND-200 liquid argon cryostat contains 64 m^3 of ultra-pure ($\sim$1 ppm impurities) liquid argon [2]. In order to improve photocoverage and maximize the probability of detecting photons from scintillation events, LEGEND-200 uses a cylindrical 138 cm diameter wavelength-shifting light reflector (WSLR) located concentric with the HGPe array. This light reflector is composed of ultra-pure Teflon coated with tetraphenyl butadiene (TPB) — TPB is a wavelength shifting material that converts the VUV light into blue wavelength light. Scintillation light is collected and read-out from two wavelength-shifting fiber barrels that are roughly 1.5 m in length. One barrel sits in the interior of the HPGe array, and one sits directly on the exterior of the array. These barrels consist of St. Gobain BCF-91A fibers that are coated in TPB to convert VUV light to blue; the optical fibers then also shift blue wavelength light to green wavelengths [99, 100]. These optical fibers are coupled to silicon photomultiplier (SiPM) photodetectors in order to read-out the scintillation light that has entered the fibers. Green wavelength light is optically matched to the optimal quantum efficiency of the SiPMs chosen for LEGEND-200 [99]. The wavelength-shifting fibers also increase the photocoverage of the liquid argon detector by increasing the geometric probability of a VUV photon being absorbed into the photon-detection system. Additionally, every

germanium detector sits on a base plate made of scintillation material called polyethylene naphthalate (PEN) that converts VUV light to blue [100]. Figure. 2.12 shows the optical emission and absorption spectra of the different materials in the scintillation light detection chain. The photon detection chain is summarized here:

Light collection chain in LEGEND-200

1. VUV scintillation light is created in the liquid argon
2. VUV scintillation light is absorbed by TPB coated material, either the WSLR or fibers, or PEN base plate
3. TPB/PEN wavelength shifts VUV to blue and re-emits it
4. The blue-wavelength photon is absorbed by the optical fibers
5. The optical fibers wavelength shift the photon to green
6. The green photon is detected by the SiPMs.

The LEGEND-200 liquid argon detector light readout consists of SiPMs mounted to the tops and bottoms of the fiber barrel. The SiPMs are 3×3 mm in size and thus allow nine optical fibers of size 1×1 mm to be optically coupled to each SiPM in order to increase the number of photocoverage of the fibers. In order to reduce the number of read-out channels, nine SiPMs are connected in parallel to form one module and one readout channel [99]. SiPM modules are placed on the top and bottom of the fibers to minimize light loss from attenuation. There are 58 SiPM channels: 18 from the inner barrel and 40 from the outer barrel. Correspondingly, the inner barrel has 162 SiPMs and the outer barrel is instrumented with 360 SiPMs. The inner barrel increases the optical coverage of the liquid argon detector, and represents a key difference between the liquid argon detector in Germanium Detector Array (GERDA) and LEGEND-200. Further details of the optical coverage of the liquid argon detector and working principles of SiPMs are deferred until section 5.2.

2.2.3 Dominant backgrounds

Radioactive backgrounds that have energy above $Q_{\beta\beta}$ can all deposit some fraction of their energy in germanium and create a background event that mimics a signal-like neutrinoless-double beta decay event at

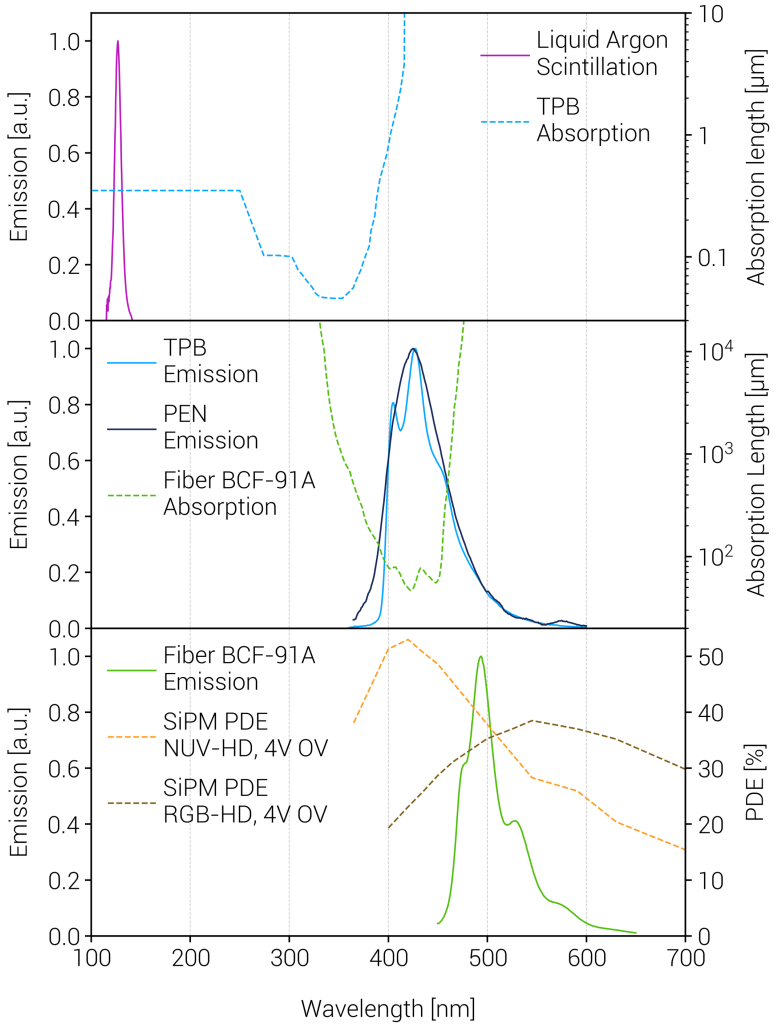

Figure 2.12: Emission and absorption spectra of wave-length shifting material in the LEGEND-200 detector, plot courtesy of Andreas Leonhard.

$Q_{\beta\beta}$. Common high energy radioactive backgrounds in LEGEND-200 include the following:

Radioactive background present at $Q_{\beta\beta}$

1. **High energy alphas:** most often come from decay products in the uranium and thorium decay chains. These isotopes are embedded in trace amounts in the HPGe array support structure, cabling, and front end electronics.

2. **High energy betas:** these can also come from decay products in the uranium and thorium decay chains. In addition, decays from ^{42}K which as an energy end point of 2.5 MeV are also a major background. The ^{42}K is populated by decays of ^{42}Ar, which is a cosmogenically activated isotope of argon present in trace amounts in the liquid argon detector.

3. **High energy gammas:** again often come from decay products in the uranium and thorium chains.

The expected amount of radioactive background near $Q_{\beta\beta}$ can be estimated using results from radioactive assays of material components and performing Monte Carlo simulations of the LEGEND-200 geometry. This results in what is called an assay-based background prediction. Background prediction is tricky business; presentation of the LEGEND-200 background model is available in other references [101].

2.2.4 Background suppression techniques

With the construction of the LEGEND-200 experiment covered, we are in the position to summarize the background suppression techniques that informed the design of the experiment.

Summary of background rejection techniques in LEGEND

1. **Radiopure material selection:** the best way to remove background is to remove it entirely from the experiment by selecting only construction materials that meet stringent purity requirements. This reduces ionizing radiation of all types.

2. **Multiplicity cut:** the closely-packed construction of the HPGe array enables removal of radioactive background events that deposit energy in multiple germanium detectors. This mostly

reduces background from gammas and cosmic muons as it is nearly impossible for a single beta or alpha decay to deposit energy in multiple detectors separated by a significant distance of liquid argon.

3. **Liquid argon anti-coincidence cut:** the scintillation light from radioactive background events that deposit energy in the liquid argon and a germanium detector enables tagging and removal of these events. This cut is effective at removing high energy beta and gamma events as they often deposit significant amounts of energy in the liquid argon before entering a detector. Alpha events are not removed by this cut as effectively as the only alpha events that enter a germanium detector do so near the point-contact where photocoverage of the scintillation light is poor. The liquid argon also offers passive shielding for all types of radiation.

4. **Water tank anti-coincidence cut:** the Cherenkov light from muons passing through the water tank allows removal of these events. The water tank also offers passive shielding for all types of radiation.

5. **Germanium pulse-shape discrimination:** the shapes of germanium waveforms can be used to remove multi-site events, events occurring near the p-type point contact, and events depositing energy in the transition layer. The MSE cut is effective at removing high energy gammas that Compton scatter. The p-type point-contact fiducilization is excellent at removing alpha events and beta events occurring near that region. The transition layer fiducilization mainly removes beta events occurring on the remaining surface of the detector. These PSD techniques are enabled by highly-customized detector geometries.

2.3 LEGEND-1000

The LEGEND-200 detector demonstrates that an array of large-mass IC detectors can be deployed and operated in a liquid argon environment; this is a step along the path towards a ton-scale experiment called LEGEND-1000. LEGEND-1000 aims to cover the allowed parameter space for Majorana neutrinos in the inverted ordering, assuming light neutrino exchange [89]. Using the range of NMEs obtained from

phenomenological methods, this corresponds to requiring a half-life discovery sensitivity of 10^{28} years. Assuming Poisson statistics, the required background index required to achieve this discovery sensitivity assuming a total exposure of 10 years$\times$ 1000 kg can be calculated as in Ref. [22] to be 1×10^{-5} ckky. This is a factor of 20 lower than the LEGEND-200 design goal. In order to reduce the background as much as possible, the design of LEGEND-1000 improves upon the LEGEND-200 iteration in many ways.

2.3.1 Experimental design

The conceptual design of LEGEND-1000 consists of 338 HPGe detectors with an average mass of 3 kg each. Every detector in LEGEND-1000 will be of the IC geometry, in contrast to the hodge-podge of the LEGEND-200 detector; this will contribute significantly to the background reduction by reducing required cabling and mitigating surface backgrounds through an increased volume-to-area ratio [89]. Unlike LEGEND-200, each string of germanium detectors will be individually deployable in order to avoid downtime during deployment. The LEGEND-1000 detector is designed to be built at LNGS; in order to further reduce cosmogenic and neutron background from the relatively shallow depth of the lab, the water tank and liquid argon veto sizes are designed to be increased and instrumented to a greater degree of photocoverage than in LEGEND-200. A conceptual rendering of LEGEND-1000 is shown in Figure 2.13.

The background in LEGEND-1000 will primarily be reduced through three key design strategies. First, the exclusive use of large-mass IC detectors will reduce the backgrounds coming from the uranium and thorium decay chains: reducing the amount of cabling required — which scales with the number of detectors — and increasing the volume-to-surface ratio helps mitigate these backgrounds [89]. Second, the uranium and thorium decay chains will be further reduced in the electronics chain through use of application-specific integrated circuits (ASICs) and lighter, flatter cabling [89]. Finally, the ^{42}K background will be reduced by utilizing argon sourced from underground wells. This **underground argon** is depleted in ^{42}Ar (by an expected factor of about 1400 [102]) because there is less cosmogenic activation of the isotope in the Earth's crust, and therefore the ^{42}K background is also reduced. Because underground argon is scarce, only a small volume of it will shield the HPGe array. This volume of underground argon is contained in what is called the re-entrant tube; the re-entrant tube sits inside a larger atmospheric liquid argon cryostat. The atmospheric liquid argon detector is necessary to help remove backgrounds coming from the re-entrant tube walls which

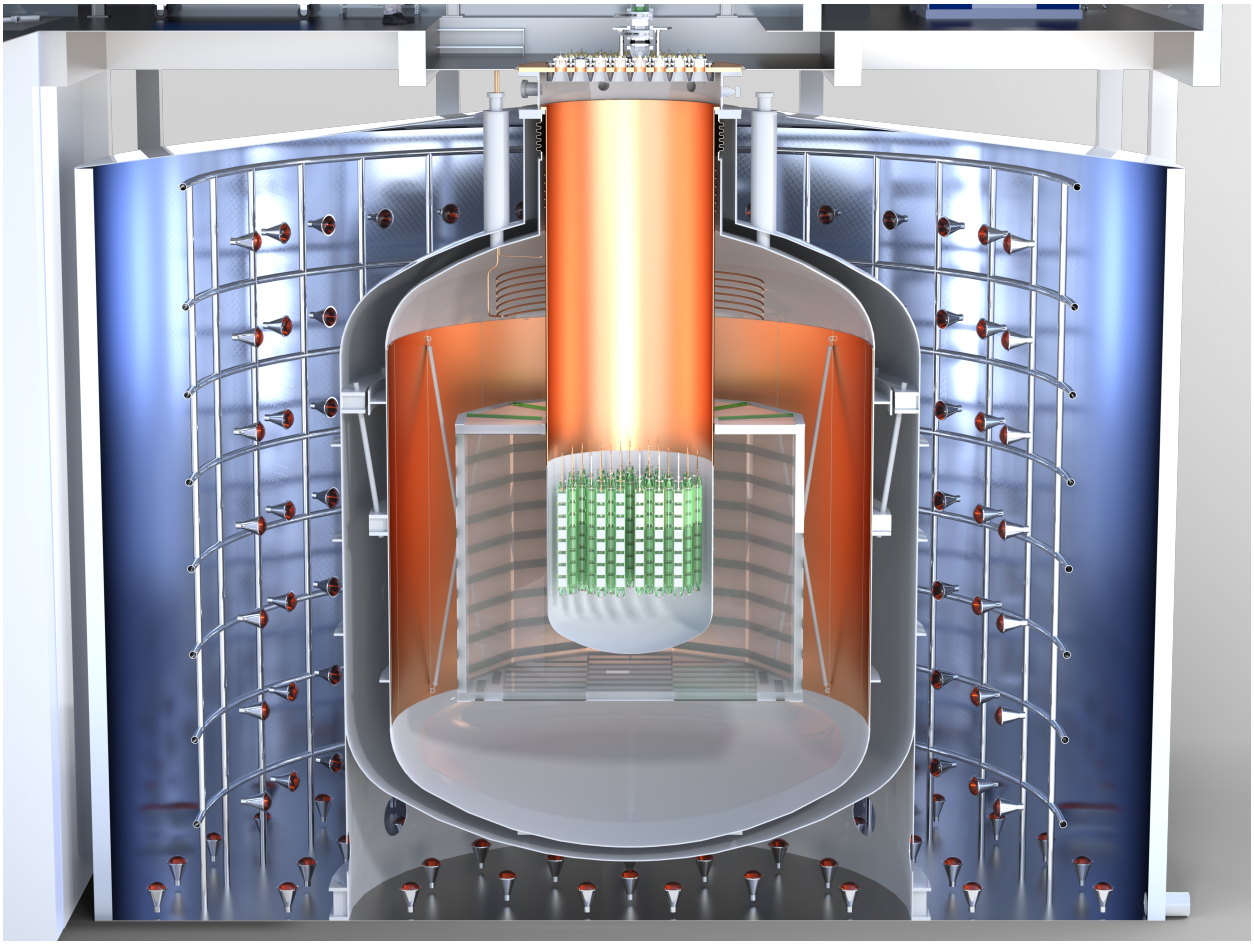

Figure 2.13: Artistic conceptual design of the LEGEND-1000 detector. The re-entrant tube is depicted by the copper volume immediately surrounding the HPGe and optical fiber array, shown in green. The re-entrant tube volume will be filled with underground sourced LAr which is depleted in ^{42}Ar. The copper volume exterior to the re-entrant tube is the outer LAr cryostat that will be filled with atmospheric sourced argon. This outer LAr cryostat is housed inside a water tank used as a muon veto.

Radioactive backgrounds	Reduction strategy
α	clean materials, PSD
β	clean materials, PSD , LAr anti-coincidence
γ	clean materials, PSD , LAr anti-coincidence
n	water tank and n moderators
Cosmogenics	LNGS overburden, water tank, LAr
ν	Irreducible

Table 2.1: Types of radioactive background and strategies to reduce them in the LEGEND experiment.

are close to the HPGe array due to the small volume of underground liquid argon. The underground liquid argon represents a significant contribution to the improvement of the background index.

2.4 Summary

The LEGEND experimental program seeks to discover neutrinoless double-beta decay, and targets a 3σ discovery sensitivity that completely covers the allowed parameter space for inverted ordered Majorana neutrinos. By using both high-purity germanium detectors enriched in germanium-76 and cleverly designed active and passive shields, LEGEND has highly optimized the experimental parameters necessary to probe the inverted ordering parameter space. A summary of the background reduction techniques used in the LEGEND experimental program are shown in Table 2.1. The LEGEND-200 detector is the first step towards this goal, and has finished its first year of physics data taking — results of the first neutrinoless double-beta decay search are presented later in this thesis. The proposed ton-scale LEGEND-1000 detector will leverage the knowledge gained from operating LEGEND-200 in order to reach a discovery sensitivity of 10^{28} years.

Chapter 3

Operations and analysis strategy for the first year of LEGEND-200

This section introduces the nomenclature associated with LEGEND-200 data taking operations (discussed more thoroughly in Ref. [103]) so that the analysis strategy for the first year of physics data can be discussed. The analysis strategy presented here is abbreviated and solely exists for the purpose of explaining how the parameters for the statistical analysis were produced. For a more in-depth discussion of the analysis strategy, the reader is kindly directed toward Ref. [104].

3.1 LEGEND-200 Operations

Construction of the LEGEND-200 experiment began in late 2020, and a test deployment of 60 kg of germanium detectors were installed in 2022 in a configuration known as LEGEND-60 that was missing the inner fiber shroud. Commissioning of the LEGEND-200 detector, as described in the previous section, and data taking both started in October of 2023. The operations of the LEGEND-200 detector and its data taking cycle are divided into the following steps:

LEGEND operation cycle

1. Major hardware changes (such as maintenance on electronics, or deploying new detectors) triggers the start of a new **period** [101].

2. Within each period, a **run** is the data collected in a full week and consists of the following:

   (a) A **calibration** dataset starts each run: data are collected for 24 hours with four ^{228}Th sources lowered into the detector in order to calibrate the energy scale and pulse shape discrimination (PSD) performance. During a calibration, the SiPMs are turned off due to the very high rate of scintillation light due to the presence of the sources.

   (b) **Physics** datasets, with the ^{228}Th sources removed, comprise the rest of the run until the next weekly calibration. In order to reduce file sizes, both physics and calibration data for a run are subdivided into **cycles** consisting of roughly one hour's worth of data.

3. Runs are continuously performed until a hardware change is proposed and then this data taking routine starts again with a new period.

During its first year of data taking, the LEGEND-200 detector had 9 periods of operation; of those, only periods 3-4 and 6-9 were deemed suitable for neutrinoless double-beta decay analysis. Detectors that were disconnected were classified as OFF during the analysis, and detectors that showed instability in energy reconstruction were used only in determining the multiplicity cut and were classified as anti-coincidence (AC) mode [105].

3.1.1 Data acquisition

The data acquisition (DAQ) strategy of the LEGEND-200 experiment aims to collect only as much necessary information to support a neutrinoless double-beta decay analysis without any extraneous data collected. To this end, the digitization window for germanium events was highly optimized. Acquired germanium waveforms were 8192 samples long, with the rising edge centered in the window; this way, enough samples of the baseline and decay tail were stored to provide accurate energy estimation. The sampling rate of the digitizer — 62.5 MHz — is also sufficient to gather PSD information from the rising edge during charge

drifts on the order of 1 μs. The waveforms acquired from SiPM channels were also the same digitization record length and sampling rate. The waveforms acquired from the water tank PMT channels are of a shorter length of 220 samples, and higher sampling frequency of 250 MHz.

The DAQ system is set to trigger on 25 keV energy events in any germanium detector, as estimated by the ADC FlashCAM system via its field-programmable-gate-array-implemented trapezoidal energy estimation filter [2, 106]. The data recording strategy varies slightly between calibration and physics data taking. During physics data taking, when the DAQ triggers, all germanium, SiPM, and PMT channels are readout and digitized simultaneously — because it is important to record the total energy deposited in the entire detector during an event in order to discriminate between signal and background events. However, in calibration data, in order to not balloon data rates during the high event rate data taking, only the germanium channels with a DAQ-estimated energy above threshold are digitized.

Additionally, two auxiliary transistor-transistor logic (TTL) channels are fed into the DAQ triggering system. The forced trigger, which sends a logic signal every 10 seconds, causes the DAQ to acquire data from the entire array. This forced trigger is helpful during physics data to acquire empty waveforms to monitor noise performance. Forced trigger events where all HPGe detectors have no energy depositions are called baseline events. However, every 20 seconds a PB-5 pulser system sends a charge signal to the front-ends of every germanium detector synchronous with a forced trigger to read out the whole HPGe array. A TTL logic pulse from the PB-5 is also digitized in an auxiliary channel to ensure that the pulser fired. This pulser mimics the deposition of energy within a germanium detector and is useful for monitoring the gain stability of the front-end electronics, and for estimating the live-time of the detector array (and for ensuring that the detector is still operational). The timing structure means that both pulser events and baseline events have a period of 50 mHz, with the type of event captured by the forced trigger alternating every 10 s.

3.1.2 Blinding

LEGEND implements a strictly blind analysis, so as to not bias the analysis cuts applied. The blinding scheme is implemented using the energy estimator supplied by the FlashCAM ADC system: events within $\pm$25keV of $Q_{\beta\beta}$ as estimated by the DAQ are removed from the raw DAQ files accessible to analysts [2]. The original unblinded DAQ files are stored in a password protected location and are only accessible once

analysis routines have been cemented and an unblinding has been agreed upon by a collaboration consensus. Because the DAQ energy estimator can be different from the optimal energy estimators produced by full offline analysis, a cross-check of event energies is performed by using the energy calibration from the first run of a period determined from calibration data. This produces a so-called blinding calibration that converts DAQ energy units to calibrated energy for a given detector. For all subsequent calibrations in a period, the blinding calibration curve is applied to the DAQ event energies, and the location of the Tl-208 FEP in the spectrum is checked to determine if it is still within $\pm$ 5 keV of the correct 2614 keV location [104]. If this peak shifts more than 5 keV then analysis is stopped and a manual check of the blinding calibration curve is performed and a new blinding calibration curve is produced if necessary using this calibration's data.

3.2 Neutrinoless double-beta decay analysis

Recall that the experimental signature of neutrinoless double-beta decay is a Gaussian peak centered at $Q_{\beta\beta}$ when measuring the summed electron energy. The statistical analysis of a neutrinoless double-beta decay experiment is therefore a peak search analysis. The analysis procedure is responsible for dictating which events enter the final spectrum and computing at what energies those events occur. Thus the two main prongs of the analysis are event selection and energy estimation. A good event selection reduces the amount of accidental radioactive background in the final spectrum and enhances the sensitivity to neutrinoless double-beta decay. This analysis procedure allows for an analysis of the number of signal counts n_s; however, in order to place a limit or estimate the rate of neutrinoless double-beta decay $\Gamma^{0\nu}_{1/2}$ through the equation

Neutrinoless double-beta decay half-rate and analysis parameters

$$n_s = \frac{\ln 2 N_a}{m_{\mathrm{Ge}}} \mathcal{E} \varepsilon \Gamma^{0\nu}_{1/2}, \tag{3.1}$$

the analysis must produce three additional pieces of information. The total exposure $\mathcal{E}$ must be computed from the data, as well as the signal acceptance efficiency ε (incorporation of systematic uncertainties is discussed more in Sec. 4.3.2). The signal acceptance efficiency is the probability that an actual neutrinoless double-beta decay event would pass all event selection criteria and be included in the final spectrum. The

analysis is also responsible for two further parameters relating to the expected width σ and location $Q_{\beta\beta} \pm \Delta$ of the (approximately) Gaussian peak from neutrinoless double-beta decay events. These parameters are discussed in detail below.

3.2.1 Analysis workflow

The analysis is performed using an automated workflow built on the SnakeMake program [107]. All steps in the analysis are performed using Python; a secondary analysis software stack was built using the Julia programming language but is not discussed here [108].

Unaltered binary files from the DAQ are first processed to the **raw** stage using the software library `daq2lh5` [109] where the files are converted to a proprietary hdf5 format called `lh5` and blinded. In order to to insure that all data adhere to pre-specified formatting conditions, the DAQ data are converted into `LGDO` [110] or legend data objects. During this stage waveforms from germanium and SiPM channels are filtered to reduce disk usage. For germanium waveforms, two copies of each waveform are created. One copy consists of the 48 microseconds of a waveform centered around the rising edge, sampled at the original sampling frequency — this saves all of the necessary PSD information. The other copy saves the sum of every 8 samples (and is thus downsampled by a factor of 8), but the entire length of the original waveform is saved — this saves all of the necessary energy estimation information. The SiPM waveforms are filtered by dropping the least significant bit. This filtering strategy reduces the overall disk usage by a factor of ~ 5.

Next, optimal filter parameters are determined for energy estimation and PSD classification using the `pygama` [111] software package on calibration data. Once these parameters are determined, they are applied to physics data using `dspeed` [112], the digital signal processing software package. Energy and PSD information for the signal from a single sensor (HPGe, SiPM, or PMT) are called **hits**. Collections of hits are grouped into **events** that collect all channels that are coincident with the trigger. The event tier is used for data selection and represents the end product of the data processing routines.

3.2.2 Partitioning

The partitioning of data for analysis does not necessarily have to follow the period- or run-level partitioning applied to the raw data on-disk; analysis partitions are formed of stable periods of data taking where the

analysis parameters do not change significantly, and thus can form sets of runs of raw data. Partitioning was performed based on both energy response stability as well as PSD calibration stability [101]:

Partitioning scheme

1. Energy calibration and gain stability. Calibration constants from the first calibration in a period were applied to subsequent runs and the position of the 2615 keV peak of ^{208}Tl was tracked. The differences between other calibrated peak locations and their literature values was also monitored.

2. Energy resolution stability. The energy resolution of the 2615 keV peakof ^{208}Tl was monitored for stability.

3. Noise stability. The root-mean square (RMS) of the baseline was monitored for fluctuations.

4. PSD stability. The location and shape of the A/E parameter for events in the 1150 - 1300 keV range were monitored for large fluctuations between runs.

Once partitions were determined, the optimal filter parameters for energy estimation and PSD classification were determined on a by-detector basis using all the data for that detector in the partition. This increases the statistics and allows improved estimation of the optimal filter parameters. The collection of all data with partition-level parameters applied to the digital signal processing (DSP) is called the partition-event tier and this data forms the basis of all analysis.

3.2.3 Energy scale

The energy scale analysis sub-task determines the optimal energy estimation for the data, as well as the expected shape and location of the neutrinoless double-beta decay peak for the statistical analysis.

Energy estimation

The energy of an event is proportional to the voltage across the feedback capacitor in the preamplifier, and the relationship between charge and voltage is determined by calibrations. The voltage across the capacitor can be estimated using several different techniques that are called energy estimators. An ideal energy estimator

is optimal, meaning it produces an energy estimate that has the lowest variance among all estimators, and is unbiased, meaning that the difference between the calibrated energy and the true energy of the event is minimal. An energy estimator having the lowest variance is equivalent to the condition that the filter has the best energy resolution possible. Therefore, an energy estimator should filter the signal to minimize unwanted contributions from electronics noise. Most energy estimator filters operate by suppressing noise (either through averaging or differentiation) while integrating the signal for voltage estimation. The calibration strategy followed for analysis is as follows [70, 104]:

Energy estimation strategy

- Deconvolve the electronics response by applying a pole-zero correction. This removes the dominant RC decay tail. However, future analysis will incorporate a double pole zero correction at this stage to correct for remaining observed shorter time scale RC decay components.
- Apply an energy filter, such as a trapezoidal filter or cusp filter, to estimate the uncalibrated energy of the event. These filters have been optimized to produce the smallest possible energy resolution.
- Apply a drift time correction to account for charge trapping in the detector. Events that drift longer in the detector encounter more charge traps, so some charge is added back to these events.
- Identify known peaks in the resulting spectrum and use those to build a calibration curve following equation 2.5.

The complete spectra of all calibrations from all partitions is shown by detector type in Figure 3.1. This figure also illustrates the different features of the calibration spectrum, broken down by event topologies occurring from the incoming Tl-208 primary γ decay. Because the primary γ can pair-produce when it enters a detector, there is some chance that the subsequent 511 keV gammas from annihilation of the positron can both escape the detector: when this happens, all of the charge produced by this event is created in a very small volume that looks single-site. Because both 511 keV gammas escape, these events are called double

escape events and collectively they produce the double escape peak (DEP) and are nearly all SSEs[1]. There is also some probability that only one of the 511 keV gammas escapes the detector: this is called single escape. The absorbed gamma deposits its charge in some other region of the detector, away from the initial site of pair production — this leads to single escape events being a population of multi-site events. The collection of single escape events leads to the single escape peak (SEP) and are all MSEs. When both 511 keV gammas do not escape the detector, these events contribute to the full energy peak (FEP). The full energy peak (FEP) is predominately multi-site due to multiple Compton scatters of the 511 keV gammas before absorption via the photo-electric effect. Single-site photoabsorption of the initial Tl-208 primary γ are also possible; this causes the FEP to be mixed in its event topology.

SiPM and PMT analysis

At the start of each period, the optimal filter for determining the number of detected photoelectrons in the SiPMs is calibrated with the following strategy. The response of a SiPM to an incident light is a well-defined two-time constant decaying exponential whose amplitude is proportional to the number of detected photons; this will be discussed more in section 5.2. Pulses and their heights are found using a peak-finding algorithm applied to the smoothed derivative of the SiPM waveforms. The RMS of the baseline is used as a threshold for the peak-finding routine. Even with this threshold, some identified peaks will originate from noise. The uncalibrated peak heights can be calibrated by taking a histogram of the peak heights and identifying the locations of the SiPM response from 1 photon and 2 photons. These peaks can be identified as the first two peaks relative to the noise peak which is centered near 0. The threshold for the number of detected photoelectrons in a SiPM is the smallest calibrated pulse amplitude above the noise peak in the histogram. The number of photons detected per SiPM waveform is then determined.

The situation is similar with the PMTs used in the muon veto system. The height of identified peaks in PMT waveforms (which must be more than 3 times the RMS of the baseline of a pulse) are recorded, because the integral of a PMT's response is proportional to the number of detected photons. The gain of each PMT, determined via calibration, is used to determine the number of photons detected in each hit PMT [114]. The total number of detected photons across all PMTs are summed at the event level in order to

[1]Some DEP events have bremsstrahlung that make them appear as MSEs, which also happens to $0\nu\beta\beta$ decay events

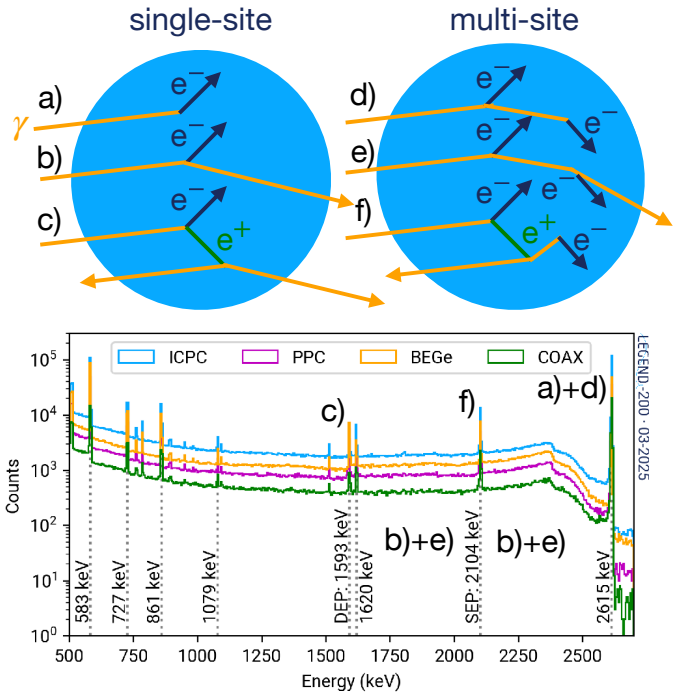

Figure 3.1: Example energy calibration spectra, as well as different γ-induced event topologies in germanium illustrated and identified in the spectra, based off of Ref. [113]. Full energy collection from the Tl-208 γ can come from single-site events a), as well as multi-site d). The incoming γ has some probability deposit a fraction of its energy and to Compton scatter out of the detector leading to the Compton shoulder, which comprised of both single-site events b), and multi-site events e). The incoming γ can also pair-produce: the immediate annihilation of the positron leads to two subsequent 511 keV gammas. These 511 keV gammas can: either both exit the detector, leading to the double escape peak c); have only one gamma escape, leading to the single escape peak f); or have neither gamma escape contributing to the full energy peak via process d). The double escape peak is comprised of single-site events as the charge deposition radius is very small, while single escape peak is inherently multi-site.

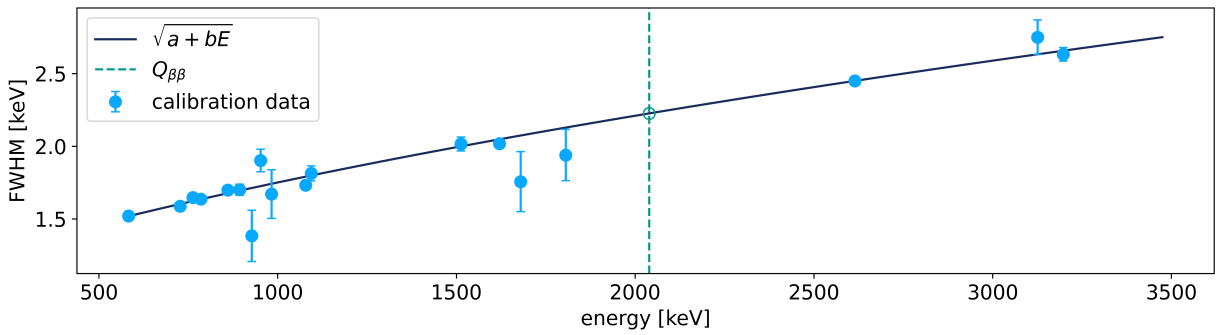

Figure 3.2: Example energy resolution calibration curve to extract the energy resolution at $Q_{\beta\beta}$ for ICPC detector V07302A from all calibration data in its first partition.

determine the anti-coincidence cut.

Energy resolution measurement

The energy resolution at $Q_{\beta\beta}$ and thus the expected shape for a peak from neutrinoless double-beta decay can be computed using the measured energy resolutions from calibration peaks. The peak shape at $Q_{\beta\beta}$ is assumed to be a Gaussian; as discussed in the next chapter, the peak could have a slightly non-Gaussian shape, such as a low-energy tail due to incomplete charge collection as seen in the MJD experiment [71]. Because the energy resolution is a function of energy given in Sec. 2.1.4 as

Energy resolution

$$\sigma(E) = \sqrt{a + bE + cE^2}, \tag{3.2}$$

the constants in this equation can be found by fitting this function to data from calibration. The energy resolution at $Q_{\beta\beta}$ was extracted using this formula with $c = 0$. An example of this is shown in Fig. 3.2.

This procedure was performed for all detectors per partition; a visualization of the energy resolutions obtained for LEGEND-200 is shown in Fig. 3.3. The ICPC detectors display fantastic energy resolution despite their large size and thus propensity for more charge trapping. The coaxial detectors have the worst energy resolution out of all detector types due to their larger capacitance and therefore larger noise. The uncertainty on the energy resolution was computed as the sum in quadrature of the statistical uncertainty, as

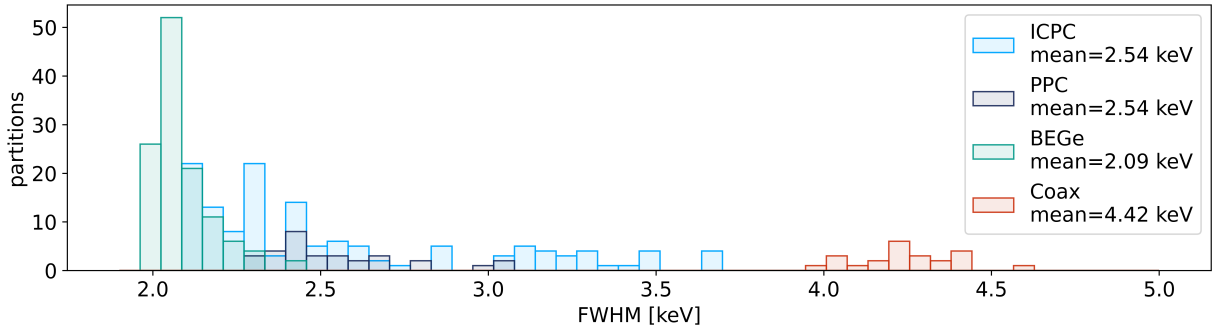

Figure 3.3: Histogram of energy resolutions at $Q_{\beta\beta}$ for all partitions, grouped by detector type. The coaxial detectors have the worst energy resolution due to their larger capacitance from the large surface-area p-type contact. These average energy resolutions are similar to those reported by GERDA and MJD collaborations [55, 56].

estimated by bootstrapping the parameters from the resolution curve fit, and the difference in the resolution at $Q_{\beta\beta}$ if the resolution curve was fit with $c \neq 0$ in the resolution model.

Energy bias measurement

The energy bias is the remaining systematic offset between the expected energy of a neutrinoless double-beta decay event and where the calibration would place its energy:

Energy bias definition

$$\Delta = E_{cal} - E_{true}. \tag{3.3}$$

The energy bias is typically non-zero due to effects such as analog-to-digital converter (ADC)-unit non-linearity causing systematic shifts in the calibration. The energy bias is approximated by computing the measured energy bias at two known gamma lines observed in calibration data, the 1593 keV DEP and 2104 keV SEP ^{208}Tl peaks, and then interpolating the energy bias at $Q_{\beta\beta}$ by fitting a line between these two points. In this procedure E_{cal} is the location of the median of the gamma peak in the spectra after calibration and E_{true} are the literature values of these gamma energies. An example of this procedure is shown in Fig. 3.4. The uncertainty on the energy bias was determined by a parametric bootstrap of the fit parameters. This procedure follows what was used for the physics analysis of the GERDA experiment [115].

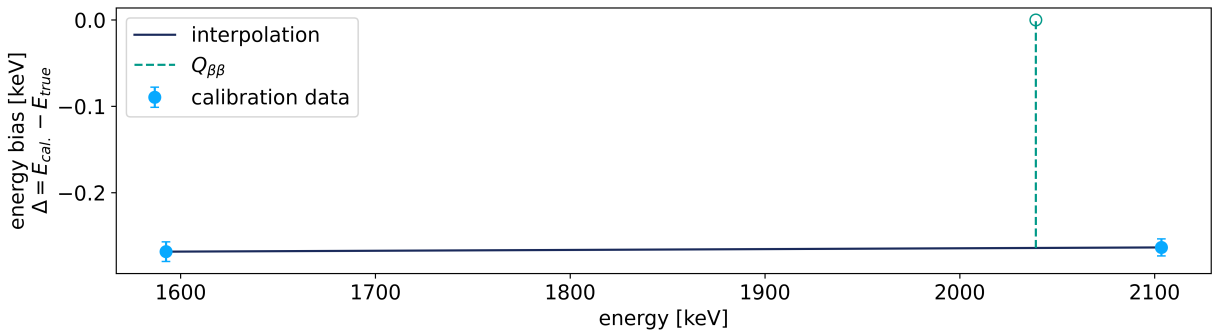

Figure 3.4: Procedure for determining the energy bias at $Q_{\beta\beta}$ by interpolating the residuals of the partition-level calibration for the SEP and DEP, shown for ICPC detector V07302A partition 0.

This procedure of estimating the energy bias was applied to all detector partitions, and the results of this are shown in Fig. 3.5 grouped by detector type. These distributions are not centered at 0, suggesting that there is some systematic effect unaccounted for in the energy calibrations, such as ADC non-linearity.

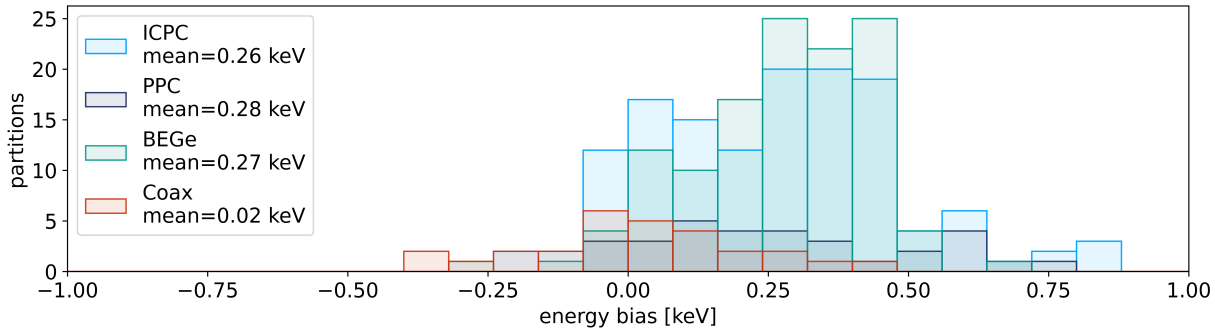

Figure 3.5: Histogram of energy biases at $Q_{\beta\beta}$ for all partitions, grouped by detector type. All detectors display roughly the same energy bias, suggesting that the calibration procedure has a systematic effect independent of detector geometry. This could be due to, for example, ADC non-linearity.

Now that the energy of energy depositions in the germanium can be determined, we now turn our attention to the process of event selection.

3.2.4 Data cleaning

The first step in event selection is removing events that do not look like valid energy depositions in germanium, which are defined as positive polarity pulses that have a flat baseline, a rising edge, and a decaying tail. This process is called data cleaning, as it removes events before further analysis cuts are computed and

applied. Events such as high voltage discharges, hard pileup (when two valid energy depositions happen in one digitization window), and retriggered events (soft pileup) are examples of events removed by data cleaning. The following cuts are applied [116]:

Quality cuts

1. Saturation cut. Events that have at least one sample at the highest or lowest register of the digitizer range are cut. Negative saturation events likely occur from high voltage discharges, after which the detector takes several hundred microseconds to recover to normal baseline. So events for 10 ms after a negative saturation event are also cut.

2. Baseline cut. Cuts on the slope and the RMS of the baseline were performed to make sure that the waveforms passing cuts have good noise performance and flat baselines.

3. Tail cuts. The RMS of the tail was also monitored and large excursions from nominal RMS values were cut. No cut was performed on the slope of the tail as that is an energy dependent parameter, and data quality cuts should be energy independent.

4. Rising edge cuts. Cuts on the rising edge were performed carefully to ensure they were not energy dependent or pulse-shape dependent. Noise bursts were cut by checking if the maximum and minimum of the windowed section of the waveform both deviated from the baseline by 300 ADC units, absolutely, and occurred within 1 μs of each other. Events with rise-times shorter than 32 ns are likely noise and removed.

A machine learning (ML) based approach to the data cleaning was also pursued and used as a cross-check for the cut-based approach described above. The ML approach has the added benefit of labeling the different types of populations that do not pass quality cuts.

Data cleaning signal acceptance efficiency

The signal acceptance efficiency for quality cuts for a detector i during one run k of a partition can be determined using the forced trigger events across all channels and the events from the FEP of ^{208}Tl during calibration. The FEP events are a good signal proxy because they have the same front-end electronics

response as a neutrinoless double beta-decay event. The probability of a FEP event passing quality cuts is the number of FEP events that pass quality cuts n_i divided by the total number of FEP events N_{FEP}, which is determined from a peak fit to that detector's FEP. The signal acceptance efficiency also requires that all other detectors must have empty waveforms that also pass analysis cuts. This number is approximated using the total number of forced triggers that have all detectors survive all quality cuts m divided by the total number of forced triggers N_{FT}:

$$P_i^k = \frac{n_i}{N_{FEP}} \frac{m}{N_{FT}}. \tag{3.4}$$

The signal acceptance efficiency for quality cuts for a detector for the whole partition is then approximated as the average over the number of runs N_R:

$$P_i = \frac{1}{N_R} \sum_{k=1}^{N_R} P_i^k. \tag{3.5}$$

Waveforms from the FEP were scaled in energy following Ref. [117] so that the signal acceptance efficiency could be determined for events at $Q_{\beta\beta}$. The uncertainty on this signal acceptance efficiency comes from propagation of Poisson errors on the number of counts. The results of the quality cut signal acceptance calculations are visualized in Fig. 3.6. The average acceptance for all detector types for all partitions is 0.97 — this number represents only a very small amount of event sacrifice and suggests the cuts are performing well.

3.2.5 Pulse shape discrimination

For detectors with a point-contact geometry (BEGe, PPC, and ICPC) a PSD cut was applied during event selection using the A/E parameter. For all detectors of these geometries a low A/E cut was applied to remove multi-site events; however, differences in passivated surface sizes and detector geometries from two different detector manufacturers, Ortec and Mirion, necessitated different approaches to surface cuts. With the exception of five ICPC detectors ordered from Mirion, all other Mirion ICPC and all BEGe detectors had a high A/E cut applied to remove surface events. Due to the larger passivated surface on ICPC Ortec detectors and all PPC detectors (as well as the five Mirion ICPCs) a late charge LQ cut was applied to remove surface events. The details of this LQ cut, and the determination of all other PSD cuts, will be

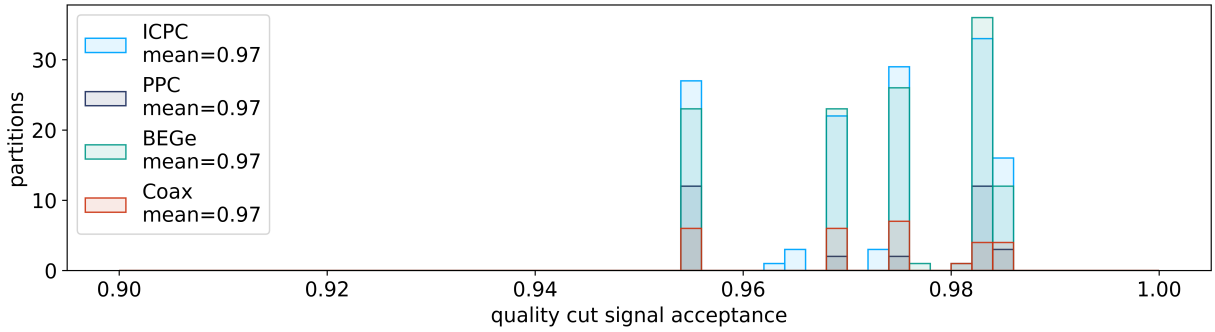

Figure 3.6: Histogram of the quality cut signal acceptance efficiency for all partitions, grouped by detector type. The quality cut acceptance is quite high, and the absence of any distinct shape in the distributions suggests that there is no systematic effects affecting the cut performance. These appear discrete with basically all detectors roughly coinciding in efficiency for a given partition. This is due to the fact that the survival fraction of forced triggers in Eqn. 3.4 is the same for all detectors per run per partition and is much smaller than the FEP survival fraction of a given detector and thus dominates the probability.

briefly described below [55, 94, 118]. The remaining coaxial detectors had a more complex pulse shape discrimination routine applied to their event selection: an artificial neural network was used to cut MSE and a rise time cut was used to remove surface events, similar to what was done in GERDA [91, 117].

The multi-site event cut is based on a calibrated $A/E_{cal.}$ classifier and its critical value $A/E_{crit.}$, which is set such that all events with a smaller classifier are identified as multi-site events and removed from the analysis. Uncalibrated A/E values are computed for calibration data using an estimate for the maximum current of the waveform A and the energy of the waveform, before charge trapping correction, E. The estimation of A varies between detector types, but in all instances involves smoothing the derivative of the charge-space waveform and taking its maximum value. The uncalibrated A/E values from all events passing quality cuts in a calibration run are then roughly calibrated: the A/E parameter is expected to follow a Gaussian on top of two exponentially modified Gaussians. This is because SSEs should have a Gaussian distribution, and MSEs and surface events cause tailing at high and low A/E values. So the uncalibrated A/E values are fit with this function and are then normalized by its μ, σ so that the resulting distribution for $A/E_{cal.}$ is centered at 1 with width 1 [119].

Two further corrections are applied to the calibrated parameter. For ICPC detectors, because they are so large, two distinct $A/E_{cal.}$ populations are visible when plotted with respect to drift-time. Recall that DEP events are more likely to occur in the corners of the detector where both gammas are more likely to escape.

This leads to a cluster of longer drift-times from events in the lobes of the detector where there is only a small change in weighting potential, and a shorter drift-time population from the corners closer to the p-type contact. Simulation confirms that drift time and A/E saturate and reproduce this bimodal behavior for DEP events [120]. A drift time correction is applied to center both of these populations at 1.

Finally, for all detectors, an energy correction is applied [2]. The energy correction is estimated by using energy windows of the Compton continuum from 900 keV to 2.3 MeV. For each 20 keV slice, the centroid and width of the $A/E_{cal.}$ distribution is fit; these centroid and widths are used to compute $\mu_{A/E}(E)$, by fitting computed values to a linear model, and $\sigma_{A/E}(E)$, by fitting it to a heuristic model of $\sqrt{a + bE^c}$. The normalized $A/E_{classifier}$ is then computed as

$$A/E_{classifier} = \left(\frac{A/E_{cal.}}{\mu_{A/E}(E)} - 1 \right) \cdot \sigma_{A/E}(E), \tag{3.6}$$

as given by Refs. [104, 119].

The critical acceptance region for the multi-site event cut was determined by using events in the double escape peak of ^{208}Th as a proxy for single site events. The DEP is a decent signal proxy as all DEP events deposit their charge in a well-localized region of the detector (although the initial charge cloud size is different compared to neutrinoless double-beta decay events). The DEP has both gammas produced during the pair-production from the full-energy 2615 keV gamma leaving the detector. The cut value is determined by computing the $A/E_{crit.}$ value for which there is a 90% survival fraction (SF) of events in the DEP. This 90% SF value allows for high signal acceptance efficiency at $Q_{\beta\beta}$ while also removing a large fraction of MSEs. The average survival fraction for events in the SEP — a proxy for MSEs — are 10%, and the survival fraction for events in the Compton continuum near $Q_{\beta\beta}$ is roughly 40%. The A/E classifier calibration constants and the A/E critical value are computed at the partition level using all calibrations during that partition. All events during physics data in a partition have the same A/E classifier calibration applied to them, and physics events with a classifier smaller than the critical value for the partition are considered background-like multi-site events.

The surface event cut was realized in Mirion detectors by using a second $A/E^{high}_{crit.} = +3$ and cutting

[2]The A/E parameter decreases as a function of energy because higher energy events create a larger charge cloud with more self-repulsion and thus a slightly lower maximum current. The variance in A/E also decreases at higher energy because of higher signal to noise ratio.

events with an $A/E_{classifier}$ greater than that critical value. This critical value is historical [117]. It cuts on average about $\mathcal{O}(1\%)$ of events not cut by the MSE cut. This cut was not used on Ortec detectors as the signal acceptance efficiency was too low.

The late charge (LQ) parameter was used as a surface event cut for the Ortec detectors instead. The LQ parameter is a measure of the rate of change of the last 20% of charge collection of a waveform: this is a good parameter because surface events have significant delayed charge collection from charge deposition in the transition layer. The uncalibrated $LQ_{uncal.}$ parameter is computed again for all data from all calibrations in a partition. A drift time correction and energy correction are also applied, see Ref. [94] for more details. The critical value for the normalized classifier $LQ_{classifier}$ are also set at $LQ_{crit.} = +3$ and physics events with a classifier larger than that value are classified as surface events and are cut.

PSD signal acceptance efficiency

To first order, the signal acceptance efficiency for neutrinoless double-beta decay events at $Q_{\beta\beta}$ is estimated using the Tl-208 DEP as a proxy for signal-like event topologies. The efficiency of a cut (MSE or surface) is determined by computing the number of events that survive the cut N_s divided by the total number of events in the DEP; this is done by fitting the (background-subtracted) DEP with a Gaussian function to both events that survive the cuts to obtain N_s as well as to the events that fail the cut to obtain the number of cut events N_c. The uncertainty on the number of events obtained from these fits σ_s, σ_c are determined from the Hessian matrix from the fit result. The cut efficiency is then

$$\varepsilon = \frac{N_s}{N_s + N_c}, \tag{3.7}$$

and its uncertainty is

$$\delta_{\varepsilon,\, stat.} = \varepsilon(1-\varepsilon)\sqrt{\left(\frac{\sigma_s}{N_s}\right)^2 + \left(\frac{\sigma_c}{N_c}\right)^2}. \tag{3.8}$$

The PSD cut signal acceptance efficiency was computed for both the MSE cut (the low A/E cut), ε_{MSE}, and the surface cuts (the high A/E or LQ cuts, depending on the detector), $\varepsilon_{Surf.}$.

The DEP is not the best proxy for signal-like events: there are volume dependent and energy dependent effects that can change the signal acceptance at $Q_{\beta\beta}$. To account for energy dependence of the signal

acceptance efficiency, two further studies were performed.

^{56}Co calibration: first, a special ^{56}Co calibration run allowed the signal acceptance efficiency to be measured across a broad range of energies, following a procedure done for MJD [55]. The ^{56}Co calibration source has seven double escape peaks at different energies: computing the signal acceptance efficiency for these peaks and then linearly fitting them allows for interpolation of the efficiency at $Q_{\beta\beta} = 2039$ keV. This information is used as a correction to the signal acceptance efficiency computed during normal calibrations by $\Delta\varepsilon_{Co} = (2039 - 1592.5)m_{fit}$, where m_{fit} is the slope of the fit to the ^{56}Co DEP signal acceptance efficiencies, and 1592.5 keV is the energy of the Tl-208 DEP used to tune the PSD during calibration. The uncertainty on this correction is the quadrature sum of the uncertainty on the slope and the residual between the predicted signal acceptance efficiency for the Tl DEP from the fit and its computed value. All detectors of a given geometry were grouped together for this analysis due to insufficient statistics in a given detector.

Scaled waveform study: second, variations in noise over time can change the energy dependence of the PSD parameters, so the impact of noise on the signal acceptance efficiency at $Q_{\beta\beta}$ was studied. This was done by scaling down DEP waveforms and adding scaled noise obtained from empty waveforms during the forced trigger, akin to what was done in GERDA [117]. The signal acceptance efficiency for each of the PSD cuts was studied at each scaled energy. The partition-to-partition variation of signal acceptance efficiency was treated as an additional uncertainty on the efficiency correction, as it was assumed that the ^{56}Co efficiency correction takes into account the average shift due to energy dependence. This yields a correction $\delta_{\varepsilon_{noise}}$ and is on the order of $0 \pm 1\%$ for most detectors.

Finally, events expected to be from $2\nu\beta\beta$ decays were used to correct the signal acceptance efficiency for volume dependent effects. Events from the Tl-208 DEP do not sample the whole germanium volume uniformly as $0\nu\beta\beta$ decay events should; instead, they tend to originate from the corners of the detectors where there is a larger chance that both gammas leave the detector without depositing any energy. Events that survive all data cleaning and anti-coincidence cuts in the 1000-1300 keV energy range are expected to come from $2\nu\beta\beta$ decays and thus represent the correct signal topology and are uniformly distributed throughout the bulk of the detector. In order to have enough statistics to perform this study, all detector types had their data grouped together and analyzed. The centroid and width of the PSD classifier distributions are different for the $2\nu\beta\beta$ population compared to the Tl-208 DEP. To understand the impact of this difference in peak

shape arising from the two populations, the (averaged) critical acceptance values for the PSD cuts were transformed as if the Tl-208 DEP PSD distribution had the peak shape of the $2\nu\beta\beta$ spectrum. The resulting difference in the survival fractions of $2\nu\beta\beta$ events obtained using the transformed and untransformed critical acceptance values yields a correction $\Delta\varepsilon_{2\nu\beta\beta} \sim \mathcal{O}(-1\%)$. This was similar to the approach taken by MJD [55]. A further uncertainty on this correction was estimated by using the background model to compute the survival fraction of $2\nu\beta\beta$ events and comparing it to the DEP survival fraction, as was done in Ref. [91].

The overall PSD signal acceptance efficiency was then calculated as

$$\varepsilon_{PSD} = \varepsilon_{MSE} \cdot \varepsilon_{Surf.} \pm \Delta\varepsilon_{Co} \pm \Delta\varepsilon_{noise} \pm \Delta\varepsilon_{2\nu\beta\beta}. \tag{3.9}$$

The uncertainties on each of these terms were added in quadrature to compute the uncertainty on the PSD efficiency $\delta_{\varepsilon_{PSD}}$. The signal acceptance efficiency of the PSD cut by detector type across all partitions is shown in Fig. 3.7. The average acceptance efficiency is below the 0.9 value targeted during the cut setting procedures due to systematic effects regarding the energy dependence and volume dependence of the Tl-208 DEP signal proxy. Further study should be performed to optimize the cut thresholds to maximize the signal acceptance while minimizing accidental background. One possible way to achieve this is through multivariate cuts: instead of applying the MSE and surface cuts sequentially, the values of the both cuts can be simultaneously optimized and applied.

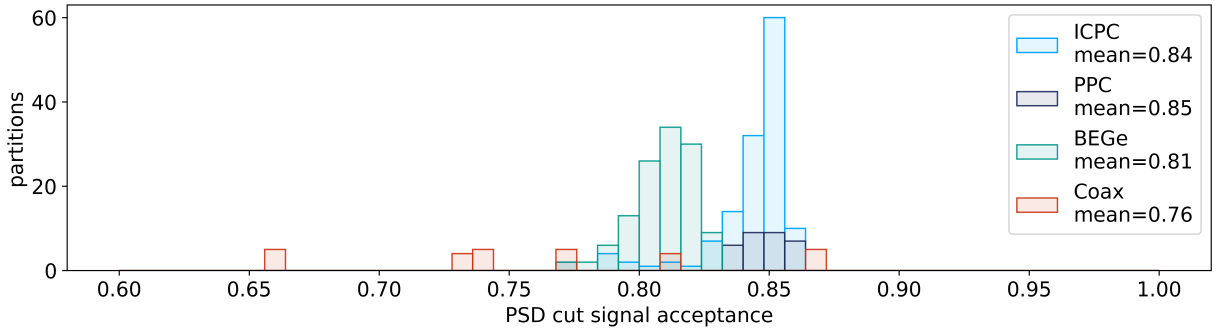

Figure 3.7: Histogram of the pulse shape discrimination cut signal acceptance efficiency for all partitions, grouped by detector type. The PSD acceptance for the coaxial detectors has a large variance due to the performance of the ML algorithm used to set the cuts. The signal acceptance efficiencies are lower than the target 0.9 during the cut setting procedures.

3.2.6 Anti-coincidence cuts

Because neutrinoless double-beta decay should deposit its energy entirely within the volume of only one germanium detector three anti-coincidence cuts are applied that look at energy depositions across different volumes: the multiplicity cut, the water tank veto, and the liquid argon veto. If an amount of energy greater than a critical threshold is determined to happen in more than one detector or detector-system the event is removed from the analysis.

Multiplicity cut: events that have significant energy deposition in more than one germanium detector are removed from the analysis. The energy threshold for this cut was set at 25 keV.

Liquid argon anti-coincidence: events that produce more than 4 total detected photons summed across all channels or have more than 4 SiPM channels each with energy above the noise threshold are classified as background and removed from analysis. The coincidence window with energy deposition in germanium was set to $[-1, 5]\,\mu$s relative to the germanium DAQ trigger time.

Water tank anti-coincidence: the water tank DAQ is triggered when a total of at least 0.5 photo-electrons are detected in any 4 PMTs directly under the LAr cryostat, 4 PMTs on the floor of the water tank, or 6 PMTs on the wall of the cryostat. The number of photo-electrons detected by the PMTs is estimated with an on-board filter on the DAQ. Two further analyses cuts were applied to remove DAQ triggers that did not arise from muons. First, DAQ triggers with a PMT multiplicity < 12 (defined as each channel having at least 0.5 p.e in it) and total energy < 30 photo-electrons are not counted as muon events. Secondly, triggers with a total number of photo-electrons less than half the number of hit PMTs are also not treated as muon triggers. These cuts remove events potentially due to scintillation of the reflective foil used in the water tank and correlated electronics noise. The identified muon trigger rate in the water tank after all analysis cuts is close to 35 mHz.

Anti-coincidence cut signal acceptance efficiencies

The signal acceptance efficiencies for the three anti-coincidence cuts can computed as follows.

Water tank anti-coincidence: the signal acceptance efficiency for this cut can be computed by calculating the number of forced triggers that survive the water tank anti-coincidence cut divided by the total number of forced triggers. It can also be estimated by computing the probability that given the muon veto

trigger rate of 35 mHz that it triggers during the digitization length of a germanium waveform of $8192 * 16$ ns for a signal acceptance efficiency of $1 - 4.6 \times 10^{-6}$. Because this probability is so close to unity, this signal acceptance efficiency is ignored during the statistical analysis.

Liquid argon anti-coincidence: the signal acceptance efficiency for this cut is approximated by computing the survival fraction of 1460 keV gammas ($\sim$10% intensity) from ^{40}K events obtained during physics data [3]. The ^{40}K background comes from trace amounts of the long-lived terrestrial isotope present in all materials used in the construction of the array. The ^{40}K events are selected as all events near 1460 keV surviving the muon, multiplicity, and quality cuts. Because 1460 keV gamma events come from ^{40}K decays via electron capture, they should deposit insufficient energy in the liquid argon to trigger the LAr coincidence cut and thus should be a good proxy for neutrinoless double-beta decay like events that only deposit their energy within a germanium detector. The survival fraction, and thus the signal acceptance efficiency were computed to be 0.93 ± 0.01. This is quite low, and is due to high levels of electrical cross-talk between HPGe and SiPM channels that the hyper-current-based SiPM analysis was sensitive to. Future work — such as template-based fitting SiPM waveforms and time-correlation-based veto conditions — is being pursued to improve the liquid argon veto acceptance efficiency

Multiplicity cut: the signal acceptance efficiency for this cut is the probability that a $0\nu\beta\beta$ decay event in one detector deposits energy (through bremsstrahlung or electrons escaping) into another detector. This probability is extremely small, and its computation is dominated by random coincidences and is negligible and taken to be unity.

3.2.7 Exposure

The exposure $\mathcal{E}$ can be calculated using the live time of each detector, T, multiplied by its mass, m. The live time is determined per partition i per detector by using the number of pulser events N_p^i divided by the pulser rate of 50 mHz:

[3] ^{40}K does not pose a background for $0\nu\beta\beta$ decay searches as the Q-values for both the β decay branch and the electron capture branch fall below the $Q_{\beta\beta}$ for germanium-76.

Exposure

$$\mathcal{E}_i = m_i \cdot LT_i = m_i \frac{N_p^i}{50 \text{ mHz}}. \tag{3.10}$$

It is important to note that LEGEND uses only the raw exposure in its analysis, and not the sensitive exposure. While the analysis cuts could be considered to impact the exposure of a detector (quality cuts reduce the live time), LEGEND instead chooses to place the impact of these cuts in the signal acceptance efficiency. In this way the exposure is a totally known parameter with no uncertainty, while all uncertainties are placed on the signal acceptance efficiency instead. This is the same as the treatment by the GERDA collaboration, although the MAJORANA DEMONSTRATOR (MJD) collaboration had uncertainties on both the exposure and signal acceptance efficiency. A plot of the exposures per partition is shown in Fig.3.8.

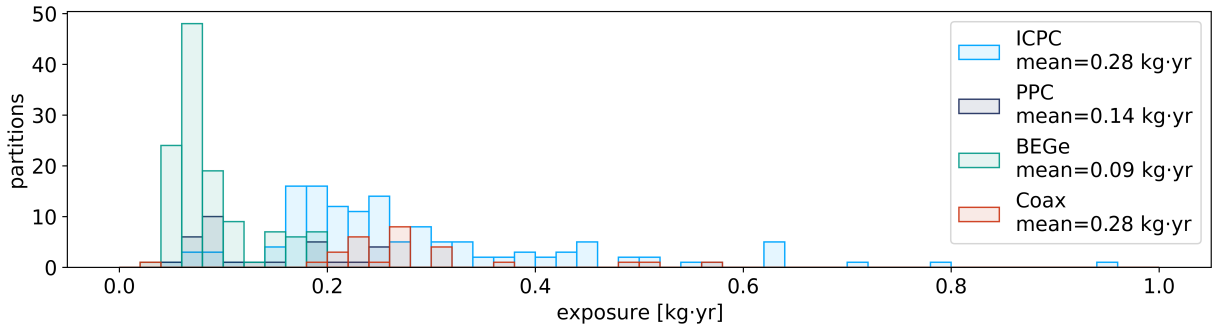

Figure 3.8: Histogram of the exposures of all partitions, grouped by detector type. Because most partitions had roughly similar live-times, the difference in exposures viewed by detector type illustrates the difference in masses between the geometries. Thus both the ICPC and coaxial detector types have the largest exposures. Similarly, the spread of exposures for a given detector type represents variations in the masses of detectors; there is a large variation in the masses of the ICPC detectors reflecting the R&D processes pursued by GERDA and MJD.

3.2.8 Active volume, containment efficiency, and enrichment fraction

There are three remaining pieces of information that influence the signal acceptance efficiency that are not determined by the data. The **active volume** ε_{AV} is computed via simulation of detector geometries and is the ratio of volume where a neutrinoless double-beta decay event has complete charge collection to the total volume of the detector. This is effectively a correction to the mass of the detector that weights the total mass

of the detector by the active volume. The active volume efficiency does not double count signal acceptance from the PSD cut. While the PSD cut *could* be thought of as an effective fiducilization of the detector volume, it is better for bookkeeping purposes to not treat it that way. If we enforce that a detectable $0\nu\beta\beta$ decay must occur in the active volume, then the DEP can be used as a proxy for $0\nu\beta\beta$ decay in the PSD efficiency, because DEP events almost always occur entirely within the active volume. There is no double counting in this procedure; the PSD efficiency is conditioned on the active volume effect. This is also easier than trying to simulate the effective volume after PSD cuts and trying to find a clean $0\nu\beta\beta$ decay proxy that occurs homogeneously throughout the active and depleted volumes of the detector.

Uncertainties on the dead layer and transition layer thicknesses lead to larger uncertainties on ε_{AV}, typically on the order $1-5\%$ [121]. The thickness of the dead layer plus that of the transition layer are called the full charge collection depth — it is the depth in the detector at which charge is collected fully. The full charge collection depth is given by the manufacturer for only some detectors. It is primarily determined through special calibrations with gamma sources: Monte Carlo simulations of the measurement with the full charge collection depth varied are performed until the simulated spectrum matches the measurement [121]. The full charge collection depth can in principle grow larger due to thermal diffusion of the Li ions used for the n-type contact. This is a highly temperature-dependent process, with more diffusion at higher temperatures. Some of the detectors were stored at room temperature after characterization and before deployment in LEGEND: these detectors had their full charge collection depths corrected by 1 mm per year, with an uncertainty of 5 mm per year. It is this uncertainty that leads to a large uncertainty on the active volume.

A plot of the active volumes for all detectors for all partitions is shown in Fig. 3.9. The active volume is on the order of $85-90\%$. It is lower for BEGe detectors because they have a larger dead layer, as determined by characterization measurements [121].

The containment efficiency is computed via simulation as the number of simulated neutrinoless double beta decays that deposit all of their energy within a germanium detector divided by the total number of simulated neutrinoless double beta decays. This efficiency is called the **containment efficiency** $\varepsilon_{cont.}$, as it computes the fraction of neutrinoless events whose entire energy is contained within one germanium detector. It is on the order of $\sim 95\%$ for most detector geometries. Full-scale simulations, starting from

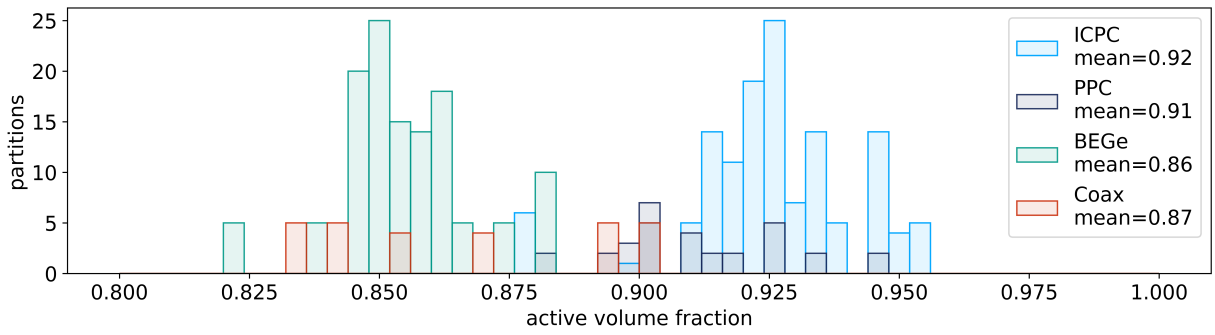

Figure 3.9: Histogram of the active volumes of all partitions, grouped by detector type. The spread in this plot represents the spread in active volumes, as each detector has the same active volume across all partitions. The active volumes for the BEGe detectors is lower due to the larger dead layer for these detectors.

GEANT4 energy depositions and charge drift simulation tools, compute the product of the active volume and containment efficiency and are compatible with the factorization described above [121].

The final part of the signal acceptance efficiency is the enrichment fraction $\varepsilon_{enr.}$, or the fraction of ^{76}Ge to all other species in the detector. This is usually supplied by the manufacturer of the HPGe detector — and verified through inductively coupled plasma mass spectrometry — and is roughly 90%, as shown in Fig. 3.10. Again, this number is essentially a correction to the mass of the detector, but is incorporated into the overall efficiency instead so that only one quantity in Eqn. 3.1 has an uncertainty on it.

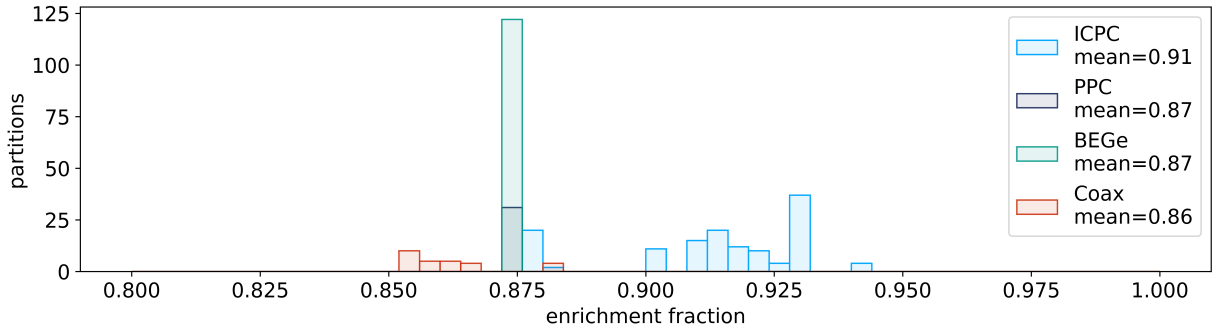

Figure 3.10: Histogram of the enrichment fraction of all partitions, grouped by detector type. The spread in this plot represents the spread in enrichment fraction, as each detector has the same enrichment fraction across all partitions. The ICPC detectors have higher enrichment fractions due to recent advancements in chemical refinement processes.

3.3 Spectrum construction

Once the data have been partitioned and all event energies are estimated from physics data, the analysis cuts can be applied to create the final event selection.

The overall signal acceptance efficiency is computed per detector per partition i as

Signal acceptance efficiency

$$\varepsilon^i = \varepsilon^i_{QC} \cdot \varepsilon^i_{PSD} \cdot \varepsilon_{LAr} \cdot \varepsilon_{cont.} \cdot \varepsilon_{AV} \cdot \varepsilon_{enr.} \tag{3.11}$$

The average signal acceptance efficiency is around 50% for BEGe and coaxial detector types, 56% for PPC, and 60% for ICPC detectors, as shown in Fig. 3.11. This is therefore equivalent to a reduction by about a factor of 2 of the active mass of the detector array. The smallest individual component of the total signal acceptance efficiency is the PSD cut acceptance efficiency, and thus it is this cut that reduces the overall efficiency dramatically. Additionally, the liquid argon acceptance (which is smaller than its target 95% because of cross-talk introducing inefficiencies) and the quality cuts (lower than optimal due to the large discharge rate inefficiency) also reduce the total signal acceptance efficiency. In order to recover active exposure and increase discovery sensitivity, it should be a high priority to optimize the PSD cut and the liquid argon cut to improve signal acceptance while minimizing the introduction of unacceptable background.

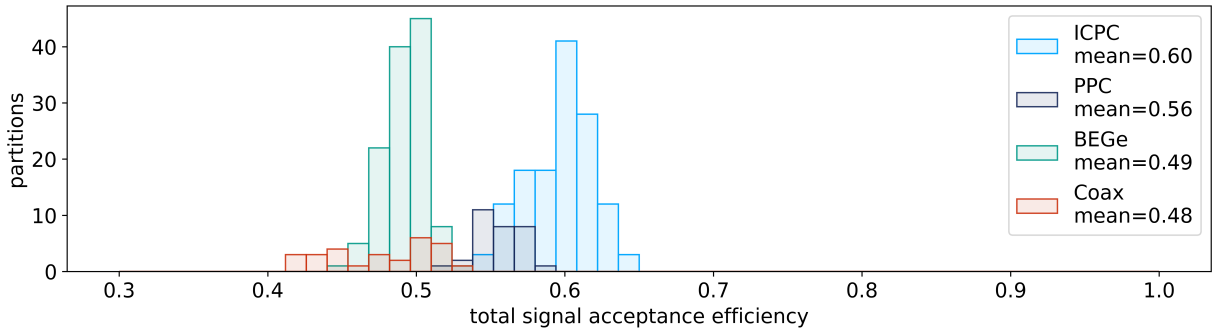

Figure 3.11: Histogram of the total signal acceptance efficiency of all partitions, grouped by detector type. The smallest individual component of acceptance efficiency for all detector types is the PSD cut acceptance efficiency. Despite ICPC and PPC detectors having similar PSD acceptances, the ICPC detectors have the largest signal acceptance efficiency due to their larger enrichment fraction, active volume, and containment efficiency.

The energy spectrum obtained after both the event energy determination and event selection procedures are performed as described above is shown in Fig. 3.12. The spectrum before cuts is shown as the gray histogram without any shading in Fig. 3.12: events at the high energy end of the spectrum likely come from surface alpha events (peaking near 5 MeV), the gamma line at 2615 keV comes from ^{208}Tl in the thorium decay chain, the gamma line at 1525 keV comes from ^{42}K decays, and the gamma line at 1460 keV comes from ^{40}K decays. The Compton shoulder is visible below the 2615 keV peak. Starting below around 2000 keV, the $2\nu\beta\beta$ decay continuum is visible, along with other β decays present on top of it. The PSD cut — shown in the gray filled histogram bins — is extremely powerful at removing alpha events, as evident by the reduction in events above 5 MeV. The PSD cut also strongly reduces the ^{40}K, ^{42}K, ^{208}Tl gamma lines. The liquid argon anti-coincidence cut — shown in the red histogram applied after the PSD cut — also strongly reduces the ^{42}K and ^{208}Tl gamma lines because these decays coincidentally release high energy betas that produce enough scintillation light to be efficiently vetoed. The ^{40}K is not efficiently vetoed by the liquid argon anti-coincidence cut because it decays via electron capture and rarely deposits enough energy in the liquid argon to trigger the veto condition. The liquid argon anti-coincidence cut highly suppresses beta decays, and thus the $2\nu\beta\beta$ decay spectral shape can be observed below 2000 keV. Simulations of the background model show that after cuts the region near $Q_{\beta\beta}$ will be essentially flat: this is an essential input to the statistical analysis that will be presented next chapter [122].

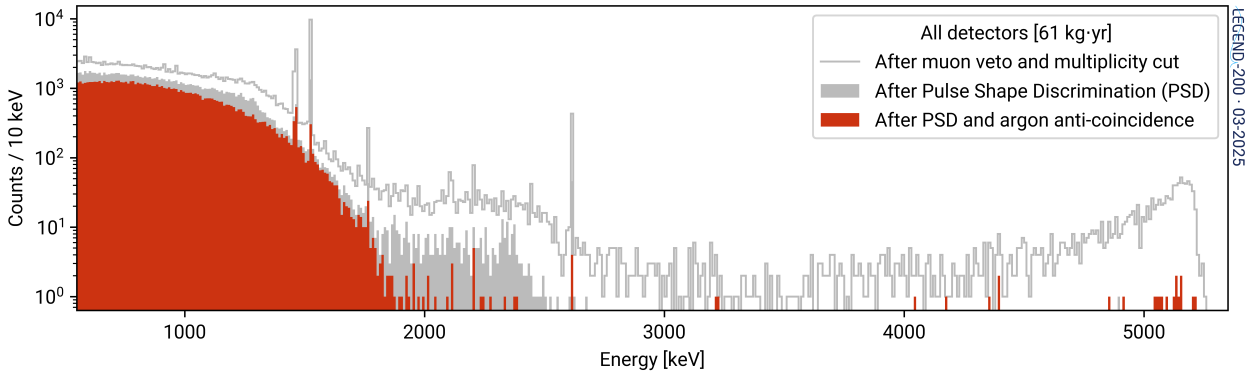

Figure 3.12: Final event spectrum for the first year of LEGEND-200 physics data taking, before and after all analysis cuts are applied.

Chapter 4

Frequentist statistical inference for the LEGEND experiment

Statistics is the primary interface between the world around us and physics. Statistics enables us to test scientific hypotheses, and to make measurements of unknown physical parameters utilizing the data obtained from experiments. We now turn our attention to the statistical inference of the first dataset from the LEGEND-200 detector. The statistical analysis parameters and the dataset are summarized first before a brief review of frequentist statistical inference is presented. Then, the first results from the LEGEND-200 experiment are detailed, along with a new software tool for performing frequentist statistical inference on data from quasi-background-free rare-event searches called `freqfit`. Finally, we summarize the unique statistical challenges and choices that quasi-background-free experiments such as LEGEND-1000 will face.

4.1 LEGEND-200 dataset

The LEGEND-200 detector took one full calendar year of stable neutrinoless double-beta decay physics data. The analysis of the first $\mathcal{O}(10$ kg·yr) was not blind; the analysis of this data was presented at the 2023 Topics in Astrophysics and Underground Physics (TAUP) conference and is called the TAUP dataset [123]. High-voltage scans were performed following collection of the TAUP dataset during period five, and stable operation resumed in period six [101]. All subsequent data were blinded, and the TAUP data was

retroactively blinded when incorporated into future analyses. Data collection continued until March 2024: the cumulative exposure of physics data amounted to 61 kg· yr.

This total exposure was analyzed as two separate datasets: the Golden dataset, from well-behaved detectors (for which traditional A/E cuts could be performed), and the Silver dataset, from detectors that needed further analysis. The Golden dataset — 48.3 kg·yr of exposure — is comprised of all BEGe detectors, all PPC detectors, all Ortec ICPC detectors, and a majority of Mirion IC detectors. The Silver dataset – 12.7 kg·yr of exposure — consists of all coaxial detectors, and five Mirion ICPC detectors. The Silver dataset required further analyses not present in analysis of the Golden dataset; machine learning was used for PSD in the coaxial detectors, and novel drift-time corrections were required for the five Mirion ICPC detectors [2, 104, 119]. Prior to unblinding, it was agreed that these two datasets would have separate background indices due to potentially different performance of analysis cuts [124]. This is well supported as the coaxial detectors were found to have a mildly higher background index from early results in GERDA [125], although the final analysis of GERDA did not see an elevated background index in the coaxial detectors [56]. Further details on the analysis of this 61 kg · yr of physics data can be found in the first LEGEND-200 result paper [2].

After unblinding, there are 13 events that survive all cuts present in the agreed-upon analysis window. The analysis window was chosen to be from [1930, 2190] keV prior to unblinding: this region is large enough to get an unbiased estimate of the background rate [124]. More importantly, the LEGEND-200 background model predicts that the spectrum in this region is flat [1] [122]. Additionally, ± 5 keV regions around two known gamma lines, 2104 keV from ^{208}Tl and 2114 keV from ^{214}Bi, are removed from analysis. The uniformity of the background in the analysis window is consistent with the GERDA experiment, which also used the exact same analysis region (with removed gamma lines) for statistical inference [56]. There are then 11 events within the analysis window that survive all cuts and are not near known gamma lines. These events are shown in Figure 4.1.

[1] Extending the analysis window to lower energies, where the spectrum starts to exponentially increase, would require a rigorous background model that includes the shape of the expected $2\nu\beta\beta$ decay spectrum. Similarly, if the analysis window were extended to higher energies it would no longer be flat at around 2381 keV due to the Compton shoulder from the 99% intensity 2614 keV Tl-208 gamma line.

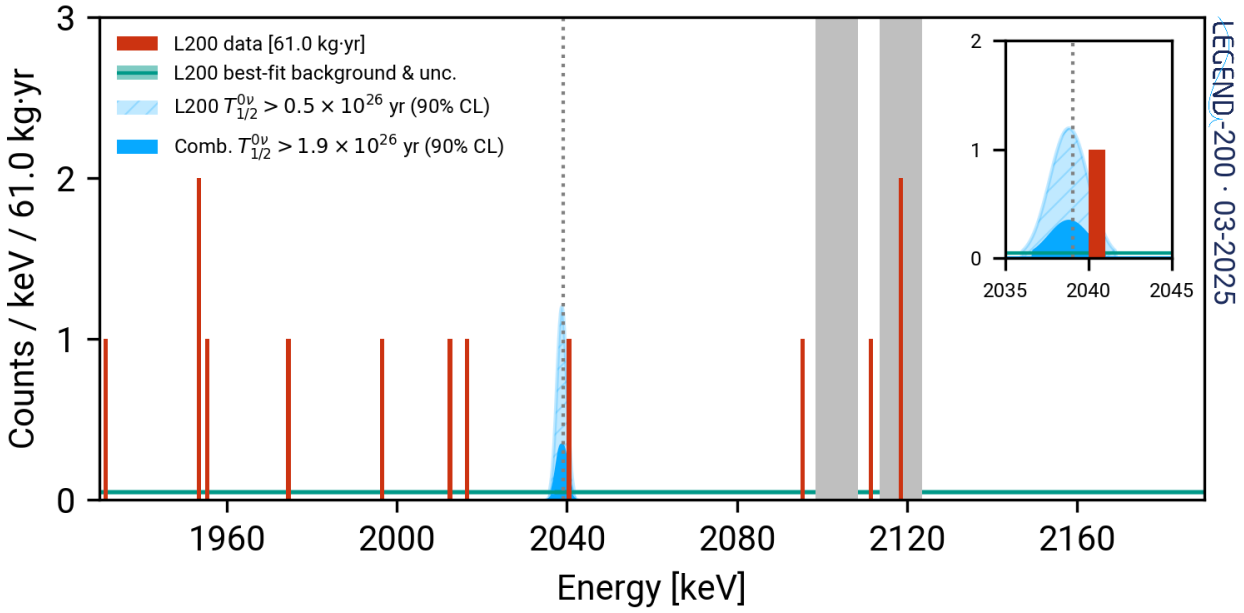

Figure 4.1: The final energy spectrum, after all analysis cuts, for the first year of physics data from the LEGEND-200 detector. Thirteen total events were observed in the analysis window, [1930, 2190] keV, although two events are removed from statistical inference because they can be attributed to a known gamma line near 2104 keV. One event surviving all cuts is within 2σ of $Q_{\beta\beta}$ for that detector's energy resolution and peak shape.

4.2 Statistical inference

We now turn our attention to interpreting this spectrum. We seek to make statistical inferences about the parameter of interest, the half-rate of neutrinoless double-beta decay, $\Gamma^{0\nu}_{1/2}$. There are two main approaches to statistical inference: the Bayesian approach, and the frequentist approach. A full description of the Bayesian approach to statistical inference falls outside the scope of this work; however, many great resources exist on explaining it [126, 127], including in the context of statistical inference on LEGEND-200 data [128]. But briefly, frequentist approaches refrain from treating the true, unknown value of a parameter of interest as a random variable, while Bayesian approaches extend the concept of probability to apply to one's degree of belief in a parameter's value [129]. In the frequentist framework, probability is interpreted as a frequency, and that the frequency converges in some long-run repetition of experiments. The Bayesian treatment, however, interprets probability as instead a degree of belief that is continuously updated and conditioned upon prior results.

The goals of statistical inference are manifold, but a small subset of questions we seek to answer can be summarized as:

Goals of statistical inference

- Point estimation: what is the best guess for the true, unknown value of neutrinoless double-beta decay half-rate Γ given some set of data?
- Hypothesis testing: does the data support the existence of neutrinoless double-beta decay?
- Interval estimation: do the data favor a range of values that are likely to contain Γ?

We will briefly discuss how each of these is handled in the frequentist treatment.

4.2.1 Point estimation

Suppose we search for neutrinoless double-beta decay in one germanium detector over a long period of time, and at the end collect our data, $\vec{x}$ (which is a list of event energies), surviving all cuts. The data we collect are random, and the random variable that described them is $\vec{X}$. We are interested in finding an estimate of the true, unknown value of the rate of neutrinoless double-beta decay Γ_{true}, using the data we have. The *point*

estimate is a method that is used to obtain an estimate of this unknown parameter. There are many methods to generate a point estimate, but we will focus on the maximum likelihood estimator. The *likelihood* is a core concept of frequentist statistics: it is the joint probability distribution of the observed data and the unknown parameters [127, 130]. If $f_X(\vec{x};\Gamma)$ is the joint probability distribution of the observed data and the neutrinoless double beta decay rate, then the likelihood $\mathcal{L}$ is a function of Γ and is given by

$$\mathcal{L}(\Gamma|\vec{x}) = f_X(\vec{x};\Gamma). \tag{4.1}$$

Note that the likelihood is a function of the parameter of interest and not the data; and unlike a probability density function (PDF) it is not necessarily normalized to one. The likelihood is our statistical model for the underlying physics process — it is up to the experimenter to cook up the correct model, and to test the validity and consistency of the data with their model.

One way to estimate the unknown value of Γ is to maximize the likelihood function, and the estimate it produces is called the maximum likelihood estimator (MLE). The MLE is used in practice because of a variety of nice properties; namely, it is consistent and thus converges in probability to the true unknown value [130]. The MLE for a parameter of interest is usually denoted by the symbol with a hat over it and is given by

$$\hat{\Gamma} = \mathrm{argmax}_{\Gamma}\, \mathcal{L}(\Gamma|\tilde{\mathrm{x}}). \tag{4.2}$$

The MLE is often called the best-fit for a parameter because its value is the best in that it maximizes the likelihood. The procedure of generating the MLE is also referred to as fitting because it searches for the best-fit value of Γ given the observed data.

It is common to instead work with the logarithm of the likelihood, called the log-likelihood, which is denoted by ℓ. The MLE of a parameter must also maximize the log-likelihood. One common approach to do so is to instead minimize the *negative* log-likelihood and use one of many available minimizer algorithms, such as `MINUIT` [131]. Minimizer algorithms typically work by stepping through the parameter space of the negative log-likelihood in a smart way — either by computing the gradient and Hessian, or using a gradient-free approach — and checking if the negative log-likelihood is converging to a minimum value.

Most minimizer algorithms allow the user to specify a tolerance, or stopping criteria, that tell the minimizer to stop searching for convergence to a minimum when the negative log-likelihood is changing less than a specified value per step.

In reality, the likelihood for $0\nu\beta\beta$ decay is not as simple as what was given in Equation 4.1 because experiments are not so simple. The one germanium detector really has a finite energy resolution, its detection efficiency is not unity, and there is the presence of persistent external radioactive background that mimics the true signal. These are examples of what are called *nuisance parameters* because they are parameters in our model of our data that affect the value of the likelihood, but are not the parameter of interest, Γ. The nuisance parameters are denoted by $\vec{\nu}$ when incorporated into the likelihood:

$$\mathcal{L}(\Gamma, \vec{\nu}|\vec{x}) = f_X(\vec{x}; \Gamma, \vec{\nu}). \tag{4.3}$$

Because we are often not interested in the values of the nuisance parameters, or because their presence complicates the computation and dimensionality of the statistical inference we want to make about the single parameter of interest, it is common to work with the *profile likelihood*[2]

$$\mathcal{L}(\Gamma|\vec{x}) = \max_{\tilde{\nu}} \mathcal{L}(\Gamma, \tilde{\nu}|\tilde{\mathrm{x}}), \tag{4.4}$$

where the profile likelihood is thus a function only of Γ that is enforced by picking the values of the nuisance parameters that maximize the likelihood for a given Γ [126, 127, 130].

4.2.2 Hypothesis testing/tests of significance

Given a set of observations $\vec{x}$ we are interested in determining if neutrinoless double-beta decay exists, i.e. that it has a non-zero half-rate. The process to test this hypothesis, that $\Gamma_{true} \neq 0$, is known as a test of significance. Frequentist hypothesis testing was pioneered by Fisher, but we will follow the approach of Neyman and Pearson which has more useful properties for the statistical inference at hand [132]. This proceeds in the following steps [126, 127, 130, 132, 133]

[2]The profile likelihood is technically not a likelihood, because there is not a joint probability density function $f_X(\vec{x}; \Gamma, \vec{\nu})$ that describes it

1. Designate a *null hypothesis* H_0, and an *alternative hypothesis* H_a. The null hypothesis is usually the hypothesis designated by the model under which current knowledge is assumed to be true by default; in searches for BSM physics like $0\nu\beta\beta$ decay, it typically corresponds to Standard Model-only processes, and is sometimes called the background-only hypothesis [127, 133]. The alternative hypothesis typically corresponds to the presence of the additional BSM process, and is called the signal-plus-background hypothesis. Hypotheses may be *simple* if they specify a single value for the parameter of interest (denoted here by Γ_0), or *composite* if multiple values of the parameter are specified by the hypothesis. It is common that the null hypothesis is simple (typically it is set to $\Gamma_0 = 0$ or $\Gamma_0 = \Gamma'$) while the alternative hypothesis is composite ($\Gamma \neq \Gamma_0$). Choices of alternative composite hypotheses lead to *one-sided* tests if Γ is allowed to be only larger or smaller than Γ_0, i.e. $\Gamma > \Gamma_0, \Gamma < \Gamma_0$, and *two-sided hypothesis* tests if Γ can be both larger or smaller than Γ_0, i.e. $\Gamma \neq \Gamma_0$ [126].

2. Define the *significance level* of the test (and optionally, set the *power* of the test). The power and significance of a test relate to different errors the analyst can make in testing hypotheses. In a type I error the analyst rejects the null hypothesis when the null hypothesis was actually true; the probability of committing a type I error in a large ensemble of experiments is called the significance level, and is denoted by α. The quantity $1 - \alpha$ is then termed the confidence level: this value is typically set as large as possible to protect against type I error, usually it is set to 90%, 95%, or 3σ (which is $\approx 99.73\%$). The analyst could also make a type II error where the null hypothesis is not rejected but the alternative hypothesis is actually true. The long-run probability of making a type II error is called β, and the quantity $1 - \beta$ is the power of a statistical test. It is called power because it is the frequency of being able to discriminate between the null and alternative hypotheses correctly. The types of error are shown in Fig. 4.2 which is known as the error matrix.

3. Select an optimal test, often through the choice of a *test statistic* denoted t. A test statistic is a scalar function of the data that summarizes the agreement between the observed data and the null hypothesis [3]. The form of the test statistic is often chosen so that large values of the test statistic reflect large

[3]For the remainder of this chapter, I will always be assuming that the test statistics are sufficient statistics. Any quantity that is dependent on the observed values of the random variable $\vec{X}$ is called a statistic; a statistic t is sufficient if the conditional distribution of $\vec{X}$ given $t = t'$ is independent of $\Gamma, \vec{\nu}$ [127]. The sufficiency of the test statistic is what makes it a powerful tool to

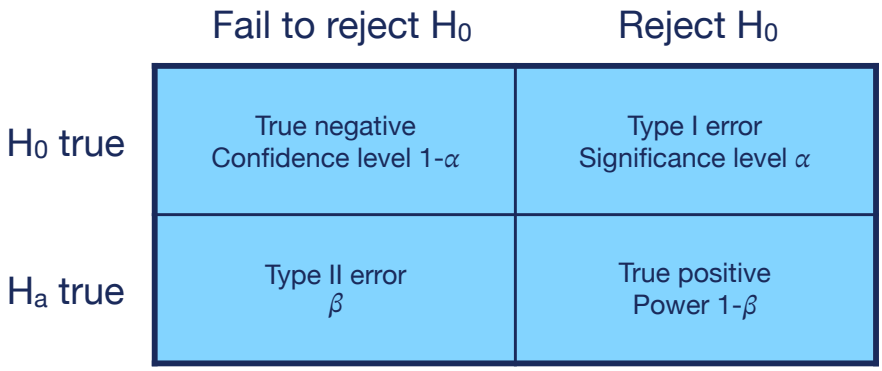

Figure 4.2: The error matrix showing possible analysis decisions that lead to different types of errors depending on which hypothesis is true.

disagreement with the null hypothesis.

Neyman and Pearson showed that the test statistic choice that optimizes the power of a simple hypothesis test for a fixed significance level is the log-likelihood ratio [126, 130]. The log-likelihood ratio can be defined as

$$t_\Gamma(\vec{x}) = -2\ln\left(\frac{\mathcal{L}(\vec{x}|\Gamma_0)}{\mathcal{L}(\vec{x}|\hat{\Gamma})}\right) = -2\left[\ell(\vec{x}|\Gamma_0) - \ell(\vec{x}|\hat{\Gamma})\right], \tag{4.5}$$

where Γ_0 is the signal rate for the null hypothesis and $\hat{\Gamma}$ is the best-fit rate in the total space of signal rate hypotheses. Note that t_Γ is positive, and that larger values of t_Γ mean greater disagreement between the data and the hypothesis being tested. The log-likelihood ratio is only zero when the hypothesis being tested is equal to the best-fit value for the data — in this simple case without nuisance parameters. The maximum log-likelihood ratio test also gives the uniformly most powerful test when H_0 is simple and H_a is composite [126]. Test statistics can also be formed out of profile likelihoods

test hypotheses. The value of the test statistic $t = t'$ made from an observation is a value obtained from $f_t(t'|\Gamma, \vec{\nu})$ and contains all of the information about the hypothesis given the model; the information obtained from the sufficient test statistic then contains all of the information about the assessment of the hypothesis given the data. Furthermore, a good test statistic will demonstrate consistency, or that the power approaches one when the number of observations grows to infinity [133].

when nuisance parameters are present in the likelihood; this is the most common form of test statistic for physics searches [134].

4. Define an acceptance region. The acceptance region is the set of t values where we do not reject the null hypothesis, and the critical region is the set of t values where we do reject the null hypothesis. The size of the two regions are set by the significance level chosen for the test. For the case of a one-sided test, which is applicable to test statistics because larger values of the test statistic imply larger disagreement with the null hypothesis (regardless if the alternative hypothesis is one-sided or two-sided!), the acceptance region is set by one critical value of the test statistic t_{crit} such that in a large ensemble of experiments where the null hypothesis is true, $1-\alpha$ of those experiments will have a test statistic value less than t_{crit}. If the analyst knows or can compute the test statistic distribution under assumption of the null hypothesis $f(t|\Gamma_0)$ which is called a *test statistic PDF* then the critical value is given by

$$\int_{t_{crit}}^{\infty} dt\, f(t|\Gamma_0) = \alpha. \tag{4.6}$$

Due to the presence of nuisance parameters, it is often not possible to compute the critical region ahead of time. Furthermore, it is extremely difficult to compute the test statistic distribution: most approaches use Monte Carlo sampling — or asymptotic formulae — to generate the PDF [133]. It is important to note that α is not a function of the observed data.

If the alternative hypothesis H_a is specified and the distribution of the test statistics under that hypothesis, $f(t|\Gamma_a)$, is computed or known then the power of the test is given by:

$$\int_{-\infty}^{t_{crit}} dt\, f(t|\Gamma_a) = \beta. \tag{4.7}$$

Figure 4.3 shows an example of test statistic distributions for the LEGEND-200 experiment testing $\Gamma_0 = 2.1702 \times 10^{-26}\,\mathrm{yr}^{-1}$, with the alternative hypothesis being $\Gamma_a = 0$, where α, β, and the critical region are indicated.

5. Perform the experiment. Note the above steps 2 and 3 must be carried out *before* seeing the data and

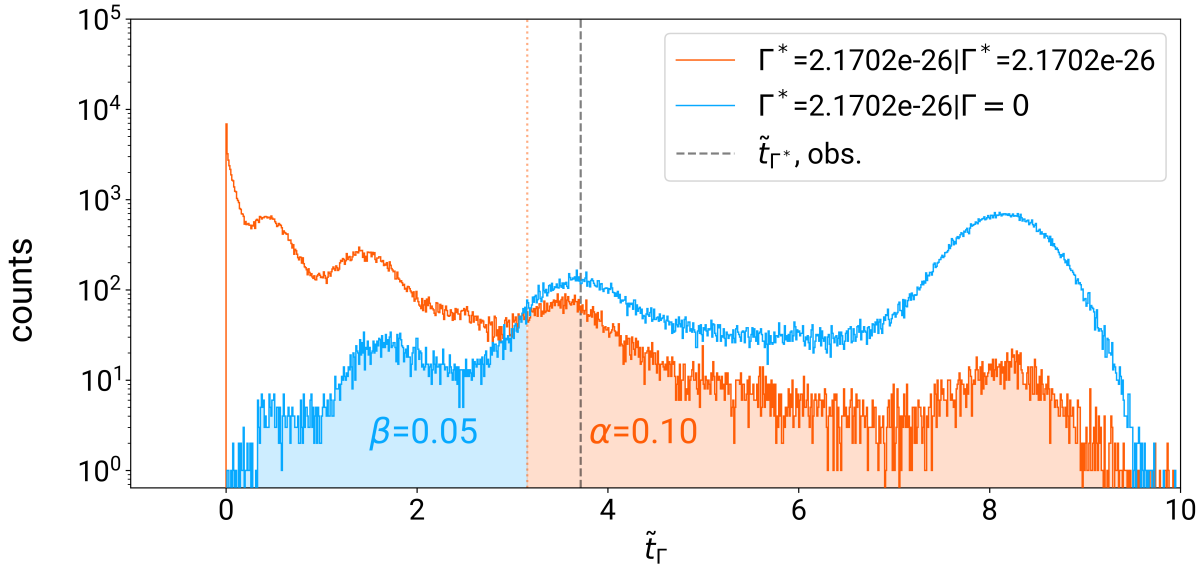

Figure 4.3: Test statistic distributions for the null and alternative hypotheses showing the significance level α and the power $1-\beta$. The critical test statistic is chosen so that the area above it is equal to the significance level α.

performing the analysis in order not to bias the statistical inference procedure [133].

6. Determine the distribution of the test statistic for the null hypothesis. For experiments with a small sample size, this must be done through Monte Carlo sampling. This proceeds by generating a new instance of the random variable $\vec{X}$ by random sampling of the distribution $f(\vec{X}|H_0)$; these are often called pseudo-experiments, or toys, because they represent another possible outcome of the experiment [126, 133]. From these random variables $\vec{x}$, one computes the value of the test statistic $t(\vec{x})$ many, many times to create a population of $\vec{t}(\vec{x})$. This ensemble of test statistics can then be turned into a PDF through normalization by the number of Monte Carlo experiments generated. When nuisance parameters are present, the distribution $f(\vec{X}|H_0)$ is more complicated, and different approaches exist on how to sample this distribution in a frequentist manner — see section 4.3.4 for more details.

7. Compute the observed value of the test statistic for the data, $t(\vec{x}_{obs.})$.

8. Make a decision. If the observed value of the test statistic is larger than the critical value $t(\vec{x}_{obs.}) > t_{crit.}$, then reject the null hypothesis because the observed data disagree significantly with the null

hypothesis. Otherwise, we fail to reject the null hypothesis. It is common at this stage to compute the p-value of the observed data, which is a measure of the observed level of significance. It can be computed using the test statistic distribution under the assumption of the null hypothesis, $f(t|\Gamma_0)$, as

$$p_0 = \int_{t_{obs.}}^{\infty} dt\, f(t|\Gamma_0) \tag{4.8}$$

Its frequentist interpretation is an ensemble of identical experiments, this fraction of experiments would obtain data that disagreed with the null hypothesis more strongly than our observed data did, assuming the null hypothesis is true.

The decision based on the p-value is related to the critical test statistic: if the p-value is less than α, then we know that the observed value of the test statistic was larger than the critical test statistic. These are two ways of looking at the same decision.

4.2.3 Interval estimation

The process of creating intervals that are likely to contain the true value of the parameter of interest is directly tied to significance testing. In frequentist statistics, the concept of a confidence interval is tricky and is often pointed out as one of the major flaws of the approach [135]. The frequentist *confidence interval* is constructed so that the interval may be different if the experiment was repeated, but in a large ensemble of identical experiments, the constructed confidence intervals will contain the true value in a specified fraction of the experiments. The frequency with which the intervals contain the true parameter is called the *confidence level* and is denoted by $1-\alpha$ [4]. This can be made mathematically rigorous following the definition in Ref. [126]

Frequentist confidence interval definition

If $\{f(x,\Gamma);\Gamma\}$ is a family of distributions and that a random sample of observations, $\vec{X}$ are taken from $f(x;\Gamma_{true})$, and if $S_{\vec{X}}$ is a subset of Γ depending on $\vec{X}$ so that $\Pr[\vec{X} : S_{\vec{X}} \supset \Gamma_{true}] = 1-\alpha$, then $S_{\vec{X}}$ is a confidence interval.

[4] Notice here how we have made the notational choice to set the confidence level using the same symbol α used to denote type I errors. This reflects the similarities between hypothesis testing and interval construction.

This reliability of interval constructions is known as *coverage*, because it is the frequency with which the intervals "cover" the true value. Frequentists seek to ensure that the coverage is equal to the stated confidence level; however, coverage can not always be guaranteed, for example when the data are discrete such as in a Poisson counting experiment; in those cases one ensures that coverage is *at least* the stated confidence level [127, 136]. Also, note that the confidence level does not apply to one specific interval, but instead a whole family of intervals.

There are several ways to build confidence intervals that all hinge on different ways to order the values of Γ to include in the confidence interval. The construction of confidence intervals was formalized by Neyman in the eponymous Neyman construction [130]. We will focus on likelihood-ratio ordering because this method yields non-empty intervals [136].

Perhaps the simplest way to think of interval estimation is as repeated hypothesis testing: if we test every allowed value of Γ as the null hypothesis and then take all the values that are not rejected at a significance level α, then we form a confidence interval with confidence level $1-\alpha$. These intervals will then have correct coverage by construction. This can be seen if we consider $A(\Gamma_0)$ to be the acceptance region for a test at significance level α of $H_0 : \Gamma = \Gamma_0$ then this proposed hypothesis test construction of the confidence interval gives $S_{\vec{X}} = \{\Gamma_0 : \vec{X} \in A(\Gamma_0)\}$. This set has the correct coverage because $\Pr[\vec{X} : S_{\vec{X}} \supset \Gamma_{true} | \Gamma_{true} = \Gamma_0] = \Pr[\vec{X} \in A(\Gamma_0) | \Gamma_{true} = \Gamma_0] \geq 1 - \alpha$ by construction of $A(\Gamma_0)$ and thus $S_{\vec{X}}$ is a confidence interval according to the above definition [126].

If we think in terms of observed p-values, where we compute $p_{obs.}(\Gamma) = \int_{t_{obs.}(\Gamma)}^{\infty} dt\, f(t|\Gamma)$, then the set of all tested Γ with $p_{obs.}(\Gamma) > \alpha$ form the confidence interval as we cannot reject their null hypothesis at the given significance level.

Now we come to the difference between one-sided and two-sided confidence intervals, where one-sided confidence intervals are sometimes called "limits". Speaking strictly statistically, one-sided and two-sided confidence intervals arise from *different* statistical tests. Two-sided intervals can arise, for example, from testing the null hypothesis $H_0 : \Gamma = \Gamma_0$ against the alternative hypothesis $H_a : \Gamma \neq \Gamma_0$ — the alternative hypothesis tests values of Γ both greater than or less than Γ_0. This can be achieved by picking a test statistic that is large when the data are more compatible with $\Gamma > \Gamma_0$ and $\Gamma < \Gamma_0$. On the other hand, one-sided intervals can arise from testing the null hypothesis $H_0 : \Gamma = \Gamma_0$ against a one-sided alternative

like $H_a : \Gamma < \Gamma_0$. In this case a test statistic is chosen so that the test statistic is large only when the data agree with $\Gamma < \Gamma_0$, and it is small when $\Gamma \geq \Gamma_0$. A one-sided test leads directly to a confidence interval that has only one bound in it because we are testing only what values the true value is likely to be smaller than.

This leads naturally to some confusion in the interpretation of two-sided and one-sided confidence intervals. Often the larger value in a two-sided confidence interval will be misinterpreted as an upper limit on the value: this is never correct! Interpreting this upper value as an upper limit will lead to mistakes in coverage, because the appropriate hypothesis has not been tested and then the confidence interval has been constructed incorrectly. Later, we will come across a test statistic $\tilde{t}_\mu$ that "naturally transitions from placing one-sided to two-sided limits" [136]. The test statistic $\tilde{t}_\mu$ is *always* testing a two-sided hypothesis test and its confidence interval is always two-sided, it just happens that some of the time the lower limit will be at 0 making it *seem* like an upper limit. But as already mentioned, to interpret this as an upper limit at the stated confidence interval is wrong, because upper limits are instead constructed under a different hypothesis test. We will return to this discussion in Sec. 4.4.5.

The hypothesis test of $H_0 : \Gamma_0 = 0$ and $H_a : \Gamma \neq 0$ is often called a *discovery* test because if the null hypothesis is rejected, then with statistical confidence neutrinoless double-beta decay is discovered. The process of creating a confidence interval for Γ is sometimes called computing the *observed limit* because both a one-sided and two-sided confidence interval give information about what values are excluded from the confidence interval.

The choice of a log-likelihood-ratio ordered test statistic guarantees certain properties of interval estimation when the number of samples is large. In particular, according to Wilks' approximation [137], because the likelihood becomes Gaussian distributed in the large N limit, the distribution of log-likelihood-ratio test statistics becomes χ^2 distributed for the null hypothesis, where the χ^2 distribution's PDF is given by

$$f(t|H_0) = \frac{e^{-t/2}}{\sqrt{2\pi t}}. \tag{4.9}$$

If Wilks' theorem is satisfied, then the critical value of the test statistic can be computed directly as

$$\int_{t_{crit}}^{\infty} dt \, \frac{e^{-t/2}}{\sqrt{2\pi t}} = \alpha. \tag{4.10}$$

For $\alpha = 0.1$, the critical test statistic value is 2.71 and is often labeled as $\Delta\chi^2 = 2.71$. Again, this approximation only works in the large N limit, and the validity of this approximation to the specific case of the LEGEND experiment is discussed in further detail in Sec. 4.4.1.

4.3 Frequentist statistical inference on the LEGEND-200 dataset

Now that we are statistically prepared, we present the frequentist statistical inference for the LEGEND-200 dataset by first defining necessary analysis parameters, then introducing a software tool `freqfit` to perform the statistical inference for a quasi-background-free rare-event search, and then focusing on results.

4.3.1 Analysis window

As stated in Section 4.1, the analysis window was defined prior to unblinding to avoid several known γ lines. These are the single-escape peak gamma line at 2103.511 keV and double-escape peak gamma line at 2118.513 keV originating from the gamma rays of 2614.511 keV from ^{208}Tl and 2204.10 keV from ^{214}Bi respectively. Regions of ± 5 keV around these γ lines for the full fit window, which is then the disjoint union $[1930.0, 2098.5] \cup [2108.5, 2113.5] \cup [2123.5, 2190.0]$ keV. The total net width of the analysis window is defined as $\Delta E = 240$ keV. The analysis window is also referred to as the fit window as it defines the region over which the likelihood is fit to the data.

4.3.2 Likelihood definition

The LEGEND-200 dataset is partitioned into stable periods of analysis parameters for each detector as described above in Sec. 3.2.2. The likelihood factorized into a product over all partitions because each partition is treated as an independent experiment. The likelihood for a single partition is given by the product of a shape term $\mathcal{L}_{shape}$ with constraints on nuisance parameters obtained from auxiliary measurements $\mathcal{L}_{aux}$ (auxiliary measurements could include for example measurements of the energy resolution obtained during calibration, or of uncertainties on the signal acceptance efficiency obtained from special calibration runs):

$$\mathcal{L}_i(\Gamma, \vec{\nu} | N_i, \vec{E}_i) = \text{Poisson}(N_i | \mu_i(\Gamma^{0\nu}_{1/2}, \vec{\nu}_i)) \prod_{j=1}^{N_i} \mathcal{L}_{shape}(\Gamma^{0\nu}_{1/2}, \vec{\nu}_i | E_{ij}) \cdot \mathcal{L}_{aux}(\vec{\nu}_i), \tag{4.11}$$

where E_{ij} are the energies of the observed events in the i-th partition, $\vec{\nu}_i$ are the nuisance parameters, N_i is the total observed number of events in the i-th partition, and μ_i is the total number of expected counts and is a function of the signal rate and nuisance parameters. This likelihood is called an *extended unbinned* likelihood; it is *extended* due to the presence of the Poisson term in the front, which yields information on the overall count rate, and it is *unbinned* because the observed energies are given to the likelihood instead of number of counts in a histogram bin. We use an extended unbinned likelihood because a binned analysis would result in a loss of information and a reduction of sensitivity to the fit parameters [138]. The Poisson term in front arises from the fact that N_i is a random variable, owing to the Poissonian nature of radioactive decays. The expected number of counts μ_i is the sum of the expected number of signal and number of background counts for the partition, $\mu_i = s_i + b_i$. In our chosen model, the number of signal counts s_i is the total number of decays obtained from the product of the signal rate and the exposure $\mathcal{E}_i$, converted into number of atoms by a factor of $\frac{\ln 2 N_A}{m_{\text{det}}} f_{76}$ where m_{det} is the average molar mass of the detector and f_{76} is its isotopic enrichment. This is then further multiplied by the detection efficiency ε_i plus some uncertainty on the detection efficiency δ_{ε_i} scaled by a strength parameter α,

Expected number of signal counts

$$s_i(\Gamma_{1/2}) = \frac{\ln 2\, N_A}{m_{\text{det}}} f_{76}(\epsilon_i + \delta_{\varepsilon_i}\alpha)\mathcal{E}_i\Gamma_{1/2} \quad \rightarrow \quad s_i(\Gamma_{1/2}) = \frac{\ln 2\, N_A}{m_{\text{det}}}(\epsilon_i + \delta_{\varepsilon_i}\alpha)\mathcal{E}_i\Gamma_{1/2}. \quad (4.12)$$

In Eqn. 4.12 we've chosen to incorporate the enrichment fraction and its uncertainty in the overall signal acceptance efficiency. In this treatment, the total exposure of the partition is exactly known and any uncertainty on it is incorporated into the signal acceptance efficiency uncertainty; therefore, the exposure is not considered a nuisance parameter. Because the signal acceptance uncertainty was estimated the same way for all partitions [5], and because the signal acceptance uncertainty represents the largest relative uncertainty among all nuisance parameters (see Table 4.1), it was decided to completely correlate the efficiency uncertainties across all partitions using a single scaling parameter α for all partitions; completely correlating the uncertainties is the most statistically conservative option [124].

[5] The signal acceptance efficiency uncertainty also contains some large terms like the containment efficiency that are calculated using the same tool (GEANT4) for all detectors [121].

The choice to completely correlate the signal acceptance efficiency uncertainties greatly reduced the computations necessary for statistical inference by reducing the number of nuisance parameters by $\mathcal{O}(50\%)$, and it also reduced analysis complexity and labor by alleviating the need for analysts to divide every uncertainty contribution into correlated an uncorrelated components. As we will see later in Sec. 4.3.6, the difference in the results of the statistical inference between a fully correlated treatment and fully uncorrelated treatment is very small, so the difference between a fully correlated treatment and the true correlation level should be negligible.

In this treatment the signal acceptance efficiency δ_{ε_i} is not treated as a nuisance parameter but α is treated as one instead. The factors of $\ln 2$, N_A, and m_{det} are dropped during the interval estimation and hypothesis testing in order to ensure that the parameter values are not too small for the minimizer; we call this *reduced* signal space or $\Gamma' = \frac{\ln 2\, N_A}{m_{\text{det}}}\Gamma$. Because they represent a proportional scaling, factors are properly restored when reporting final intervals.

The number of expected background counts is simpler; it is just the product of the background index BI_i, the analysis region width ΔE, and the exposure,

Expected number of background counts

$$b_i = BI_i\, \mathcal{E}_i\, \Delta E. \tag{4.13}$$

There are two background indices present in the analysis: BI_1 for detectors from the Golden dataset, and BI_2 for detectors from the Silver dataset — so $BI_i = BI_1$ if i is a Golden dataset partition, and $BI_i = BI_2$ if i is a Silver dataset partition. This was decision was made based on the possibility that the two datasets could have different background performances after analysis cuts. For each dataset, BI_i was chosen to be shared among all detectors because, even though in principle we expect the background indices to be different for each detector, we see no statistically significant variation in the observed rates among detectors.

The constraints on the nuisance parameters come from auxiliary measurements of the central value (also called nominal value) of each nuisance parameter $\tilde{\nu}$ and its uncertainty δ_ν, and are written as Gaussian

constraints (also called penalty terms [6]),

$$\mathcal{L}_{aux}(\nu) = \frac{1}{\sqrt{2\pi}\delta_\nu} \exp\left(-\frac{(\nu-\tilde{\nu})^2}{2\delta_\nu^2}\right). \tag{4.14}$$

For LEGEND-200, the shape term of the likelihood consists of a Gaussian signal shape on top of a uniform/flat background. We will investigate the impact of the background shape on the statistical inference in a later section 4.3.6, but to spoil the punchline, no variation in the statistical inference is observed under different background models — a flat background is a valid assumption, at least for this dataset. The expected signal shape is chosen because, to first order, a germanium detector's response to a monoenergetic energy deposition from a neutrinoless-double beta decay event should be Gaussian [57]; however, fits performed on calibration data preferred an extended Gaussian peak shape model with a significant low energy tail, so this peak shape should be investigated in future analyses. For LEGEND-200, the PDF or shape part of the likelihood is given by

$$\mathcal{L}_{i,\,\mathrm{shape}}(E_j) = \frac{s_i}{s_i+b_i}\frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(E_j - Q_{\beta\beta}+\Delta_i)^2}{2\sigma_i^2}\right) + \frac{b_i}{s_i+b_i}\frac{1}{\Delta E}, \tag{4.15}$$

where σ_i is the detector's energy resolution and Δ_i is the energy bias correction estimated at $Q_{\beta\beta}$. The energy bias correction reflects potential differences in the estimation of the energy scale via calibrations (due to energy non-linearity among other effects) and the true energy of an event: it is given by $E_{cal} - E_{true}$. The uncertainties on the energy resolution and energy bias correction are denoted as $\delta_{\sigma_i}, \delta_{\Delta_i}$ respectively. Auxiliary measurements of these nuisance parameters, called the *shape nuisance parameters* because they relate to the shape of the PDF, are performed as described in Sec. 3.2, and are incorporated as Gaussian penalty terms in the likelihood.

To summarize, the nuisance parameters for each partition are as follows: the energy bias Δ_i and energy resolution σ_i, which are typically partition specific; the background index BI_i, which is either BI_1 or BI_2 depending on which dataset the partition belongs to; and the signal acceptance efficiency scaling parameter α which is common to all partitions. The background indices are the only nuisance parameters without

[6] They are called penalty terms because they take on larger values in the log-likelihood when nuisance parameters are pulled away from their nominal value, and could potentially move the log-likelihood away from its minimum when data constrain the nuisance parameters to far-away values.

auxiliary measurements — we rely entirely on the observed data to constrain these nuisance parameters. The nuisance parameters with available auxiliary measurements are the energy bias, energy resolution, and the signal acceptance efficiency uncertainty scaling parameter α (centered at $\tilde{\alpha} = 0$ with uncertainty $\delta_\alpha = 1$) and thus the overall auxiliary term of the total likelihood can be written as

$$\mathcal{L}_{\text{aux}}(\vec{\nu}) = \prod_{i}^{N_{\text{part.}}} \prod_{\nu=\sigma_i,\Delta_i} \left[\frac{1}{\sqrt{2\pi}\delta_\nu} \exp\left(-\frac{(\nu-\tilde{\nu})^2}{2\delta_\nu^2} \right) \right] \times \frac{1}{\sqrt{2\pi}\delta_\alpha} \exp\left(-\frac{(\alpha-\tilde{\alpha})^2}{2\delta_\alpha^2} \right). \tag{4.16}$$

The overall likelihood for the LEGEND-200 dataset is therefore

LEGEND-200 likelihood definition

$$\begin{aligned}\mathcal{L}_{L200} = \prod_{i}^{N_{\text{part.}}} \text{Poisson}(N_i|s_i+b_i) \prod_{j=1}^{N_i} \frac{1}{s_i+b_i} \left[\frac{s_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(E_{ij}-Q_{\beta\beta}+\Delta_i)^2}{2\sigma_i^2} \right) + \frac{b_i}{\Delta E} \right] \\ \times \prod_{\nu=\sigma_i,\Delta_i} \left[\frac{1}{\sqrt{2\pi}\delta_\nu} \exp\left(-\frac{(\nu-\tilde{\nu})^2}{2\delta_\nu^2} \right) \right] \times \frac{1}{\sqrt{2\pi}\delta_\alpha} \exp\left(-\frac{(\alpha-\tilde{\alpha})^2}{2\delta_\alpha^2} \right).\end{aligned} \tag{4.17}$$

For the LEGEND-200 data set there are 316 partitions, and a total of 2216 input parameters to the likelihood: each partition has seven parameters ($\mathcal{E}_i, \varepsilon_i, \delta_{\varepsilon_i}, \sigma_i, \delta_{\sigma_i}, \Delta_i, \delta_{\Delta_i}$) and then the signal rate Γ, the two background indices BI_1, BI_2, and finally the signal acceptance efficiency uncertainty scaling α. Because the relative uncertainty on the energy resolution is extremely small, and much smaller than the relative uncertainty on the other peak shape nuisance parameter Δ, it is a good approximation to keep this nuisance parameter at its nominal measured auxiliary value in the likelihood [124]. Then, of the 2216 total input parameters, only 320 are fit parameters: this comes from 316 Δ terms that are fit, the two background indices BI_1, BI_2, the signal rate Γ, and the efficiency scaling α. The analysis parameters and their exposure-average values are summarized in the next section and in Table 4.1. The values of all analysis parameters per partition are available in the supplemental file published with Ref. [2].

Likelihood and partitions summary

The experiment parameters, per detector per partition, required for the neutrinoless double-beta decay analysis are summarized as follows:

Analysis parameters per partition

1. partition exposure $\mathcal{E}$ in kg·yr, corresponding to the mass of the detector multiplied by its live-time
2. peak shape parameters
    - energy bias correction at $Q_{\beta\beta}$, $\Delta = E_{\text{true}} - E_{\text{cal}}$ (keV), and its uncertainty δ_Δ
    - energy resolution σ (keV) and its uncertainty δ_σ
3. signal acceptance efficiencies and their uncertainties, taken to be independent
    - active volume fraction and its uncertainty
    - containment efficiency and its uncertainty
    - ^{76}Ge enrichment fraction and its uncertainty
    - data quality cut efficiency and its uncertainty
    - PSD cut efficiency and its uncertainty
    - muon anticoincidence cut efficiency and its uncertainty
    - LAr cut efficiency and its uncertainty
4. event energies E and partitions to which they belong (i.e. timestamps)

4.3.3 Test statistic definition

LEGEND uses the two-sided test statistic, with a physical boundary set at $\Gamma^{0\nu}_{1/2} \geq 0$, for both tests of discovery and exclusion/interval-estimation. This test statistic is often called $\tilde{t}_\mu$, but in this context it will be called $\tilde{t}_\Gamma$ because Γ is our parameter of interest [134]. The tilde is introduced because the imposition of

	LEGEND-200
# partitions	316
total exposure (kg·yr)	61.0
energy bias correction (keV), Δ	0.24 ± 0.31
energy resolution (keV), σ	1.17 ± 0.01
signal acceptance, ϵ	0.56 ± 0.03

Table 4.1: Statistical analysis parameters, reported as the exposure-weighted average across partitions. Uncertainties are treated as correlated, i.e. also exposure-weighted averaged. Signal acceptance includes all efficiencies, including enrichment.

a physical boundary condition distinguishes it from the test statistic given by Eqn. 4.5. This test statistic is a negative log-likelihood ratio that parametrizes how much the observed data disagree with the signal rate hypothesis Γ that we are testing,

$\tilde{t}_\Gamma$ test statistic definition

$$\tilde{t}_\Gamma = \begin{cases} -2\ln\frac{\mathcal{L}(\Gamma,\hat{\hat{BI}},\hat{\hat{\vec{\nu}}})}{\mathcal{L}(\hat{\Gamma},\hat{BI},\hat{\vec{\nu}})}, & \hat{\Gamma} > 0 \\ -2\ln\frac{\mathcal{L}(\Gamma,\hat{\hat{BI}},\hat{\hat{\vec{\nu}}})}{\mathcal{L}(0,\hat{BI}(\Gamma=0),\hat{\hat{\vec{\nu}}}(\Gamma=0))}, & \hat{\Gamma} \leq 0 \end{cases}, \tag{4.18}$$

where $\hat{\hat{\vec{\nu}}}$ are the values of the nuisance parameters that maximize the profile likelihood, and $\hat{\Gamma}$, $\hat{\vec{\nu}}$ are the best-fit values that maximize the likelihood. A large value of the test statistic indicates that the best fit signal rate is incommensurate with the signal rate hypothesis we are testing. Because we impose a physical boundary on $\Gamma^{0\nu}_{1/2} \geq 0$, if our best fit signal rate $\hat{\Gamma}$ is negative, then the best fit to the data must be $\hat{\Gamma} = 0$. It is important to note that large test statistics can be generated if the best fit signal rate is *either* higher or lower than the signal rate hypothesis to be tested — the alternative hypothesis is two sided. This is in direct contrast with another test statistic that could be chosen called $\tilde{q}_\Gamma$,

$\tilde{q}_\Gamma$ test statistic definition

$$\tilde{q}_\Gamma = \begin{cases} -2\ln\frac{\mathcal{L}(\Gamma,\hat{\vec{B}}\hat{I},\hat{\vec{\nu}})}{\mathcal{L}(0,\vec{B}\hat{I}(\Gamma=0),\hat{\vec{\nu}}(\Gamma=0))}, & \hat{\Gamma} \leq 0 \\ -2\ln\frac{\mathcal{L}(\Gamma,\hat{\vec{B}}\hat{I},\hat{\vec{\nu}})}{\mathcal{L}(\hat{\Gamma},\vec{B}\hat{I},\hat{\vec{\nu}})}, & 0 < \hat{\Gamma} < \Gamma \\ 0, & \hat{\Gamma} > \Gamma \end{cases}, \tag{4.19}$$

where data with best fit signal rate larger than the testing hypothesis are deemed to be in agreement with the hypothesis [134]. Because $\tilde{q}_\Gamma$ uses a one-sided hypothesis as the alternative hypothesis, it cannot be used to test for discovery. The differences between $\tilde{q}_\Gamma$ and $\tilde{t}_\Gamma$ will be explored further in section 4.4.5.

4.3.4 The `freqfit` python package for unbinned frequentist statistical inference

Although there are mature tools to do frequentist statistical inference on binned datasets [139], there are few software tools to do unbinned analyses that handle toy generation appropriately. To that end, we developed a Python-based software package called `freqfit` that is capable of performing unbinned frequentist statistical inference, with specific focus on data from quasi-background-free rare-event searches [140].

Likelihood handling

This package uses the `iminuit` package to handle likelihood instantiation and minimization [141]. Minimizations are sped up through the use of the `numba` package [142]: each term in the likelihood is just-in-time compiled, which enables a large performance boost. Each shape term $\mathcal{L}_{shape}$ in the likelihood is created from an `iminuit ExtendedUnbinnedNLL` object that holds that partition's data as well as PDF. Each PDF calls an underlying numba-wrapped model class, specific to that experiment's shape terms; `iminuit` requires that this PDF returns both the total normalization $s_i + b_i$, as well as the PDF evaluated at the data. `iminuit` uses the negative log-likelihood for minimizations, and handles the extended part of the negative log-likelihood on its own:

$$-2\ln(\mathcal{L}_{extended,shape}) = -2\ln(\frac{N^N e^{-N}}{N!}\prod_{i=1}^{N}\frac{1}{N}\mathcal{L}_{shape}) = 2N - \sum 2\ln\mathcal{L}_{shape} + 2\ln(N!). \tag{4.20}$$

The constant term $2\ln(N!)$ is dropped from computation because it is not a function of the parameters and thus doesn't affect the minimization. The test statistic remains a sufficient statistic even after dropping the term. The auxiliary terms in the likelihood are instantiated as `Constraint` objects that are Gaussian penalty terms on the likelihood. These `ExtendedUnbinnedNLL` objects and `Constraint` objects can be summed programatically to create the overall negative log-likelihood for minimization. To compute the test statistic, this negative log-likelihood is minimized twice: once with the Γ parameter free to float to find the best fit, and once with Γ fixed to the hypothesis value in order to compute the profile likelihood. Because `iminuit` computes the negative log-likelihood, the test statistic is simply the difference between the values that `iminuit` calculates for the negative log-likelihood in these instances. The physical boundary in the test statistic $\tilde{t}_\Gamma$, $\Gamma \geq 0$, is handled using only the `iminuit` interface: a limit on the signal rate at $\Gamma = 0$ ensures that if the best-fit signal rate would be negative, it is sent to 0 instead.

Toy generation

In order to compute p-values for hypotheses under test, `freqfit` needs to generate distributions of test statistics. These test statistic PDFs come from computing the test statistic for a large ensemble of pseudo-experiments, also called toys. Toy generation proceeds in two steps. First, draw new energy data for each partition using the profile likelihood's best-fit value of nuisance parameters $\hat{\hat{\vec{\nu}}}$; this is done first by computing the number of signal- and background- like events to draw for each partition from Poisson distributions with means $s_i(\Gamma), b_i(\hat{\hat{BI}}_i)$, and then the energies of these events are sampled from the shape PDF using the shape nuisance parameters $\hat{\hat{\vec{\nu}}}_i$. Second, vary the central value of each of the auxiliary constraint terms based on a Gaussian random variable, which is drawn from a Gaussian centered at the nuisance parameter's profiled best fit value and with width equal to the uncertainty from the auxiliary measurement. This second step is a valid frequentist treatment of auxiliary measurements and is called the "unconditional ensemble" [143–147]; an equally valid procedure is to simply let the auxiliary measurements be centered on their measured value which is called the "conditional ensemble". The unconditional ensemble has the correct asymptotic behavior that reproduces Wilks' theorem and thus was used to generate the statistical inference for LEGEND-200 [143]. Mathematically, `freqfit` follows this procedure to compute test statistics and to generate toys at a hypothesis of signal rate Γ:

Toy generation procedure

1. Define the likelihood for the experiment

$$\mathcal{L}(\Gamma, BI, \vec{\nu}|\vec{E}) = \prod_i \text{Poisson}(s_i + b_i) \prod_{j=1}^{N_i} \frac{1}{s_i + b_i} \left[s_i S(E_{ij}) + b_i B_i \right] \tag{4.21}$$

2. Minimize the negative log-likelihood to obtain $\hat{\Gamma}, \hat{BI}, \hat{\vec{\nu}}$

$$-2 \ln \mathcal{L}_{bestfit}(\hat{\Gamma}, \hat{BI}, \hat{\vec{\nu}}|\vec{E}) \tag{4.22}$$

3. Minimize the negative log-likelihood, with the signal rate profiled at the hypothesis of Γ. Obtain $\hat{\hat{BI}}, \hat{\hat{\vec{\nu}}}$

$$-2 \ln \mathcal{L}_{profile}(\Gamma, \hat{\hat{BI}}, \hat{\hat{\vec{\nu}}}|\vec{E}) \tag{4.23}$$

4. Compute the observed test statistic $\tilde{t}_{\Gamma,obs.}$

$$\tilde{t}_{\Gamma,obs.} = -2 \ln \mathcal{L}_{profile}(\Gamma, \hat{\hat{BI}}, \hat{\hat{\vec{\nu}}}|\vec{E}) + 2 \ln \mathcal{L}_{bestfit}(\hat{\Gamma}, \hat{BI}, \hat{\vec{\nu}}|\vec{E}) \tag{4.24}$$

5. For each partition, generate a random number of signal-like events and background-like events drawn from Poisson distributions, where $s_i = \Gamma \mathcal{E}_i(\varepsilon_i + \hat{\hat{\alpha}}\delta_{\varepsilon_i})$ and $b_i = \hat{\hat{BI}}_i \mathcal{E} \Delta E$ — note that toys are generated with $\hat{\hat{\vec{\nu}}}, \hat{\hat{BI}}$

$$Pois(s_i), Pois(b_i) \rightarrow n_{s_i}, n_{b_i} \tag{4.25}$$

6. For each event, generate a random event energy E'_{ij}. For signal-like events draw from the peak shape (using $\hat{\hat{\sigma}}, \hat{\hat{\Delta}}$), and for background-like events draw from a uniform distribution.

7. For each auxiliary measurement ν, generate a Gaussian constraint $\tilde{\nu}' + \delta_\nu$ by grabbing a Gaussian random variable from a Gaussian centered at the profile fit value with width equal to the

auxiliary measurement uncertainty:

$$G(\hat{\tilde{\nu}}, \delta_\nu|\nu) \rightarrow \tilde{\nu}' \pm \delta_\nu \tag{4.26}$$

8. Re-compute the likelihood for this toy by summing the negative log-likelihoods for each partition and the Gaussian constraints

$$\mathcal{L}'(\Gamma, BI, \vec{\nu}|\vec{E}') \tag{4.27}$$

9. Minimize the negative log-likelihood to obtain $\hat{\Gamma}', \hat{BI}', \hat{\vec{\nu}}'$

$$-2\ln \mathcal{L}'_{bestfit}(\hat{\Gamma}', \hat{BI}', \hat{\vec{\nu}}'|\vec{E}') \tag{4.28}$$

10. Minimize the negative log-likelihood, with the signal rate profiled at the hypothesis of Γ. Obtain $\hat{\hat{BI}}', \hat{\hat{\vec{\nu}}}'$

$$-2\ln \mathcal{L}'_{profile}(\Gamma, \hat{\hat{BI}}', \hat{\hat{\vec{\nu}}}') \tag{4.29}$$

11. Compute the toy test statistic $\tilde{t}'_{\Gamma,obs.}$

$$\tilde{t}'_{\Gamma,obs.} = -2\ln \mathcal{L}'_{profile}(\Gamma, \hat{\hat{BI}}', \hat{\hat{\vec{\nu}}}') + 2\ln \mathcal{L}'_{bestfit}(\hat{\Gamma}', \hat{BI}', \hat{\vec{\nu}}') \tag{4.30}$$

In earlier versions of `freqfit`, an approximation was made at the step with Eqn. 4.26 by using the the nominal value of the nuisance parameter from auxiliary measurements instead of the profile fit value. This eased computation initially, but it was eventually removed from the code.

Limit generation

Toys need to be generated at different true values of the signal rate to test different hypotheses. `freqfit` allows for many different toy generation routines that are useful for answering different statistical questions. For performing discovery tests, toys are generated with null hypothesis $H_0 : \Gamma = \Gamma_0 = 0$ (using the two-

sided test statistic $\tilde{t}_\Gamma$ which has alternative hypothesis $H_a : \Gamma \neq \Gamma_0$) and tested against the null hypothesis to build up the test statistic distribution $f(\tilde{t}_{\Gamma_0}|\Gamma_0)$. The p-value for discovery is then computed using the observed value of the LEGEND-200 test statistic computed under the zero-signal hypothesis, $\tilde{t}^{\text{obs.}}_{\Gamma_0}$. The procedure for generating the confidence interval is the same, except that toys are generated at the null hypothesis $H_0 : \Gamma = \Gamma'$ (again, using the two-sided test statistic $\tilde{t}_\Gamma$ which has alternative hypothesis $H_a : \Gamma \neq \Gamma'$) and tested against the null hypothesis to build up the distributions $f(\tilde{t}_{\Gamma'}|\Gamma')$. The p-value for each hypothesis under the observation of LEGEND-200 data is again computed using Eqn. 4.8 and the confidence interval is constructed via the union of all values at which $p^{\text{obs.}} \geq 0.1$ for a 90% confidence level (CL) interval. We use linear interpolation between discrete Γ' test points to find the end points of these intervals in order to reduce computation time.

The toy generation procedure is performed 100,000 times for each signal hypothesis Γ'. This number was chosen so that when p-values are computed — by taking the number of test statistics greater than or equal to the observed test statistic and dividing by the total number of test statistics — the binomial uncertainty on this value will be small for 3σ discovery p-values of $p \approx 0.0027$:

$$\delta p \approx \sqrt{(0.0027)(1 - 0.0027)/(100000)} \approx 0.00016. \tag{4.31}$$

We compute p-values on a grid of 100 equally-spaced points in reduced-signal space, starting at $\Gamma = 0$. The upper range of this grid is determined first by-eye using Wilks' approximation to guess where the 90% CL upper limit should be using the observed values of the test statistics (by finding the intersection of the observed test statistics and $\tilde{t}_{\Gamma,90\%\text{crit.}} \approx 2.71$), and then through stepping up the signal rate until the p-value of the largest signal rate hypothesis has a p-value of $\sim 10^{-3}$.

When computing the median exclusion sensitivity, toys are also generated with the null hypothesis $\Gamma = 0$ and test statistics are computed at $\Gamma = \Gamma'$ so that the distribution $f(\tilde{t}_{\Gamma'}|\Gamma_0)$ is determined. These are therefore toys generated under the background-only hypothesis. We compute the median test statistic of this distribution $\tilde{t}^{\text{med.}}_{\Gamma'|\Gamma_0}$ and evaluate the p-value of this test statistic assuming $H_0 : \Gamma = \Gamma'$ using the distribution $f(\tilde{t}_{\Gamma'}|\Gamma')$. The set of all Γ' for which the median p-value of the background-only toys is greater than 0.1 constructs the 90% CL confidence interval for what is called the median *exclusion sensitivity*. This computation is called such because it is the median 90% CL confidence interval reported by those

Purpose of test	Value to generate toys at	Values to test toys at
Confidence interval	$\Gamma', \hat{\hat{\nu}}$	Γ'
Median exclusion	$\Gamma_0, \hat{\hat{\nu}}(\Gamma_0)$	Γ'
Median discovery	$\Gamma', \hat{\hat{\nu}}$	Γ_0
Background misevaluation	$\Gamma', BI', \hat{\hat{\nu}}(\Gamma', BI')$	Γ'
Joint profile	$\Gamma', BI', \hat{\hat{\nu}}(\Gamma', BI')$	Γ', BI'

Table 4.2: Summary of toy generation and test statistic computation for various purposes, such as computing confidence intervals, or determining the median exclusion or discovery sensitivities. Background misevaluation refers to tests where the toys are generated with a background index that is different from its profiled value for an experiment, but the test statistic is computed using the correct profile likelihood treatment.

experiments with a true signal rate of $\Gamma = 0$. In addition to the median p-value, values of median p-value $\pm 1\sigma, \pm 2\sigma$ are constructed as well by computing quantiles around the median, i.e. critical values of the test statistic $\tilde{t}_{\Gamma'|\Gamma_0}$ are found such that $\int_{\tilde{t}^{crit.,low}_{\Gamma'|\Gamma=0}}^{\tilde{t}^{crit.,high}_{\Gamma'|\Gamma=0}} d\tilde{t}_{\Gamma'} f(\tilde{t}_{\Gamma'}|\Gamma = 0) = \Phi(n) - \Phi(-n)$, where Φ is the cumulative distribution function (CDF) of the normal distribution and n is the number of sigma. These uncertainty bands highlight the range of upper limits that an experiment would place due to statistical variation. Note that the median exclusion sensitivity is also computed using a two-sided test statistic, although a one-sided test statistic would also be valid for constructing a median upper limit.

The median discovery sensitivity is computed similarly, except that toys are generated with the null hypothesis $\Gamma = \Gamma'$ and test statistics are computed at $\Gamma = \Gamma_0$ so that the distribution $f(\tilde{t}_{\Gamma_0}|\Gamma')$ is determined. The median test statistic of this distribution is then determined, and the p-value of obtaining this median test statistic given that the null hypothesis $H_0 : \Gamma = 0$ is computed using the distribution $f(\tilde{t}_{\Gamma_0}|\Gamma_0)$. The median 3σ discovery sensitivity is therefore the signal rate at which the p-value of the median "discovery" test statistic $\tilde{t}_{\Gamma_0|\Gamma'}$ is equal to $p = 0.0027$. Table 4.2 summarizes these procedures for generating toys and computing test statistics.

Because statistical inference requires roughly 10 million toys for each experiment (100000 toys per signal hypothesis times 100 signal hypothesis) the whole process is CPU limited and computationally expensive. We parallelize pseudo-experiment generation at two levels in `freqfit` in order to utilize high performance computing clusters — such as CENPA-rocks or Perlmutter at NERSC — to their full capabilities: each signal hypothesis is assigned a series of 10 nodes, and then each node generates 10000 toys on all available cores using the Python `multiprocessing` module. This dramatically improves throughput. It

takes roughly 1000 core-hours to perform statistical inference on the LEGEND-200 dataset.

Tips and tricks

The `freqfit` package also employs a number of tricks to speed-up performance and improve convergence of likelihood minimizations. The main trick to speed-up minimizations is to reduce the parameter space: because LEGEND has numerous datasets that contain no events, it is possible to combine these "empty datasets" into one exposure-weighted partition. Because these datasets have no events to constrain shape nuisance parameters such as σ and Δ, the best-fits to these nuisance parameters will be their central values, and thus the auxiliary constraint terms contribute a constant value to the overall likelihood. This trick analytically profiles out the nuisance parameters in the empty partitions.

The nuisance parameters in the number of signal counts s_k and background counts b_k, however, do contribute to the overall likelihood through the extended term. Because `iminuit` handles the extended part of the likelihood separately, it is sufficient to compute the sum of the expected number of signal and background counts for all of these empty partitions:

$$-2n_{empty} = -2(s_{empty} + b_{empty}) = -2\sum_{i=1}^{n_{empty}}(s_i + b_i) = -2\sum_{i=1}^{n_{empty}}(\Gamma\mathcal{E}_i(\varepsilon_i + \alpha\delta_{\varepsilon_i}) + BI_i\mathcal{E}_i\Delta E). \quad (4.32)$$

The common term of $\mathcal{E}_i$ suggests that we can pull out the overall sum exposure of all the empty datasets, called the total empty exposure $\mathcal{E}_{\text{empty}} \equiv \sum_j^{n_{empty}} \mathcal{E}_j$, and be left with the exposure-weighted average of the nuisance parameters:

$$-2n_{empty} = -2\sum_{i=1}^{n_{empty}} \mathcal{E}_{\text{empty}} \left(\Gamma\mathcal{E}_i \frac{(\varepsilon_i + \alpha\delta_{\varepsilon_i})}{\mathcal{E}_{\text{empty}}} + \mathcal{E}_i \frac{BI_i\Delta E}{\mathcal{E}_{\text{empty}}}\right). \quad (4.33)$$

This can then be expressed in terms of the exposure weighted averages of the nuisance parameters, $\bar{\varepsilon} = \frac{1}{\mathcal{E}_{\text{empty}}}\sum \mathcal{E}_i\varepsilon_i$ and $\bar{\delta}_\varepsilon = \frac{1}{\mathcal{E}_{\text{empty}}}\sum \mathcal{E}_i\delta\varepsilon_i$. To improve code readability and computational transparency, we decided that instead of computing the exposure weighted average background index to instead split this sum

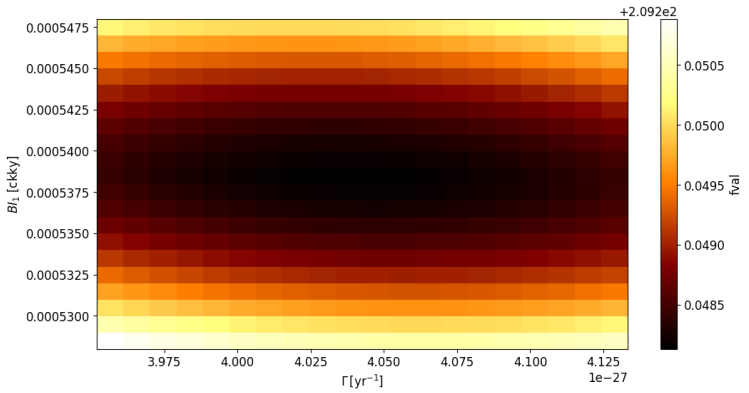

Figure 4.4: The LEGEND-200 negative 2 log likelihood in the vicinity of the global best-fit, when other nuisance parameters are profiled out, shown in a two-dimensional space in terms of the neutrinoless double-beta decay half-rate and the Golden dataset background index. The change in negative 2 log likelihood values are extremely small near the global best fit and thus the minimum is shallow — numerical issues due to instability in finding the true minimum abound.

into the sum of exposure weighted averages per background index:

$$-2n_{empty} = -2\mathcal{E}_1\Gamma(\bar{\varepsilon}_1 + \alpha\bar{\delta}_{\varepsilon_1}) - 2\mathcal{E}_1 BI_1\Delta E - 2\mathcal{E}_2\Gamma(\bar{\varepsilon}_2 + \alpha\bar{\delta}_{\varepsilon_2}) - 2\mathcal{E}_2 BI_2\Delta E, \tag{4.34}$$

where the subscripts 1 and 2 refer to the Golden and Silver datasets respectively, $\mathcal{E}_1, \mathcal{E}_2$ are the total exposure of all empty datasets, BI_1 and BI_2 are the background indices, and $\bar{\varepsilon}_1, \bar{\varepsilon}_2, \bar{\delta}_{\varepsilon_1}, \bar{\delta}_{\varepsilon_2}$ are the exposure-per-background-index weighted averages of these parameters.

Combining all of the empty-datasets into one exposure-weighted dataset significantly reduces the dimension of the parameter space from 320 fit parameters to 15 for [7] for LEGEND-200 at no penalty in computing and minimizing the likelihood — again, these shape nuisance parameters for each empty dataset would be fit to their central value anyway.

Because the data are so sparse, the likelihood is extremely flat (see Fig. 4.4) and thus in addition to implementing analytic solutions where possible (as described above), some other tricks are needed to improve minimization convergence.

[7] These 15 come from the Δ terms in the 11 partitions with events in them, the two background indices BI_1, BI_2, the signal rate Γ, and the efficiency scaling α.

Tricks used by `freqfit` for frequentist statistical inference

- **Initial guesses:** because the likelihood is extremely flat, minimizations can terminate early before the global minimum is identified due to stopping criteria in the minimizer. Giving a good initial guess can start the minimization closer to the global minimum and improve fit stability. The signal and background rate dominate the likelihood in terms of numerical values; the nuisance parameters play essentially no role in the guess (also, starting them at their best fits in the auxiliary terms is always a good starting point). For LEGEND-200, it is sufficient to guess a signal rate equal to the number of counts within ± 3 keV of $Q_{\beta\beta}$ times the exposure-weighted efficiency, and a background rate equal to the number of remaining counts outside the ± 3 keV of $Q_{\beta\beta}$ divided by the exposure and analysis window width. The initial guess value of the nuisance parameters are set to their best fit from the profile.

- **Instabilities near $\Gamma = 0$:** the minimizer can have some difficulty reaching the actual physical boundary of $\Gamma = 0$ before terminating. If the minimization terminates early, and the true minimum is at $\Gamma = 0$, then the values of the nuisance parameters are not at their true best fit values. If the toy is then profiled at or near $\Gamma = 0$, then negative test statistics can occur quite often because the profile fit will obtain the true best fit values of the nuisance parameters. This issue is solved by forcing a profile at $\Gamma = 0$ if the best fit for Γ is much smaller than our sensitivity, $\Gamma < 10^{-10}$.

- **Minimizer termination parameters:** in order to ensure that the minimization routine can reach the true global minimum, we optimized the termination criteria. For `iminuit` using the `Minuit` minimizer this corresponds to a tolerance of 10^{-5} and a precision of 10^{-10}. The tolerance sets the minimum difference to look for in the stopping criteria; for `Minuit`, the minimization stops when the estimated vertical distance to the minimum falls below what is set by the tolerance [131].

- **Re-minimizing non-converging fits:** some fits (<10 per 100,000 toys) will fail to converge properly and will thus result in a negative test statistic for that toy. This arises from the best-fit

not converging to the true global minimum, whence it is then possible for the profile likelihood to have a smaller test statistic value. For toys with negative test statistics, the best-fit is performed again, this time using the best fit value of the parameters from the profile likelihood of the negative-test-statistic toy as the initial guess.

- **Parameter grid:** to ease numerical instability from floating point errors during computation of the negative log-likelihood (especially in computing the difference between the profile and the best-fit negative log-likelihoods) parameter values are placed on a grid via rounding them to the decimal precision to which they can be accurately known. For example, σ is rounded to 4 decimal places because its uncertainty can be measured on that order. Rounding parameter values fixed a floating point error that was noticed during debugging; this is discussed in the cross-checks section in more detail in section 4.4.1. The rounding scheme for LEGEND-200 is as follows: Δ was rounded to 4 decimal places, σ was rounded to 2 places, the background indices and signal rate were rounded to 6 places, and the global efficiency uncertainty scaling was rounded to 2 places.

There were also things that we experimented with but determined did not improve fit convergence:

Summary of abandoned approaches to improving fit convergence

- **Removing the physical boundary on Γ:** we investigated if letting Γ fit negative during the best-fit and then imposing $\Gamma \geq 0$ by-hand would improve fit stability. This did not work, as large under-fluctuations in the number of signal events drove the signal rate so negative that the number of expected counts became negative and `iminuit` would error out. We tried solving this by then placing a discontinuous physical boundary in the PDF, but this made convergence issues worse. Placing a limit on the parameter during the fit was determined to be the best solution.

- **Truncating PDFs and numerical approximations to PDFs:** we investigated if truncating the Gaussian distribution past 5σ or using a Taylor approximation for the Gaussian tail past

5σ would improve numerical stability. We found no difference in the case where nuisance parameters are floated in the fit. It is possible that some numerical instability is introduced by small floating point values of the PDFs, but these instabilities manifest on a scale much smaller than the scale at which the test statistic is smeared out by the presence of nuisance parameters.

- **Minimizer:** we investigated if choice of minimizer algorithm could improve convergence of fits. Without a good initial guess strategy, minimizer choice does have some impact on obtaining the global minima: we found the Powell algorithm implemented by the `scipy.minimization` package converges to the boundary at $\Gamma = 0$ more robustly. However, once an initial guess strategy was implemented, we obtained the same results with the Powell and Minuit minimization routines.

4.3.5 Cross-checks

In order to ensure that the `freqfit` toy generation and minimization routines were working correctly, numerous cross-checks were performed. First, upper limits were generated for a one-bin Poisson counting experiment with a known number of background counts b. The upper limits generated at different true values of the signal rate agree with those obtained by Feldman-Cousins in Ref. [136]. Next, we checked the performance of `freqfit` in reproducing the final result of MAJORANA DEMONSTRATOR using the published full dataset [55]. The likelihood for MAJORANA DEMONSTRATOR is slightly different than for LEGEND-200: the peak shape used in the signal PDF takes the form of an extended Gaussian,

$$\begin{aligned} \mathcal{S}_k(E_i) = \frac{1-f_k}{\sqrt{2\pi}\gamma_k\sigma_k}\exp\left(-\frac{(E_i - Q_{\beta\beta} + \Delta_k)^2}{2(\gamma_k\sigma_k)^2}\right) \\ + \frac{f_k}{2\gamma_k\tau_k}\exp\left(\frac{\sigma_k^2}{2\tau_k^2} + \frac{E_i - Q_{\beta\beta} + \Delta_k}{\gamma_k\tau_k}\right)\operatorname{erfc}\left(\frac{\sigma_k}{\sqrt{2}\tau_k} + \frac{E_i - Q_{\beta\beta} + \Delta_k}{\sqrt{2}\gamma_k\sigma_k}\right). \end{aligned} \tag{4.35}$$

This extended Gaussian has a low-energy-side tail that is described by its fraction f_k, decay parameter τ_k, and width γ_k; this tail is present in the peak shape due to significant charge trapping [55, 148]. The analysis window for the MAJORANA DEMONSTRATOR dataset is also different from that for LEGEND-200: the analysis window starts at 1950 keV and extends to 2350 keV, with both the same γ lines removed as in

LEGEND-200, and with an additional ± 5 keV removed from the γ line at 2204.1 keV from ^{214}Bi.

Figure 4.5 shows the frequentist statistical inference we obtain for the published MAJORANA DEMONSTRATOR partitions. The points in solid black are the p-values obtained from comparing the observed test statistic to the Monte Carlo generated test statistic PDFs. The observed 90% CL lower limit is $T^{0\nu}_{1/2} > 8.41 \times 10^{25}$ years; the published result from MAJORANA DEMONSTRATOR is $T^{0\nu}_{1/2} > 8.3 \times 10^{25}$ years. One major difference between the analysis implemented in `freqfit` and the published final MJD result is the handling of nuisance parameters. The unconditional ensemble approach taken by `freqfit` means that toys are generated with values of the nuisance parameters at their best-fit from profiling the LEGEND-200 data, and only the auxiliary measurements are randomized; the approach taken by MJD was to vary the value of the nuisance parameters and *then* to generate toys at those new values, with the constraint terms kept at their nominal value [149]. The latter option, while still a valid statistical approach, is not a strict frequentist treatment of nuisance parameters in the profile likelihood method — but this can account for the fact that the observed limits for the two treatments disagree slightly. The median exclusion sensitivity is not sensitive to the differences in toy generation because the majority of toys generated with no signal have no signal counts to pull the shape nuisance parameters and the median of this distribution is unaffected. Thus, the values obtained from `freqfit` agree with the published MAJORANA DEMONSTRATOR results with a median exclusion sensitivity 90% CL lower limit of $T^{0\nu}_{1/2} > 8.1 \times 10^{25}$ years.

Finally, we performed a cross-check to the final results of GERDA. The full dataset and nominal values of nuisance parameters and their uncertainties are now available from the GERDA collaboration [56]. This dataset consists of 387 partitions and five background indices: four from phase I of GERDA, and one from phase II. The peak shape, likelihood, and analysis window are identical to that used in LEGEND-200.

Figure 4.6 shows the observed limit and exclusion sensitivity obtained using `freqfit` for the published GERDA partitions. The observed 90% CL lower limit is identical to the published value: $T^{0\nu}_{1/2} > 1.8 \times 10^{26}$ years. This value coincides with the median experimental sensitivity.

4.3.6 LEGEND-200 statistical results

The `freqfit` package is able to reproduce published results and passes a number of validity cross-checks. Now we are able to use it to produce the first frequentist statistical inference for the LEGEND-200 experi-

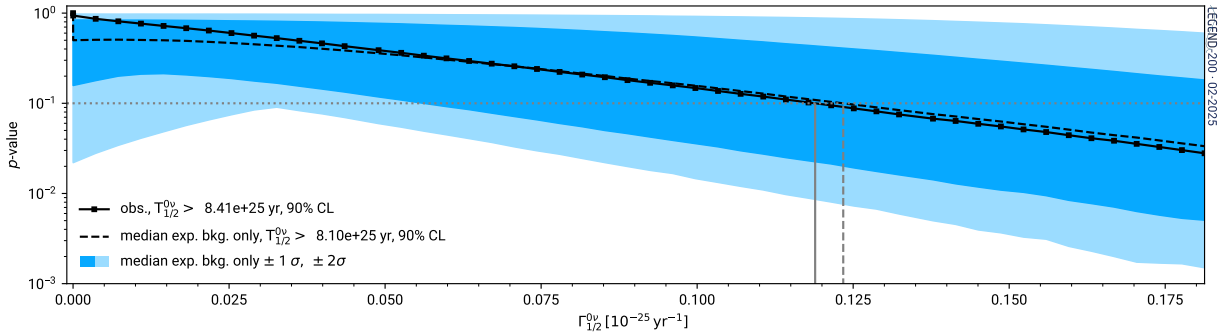

Figure 4.5: Sensitivity plot for the MAJORANA DEMONSTRATOR's final dataset obtained using the `freqfit` frequentist inference package. The observed p-values for the experiment are shown in the solid black line with square markers. The median exclusion sensitivity, or the median p-value expected for background only experiments, is shown in the dashed black line. The $\pm 1, 2\sigma$ uncertainty bands on the median exclusion sensitivity are shown in shades of blue. The median exclusion sensitivity matches published results in Ref. [55], while the observed limit deviates slightly due to differing treatments of nuisance parameters.

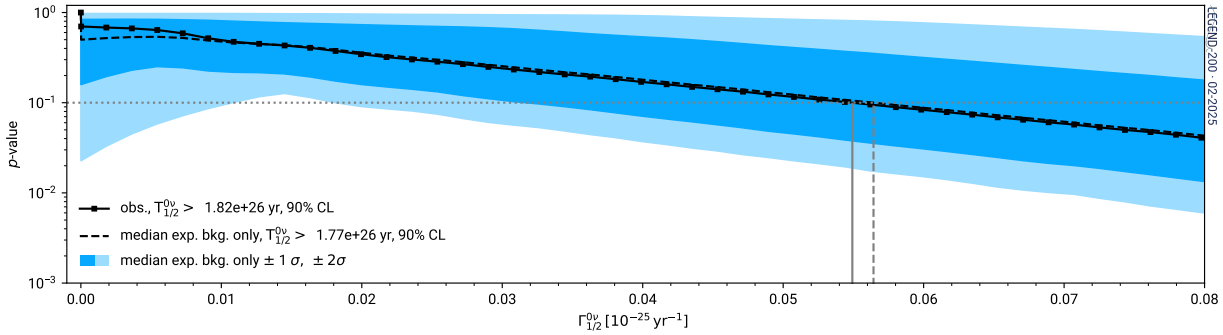

Figure 4.6: The exclusion sensitivity and observed limit for the final GERDA dataset obtained using the `freqfit` frequentist inference package. Both the median exclusion sensitivity and the observed limit agree with the published values in Ref. [56].

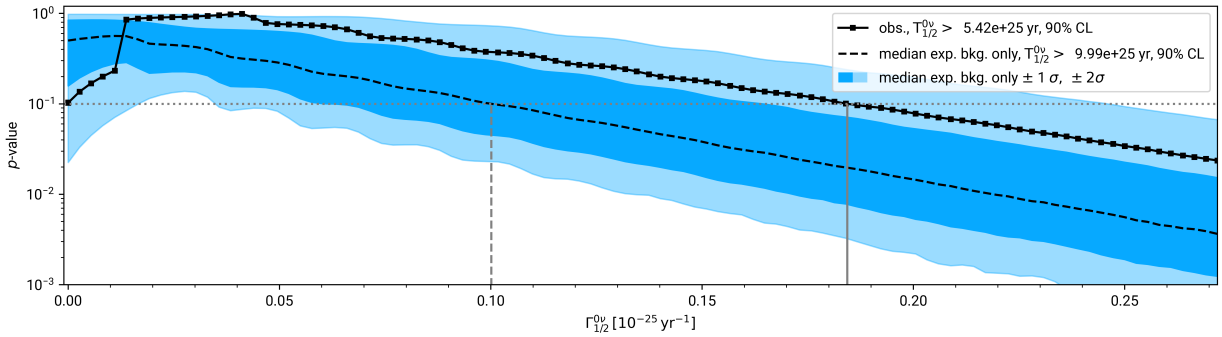

Figure 4.7: Observed limit and exclusion sensitivity for neutrinoless double-beta decay with LEGEND-200. The observed limit falls outside the 1σ confidence bands for the median exclusion sensitivity, but is contained within the 2σ bands — this is due to the presence of an event near $Q_{\beta\beta}$. The p-value for the $\Gamma = 0$ hypothesis is 0.1.

ment.

Observed limit and median exclusion sensitivity

The $\Gamma = 0$ hypothesis was determined to have a p-value of $p = 0.100$, which is above the 3σ discovery critical value of $p = 0.0027$. As such, we are unable to reject the $\Gamma = 0$ hypothesis and have not discovered neutrinoless double-beta decay. The sensitivity plot for LEGEND-200 is shown in Figure 4.7. The 90% CL lower limit on the half-life is at $T^{0\nu}_{1/2} > 5.4 \times 10^{25}$ years [8], and the median exclusion sensitivity is $T^{0\nu}_{1/2} > 1.0 \times 10^{26}$ years. The observed limit deviates significantly from the median sensitivity due to the presence of an event near $Q_{\beta\beta}$. An event surviving all cuts was observed at 2040.26 keV in the PPC detector P00661C: its energy is $1.3\sigma_{\mathrm{P00661C}}$ away from $Q_{\beta\beta}$, accounting for the energy bias correction, where $\sigma_{\mathrm{P00661C}}$ is the energy resolution of the detector during the partition in which the event occurred. Its A/E value for this event is known to be rather high; however, no high A/E cut was applied to the PPC detectors because it was determined to be more likely to generate inefficiency than to remove background, and thus its application was expected to hurt sensitivity [2]. The presence of this event near $Q_{\beta\beta}$ causes the nuisance parameters for this detector to be pulled during the profile likelihood treatment. The pull plot for the nuisance parameters in LEGEND-200 is shown in Fig. 4.8.

It is critical to note that because $\tilde{t}_\mu$ is a two-sided test statistic, we *should* be reporting both the lower

[8] If the approximation described in toy generation in step 7 is used, the limit is reduced to 5.3×10^{25} years. The approximation of using the nuisance nominal values instead of profile values to vary the auxiliary constraint is thus a conservative approach.

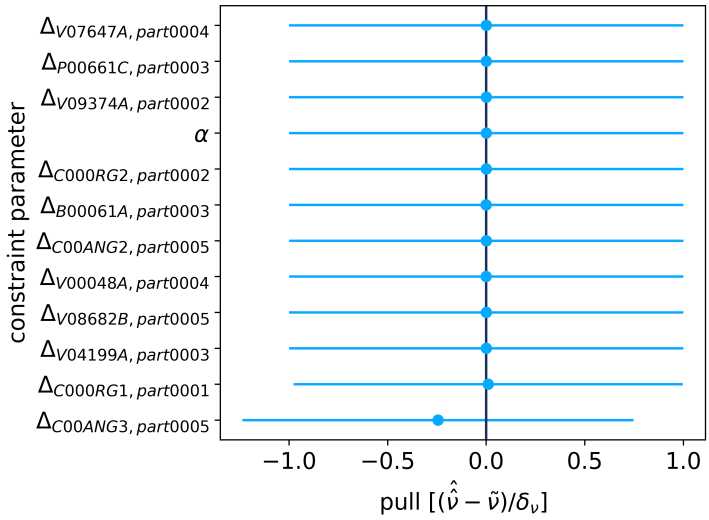

Figure 4.8: Pull plot showing the number of standard deviations each nuisance parameter is pulled from its nominal value during the best-fit to the dataset. Most constraints stay centered at their nominal value because there is no data near $Q_{\beta\beta}$ to pull them. The exceptions to this are the $\Delta_{P00661C,part0003}$ energy bias constraint and α, the efficiency uncertainty scaling. These are pulled because of the PPC that has an event near $Q_{\beta\beta}$: it pulls the energy bias so that the event energy is closer to the peak in the best-fit.

and upper limit for the observed limit. If the p-value at $\Gamma = 0$ were any lower, then it would be necessary to report a two sided limit. But because $\Gamma = 0$ is the lower value in the two-sided signal rate confidence interval, then it is understood that infinity is the upper value of the two-sided limit on the half-rate. This two-sided nature of the test statistic can lead to some confusion — especially with regards to coverage — and is discussed further in section 4.4.5.

The bumpy shape of the observed limit comes from the Poisson-like nature of the test statistic PDFs, which in turn arises from the extremely low background index and observed number of counts, see Fig. 4.10 for an example test statistic distribution. The shoulder-like features in the observed limit come from the relative motion between the observed test statistic (which follows a parabolic trend in value, see Fig. 4.11 for the observed value of the test statistic) and the complex motion of the test statistic PDFs as some bumps move rightward and some move leftward as the signal hypothesis being tested increases. This is seen more clearly in Figure 4.9. The test statistic PDFs deviate significantly from the asymptotic approximations to them from Ref. [134] due to the fact that the small number of counts creates this Poisson-like structure, Fig.

4.10 shows the asymptotic distribution given by

$$f(\tilde{t}_{\Gamma'}|\Gamma') = \begin{cases} \frac{1}{\sqrt{2\pi\tilde{t}_{\Gamma'}}} \exp\left(-\frac{\tilde{t}_{\Gamma'}}{2}\right) & \tilde{t}_{\Gamma'} \leq \Gamma'^2/\sigma^2, \\ \frac{1}{2\sqrt{2\pi\tilde{t}_{\Gamma'}}} \exp\left(-\frac{\tilde{t}_{\Gamma'}}{2}\right) + \frac{1}{\sqrt{2\pi}(2\Gamma'/\sigma)} \exp\left(-\frac{1}{2}\frac{(\tilde{t}_{\Gamma'}+\Gamma'^2/\sigma^2)^2}{(2\Gamma'/\sigma)^2}\right) & \tilde{t}_{\Gamma'} > \Gamma'^2/\sigma^2, \end{cases} \tag{4.36}$$

in the grey curve, where σ is the error on the signal rate Γ' estimated using the MINOS errors available in `iminuit`.

Discovery sensitivity

The median discovery sensitivity for the LEGEND-200 dataset can be obtained by computing the p-value for the median test statistic of toys generated and tested against signal rate Γ. This is done by comparing the median test statistic of toys generated and tested against signal rate Γ against the test statistic PDF for toys generated at true signal Γ and tested at $\Gamma = 0$. The median 3σ discovery sensitivity is shown in Fig. 4.12 and is computed to be 7.9×10^{25} years.

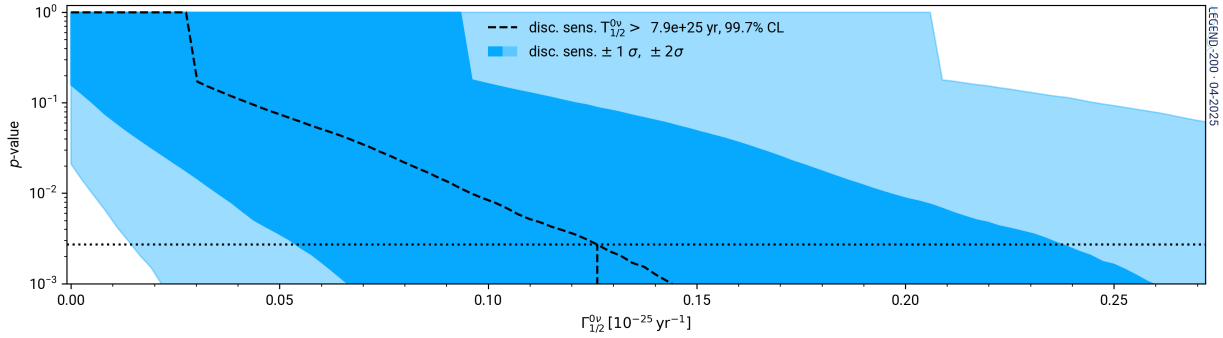

Figure 4.12: Median 3σ discovery sensitivity for the first year of physics data taking with the LEGEND-200 detector. The discovery sensitivity is highly impacted by the width of the peak shape and the energy bias uncertainty.

Coverage and power

The coverage can be computed by determining the frequency of toys that have the true signal rate (the signal rate the toys were generated at) falling within their confidence interval. Again, because $\tilde{t}_\mu$ is a two-sided test statistic, it is important to make sure the true value falls between *both* the lower and upper bounds. By

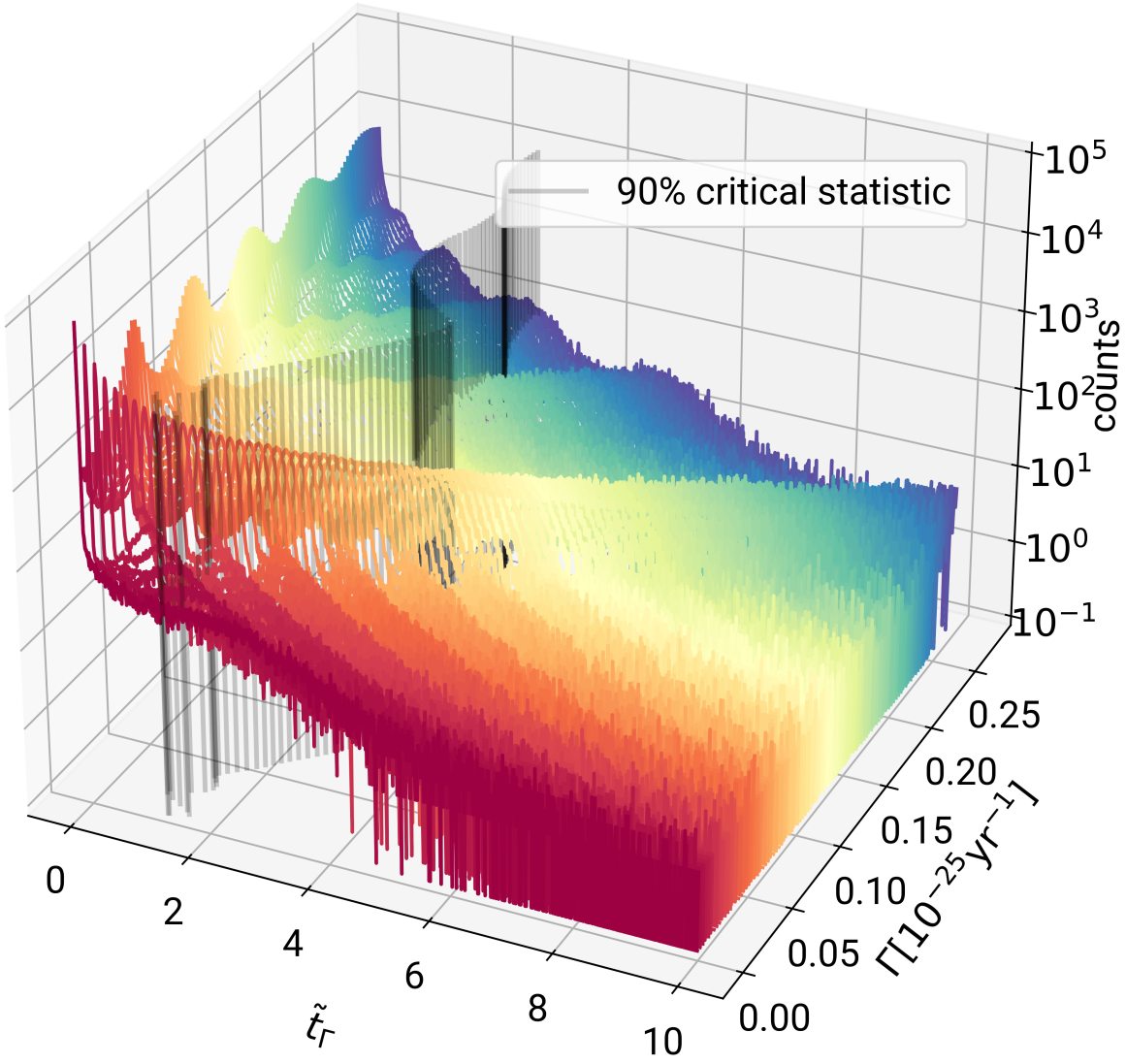

Figure 4.9: Evolution of test statistic distributions for the LEGEND-200 experiment's first statistical analysis. Each distribution has 10^5 pseudo-experiments generated at a specified Γ value on the y-axis, and each pseudo-experiment is tested at the same Γ value signal rate hypothesis using the two-sided test statistic $\tilde{t}_\Gamma$. Coloring is for visualization of depth only — starting from red at $\Gamma = 0$ to blue at large Γ. The large delta spike present in the $\Gamma = 0$ test statistic distribution has a component that splits out and moves to higher test-statistic values and broadens; this corresponds to toys that have no counts near $Q_{\beta\beta}$. Other Gaussian bumps become present at larger signal values that correspond to different number of counts present near $Q_{\beta\beta}$.

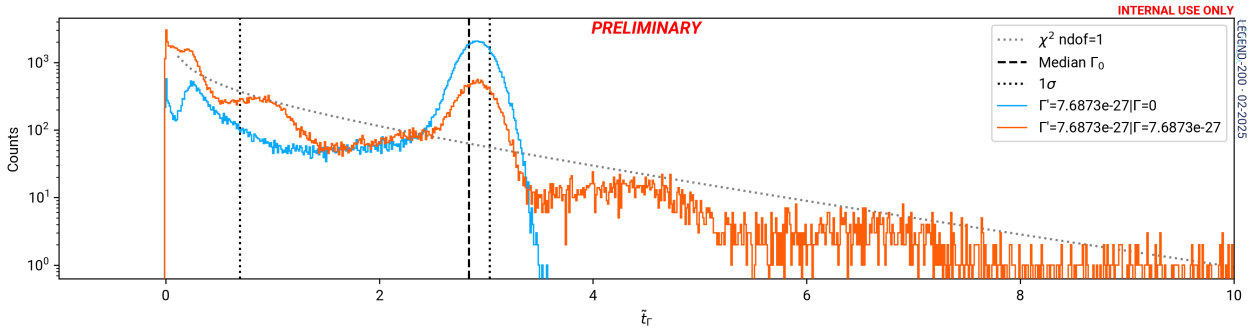

Figure 4.10: Example test statistic distributions from the statistical inference on the LEGEND-200 dataset. The two distributions each show 10^5 toys generated under different signal hypotheses, but with test statistics computed at the same signal rate hypothesis. The orange histogram generated and testing the signal rate $\Gamma = 7.6873 \times 10^{-27}$ yr^{-1} is a slice in Fig. 4.9; while the blue curve is generated at $\Gamma = 0$ and is used to compute the median exclusion sensitivity. The distributions of the test statistics follow a Poisson-like structure due to the limited number of counts being treated as signal-like due to the low background index and signal test points. The structure deviates highly from the asymptotic formula for the test statistic presented in Ref. [134] which is shown in the grey curve.

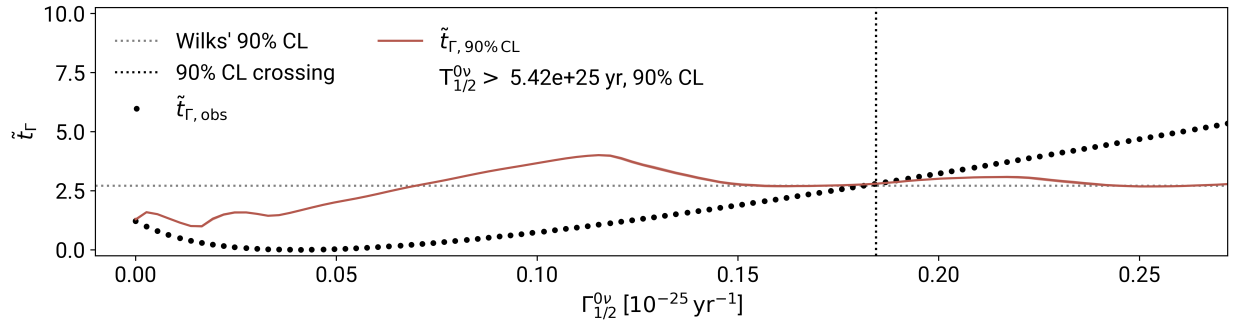

Figure 4.11: Plot of the critical and acceptance regions for the LEGEND-200 dataset, determined by the crossing of the 90% CL critical test statistic, $\tilde{t}_{\Gamma,90\%\text{CL}}$ in red, and the observed test statistic curves, $\tilde{t}_{\Gamma,\text{obs.}}$ in black. Note that the 90% CL critical test statistic deviates from the Wilks' approximation value for it of 2.71. This is due to the extremely small size of the LEGEND-200 dataset.

construction, there is exact coverage for toys that are generated at the profiled values of the test statistic. If we were to only always report the upper value of the limit, even in cases where the lower confidence interval values deviates from 0 and compute coverage this way, we will run into overcoverage. Figure 4.13 shows the coverage for the LEGEND-200 experiment if only an upper limit from the two-sided confidence interval is reported: there is significant over-coverage at the 95% level. This over-coverage comes from the toys where the true value of the signal rate is extremely small, and the two-sided confidence interval has a lower component larger than this signal rate, but if we only use the upper value as the confidence interval, we accidentally include this value that should not have been included.

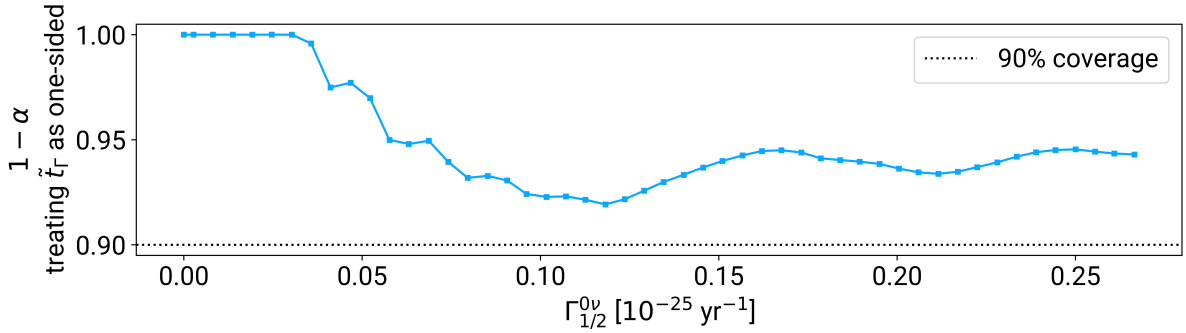

Figure 4.13: Coverage of the LEGEND-200 dataset if the lower value of the two-sided confidence interval was not reported. Note that there is significant overcoverage because toys where the true value would have fallen low outside a two sided confidence interval are now incorrectly covered. This shows the importance of always reporting both sides of the two-sided confidence interval.

The statistical power for the discovery test for the LEGEND-200 dataset can also be computed by determining the frequency of toys where a 3σ discovery is made. This curve should start at the 3σ p-value because toys generated at 0 signal should only make a 3σ discovery at the confidence level we set for the statistical test. The curve should also rise to 1 at some signal rate where we are sensitive enough to make a discovery. For a well-designed experiment, this curve should rise to 1 as fast as possible and at as small values of the signal rate as possible. Fig. 4.14 shows these features for LEGEND-200. The statistical power of the exclusion test can also be computed by taking the 90% critical value of the test statistic for background-only toys $f(\tilde{t}_\Gamma|\Gamma_0)$ and computing the power directly using this test statistic against the alternative hypothesis of toys generated with true signal rate $f(\tilde{t}_\Gamma|\Gamma)$. The power of the exclusion test should start at 0.1, and it does indeed start there in Fig. 4.15, which shows a plot of the exclusion power for the LEGEND-200 dataset.

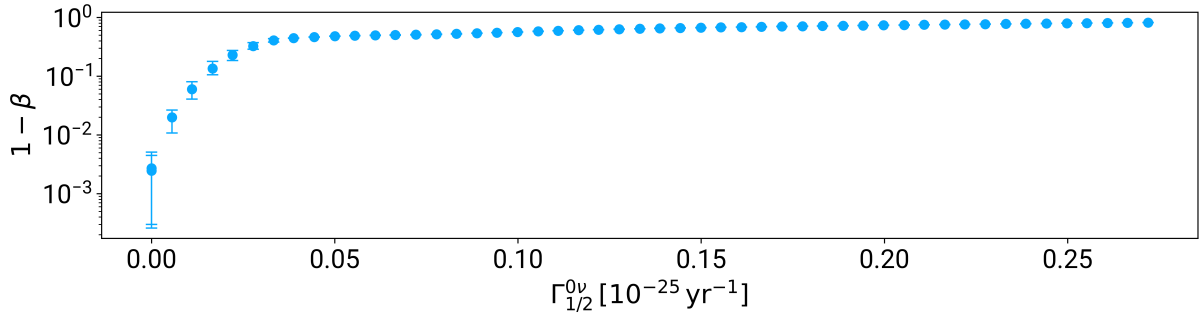

Figure 4.14: Statistical power of the discovery test for the LEGEND-200 dataset.

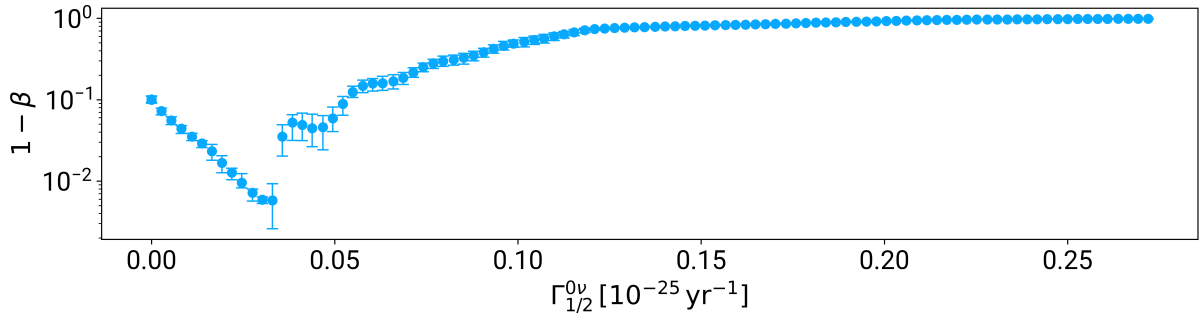

Figure 4.15: Statistical power of the exclusion test for the LEGEND-200 dataset. At $\Gamma = 0$ both the null and alternative hypotheses are $\Gamma = 0$, so the power should be equal to the significance level of the test, 0.1. The exclusion power then decreases until where the best-fit signal rate roughly occurs, $\Gamma \approx 0.04 \times 10^{-25}$ yr^{-1}. Because of the event present near $Q_{\beta\beta}$ the best-fit signal rate is non-zero and this causes the exclusion power to be weaker where the data prefer a count.

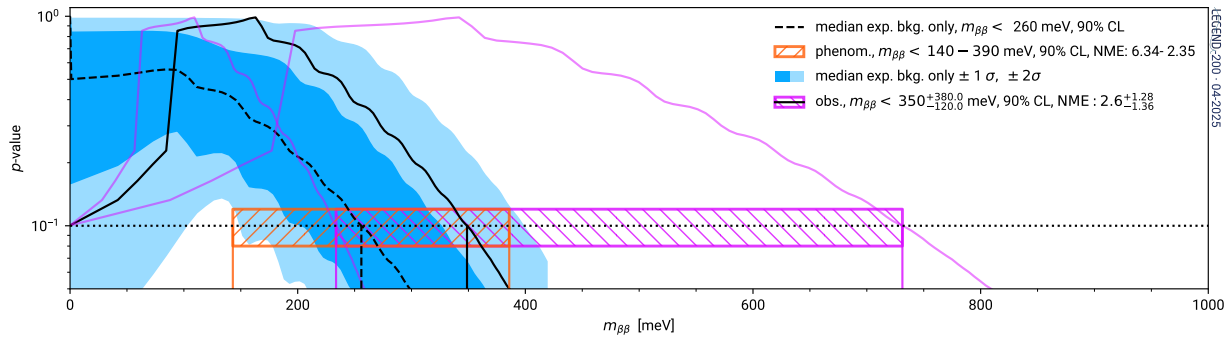

Figure 4.16: Limits on the effective Majorana mass using LEGEND-200 data, and both the phenomenological range of NME values taken from the U.S. Nuclear Physics Long Range Plan [150] shown in orange, as well as a hybrid-frequentist-Bayesian treatment using the symmetrized theory uncertainty from Ref. [37] shown in magenta. The central value of the *ab-initio* NME is used to plot the observed limit and median exclusion sensitivity. The observed limits from the phenomenological range of NMEs fall within the uncertainty on the median exclusion sensitivity; however, a large part of the parameter space from the very small allowed values of the *ab-initio* NME falls outside this uncertainty band. The uncertainties on the observed limit, in purple, come from the hybrid treatment of the uncertainty on the NME: the upper and lower uncertainties correspond to the limit obtained using the central value of the NME $\pm 1\sigma$ to produce a limit.

Limits on $m_{\beta\beta}$

The limit on the neutrinoless double-beta decay half-rate can be converted to a limit on the effective Majorana mass using Eqn. 1.47. The limit is sensitive to the value of the nuclear matrix element used. We use the range of phenomenological values from the U.S. Nuclear Physics Long Range Plan[150] 2.35 [151]-6.34 [152] to convert the observed half-life limit to a range of limits on the effective Majorana mass. In addition, we employ a hybrid-frequentist-Bayesian approach to leverage a new NME available for ^{76}Ge with quantified theoretical uncertainties [37]: this method is hybrid because the theory uncertainties are computed using Bayesian statistics. We take the central value of the NME posterior and its $\pm 1\sigma$ errors from Ref. [37] of $2.60^{+1.28}_{-1.36}$. Figure 4.16 shows the limits on $m_{\beta\beta}$ obtained using this range of values for the NME shown in the orange and magenta boxes on the plot. The central value of the *ab-initio* NME from Ref. [37] is used to draw the observed limit and median exclusion sensitivity — and its uncertainty — in terms of the effective Majorana mass.

Background index

In addition to the limit on the signal rate, another important statistical result from the inference procedure is a point estimate for the background indices and their uncertainties. The background indices are treated as nuisance parameters in the fit to the signal rate, but we can also perform statistical inference on the background indices with the signal rate treated as a nuisance parameter. We use the same test statistic $\tilde{t}_\mu$ again with a physical boundary on the background indices $BI \geq 0$. Because we have two background indices BI_1, BI_2 the correct frequentist treatment is to create a joint profile of these parameters, with the signal rate profiled out. Test statistic PDFs were generated on a two-dimensional Cartesian grid of values for the background indices, $[0, 1.4 \times 10^{-3}\,\text{ckky}] \times [0, 3.5 \times 10^{-3}\,\text{ckky}]$ each linearly spaced with 20 points. The p-value of the observed test statistic at each of these points was computed, and the contours of 1σ and 2σ p-values are indicated in Fig. 4.17. We see that the two background indices are almost completely uncorrelated; if, however, the Silver dataset had an event near $Q_{\beta\beta}$, then these background indices would become correlated as the fit tries minimize both background indices to optimize for the profiled signal.

Because the background indices are uncorrelated, it is a valid approximation to profile out one background index at a time to obtain uncertainties and central values. We are primarily interested in BI_1 from the Golden dataset because it contains the largest fraction of ICPC detectors, the detector technology that will be used exclusively in LEGEND-1000. Profiling out BI_2, Γ, and the remaining nuisance parameters we can generate toys at different values of BI_1 and compute the p-value of the observed test statistic from LEGEND-200 data. We can alternatively compute the critical value of the test statistic, or the value of the test statistic where the p-value corresponds to 1σ. The confidence interval is the where this critical test statistic crosses the observed values of the test statistic. This is shown in Fig. 4.18 for BI_1, which has a value of

BI_1 for the Golden dataset detectors

$$BI_1 = 5.4^{+2.7}_{-2.0} \times 10^{-4} \quad \text{counts}/(\text{keV} \cdot \text{kg} \cdot \text{yr}). \tag{4.37}$$

Carrying out the same process for BI_2 and profiling out the rest of the nuisance parameters yields

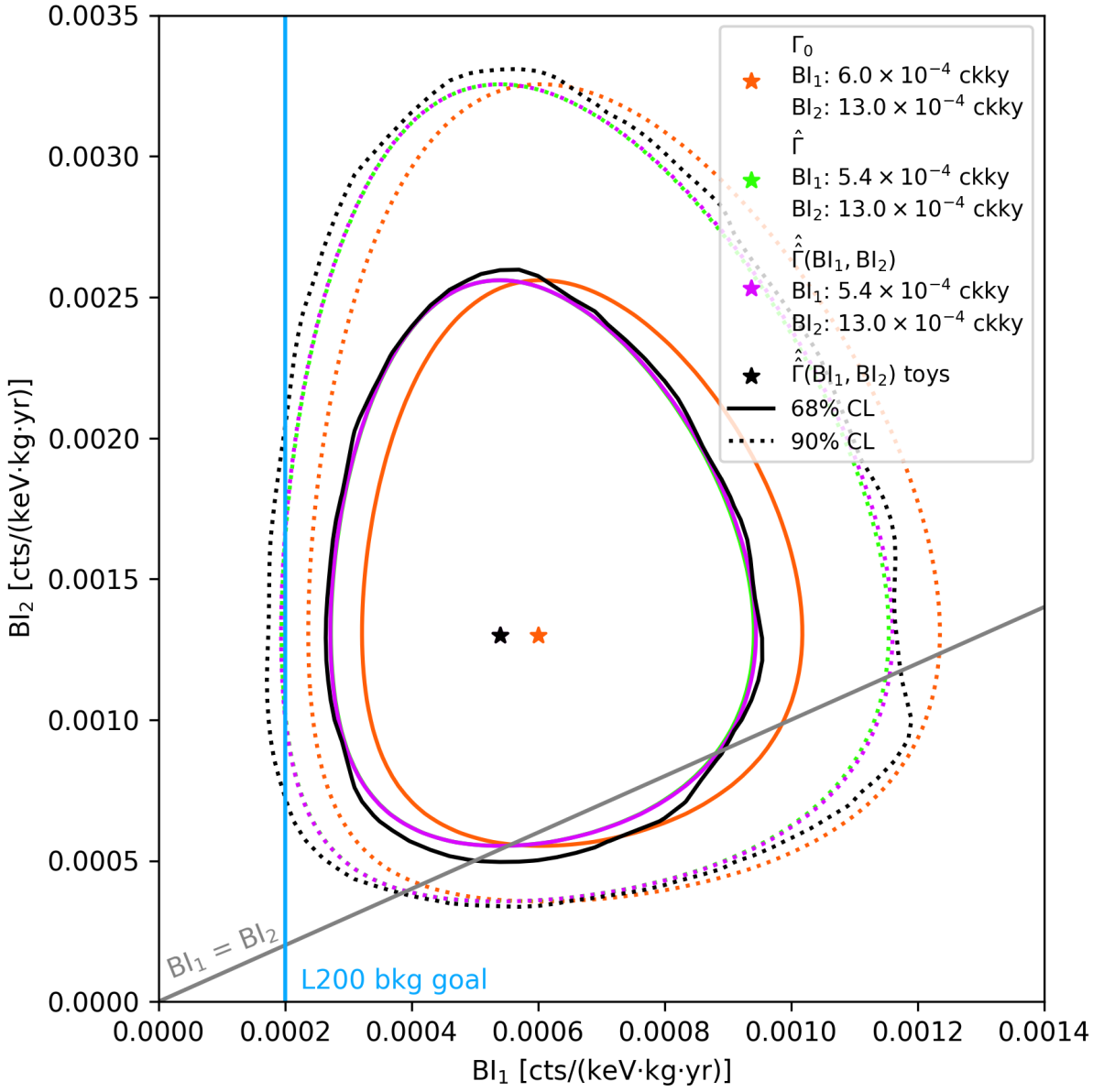

Figure 4.17: Joint profile of the two background indices in the LEGEND-200 dataset, corresponding to the Golden and Silver detectors' background indices BI_1, BI_2. The lack of tilt of the major axis of the confidence region indicates that these two parameters are largely uncorrelated.

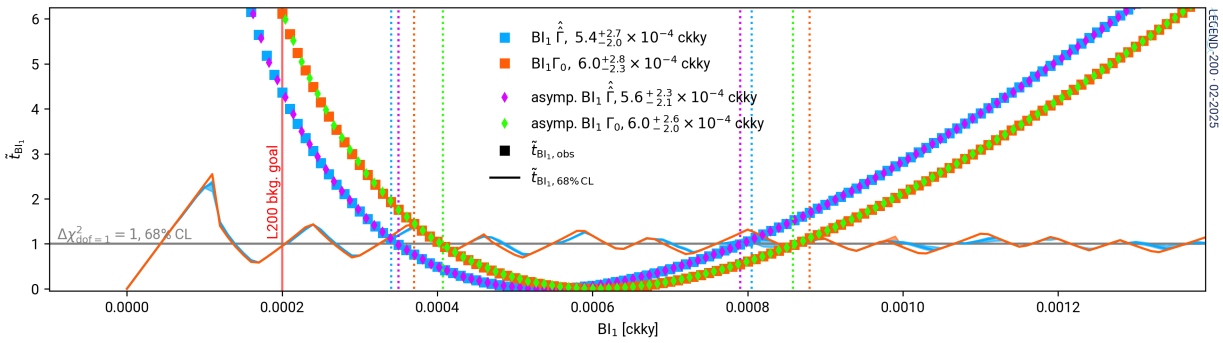

Figure 4.18: Profile likelihood and interval estimation for the Golden dataset's background index, BI_1. The interval estimation is performed in three ways. The first, and most correct treatment, is done by profiling out all other nuisance parameters to their best fit values and generating toy distributions; the confidence interval is built by checking where the 68 % critical test statistics (shown as the blue line) cross the observed test statistic (shown as blue squares). The second way is to use Wilks' approximation and the profile likelihood; the confidence interval is where the profile likelihood (the purple diamonds) crosses Wilks' critical value of 1, shown as a grey horizontal line. These two methods yielded similar results. The third estimation technique is to profile the likelihood with $\Gamma = 0$, which is a hybrid-Bayesian method. The toys-based method with $\Gamma = 0$ is shown in the orange, while the Wilks'-based method with $\Gamma = 0$ is shown in the green. These both yield a higher point estimate for the background index.

BI_1 for the Silver dataset detectors

$$BI_2 = 13.0^{8.0}_{-5.4} \times 10^{-4} \quad \mathrm{counts}/(\mathrm{keV} \cdot \mathrm{kg} \cdot \mathrm{yr}), \tag{4.38}$$

which is obtained from Fig. 4.19.

The LEGEND-200 background index goal of 2×10^{-4} counts/(keV · kg · yr) falls within the 2σ uncertainty intervals for both background indices. The Silver dataset background index BI_2 is larger than that of the Golden dataset due to worse PSD background-rejection performance of the coaxial detectors, where machine learning is required to gain any background-rejection capabilities. Due to the elevated background index BI_1 compared to the LEGEND-200 goal — as well as an excess of events before analysis cuts based on the background model — a series of radioassays and background modeling was undertaken to identify and remove any hot spots in the detector array. Short periods of non-physics data were taken and showed reduction in the background from uranium and thorium decay chains [128].

This inference on the background indices was performed with the signal rate profiled as a nuisance parameter, which is the correct frequentist treatment; however, because the best fit signal rate is non-

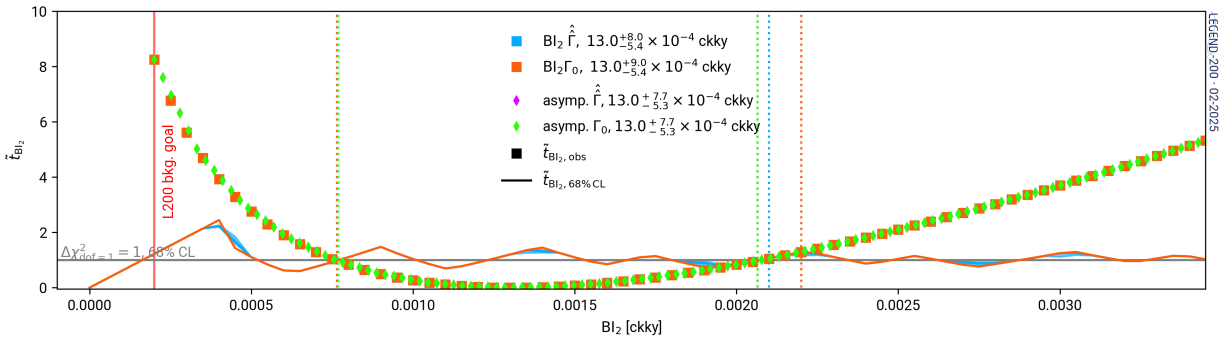

Figure 4.19: Profile likelihood and interval estimation for the Silver dataset's background index, BI_2. The interval estimation is performed in the same three ways as was done for BI_1. Again, the estimates obtained from the full frequentist treatment and from Wilks' approximations agree; this time, the estimate obtained from profiling at $\Gamma = 0$ also gives the same value. This is because there were no events near $Q_{\beta\beta}$ in detectors from the Silver dataset.

zero for this dataset, this dramatically alters the best-fit values of the background indices. In a hybrid frequentist-Bayesian treatment, where we interpret the one event near $Q_{\beta\beta}$ to originate from background and not neutrinoless double-beta decay because the $\Gamma = 0$ p-value is above the discovery threshold, we could instead assign this count to background and profile with $\Gamma = 0$ to obtain the background indices. Under the $\Gamma = 0$ hypothesis we find slightly higher background indices for the dataset with the event in it $BI_1 = 6.0^{+2.8}_{-2.3} \times 10^{-4}$ counts/(keV · kg · yr).

Background PDF shape

The main LEGEND-200 frequentist statistical inference used a uniform PDF to describe the background based on the background model; we also performed full statistical analyses of the LEGEND-200 dataset using a linear and an exponential model to describe the background PDF. The (un-normalized) linear background PDF was described by

$$B_k(E, BI, a) = BI\mathcal{E}_k\Delta E\left(\frac{2a}{(E_5 - E_0)^2}\left(E - \frac{E_0 + E_5}{2}\right) + \frac{1}{E_5 - E_0}\right), \tag{4.39}$$

where E_0, E_5 are the start and end point energies of the analysis window (with $[E_1, E_2]$ and $[E_3, E_4]$ forming the removed gamma line windows), and a is a shape parameter controlling the slope of the line and $a \in (-\infty, \infty)$. The background index BI and exposure $\mathcal{E}$ control the overall scaling of the line. This PDF was

properly normalized by the total integral on the full analysis window such that $\int B_k = BI\mathcal{E}_k\Delta E$.

The (unnormalized) exponential model with a constant offset was given by

$$BI_k(E, BI, a) = BI\mathcal{E}_k\Delta E\left(\frac{1}{a}\exp\left(-a(E-E_0)\right)+1\right), \tag{4.40}$$

where again a is a shape parameter controlling the decay of the exponential and was allowed to take on any value, while the background index and exposure controlled the overall scaling. Again, this PDF was properly normalized by the total integral on the full analysis window such that $\int B_k = BI\mathcal{E}_k\Delta E$.

Toys were generated with a background model and tested with the same background model — no studies were performed of background mismodeling where toys and test statistics were computed using different background models. The two background indices BI_1, BI_2 shared shape parameters in the fit, but had different scaling constants. We observed no significant difference in observed limits or in the p-value of the $\Gamma = 0$ hypothesis using these different background models. These results are summarized in Table 4.3.

	exclusion sensitivity 90% CL [$\times 10^{25}$ yr]	observed 90% CL limit [$\times 10^{25}$ yr]	p-value of $\Gamma = 0$
uniform background	10.0	5.43	0.101
linear background	10.0	5.45	0.114
exponential background	9.97	5.37	0.092

Table 4.3: Results summary for the $0\nu\beta\beta$ decay half-life assuming different background models

Coverage under background misevaluation

Because the background indices are nuisance parameters without auxiliary constraints in the fit, we are interested in seeing how sensitive the statistical inference results are to changes in the estimation of the background index. This is in the spirit of the recommendation from Ref. [153] to perform coverage checks if the true values of nuisance parameters differs from the values obtained during profile construction of the confidence interval. We generated toys at $\pm 1\sigma$ of the nominal background index for the Golden dataset BI_1 and then allowed the background index to float in the fit. We observed no major change in the observed limit, or the p-value of the $\Gamma = 0$ hypothesis in these two cases, as shown in Fig. 4.20.

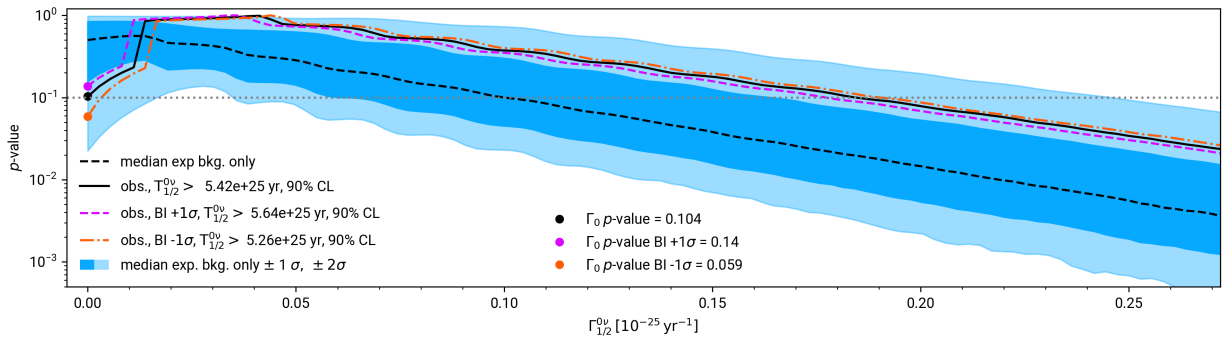

Figure 4.20: Sensitivity of the LEGEND-200 dataset under background misevaluation hypotheses for the Golden dataset background index. These limits were placed by generating toys with the background index at $\pm 1\sigma$ of its global best-fit value. No major variation in the limit or the significance of the $\Gamma = 0$ hypothesis were observed.

Coverage of ensembles

We also compute the coverage of the true signal rate for ensembles of pseudo-experiments generated with different true values of the background indices. The background indices are the primary nuisance parameter of concern because, again, they lack auxiliary measurements, and have the highest relative uncertainty on their point estimates. For a given true value of the background index, the profiled background index fluctuates toy-by-toy due to the Poissonian nature of the number of background counts; the profiled background index is then used to create test statistic distributions for hypothesis testing on a per-toy basis. Coverage is only guaranteed when the true value of the nuisance parameters are equal to the profiled value. We are interested in computing the coverage for ensembles of experiments treated with this profile likelihood ratio procedure where the background index is profiled out of the data and used to generate toys.

We are more interested in the coverage of BI_1, because its detectors are more similar to those that will be installed in LEGEND-1000; and because BI_1 and BI_2 are nearly uncorrelated, it is a good approximation to profile out BI_2 — this also makes things easier to visualize.

To perform this coverage check, we generated toys on a (Γ, BI_1) Cartesian grid of $[0, 0.01] \times [0, 1 \times 10^{-3}]$ with each linear dimension sampled at 20 equally spaced points. Each toy was generated with true signal rate Γ' and background index BI_1', and the test statistics were computed allowing the background index to float during the fit. In this way, the test statistic distributions $f(\tilde{t}_{\Gamma'}|\Gamma', BI_1')$ were generated and the 90% CL critical test statistics were computed for each point on the grid. For each point on the grid 10,000 more toys

were generated and the profiled best-fit of the background index was also recorded, giving $\tilde{t}_{\Gamma',\hat{BI}|\Gamma',BI'}$. We are interested in checking the coverage generated by the procedure of the profile likelihood, so we are interested in checking the coverage of the toys as if the profiled value of the background index $\hat{BI}_1$ were the true value of the background index instead of BI'_1 — this is what we mean by background misestimation. This would ideally mean computing the p-value of obtaining $\tilde{t}_{\Gamma',\hat{BI}|\Gamma',BI'}$ given the distribution of test statistics we would have built for the profiled toy using its profiled background index $f(\tilde{t}_{\Gamma'}|\Gamma',\hat{BI})$. This would be very computationally expensive; instead, we use the existing critical test statistics generated from the distributions $f(\tilde{t}_{\Gamma'}|\Gamma',BI'_1)$ and use a cubic spline interpolation to get the best estimate of the 90% CL critical test statistic for the background index $\hat{BI}_1$. If the observed test statistic for the toy $\tilde{t}_{\Gamma',\hat{BI}|\Gamma',BI'}$ is greater than the interpolated critical test statistic $\tilde{t}^{90\%\,\mathrm{crit.}}_{\Gamma'|\Gamma',BI'}(BI' = \hat{BI})$ then this toy does not cover the true value of Γ because we would reject the null hypothesis at 90% confidence. We repeat this process for 10000 extra toys at each point in the (Γ, BI_1) grid in order to compute the coverage for an ensemble of pseudo-experiments.

We only need to run 10000 extra toys at each grid point because the binomial uncertainty on the nominal 90% coverage is smaller than the precision on the confidence level: $\delta_p \approx \sqrt{\frac{(0.1)(1-0.1)}{10000}} = 0.003$, so a 0.3% uncertainty on the coverage values we compute. The coverage of ensembles is shown in Fig. 4.21 — the coverage at every value of the BI_1 is close to the nominal coverage of 90%. The full frequentist interval construction is robust in this case to errors in the point estimates of nuisance parameters.

Joint signal rate and background index profile

We can also perform a joint profile on both the signal rate and the background indices. We restrict our attention to BI_1 again because it is the nuisance parameter of primary interest. Fig. 4.22 shows the results of the joint profile for the signal rate and background index BI_1. Toys were generated and tested against the values Γ', BI'_1 and the observed p-value for the LEGEND-200 dataset was computed by fixing these parameters and profiling the remaining nuisance parameters. Again, these parameters are essentially uncorrelated, and the separate profile treatment of their confidence intervals is a very good approximation to the full frequentist confidence region. The contour for $p = 0.1$ is quite large in the parameter space because we are using a limited amount of data to fit two parameters with rather large uncertainty.

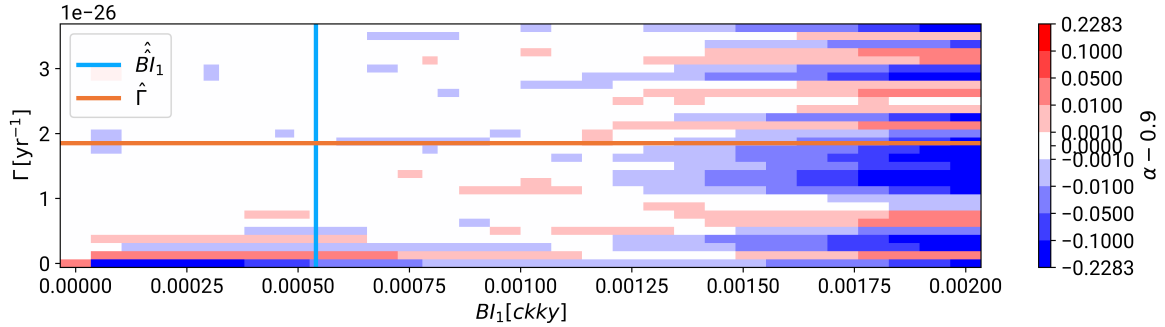

Figure 4.21: Coverage of ensembles of LEGEND-200 experiments generated at a variety of true background indices. The color bar represents the computed coverage the minus stated 90% coverage. Recall that $1 - \alpha$ is the confidence level of the test, and the significance level α is the probability of making a type I error. The coverage is below the stated 90% CL for toys generated with 0 true signal rate for a wide range of background indices close to the best-estimate for LEGEND-200. For each toy generated at true signal and true background BI, the coverage was checked against the test statistic distribution generated at the toy's profiled background index — as it would be done for a real experiment in the frequentist treatment.

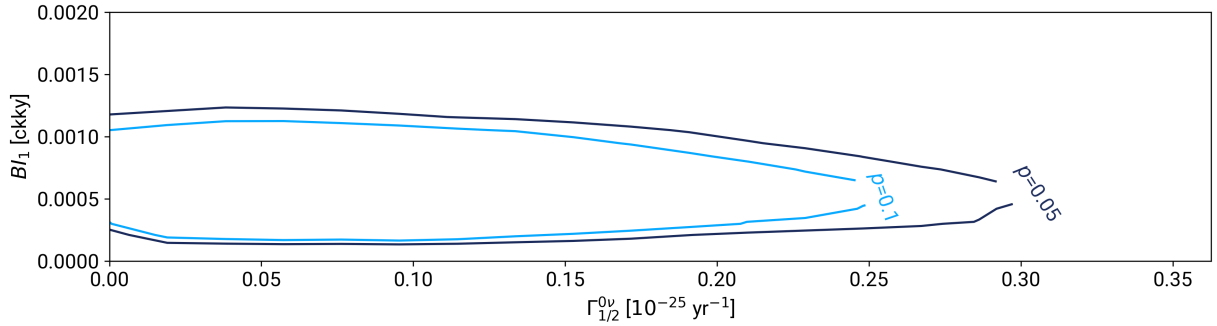

Figure 4.22: Joint confidence region for the neutrinoless double-beta decay half rate and the Golden dataset background index for LEGEND-200 data.

Scenario	Observed limit [10^{25} yr]	p_0
$\Delta_{\mathrm{P00661C,part003}}$	5.4	0.1
$\Delta_{\mathrm{P00661C,part003}} + 1\delta_{\Delta_{\mathrm{P00661C,part003}}}$	5.5	0.11
$\Delta_{\mathrm{P00661C,part003}} - 1\delta_{\Delta_{\mathrm{P00661C,part003}}}$	5.3	0.09

Table 4.4: Impact of the energy bias term for PPC00661C, partition 3 on the observed limit and p-value.

Impacts of other nuisance parameters on the limit

Besides the background indices, we are interested in how other nuisance parameters can impact the observed limit — this gives important information on how to change analysis strategies for the next physics result. Because one event was observed near $Q_{\beta\beta}$ in only one detector, the values of the nuisance parameters in the signal shape PDF for that one detector will have the largest impact on the observed limit for LEGEND-200 data. The impact of the energy bias on the limit is shown in Table 4.4 where the observed limit is reported for several values of the energy bias uncertainty of the PPC detector with the one event near $Q_{\beta\beta}$. The relative uncertainty on the energy resolution σ is the smallest for all nuisance parameters (see Table 4.1): this analysis parameter is highly optimized and can be constrained by auxiliary measurements very well and therefore has little to no impact on the limit. The energy bias term, however, is not highly optimized, and its relative uncertainty is extremely large because auxiliary measurements are difficult to perform. The values of the energy bias are large and contribute to an increased spread in the expected peak shape, reducing the discovery sensitivity. The energy bias uncertainty is also quite large in LEGEND-200 because it was decided to add an additional 0.3 keV uncertainty arising from ADC non-linearity effects in quadrature from the uncertainty estimated from the standard energy scale interpolation. The observed limit for LEGEND-200 without this additional 0.3 keV energy bias uncertainty is 5.5×10^{25} years, which is only a 2% change in the observed limit. Future neutrinoless double-beta decay analyses for LEGEND-200 will improve the energy bias uncertainty by applying an energy non-linearity correction to the data.

The final feature of this frequentist statistical treatment to note is the handling of the complete correlation of the efficiency uncertainties. The observed limit if the efficiency uncertainties are completely uncorrelated (where each partition gets its own efficiency scaling parameter α_i) is 5.7×10^{25} years, and the median exclusion sensitivity is unchanged at 1×10^{26} years. This confirms that the choice to correlate all the efficiency uncertainties was the more conservative option as it places the weaker limit. The decision to

correlate all efficiency uncertainties was made prior to unblinding due to dominance of a PSD efficiency systematic uncertainty common to all detectors. Future work could be done to determine the correlation matrix for the efficiency uncertainties and incorporate it into the statistical inference.

4.3.7 Statistical inference on LEGEND-200 + MJD + GERDA

We can place a limit on neutrinoless double-beta decay using recent datasets from the three most-recent ^{76}Ge $0\nu\beta\beta$ decay experiments: combining data from MAJORANA DEMONSTRATOR, GERDA, and LEGEND-200 increases the total exposure and sensitivity to $0\nu\beta\beta$ decay. Because all three experiments used very similar analysis techniques, the same method for estimation of the PSD cut efficiency uncertainty (checking the acceptance of $2\nu\beta\beta$ decay events represents the largest systematic uncertainty, see Sec. 3.2.5 for more details) [2, 55, 56, 148, 154], and the same simulation tools to calculate some efficiencies (such calculating the containment efficiency with GEANT4), statistically conservative treatment is to completely correlate the efficiency uncertainty between all three experiments. This is achieved by using one global efficiency uncertainty scaling parameter α for all three experiments. Again, this procedure was decided upon prior to unblinding, and future work could be done to determine the complete correlation matrix. The likelihood can be written as a product of the three individual likelihoods as well as the Gaussian auxiliary constraint term from this efficiency uncertainty:

$$\mathcal{L}_{combination} = \mathcal{L}_{MJD} \cdot \mathcal{L}_{GERDA} \cdot \mathcal{L}_{L200} \cdot \mathcal{L}_{aux.} \tag{4.41}$$

This likelihood is fed into the `freqfit` package to perform statistical inference on the combined data of these three experiments. Table 4.5. summarizes the key analysis parameters of this dataset.

Observed limit and median exclusion sensitivity

The p-value for the $\Gamma = 0$ hypothesis was computed to be $p = 0.27$, well above 3σ discovery critical value. The lower limit on the neutrinoless-double beta decay half-life was then computed to be $T^{0\nu}_{1/2} > 1.9 \times 10^{26}$ years. This is the strongest lower limit in ^{76}Ge to date and represents the combined efforts of three different experimental collaborations. The median exclusion sensitivity is calculated to be $T^{0\nu}_{1/2} > 2.8 \times 10^{26}$ years: this is the strongest half-life sensitivity ever achieved for $0\nu\beta\beta$ decay of ^{76}Ge. The results of the statistical

	MJD original	GERDA	LEGEND-200
# partitions	14	387	316
total exposure (kg·yr)	64.5	127.2	61.0
energy bias correction (keV), Δ	0.0 ± 0.15	-0.06 ± 0.16	0.24 ± 0.31
energy resolution (keV), σ	1.07	1.46 ± 0.06	1.17 ± 0.01
signal acceptance, ϵ	0.68 ± 0.03	0.56 ± 0.03	0.56 ± 0.03
tail fraction, f	0.17	-	-
tail scale (keV), τ	1.6	-	-
peak shape scale, γ	1 ± 0.03	-	-

Table 4.5: Statistical analysis parameter comparison between experiments, reported as the exposure-weighted average across partitions. Efficiency uncertainties are treated as correlated, i.e. also exposure-weighted averaged. Note that MJD used a slightly different accounting and incorporated some efficiencies into the exposure, hence the difference in signal acceptance compared to GERDA and LEGEND-200. The energy bias uncertainty is less constrained in LEGEND-200 due to the presence of an ADC non-linearity that was uncorrected in the first analysis but is now under study, see Sec. 4.3.6 for details.

inference on the combined datasets are shown in Fig. 4.23. Again, the observed limit falls within the $\pm 2\sigma$ uncertainty band on the median exclusion limit due to the presence of a few counts near $Q_{\beta\beta}$ within the combined datasets.

Discovery sensitivity

The median 3σ discovery sensitivity of the combination of experiments was also computed and was determined to be $T^{0\nu}_{1/2} = 1.8 \times 10^{26}$ years. The discovery sensitivity is shown in Fig. 4.24.

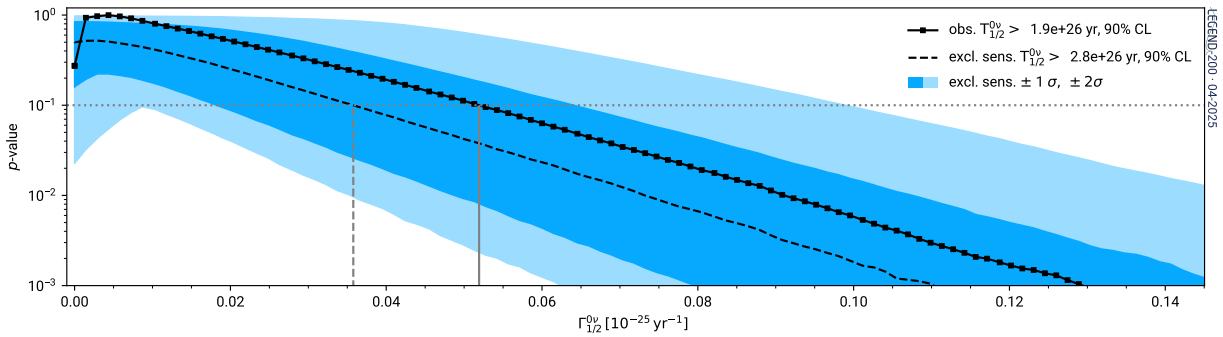

Figure 4.23: Observed limit and median exclusion sensitivity for the combination of likelihoods from GERDA, MAJORANA DEMONSTRATOR, and LEGEND-200.

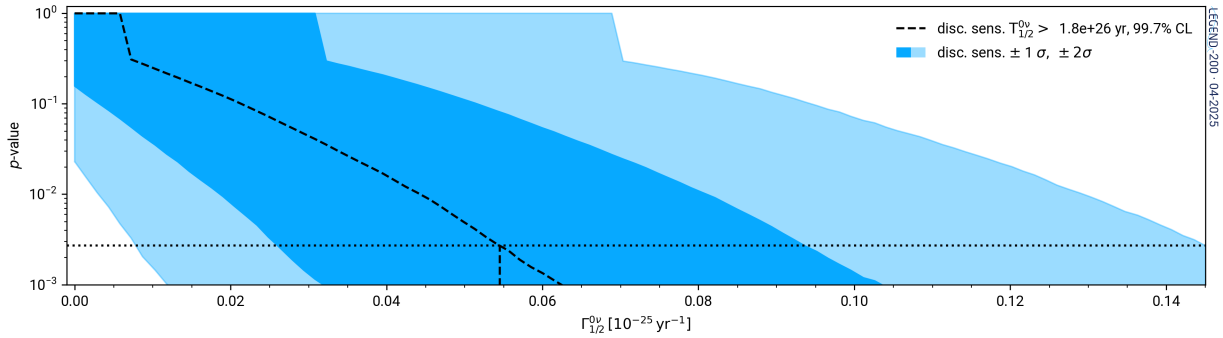

Figure 4.24: Median 3σ discovery sensitivity for the combination of likelihoods from the three most recent germanium-76 neutrinoless double-beta decay experiments.

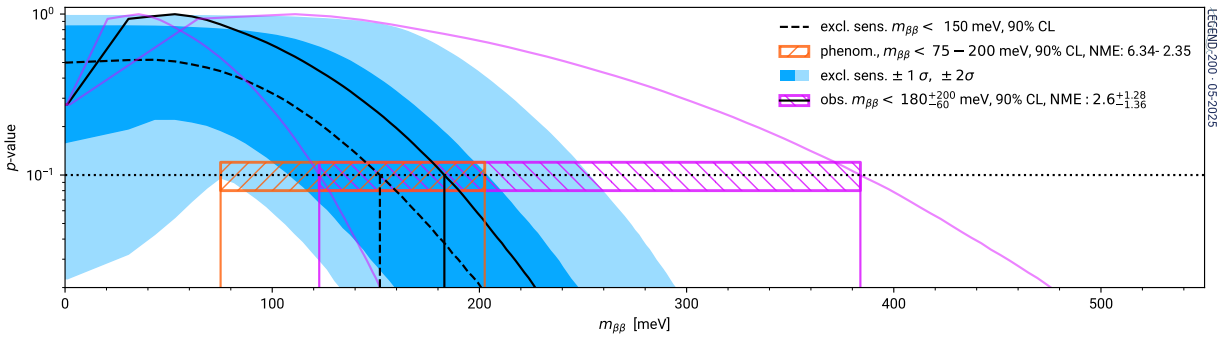

Figure 4.25: Limits on the effective Majorana mass using the combination of likelihoods from GERDA, MAJORANA DEMONSTRATOR, and LEGEND-200. Limits obtained from both the strict-frequentist phenomenological NME range and the hybrid-Bayesian-frequentist method using the theory uncertainty from Ref. [37] are shown.

Limits on $m_{\beta\beta}$

We can now convert the lower-limit on the neutrinoless half-life to a lower limit on the effective Majorana mass, again using Eqn. 1.47. We report both the range of lower-limits obtained using the range of phenomenological values of the NME [150], as well as the lower-limit and its uncertainties obtained using the hybrid-frequentist interpretation of the uncertainty-quantified NME [37]. Fig. 4.25 shows the ranges of these limits on the effective Majorana mass. These are the most stringent limits on the effective Majorana mass obtained from ^{76}Ge data to-date. The uncertainty quantified NME from Ref. [37] can be converted into a limit on the effective Majorana mass using a completely Bayesian statistical inference method; this gives $m_{\beta\beta} < 320$ meV, 90% credible interval for a uniform prior on the signal half-rate [2, 155].

The race to cover the inverted ordering

We can plot the experimental lower limits for the effective Majorana mass on top of the allowed parameter space for neutrinoless double-beta decay, as discussed in Sec. 1.7.2. Figure 4.26 shows this comparison; the limits on the effective Majorana mass from ^{76}Ge data are still far away from the flat part of the inverted ordering allowed parameter space in the $m_{lightest} \rightarrow 0$ limit, which begins at $\mathcal{O}(50$ meV$)$. The top of the inverted ordering has been excluded under some values of the NME for data collected by KamLand-Zen 800 with xenon-136 as the neutrinoless double-beta decay isotope [1]. The ultimate exclusion sensitivity of LEGEND-200 should reach this parameter space after more data taking. An exclusion sensitivity of 1×10^{27} years would correspond to a limit on the effective Majorana mass — using 6.34 [152] value of the NME — of 50 meV. If LEGEND-200 continues to take data using the same active exposure as the first physics deployment, and with the same background index, it will take 6 years to reach this exclusion sensitivity. This value can be computed using asymptotic approximations to the Poisson distribution found in Ref. [22]. The run time to obtain this exclusion sensitivity can be dramatically reduced if the active exposure is increased through better live-time and/or better PSD acceptance.

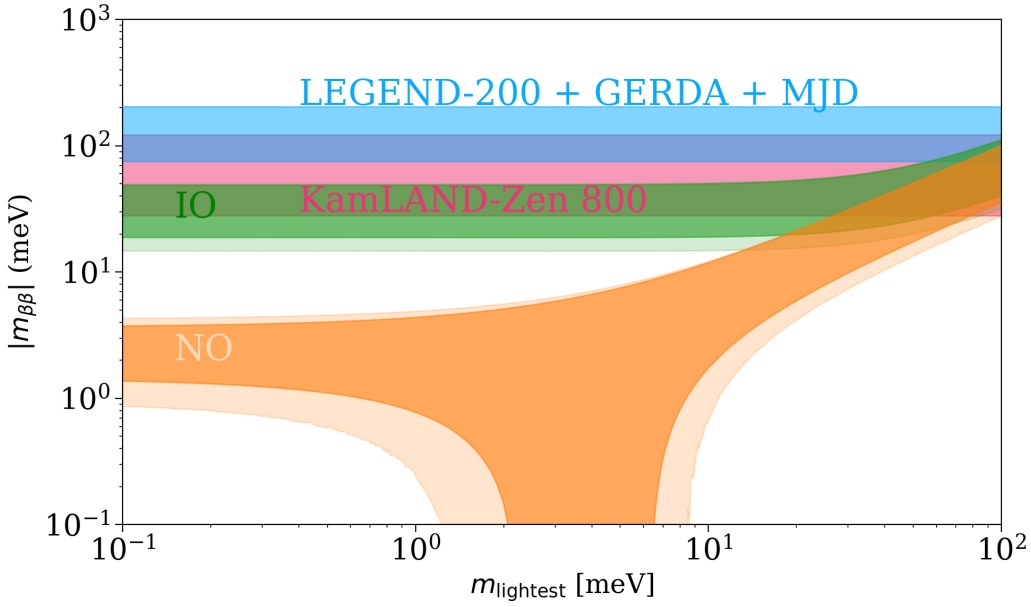

Figure 4.26: Allowed parameter space for the effective Majorana mass and two recent exclusion limits from LEGEND-200 [2] and KamLAND-Zen 800 [1].

4.4 Statistical methods for quasi-background-free experiments

As neutrinoless-double beta decay experiments continue to reduce their background index, the expected number of background counts in the vicinity near $Q_{\beta\beta}$ will also decrease. Experiments that have less than one expected background count within a 2σ region of $Q_{\beta\beta}$ are known as *quasi-background-free* (QBF) experiments. LEGEND-1000 aims to reach this quasi-background-free regime: with a projected background index of 10^{-5} counts/keV/kg/yr after analysis cuts, only 0.44 counts are expected within 2σ of $Q_{\beta\beta}$ after 10000 kilogram-years of exposure [89]. Quasi-background-free experiments enjoy a nearly linear scaling in the discovery sensitivity as a function of exposure throughout their experimental program. As the total number of counts in the analysis window decreases, it will become even more important to perform the full frequentist statistical treatment with pseudo-experiment generation, as the test statistic distributions for these experiments will become even more non-asymptotically distributed. This section marks a step forward in the understanding of extended unbinned frequentist statistics in this low-count regime, and numerous suggestions are made for better and faster handling of data from these quasi-background-free experiments.

We utilized the `freqfit` package on a mock experiment, modeled after the proposed analysis goals of LEGEND-1000 [89]. The mock experiment consists of 100 partitions, each with a Gaussian-distributed exposure (centered at 100 kg · yr with an uncertainty of 1 kg · yr) so that the total exposure amounts to 10000 kg · yr. Each partition has the same energy resolution of 1.1 keV with an uncertainty of 0.1 keV, where the uncertainty is based off of LEGEND-200. The energy bias for each partition is a Gaussian random variable sampled from a distribution centered at 0.3 keV and with width of 0.3 keV, again based on LEGEND-200. The signal acceptance efficiency was set at 0.69 for all partitions, with the uncertainty on the signal acceptance efficiency equal to 0.05 for all partitions. All partitions were given the same background index BI and efficiency uncertainty scaling α.

4.4.1 Deviation from asymptotic behavior

As mentioned previously in section 4.3.6, both the distribution of test statistics and the critical value of the test statistics deviate from their asymptotic values under Wilks' approximation. The test statistics follow a clearly non-chi-square distribution as seen in Fig. 4.10, and the critical test statistics deviate from the predicted $\Delta\chi^2 = 2.71$ in Fig. 4.11. Because the test statistic behaves non-asymptotically for LEGEND-

200, it is worth investigating if other asymptotic approximations break down. One of particular interest is the approximation to the discovery and exclusion sensitivities that are used to quickly project future experiments' sensitivity to new physics as presented in Ref. [22].

In order to compare against this approximation, we created ensembles of pseudo-experiments — similar to what was done in Sec. 4.3.6 — in order to compute the required median discovery and exclusion sensitivities. For a given value BI_{true} it was necessary to generate toys at many different true Γ signal rates to determine the median sensitivities. Then at a particular $(\Gamma_{true}, BI_{true})$ point, toys are generated. But because the background index is estimated from data, and the data have Poisson fluctuations, the profiled value of the background index $\hat{\hat{BI}}$ varies from toy-to-toy. Therein lies the rub: the profile likelihood ratio procedure then dictates that the observed toy test statistic must be compared against a test statistic distribution generated with *the same background index* that the toy was profiled at, $\hat{\hat{BI}}$. This causes the number of required Γ, BI pairs to grow rapidly. Computational requirements were ameliorated by creating toys on a fine grid of BI — for a fixed Γ — and then computing the critical test statistic for each grid point BI. The critical test statistic as a function of BI was then interpolated with a cubic spline to yield $\tilde{t}_{crit.}(BI)$. The observed test statistic for a toy generated at a $(\Gamma_{true}, BI_{true})$ point was then compared against the critical test statistic determined using the cubic spline at the toy's profiled background index, $\tilde{t}_{crit.}(\hat{\hat{BI}})$, to determine if the toy makes a discovery or exclusion — depending on what hypothesis was being tested. For a given BI_{true} we scan a grid of Γ_{true} with this procedure until the fraction of toys making a discovery/exclusion is equal to 50%.

In Fig. 4.27 we plot the exclusion and discovery sensitivities for the mock LEGEND-1000-like experiment using the full, extended unbinned likelihood treatment presented in this chapter. For comparison, we also plot the estimates of these sensitivities using the formulae in Ref. [22], as well as the exact Poisson trends they approximate. We see that there is large deviation from the predicted sensitivities. This is not surprising as the full likelihood treatment including the peak shape will give different limits than one-bin Poisson counting. For the LEGEND-1000 background index goal, it will require 4 counts to make a 3σ discovery. The exclusion sensitivity is better than the approximate predictions because the toy construction has exact coverage, while the coverage obtained using the approximation is conservative.

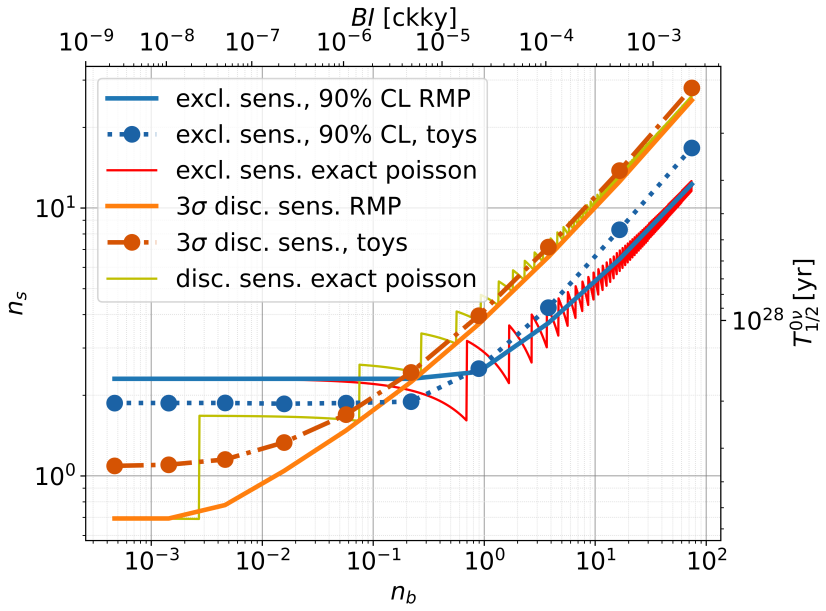

Figure 4.27: Number of signal counts required for a median 3σ discovery and the exclusion sensitivity as a function of background counts in the region of interest for a LEGEND-1000-like experiment. The solid orange and blue lines are obtained by using approximate formulae found in Ref. [22], and are referred to in the plot as RMP. These formulae are analytic approximations to the discrete one-bin signal-plus-background Poisson model, shown in the solid yellow and red lines. The darker blue and orange dashed lines with markers were computed using pseudo-experiment ensembles treated with the full profile likelihood ratio treatment using an extended unbinned maximum likelihood. A Gaussian signal plus flat background was used for the PDF, and importantly, the background was treated as an unknown nuisance parameter without any constraints. For experiments the in quasi-background-free regime near 1 expected count in the region of interest, there is a slight deviation from the approximated behavior.

4.4.2 Approximate test statistic distributions for quasi-background-free searches

When investigating the structure of the test statistic distributions for LEGEND-200 we performed the full statistical inference with all the nuisance parameters, except the background index, set to their central value and not included in the fit. This resulted in the test statistic distribution shown in Fig. 4.28, which is overlaid on the test statistic distribution when the nuisance parameters are treated in the full frequentist method. The structure of the distribution is jarringly different: there are large spikes in the test statistic instead of semi-Gaussian peaks. This was a rather surprising result and we wanted to make sure that the feature was real and that `freqfit` was not giving incorrect fit results that produced this spike in error.

The structure in the distribution is indeed a Dirac-delta-like spike centered at one value of the test

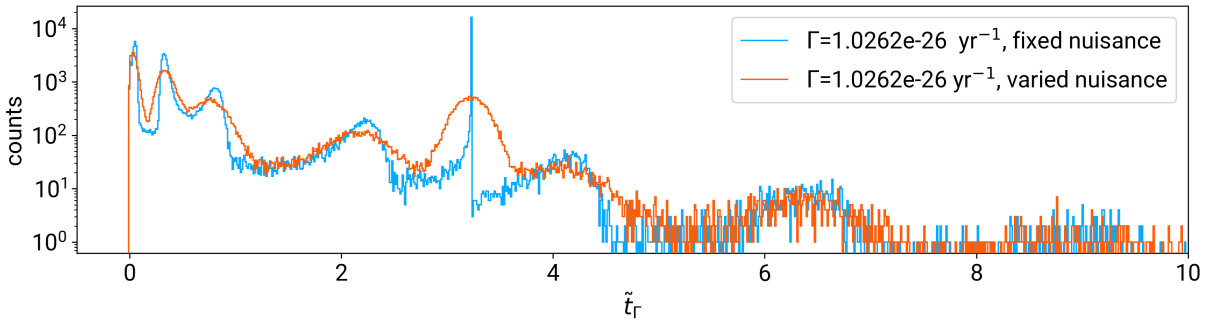

Figure 4.28: Test statistic distributions for LEGEND-200 when nuisance parameters are allowed to vary and when nuisance parameters besides the background index are fixed to their nominal values and not fit. There is a marked difference in the distributions near the median, where the varied nuisance parameters have a Gaussian peak structure, while the fixed nuisance parameter toys have a large spike in the test statistic value.

statistic. It is generated by toys in which no events are classified as signal-like and the best-fit signal rate is zero; furthermore, the background index global and profiled best-fits coincide because there are no signal-like counts. In the case of no nuisance parameters, this means that the test statistic is a single value directly proportional to the signal rate. See Appendix A.1 for more details. Following the derivation shown in Appendix A.1, the location of the test statistic for toys with no signal-like counts can be computed as

Test statistic for toys without counts near $Q_{\beta\beta}$

$$\tilde{t}_\Gamma^{\text{bkg.only}} = 2\Gamma \sum_{i=0}^{M} (\epsilon_i + \delta_{\epsilon_i}\tilde{\alpha})\mathcal{E}_i - \Gamma^2 \delta_\alpha^2 \left(\sum_{j=0}^{M} \delta_{\epsilon_j} \mathcal{E}_j \right)^2 . \tag{4.42}$$

With the nuisance parameters incorporated into the frequentist treatment, Eqn. 4.42 yields both the location and the width of the Gaussian peak from toys with no signal-like events arising in the test statistic distribution. The number of toys with no signal-like events can be approximated by computing the number of toys without any counts within $\pm 3\bar{\sigma}$ of $Q_{\beta\beta}$, where $\bar{\sigma}$ is the exposure weighted average energy resolution. This approximation works because the signal PDF density is very small at $\pm 3\bar{\sigma}$ and can be neglected without much loss of accuracy. Fig. 4.29 shows a test statistic distribution along with the location and width of the Gaussian peak computed using Eqn. 4.42.

Investigations of the `freqfit` results initially showed that there was some small variation, on the order

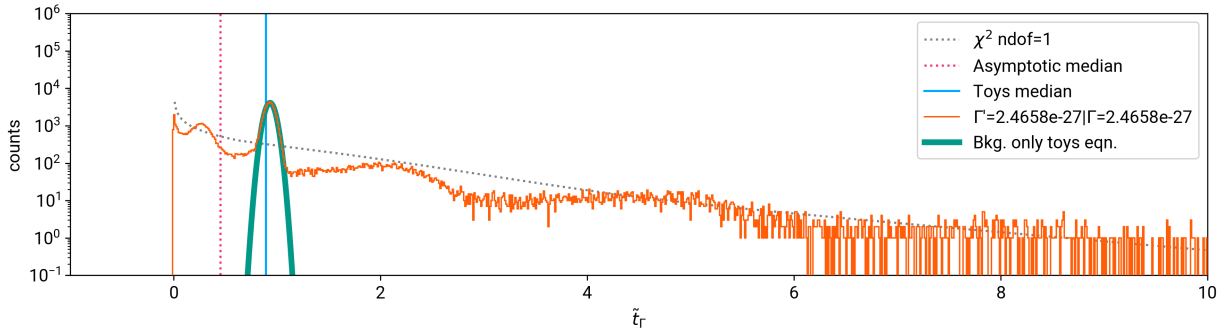

Figure 4.29: Test statistic distribution generated with nuisance parameters in the unconditional ensemble with an overlay of the contribution to the PDF from toys with no counts near $Q_{\beta\beta}$ computed exactly.

of 10^{-15} (!), in the computed value of this spike for toys with fixed nuisance parameters — see Fig. 4.30. The origin of this variation was traced to floating point instabilities in the computation of the difference in the numerator and denominator of the test statistic, which was kept at double floating point precision for all calculations. Putting the parameters on a grid effectively puts the test statistic on a grid that is larger than this floating point precision and yielded more stable numerical subtraction. Putting parameters on a grid and rounding the test statistic to the 10^{-10} level yielded a correct delta spike, shown in Fig. 4.31.

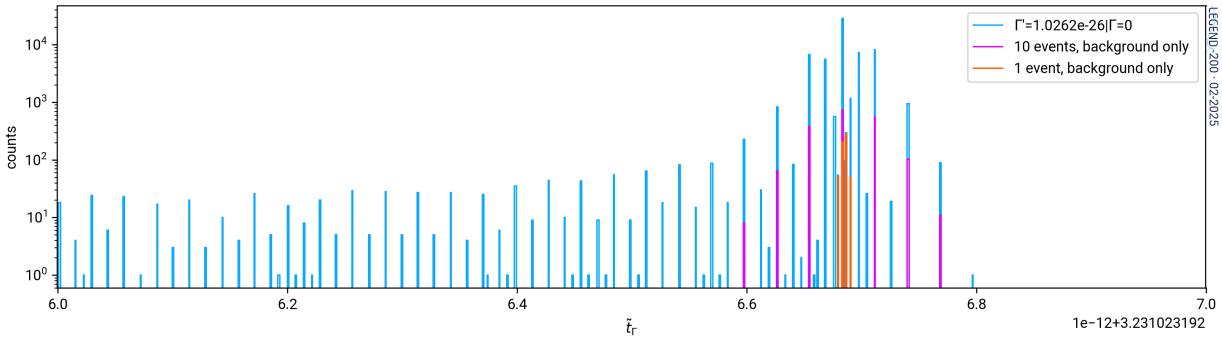

Figure 4.30: PDF of the test statistic for LEGEND-200 with all nuisance parameters fixed for toys where all events fall outside a 30 keV window of $Q_{\beta\beta}$. This figure is zoomed in on the blue spike located near $\tilde{t}_\Gamma \approx 3.23$ in Figure 4.28. The x-axis in this plot can be read by adding 3.231023192 to the abscissa value multiplied by 10^{-12}. According to Eqn. 4.42, these background-like toys should all have the same test statistic; due to machine precision errors, they do not.

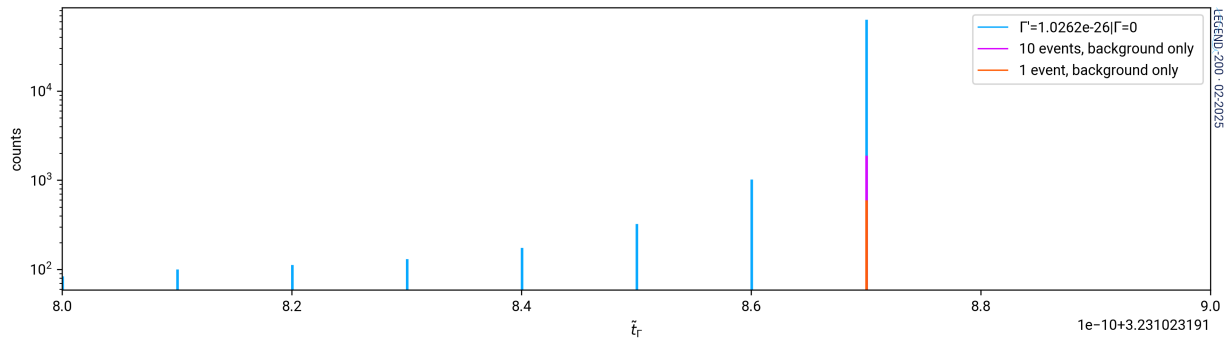

Figure 4.31: PDF of the test statistic for LEGEND-200 with all nuisance parameters fixed, but with the test statistic rounded at the 1×10^{-11} level, for toys where all events fall outside a 30 keV window of $Q_{\beta\beta}$. Rounding the test statistic ensures these events fall in a delta spike given by Eqn. 4.42.

4.4.3 Handling of the background index in the quasi-background-free regime

Some future tonne-scale $0\nu\beta\beta$ decay experiments, such a LEGEND-1000 and CUORE with Particle Identification (CUPID), will be able to treat their background after analysis cuts as uniformly distributed [89, 156]. Moreover, because these experiments often lack a sufficient background model after analysis cuts, the extremely low background rate in these experiments is treated as a nuisance parameter of the statistical inference without any auxiliary measurement to constrain it. The experiments thus take as wide an analysis window as possible in order to determine the background rate as precisely as possible; although, due to limited statistics, large uncertainty remains for point estimates of the background rate. This treatment of the background rate causes some unique statistical problems, such as under/over coverage in the event of background misevaluation, and as experiments start to reach the quasi-background-free regime, these problems will only be exacerbated. While some previous investigations have shown that low background rates — and thus a small number of counts due to background near $Q_{\beta\beta}$, n_b — can cause deviation from the asymptotic limit for critical values of the test statistic [157], there has been limited investigation into statistical inference when the background rate is *this* low. It is particularly important to check the coverage of the experiment in this case, just as we did above for LEGEND-200.

Figure 4.32 shows the results of checking coverage for ensembles of LEGEND-1000-like pseudo-experiments with different true backgrounds. At very low background rates there are modest amounts of undercoverage, but, by-and-large, the coverage is exceedingly good.

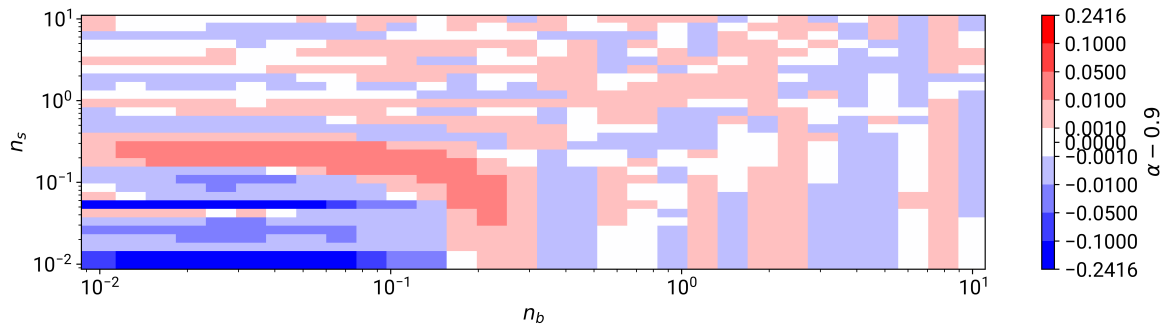

Figure 4.32: Coverage of ensembles of LEGEND-1000-like experiments generated at a variety of true background indices. See the text for further information about how the computation was performed. There are severe amounts of undercoverage for experiments in the background-free regime; and there is only modest overcoverage for experiments that are quasi-background-free.

4.4.4 The unconditional vs. conditional handling of nuisance parameters

It is the α parameter that contributes the major difference between the observed limits obtained using the "unconditional" or the "conditional" ensemble treatment of nuisance parameters. Table 4.6 shows the limits obtained using these two valid frequentist statistical treatments on LEGEND-200 data. Although the "unconditional" ensemble yields a lower value for the upper limit, we follow the recommendations from the ATLAS statistics forum for its use due to the fact that the "unconditional" ensemble reproduces the correct asymptotic behavior [143].

	Obs. limit 90% C.L. [10^{25} yr]	Med. exc. sens. 90% C.L. [10^{25} yr]
Unconditional treatment	5.42	9.99
Conditional treatment	5.61	9.88

Table 4.6: Comparison of frequentist statistical inference on LEGEND-200 data with conditional and unconditional treatment of the nuisance parameters.

4.4.5 Choice of test statistic

Although LEGEND-200 used the two-sided $\tilde{t}_\mu$ test statistic, it is worth thinking about what scenarios this test statistic is good for, and what other test statistic options exist. There has been some debate within the rare-event search community over which test statistic to use. All neutrinoless double-beta decay experiments have used the $\tilde{t}_\mu$ test statistic to report upper limits and to project sensitivities. Some experiments in the dark matter community however have used the $\tilde{q}_\mu$ test statistic for exclusion tests and upper limits, but there has

recently been an agreement to switch to use of $\tilde{t}_\mu$ to avoid allegations of flip-flopping [153, 158].

The idea of *flip-flopping* was investigated by Feldman and Cousins in their paper where they re-introduced the profile-likelihood ratio ordering of the test statistic [136]. They noted that statistical coverage was not guaranteed if one switched from setting a confidence interval to an upper limit if the significance of discovery was below 3σ. This problem is easy to see in Fig. 4.13 which shows the coverage of LEGEND-200 computed if only the upper side of the two-sided confidence interval were reported: there are some toys where the lower band of the correct 90% confidence interval is *not* at zero rate, but instead is set to 0. It would then be a mistake in these cases to only report an upper limit instead of a two-sided confidence interval — the coverage would be incorrect. Flip-flopping arises when the statistical exclusion test is conditioned on the outcome of the discovery test, which is to say that because 0 rate was not excluded at 3σ then it shouldn't be part of the 90% confidence interval. Their solution to the problem of flip-flopping was to introduce a profile likelihood ratio ordered test statistic that "lifts off" naturally from an upper limit to a two sided confidence interval. This profile likelihood ratio test statistic is equivalent to $\tilde{t}_\mu$ without nuisance parameters.

One possibility for rare-event search experiments is to use $q_0 = \tilde{t}_{\Gamma=0}$ for discovery and perform two different hypothesis tests depending on the result of it: if $p(q_0)$ falls below the threshold for a 3σ discovery, use $\tilde{q}_\mu$ for exclusion/upper limits; and if $p(q_0)$ is a 3σ discovery, use $\tilde{t}_\mu$ to report a two-sided confidence interval for a measurement. Although this has the *appearance* of flip-flopping, it is not an example of it. The crux of the issue is the interpretation of the probability content of a one-sided upper-limit and the upper value of a two-sided confidence interval (even where the lower value of this two-sided confidence interval may be at 0!). The test statistic $\tilde{q}_\mu$ is a one-sided hypothesis test and is built for upper limits; it has proper coverage by construction. The interpretation of the two-sided confidence intervals placed by use of $\tilde{t}_\mu$ should be exactly that, two-sided; it also has proper coverage by construction. These two test statistics have different hypotheses and answer different and valid statistical questions. If we interpret $\tilde{q}_\mu$ as a one-sided upper limit and $\tilde{t}_\mu$ as a two-sided limit, and because $\tilde{t}_\mu$ and $\tilde{q}_\mu$ both have proper coverage due to their construction, this is a valid approach.

One drawback of this approach is that one experiment could report a two-sided interval, but then a subsequent experiment fails to reject the zero-signal hypothesis and then reports a one-sided upper limit.

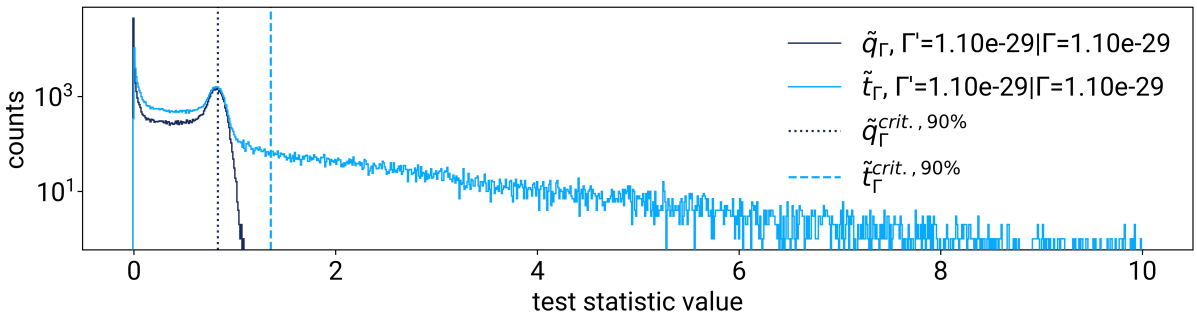

Figure 4.33: Comparison of test statistic probability distributions generated using the two-sided $\tilde{t}_\mu$ and one-sided $\tilde{q}_\mu$ test statistics. For $\tilde{q}_\mu$, toys that have a best-fit signal rate that is larger than the test value get their test statistic set to zero. This results in a distinct lack of Gaussian bumps from Poisson fluctuations near $Q_{\beta\beta}$ at high test statistic value compared to those obtained by using $\tilde{t}_\mu$.

This may be embarrassing, but it is statistically valid. Finally, one other way to get around flip-flopping allegations is to provide with the full likelihood function that allows a reader to compute intervals ordered with $\tilde{t}_\Gamma$ or $\tilde{q}_\Gamma$, whichever they need to suit their analytical needs. In that case, what gets reported in a paper is just a matter of convention, and does not affect the coverage of either set of intervals taken individually. CERN analyses recommend use of q_0 for discovery tests, and $\tilde{q}_\mu$ for upper limits [159]. This overall approach is statistically valid and not treated as flip-flopping.

We have generated median exclusion sensitivity under the LEGEND-1000 target background index using both $\tilde{q}_\mu$ and $\tilde{t}_\mu$ test statistics. We note several differences between these constructions. Because $\tilde{q}_\mu$ is designed to construct one-sided upper limits by setting the test statistic to 0 when the best-fit signal rate is higher than the hypothesis under test, the structure of its test statistic distributions is different from those generated by $\tilde{t}_\mu$. Figure 4.33 shows an example comparing the two test statistic distributions for the example experiment generated at the same signal rate and nuisance parameters and tested against a 1.1×10^{-29} yr^{-1} signal rate. Because the signal rate hypothesis is low, and the true signal rate used to generate these toys was high, there are a large population of toys where $\tilde{q}_\mu$ was set to 0 because of a large best-fit rate.

The differences between the median exclusion sensitivity are apparent in Fig. 4.34 which shows the obtained result for the LEGEND-200 dataset. The median exclusion sensitivity at 90% CL increases from 1.00×10^{26} years to 1.16×10^{26} years, and the observed one-sided limit is 6.6×10^{25} years compared to the two-sided interval 5.4×10^{25} years. The structure of the median p-values for $\tilde{q}_\mu$ is also distinct: the

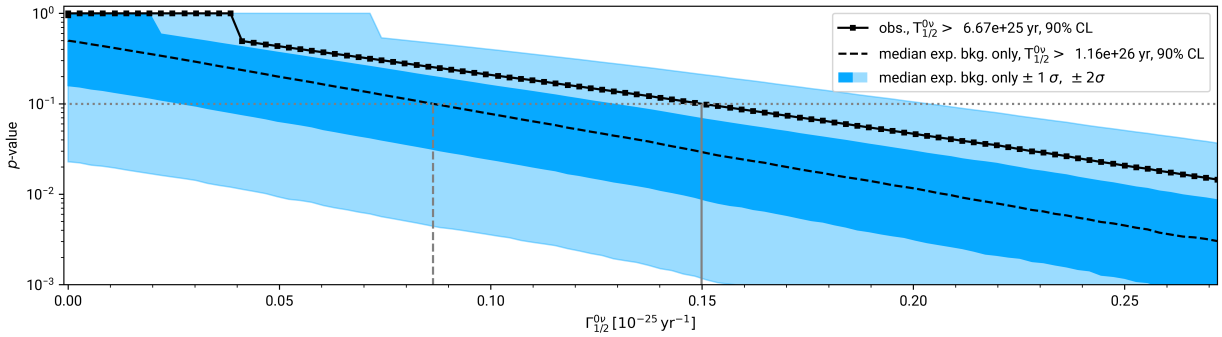

Figure 4.34: Exclusion sensitivity and observed limit of LEGEND-200 data with the $\tilde{q}_\mu$ test statistic.

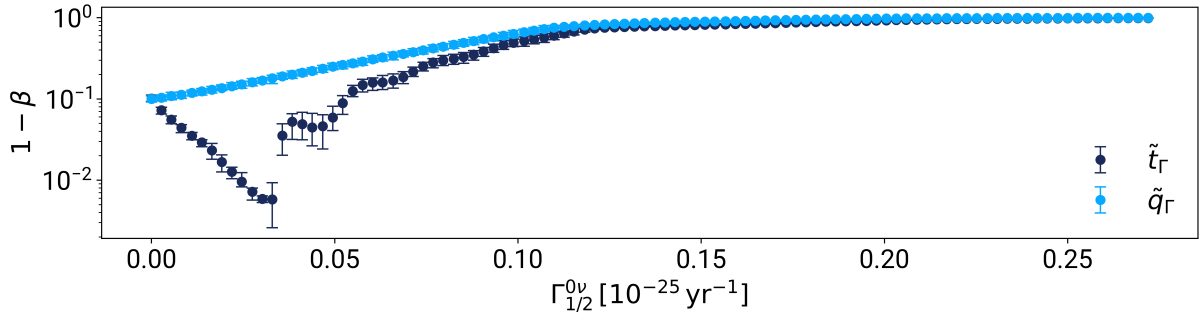

Figure 4.35: Exclusion power of $\tilde{t}_\mu$ and $\tilde{q}_\mu$ when applied to LEGEND-200 data. The exclusion power for $\tilde{q}_\mu$ goes to 1 much more rapidly than $\tilde{t}_\mu$. This is desirable for a statistical test for exclusion.

p-values at small signal rates are all close to 1 because the test has no discovery power, it only can compute upper limits. The superiority of $\tilde{q}_\mu$ over $\tilde{t}_\mu$ for setting upper limits can also be seen in Fig. 4.35 which shows the difference in exclusion power of these two test statistics. The value of $1-\beta$ is computed by determining the fraction of toys generated with true signal equal to zero and tested against rate Γ' that are rejected by the critical value of the test statistic determined from toys generated and tested at Γ' (see Sec. 4.2.3 for discussion of one-sided tests and hypothesis testing). Across a wide range of true signal and true background values, $\tilde{q}_\mu$ has much larger exclusion power.

It is our recommendation to use $\tilde{q}_\mu$ over $\tilde{t}_\mu$ for setting upper limits: $\tilde{q}_\mu$ gives stronger upper limits due to its one-sided construction at no cost to exclusion power or coverage. As others have pointed out, there is about a 20% increase in statistical sensitivity obtained in using $\tilde{q}_\mu$ over $\tilde{t}_\mu$ as a test statistic [153].

4.5 Summary and outlook

As experiments head further into the quasi-background-free regime, their statistical tests will deviate even further from asymptotic approximations. As is, the `freqfit` package is deployable for statistical inference on LEGEND-1000 data — as shown above in Section 4.4 where results from a simulated LEGEND-1000-like experiment are shown. As the number of datasets grows, the timing performance of `freqfit` will suffer due to a large amount of random numbers being generated in pseudo-experiment generation. GPU-based acceleration, or smart parallelization, of toy data draws would speed up performance by many factors. Because the background index in LEGEND-1000 will be so low, many datasets will be empty and without events in them; continuing to combine empty datasets will be of critical importance for speed in LEGEND-1000 statistical inference. Not combining empty datasets in the simulated LEGEND-1000 experiment causes the run time to increase by more than a factor of $\mathcal{O}(10)$[9]. Finally, if the dimension of fit during likelihood minimization increases to order 100 parameters (most likely through a multivariate treatment of the analysis), it would be beneficial to GPU-accelerate the minimization. This can be done with a drop-in replacement of the `numba` kernel set to target a GPU instead of a CPU; auto-differentiation could also improve minimization performance if need be. Further exploration of analytical treatments for specific aspects of the calculation could also yield large improvements.

We have shown that the `freqfit` Python package passes robust statistical cross-checks, and then used the software to generate frequentist statistical limits for the first LEGEND-200 data release. By combining the LEGEND-200 dataset with those of GERDA and MAJORANA DEMONSTRATOR, we obtain the best half-life limit on neutrinoless double-beta decay in germanium-76 to-date: 1.9×10^{26} years. However the corresponding limit on the effective Majorana mass is still far away from the allowed parameter space for the inverted ordering. LEGEND-200 should be able to enter this parameter space once it collects more data in the coming years, and LEGEND-1000 will have a median 3σ discovery sensitivity that completely covers the inverted ordering.

[9] Tested using LEGEND-200 data on a Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80GHz CPU

Chapter 5

Cryogenic characterization of silicon photomultipliers

5.1 Light readout and background rejection in LEGEND

A majority of the results presented in this section were reported previously in Ref. [160]. The detection of liquid argon (LAr) scintillation in coincidence with a germanium detector energy deposition is used to tag and remove events that do not resemble $0\nu\beta\beta$ decay, which should be contained entirely within the germanium. Silicon photomultipliers were chosen as the main light readout technology for LEGEND-200 over photomultiplier tubes due to their low radioactive background, low operating voltage, and comparably high quantum efficiency [161–163]. The liquid argon scintillation, peaked at 128 nm in the vacuum ultraviolet (VUV), is wavelength-shifted and guided to several arrays of SiPMs by a shroud of tetraphenyl butadiene (TPB) coated, wavelength-shifting (WLS) optical fibers [161]. There are two wavelength conversions involved in this collection process: the TPB first shifts the VUV scintillation light to approximately 430 nm [164], and then the WLS optical fibers shift the light to green wavelengths, approximately 460 nm to 580 nm. These green light wavelengths (460 nm to 580 nm) were chosen because they are well matched to the spectral response of the KETEK PM33100T SiPMs used for light readout in LEGEND-200 [165], as well as other models of SiPMs such as the RGB-HD from Fondazione Bruno Kessler, the S13360 from Hamamatsu, or any other commercially available visible light sensitive SiPMs. This method of collecting

the VUV scintillation offers a large area of exposure to scintillation as opposed to using VUV-sensitive SiPMs; in addition, the wavelength-shifting also better matches the spectral response of commercially available visible-light sensitive SiPMs. However, the total collection efficiency of the liquid argon detector — without accounting for shadowing from the LEGEND geometry — is expected to be $\sim 0.1\%$ of the total number of VUV photons produced [166].

Characterizing the photon detection efficiency (PDE) of SiPMs at cryogenic temperatures is important for modeling the LAr scintillation detector response of LEGEND-200 — the current phase of LEGEND that is actively taking data. The PDE is the probability of a SiPM detecting a single incident photon and converting it into a useful signal. The overall light collection efficiency of the LAr detector is the product of individual collection, absorption, transmission, emission, and detection probabilities in the chain of steps a VUV photon takes from scintillation to detection in a SiPM; the PDE is the final link in this concatenation. Cryogenic SiPM PDE measurements will also inform the design of the future phase of LEGEND with 1000 kg of active isotope which aims to reach a discovery sensitivity for $0\nu\beta\beta$ decay beyond 10^{28} years. Knowledge of the SiPMs' PDE at liquid argon temperature (87 K) from experimental characterizations is important for optimizing LEGEND-1000's LAr light readout, and for accurately projecting the background [89].

In this chapter, we describe the design of a test stand capable of measuring the PDE at cryogenic temperatures that was first introduced in Ref. [113]. We report a characterization of the cryogenic PDEs of two commercially available SiPMs — a KETEK PM3325-WB-D0 SiPM and a Hamamatsu S13360-3050CS SiPM — near liquid argon temperature to determine potential effects on the efficiency of the LEGEND liquid argon scintillation readout system. LEGEND currently uses KETEK PM33100T SiPMs; the KETEK PM3325-WB-D0 was one of the latest available SiPM models from KETEK before the acquisition of SiPM assets by BROADCOM®. The Hamamatsu SiPM, in contrast, has a higher PDE at the peak emission wavelength of the wavelength-shifting fibers currently used in LEGEND and is of R&D interest for LEGEND-1000. We find that the cryogenic PDEs at 562 nm for these two devices exhibit a greater than 20% decrease from their room temperature PDE values across a broad range of overvoltages.

5.2 SiPM operating principles

Silicon photomultipliers (SiPMs) are semiconductor Geiger-Müller photodetectors with single-photon resolution and can be used to determine the arrival time of individual photons with sub-nanosecond precision due to their good single photon time resolutions and good photon detection efficiencies [167–171]. This makes SiPMs ideal candidates for low-light detection applications [167, 171–175].

5.2.1 Avalanche Photodiodes

A SiPM consists of an array of single-photon avalanche photodiodes (SPADs) connected in parallel. Each SPAD is a silicon $p - n$ semiconductor junction with a large enough reverse bias applied such that the junction is above its breakdown voltage. The high electric field in the depletion region makes it so that charge carriers excited by energy depositions are accelerated enough to sustain continuous avalanche multiplication. The model of avalanche multiplication is understood through a process akin to the Townsend effect seen in gasses in which an accelerated charge carrier excites one or more further charge carriers through impact ionization with the crystal lattice [58, 176]. Because the number of ionized carriers quickly diverges, the average number of avalanche charge carriers is independent of the number of initial charge carriers and thus deposited energy. When operated this way, the junction is said to be in Geiger mode: as long as an incident photon has an energy greater than the band gap of silicon, it has some probability of exciting a charge carrier and creating an avalanche, regardless of the initial energy of the photon. The initial energy, however, does influence how likely the photon is to both be absorbed in the depletion region and how likely an avalanche will occur.

The geometry of a SPAD is shown in Fig. 5.1, although the ordering of the junction depends on the desired wavelength sensitivity. The avalanche triggering probability is higher for electrons than it is for holes [177], so it is beneficial to place the n junction for electron collection as far away as possible from the average depth of the desired interaction. It is common for infrared (IR) sensitive devices to have an $n - p$ structure with the n region exposed to incident light: IR light is absorbed deeper in the device, so it is advantageous to have the n junction near the top. For visible-light and ultraviolet (UV)-sensitive SPADs, it is common to have a $p - n$ structure because the photons are more often absorbed in the top of the device. Recent advancements in entrance window coatings and electric field profiles have allowed for VUV sensitive

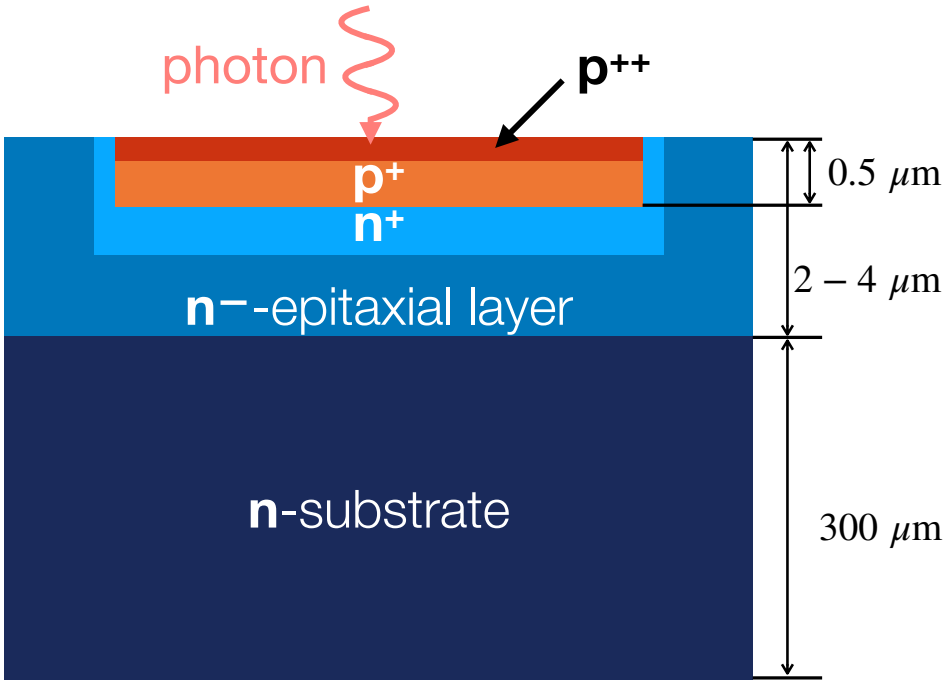

Figure 5.1: Geometry of a single photon avalanche diode, figure modified from Ref. [171]. This SPAD is an example of the p-on-n structure. UV light is absorbed closer to the surface of the SPAD so that electrons generated in the p-type part of the depleted region drift over a longer path to the n-type side of the junction. The longer drift path increases the avalanche triggering probability. The depleted region extends into the p^+ and n^+ doped regions, and is around $\mathcal{O}(1\ \mu\text{m})$ thick [179].

SPADs as well [178]. It is common that the first dopant region is on the order of $0.5\ \mu$m and the other dopant region is on the other of $1-2\ \mu$m [171] — both regions are heavily doped on the order of $> 10^{16}$ cm^{-3} [171, 179]. The depletion width is typically on the order of $1-2\ \mu$m [180, 181]. Above breakdown voltage, the electric field has a maximum strength on the order of 10^5 V/cm [171, 172].

Because the avalanche process proceeds exponentially [1] it is necessary to reset the SPAD so that it can continue to detect photons. This is achieved by adding a large quenching resistor in series with the diode and the bias supply. This way, when the current rapidly increases, a voltage drop across the quenching resistor causes the voltage across the SPAD to fall below the breakdown voltage. Once the SPAD is below its breakdown voltage, avalanche multiplication terminates and the remaining charge is collected and the device is reset.

The onset of the breakdown of the SPAD can be clearly seen in the reverse current-voltage character-

[1] roughly, the avalanche can also terminate due to space charge effects [58]

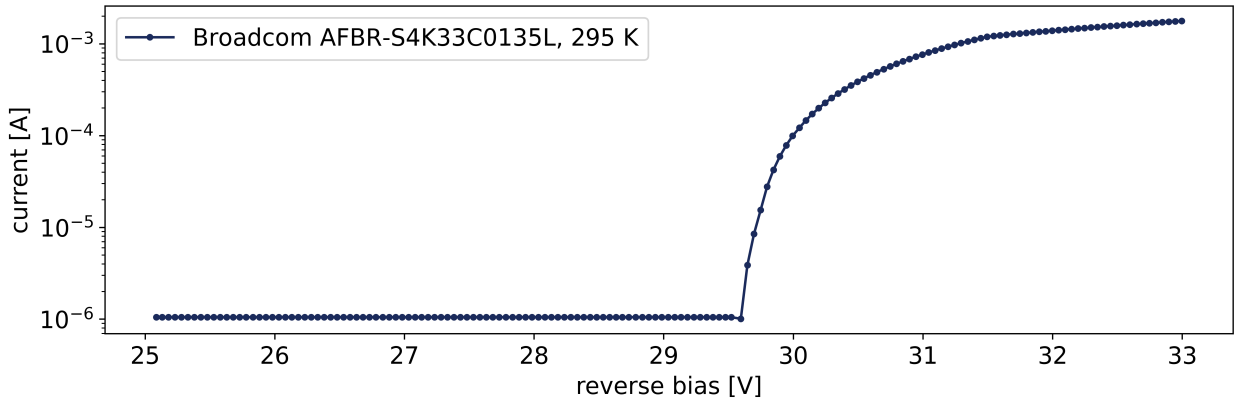

Figure 5.2: Current versus reverse voltage for a Broadcom SiPM at room temperature. The onset of avalanche breakdown is visible where the current rapidly increases by many orders of magnitude.

ization of a device, an example of which is shown in Fig. 5.2. The voltage at which the leakage current I suddenly and sharply increases is the location of the breakdown voltage V_{bd}. Immediately prior to the breakdown voltage, avalanche multiplication still occurs but does not rapidly run away and become self-sustaining. In this mode of operation, the response of the SPAD is linear. This means it is sensitive to the number of impinging photons. In linear mode the gain is on the order of 10^3, which is often too low for single-photon detection. Above the breakdown voltage, however, the gain increases by many orders of magnitude.

5.2.2 Silicon Photomultipliers

A SiPM is an array of SPADs connected in parallel, with each SPAD connected in series with a quenching resistor. SiPMs are operated a few volts above the breakdown voltage, and the amount of voltage exceeding the breakdown is referred to as the overvoltage, V_{OV}. This enables counting of multiple photons for low light intensities: the number of SPADs that undergo an avalanche is equal to the number of incident photons [2]. Because the response of each SPAD is roughly identical, the pulse height of a SiPM signal is proportional to the number of firing SPADs and thus the number of detected photons. Each individual SPAD is sometimes called a microcell [172]. The size of each SPAD is called the pitch — for most commercial visible light-sensitive SiPMs the pitch is on the order of 50 μm. The total area of most SiPMs is $\mathcal{O}(10\,\mathrm{cm}^2)$; however,

[2] Assuming some conditions to ensure linearity, such as that the rate of incident photons needs to be low enough that the probability of getting more than one photon per SPAD within the recovery time is small [171].

due to trenches and guard rings, the active area for photon absorption is smaller. The fraction of active area to total area is known as the fill factor, which is typically around 80%. The number of SPADs per device can be found by dividing the active area by the pitch and can range $\mathcal{O}(10^2 - 10^4)$.

5.2.3 Equivalent circuit and expected signal shape

To understand the expected signal shape from a SiPM it helpful to create a model equivalent electrical circuit of a SiPM, shown in Fig. 5.3 from Ref. [182]. The circuit is divided into three major loops: the primary SPAD that is discharging, the load effect of having unfired SPADs in parallel, and the capacitance from the metal grid that connects the SPADs. Each SPAD has a diode capacitance C_d due to the depleted region and a quenching resistor R_q in series. The quenching resistor is modeled to have a parasitic capacitance C_q; likewise, the metal frame used to electrically connect the SPADs contributes C_g in parallel [182]. Additionally, a shunt resistor R_s is added for for readout of the signal, although it is not shown in Fig. 5.3. The transfer function of this circuit predicts that in response to a delta-spike current avalanche, the resulting signal is an exponential with two time constants. The long time constant is associated with the current flowing through the quenching resistor of the firing SPAD in order to recharge the diode and is given by [183]

$$\tau_{long} = R_q \cdot (C_d + C_q), \tag{5.1}$$

while the short time constant is caused by capacitive coupling between the microcells [172] and is given by the product of the shunt resistance and the total capacitance $C_T = C_g + N(C_d \parallel C_q)$ (where $\parallel$ represents addition in parallel, so that $C_d \parallel C_q = (C_d^{-1} + C_q^{-1})^{-1}$) [183],

$$\tau_{short} = R_S \cdot C_T. \tag{5.2}$$

The values of the short and long time constants depend on the resistors used in the device, and on the coupling to readout electronics, but at room temperature are on the orders of 100 ns [183]. Typical quenching resistors used in SiPMs are on the order of a few hundred kilohms [172], and values for specific devices can be measured using the forward bias region of the current-vs-voltage (IV) curve [174]. The value of the quenching resistor can increase by a factor of 5 or more at cryogenic temperatures due to material

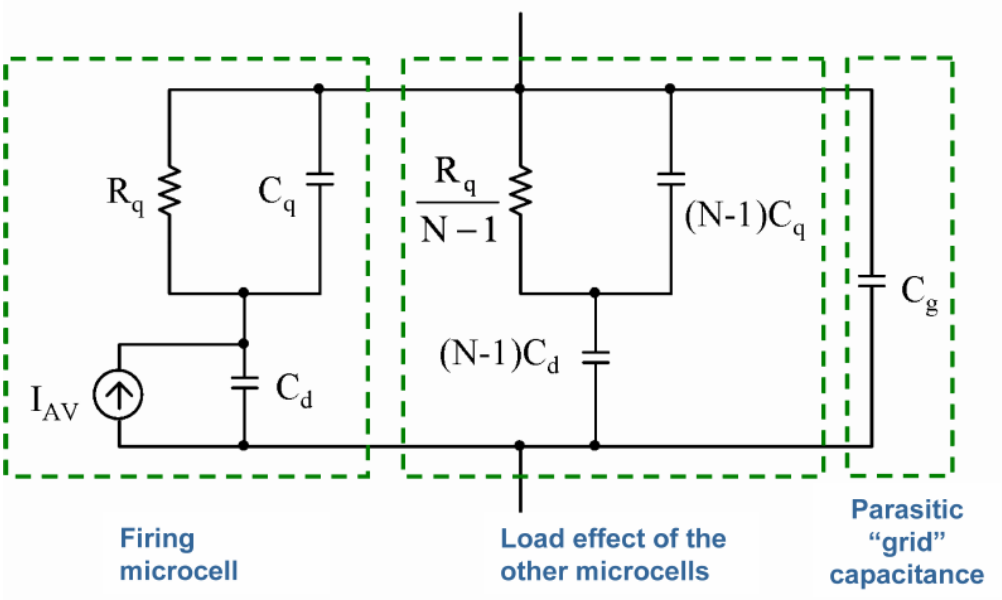

Figure 5.3: Equivalent circuit of a silicon photomultiplier, from Ref. [182]

properties of the chosen resistor material in a device [184]. This leads to substantially longer recovery times at cryogenic temperatures.

The recovery times are independent of the overvoltage at which the device is operated; however, the amplitude of the response to one detected photon does depend on the overvoltage and is discussed in Sec. 5.2.4. The signal response to one detected incident photon is called the one photoelectron (1 p.e.) waveform in analogy to vacuum photomultiplier tubes [174]. At a fixed overvoltage, the amplitude of a signal is proportional to the number of firing SPADs. The assignment of pulse heights to number of firing SPADs is done through either of two complementary methods. First, the heights of each pulse can be aggregated into a histogram called the pulse height spectrum (PHS) [3]. The PHS is a sum of Gaussians, where each Gaussian corresponds to the pulse amplitude from an integer number of SPAD firings convolved with noise (from statistical and electronics contributions). The presence of the pedestal peak, or the peak from zero firing SPADs, then allows integer labels to be assigned to the peaks: the first peak after the pedestal must come from one SPAD firing and so on. This method works well for low rates of SPAD firing, but can be computationally expensive in analysis because complex peak finding routines need to be run. It is possible to also do a variation of this technique called the hypercurrent method, as was done in the LEGEND-200

[3] Also colloquially known as a finger plot.

first physics analysis, using the heights of the derivatives of the waveforms. The second method relies on the full integration of pulses and the building of the (SiPM) charge spectrum (QDC). This method is computationally much simpler — because integrating a waveform is a faster algorithm than peak finding — and works also in high rate scenarios. The shape of the QDC is very similar to that of the PHS, but the peaks have a dominant Gaussian component (due to fluctuations in the pulse height) as well as an extended tail from the presence of correlated noise, discussed in more detail below in Sec. 5.2.8. Again, the pedestal is used to determine assignment of peaks.

5.2.4 Gain

The gain of a SiPM is the number of electrons produced per quenched avalanche, and thus the pulse height is dependent on the gain. The total charge produced during an avalanche is found by noting that the quenching resistor causes a voltage drop equal to V_{OV}. This change in voltage discharges the capacitors such that

$$Q = C_{tot}\Delta V = (C_d + C_q)V_{OV}. \tag{5.3}$$

The gain is the total charge divided by the electron charge e and is therefore

SiPM gain

$$G = \frac{(C_d + C_q)(V - V_{bd})}{e}. \tag{5.4}$$

This a linear function of the overvoltage, as shown in Fig. 5.4. Gains for SiPMs typically range from $10^5 - 10^7$ [169, 172]. The gain may be slightly temperature dependent due to the temperature dependence of the diode capacitance, although the magnitude of this term is much smaller than the magnitude of the voltage term [58].

5.2.5 Breakdown voltage

The breakdown voltage is the minimum voltage such that the internal electric field is strong enough to generate an avalanche. Typical breakdown voltages range from 30-60 V depending on the device and man-

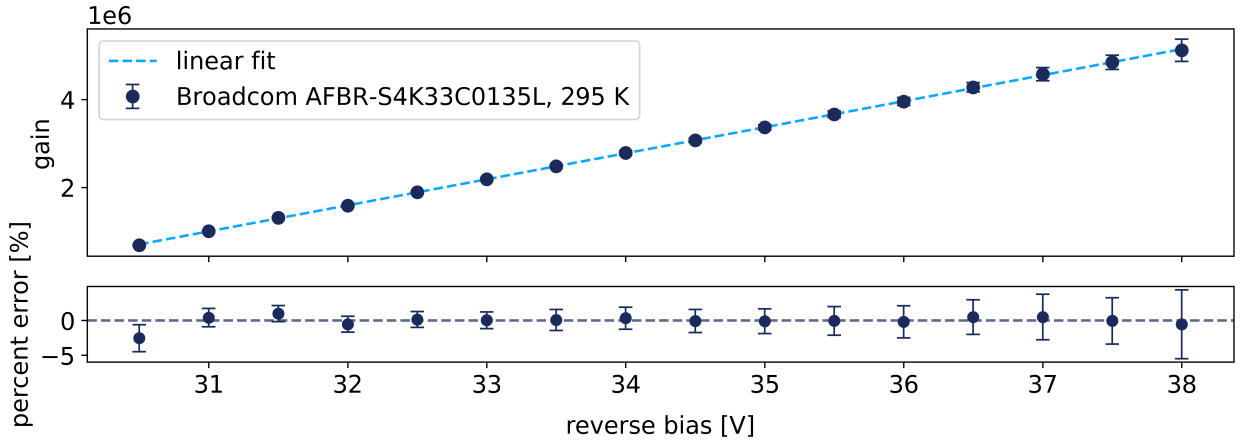

Figure 5.4: Plot of the gain of a Broadcom SiPM at room temperature as a function of applied voltage. The gain is a linear function of the applied voltage, and the y-intercept gives the breakdown voltage. The slope of the line yields the total capacitance of the diode and quenching parasitic capacitance.

ufacturer. The breakdown voltage increases as a function of temperature, as shown in Fig. 5.5. This is because the ionization rate is lower at higher temperatures due to increased energy loss into optical phonon modes via scattering [185]; a stronger electric field is needed to give the carriers sufficient energy to create a charge avalanche. At lower temperatures the breakdown voltage decreases due to increased ionization rates — a lower field is necessary to sustain an avalanche.

5.2.6 Photon detection efficiency

The photon detection efficiency (PDE) of a SiPM is the probability that an incident photon is converted into a usable signal pulse, and can be broken down into three components [172],

Photon detection efficiency, model 1

$$\mathrm{PDE}(V_{OV}, \lambda, T) = QE(\lambda, T) \times P_t(V_{OV}, \lambda, T) \times FF_{\mathrm{eff.}}(V_{OV}, \lambda)\,. \tag{5.5}$$

$QE(\lambda, T)$ is the quantum efficiency, $P_t(V_{OV}, \lambda, T)$ is the avalanche triggering probability, and $FF_{\mathrm{eff.}}(V_{OV}, \lambda)$ is effective fill factor. Each of these can depend on the wavelength (λ) of incoming light; the temperature (T) of the SiPM; and/or the overvoltage (V_{OV}), which is the difference between the operating reverse bias and the breakdown voltage. Both the quantum efficiency and the avalanche triggering probability are functions

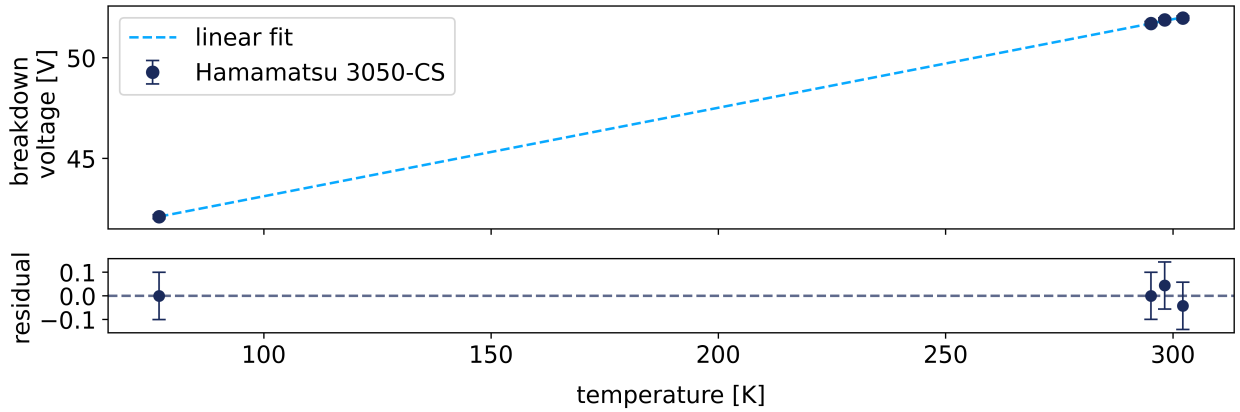

Figure 5.5: Temperature dependence of the breakdown voltage as measured for a Hamamatsu 3050CS SiPM. The breakdown voltage was measured at 3 points within a 30 K range of room temperature by cooling it down with a Peltier cooler. The breakdown voltage was also determined at 77 K by dunking the SiPM in liquid nitrogen.

of the wavelength of incident light due to the wavelength-dependent absorption depth: the absorption depth determines how likely a photon is absorbed, and also how likely an absorbed photon at that depth triggers an avalanche. The avalanche triggering probability is depth dependent due to the structure of the depletion region [179]. Because both these parameters depend on the geometry of the depletion region, these are both device-dependent quantities. In principle, both of these parameters could also be temperature dependent because the absorption depth has a known temperature dependence [58]. The fill factor is a predominately geometrical factor describing the area of a SiPM that is active and not filled with passive materials such as trenches and quenching resistors. The fill factor is voltage and wavelength dependent due to electric-field-based effects creating an inactive region that extends some depth into the device near the border of a microcell.

The PDE has a well-investigated behavior as a function of overvoltage: the PDE increases to a saturation value as the overvoltage increases, as shown in Fig. 5.6. The rate of increase in the PDE is device-dependent. Likewise, the functional relationship between the photon wavelength and PDE has also been studied: the PDE has a peak spectral response, and then decays at wavelengths higher and lower than this, as shown in Fig. 5.7. The peak responsivity wavelength is determined by the doping profile and geometry of the device: these parameters set the location of the depletion region which in turn sets the wavelengths of light absorbed due to the wavelength-dependence of the absorption lengths of photons. The temperature

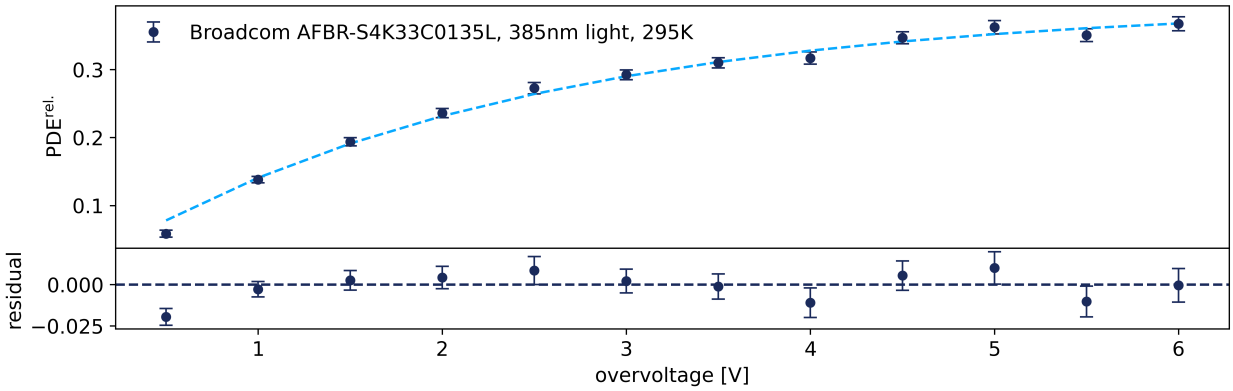

Figure 5.6: Photon detection efficiency as a function of overvoltage, for a Broadcom SiPM illuminated with 385 nm light at room temperature. The saturation of the PDE corresponds to the Geiger triggering probability approaching unity.

dependence of the PDE is less studied, but it seems the PDE decreases at cryogenic temperatures below 100K [186, 187], although there is some contention at the behavior at intermediate temperature ranges.

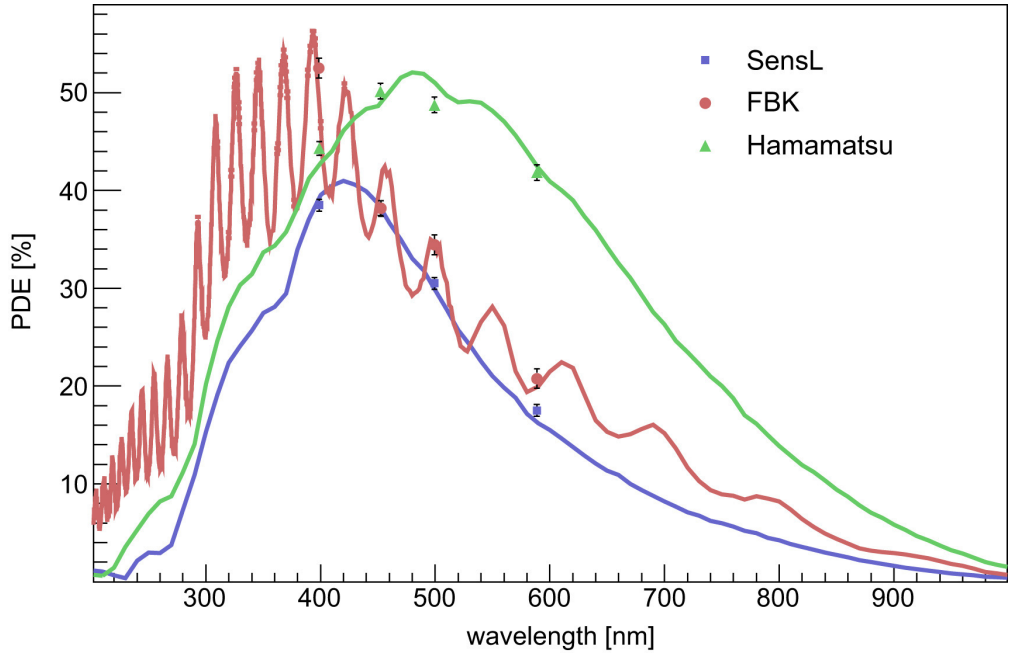

Figure 5.7: Photon detection efficiency as a function of wavelength, for a SensL 41 μm pitch J-Series, a Hamamatsu 3050-CS, and a FBK NUV-HD SiPM from Ref. [188]. The oscillations in the FBK SiPM PDE are due to thin film interference between the passivated silicon-dioxide entrance window and the silicon substrate. This effect is well studied in Ref. [189].

In the parametrization of Eqn 5.5, the quantum efficiency represents the probability for the incoming

photon to create an electron-hole pair, irrespective of where in the junction this pair is created — this way, the QE is only dependent of the incoming energy of the photon. The avalanche, or Geiger, triggering probability is then the probability that this carrier pair starts a sustained avalanche. The triggering probability is thus dependent on the position of the initial carrier creation and therefore also the energy of the incoming photon; it also depends on the strength of the electric field and thus the voltage applied. The effective fill factor also accounts for absorption in the active volume of the detector, and thus depends on the dead layer width — determined by the field — and the site of the energy deposition — determined by the initial photon energy.

While this model is a conceptually simple partitioning of the PDE, as others have pointed out in Ref. [179], it hides the importance of the position-dependence of the absorption of the initial photon. The authors of the excellent work presented in Refs. [179] and [189] advocate for the following model:

$$PDE(V_{OV}, \lambda, T) = FF \cdot \epsilon_{\text{transmission}}(\lambda, T) \cdot \int_{d_P^*}^{d_W^*} dP_{\text{absorption}}(x, T, \lambda) P_t(x, V_{OV}, T). \tag{5.6}$$

Now the fill factor is only the photo-sensitive area and is no longer a function of V_{OV} and λ. Importantly, the avalanche triggering probability is no longer wavelength dependent as it was in Eqn. 5.5. There is no analytic form of $P_t(x, V_{OV}, T)$, but it can be simulated as was done in Ref. [179]. The term $\epsilon_{\text{transmission}}$ is the probability that a photon is transmitted through the passivated silicon-dioxide layer on top of the SiPM. Here for a p on n SiPM with depletion region starting at d_P and ending at d_W the parameter d_P^* (d_W^*) represents the effective width that electrons (holes) are also able to drift into the depletion region and start an avalanche. Following Ref. [179], a step function is used to simplify the avalanche triggering probability: this means $P_t(x, V_{OV}, T) = P_e(d_P)$ if the photon is absorbed anywhere in the p-side up to the depletion region, and $P_t(x, V_{OV}, T) = P_h(d_W)$ if it is absorbed in the depletion region and the n-side. This leads to a model

Photon detection efficiency, model 2

$$\mathrm{PDE}(V_{OV},\lambda,T) = PDE_{max}(\lambda,T)\Big(P_e(d_P;V_{OV},T)\cdot f_e^*(\lambda) + P_h(d_W;V_{OV},T)\cdot(1-f_e^*(\lambda))\Big), \tag{5.7}$$

where $P_e(d_P)$ is the probability an electron triggers an avalanche at $x = d_P$ and $P_h(d_W)$ is the probability of a hole-driven avalanche at $x = d_W$, PDE_{max} is a mostly voltage-independent quantity that is a function of the photon attenuation length and effective depletion region, and f_e^* is the fraction of electron driven avalanches and is a mostly voltage-independent quantity that is a function of the photon wavelength and effective depletion region. The electron avalanche triggering probability is treated empirically as in Ref. [188] to reproduce the asymptotic form of PDE measurements:

$$P_e(d_P) = 1 - \exp(-k_e \cdot V_{OV}), \tag{5.8}$$

where k_e can be determined from fits to data. The hole-triggering probability is then related to the electron triggering probability with a free parameter k that represents the ratio of hole and electron impact-ionization coefficients:

$$P_h(d_W) = 1 - (1 - P_e(d_P))^k. \tag{5.9}$$

For wavelengths of light — such as UV — that have a small attenuation length and are thus likely to be absorbed in the p side of the junction, it can be assumed that $f_e^* \approx 1$ and thus Eqn. 5.7 can be expressed as

Photon detection efficiency for UV light

$$\mathrm{PDE}(V_{OV},\lambda_{UV},T) = PDE_{max}(V_{OV},\lambda_{UV},T)\Big(1 - \exp(-k_e(T)\cdot V_{OV})\Big). \tag{5.10}$$

This equation is helpful for fitting data and determining properties of a SiPM being characterized.

5.2.7 Dark counts: primary noise

In addition to avalanches triggered by incident photons, there is also some probability that the thermal generation of charge carriers in the depletion region will trigger an avalanche. Avalanches coming from thermal carriers are called **dark counts** because they will occur even when the SiPM is not illuminated. The dark counts are an irreducible source of noise for SiPM measurements; they perfectly mimic signals from real incident photons. It is thus desirable to have as low a dark count rate (DCR) — not to be confused with the delayed charge recovery (DCR) parameter for pulse shape discrimination introduced earlier — as possible. Common DCRs at 5 V overvoltages at room temperature are $\mathcal{O}(1000\,\mathrm{kHz})$; it is common to express the DCR per unit of active area leading to DCRs of $\mathcal{O}(100\,\mathrm{kHz/mm^2})$. A high DCR for LEGEND would mean that the signal acceptance of the LAr veto would go down due to random accidental coincidences — this undesirably reduces the sensitivity to neutrinoless double-beta decay.

Before looking at the functional dependence of the DCR on voltage and temperature, it is helpful to first review mechanisms responsible for carrier generation in reverse biased $p-n$ junctions.

Carrier generation mechanisms in SiPMs

1. **Band-to-band generation:** this is the standard promotion of a valence band electron to a conduction band electron through the indirect band gap. Thermal fluctuations at room temperature (on the scale of meV) are much too low to excite an electron this way. This process is mostly driven by the photoelectric effect, or by impact ionization/Auger process [58].

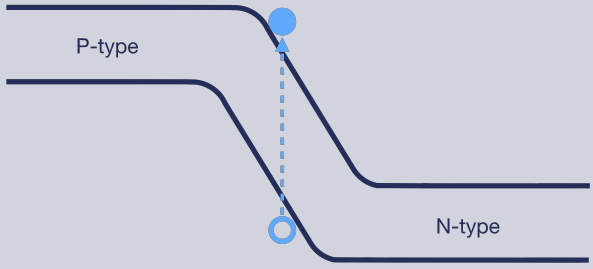

Figure 5.8: Diagram showing charge carrier generation through direct excitation across the band gap. This sketch shows a simplified one-dimensional $p-n$ semiconductor junction where the x-axis represents spatial extent, and the y-axis represents the energy of the bands.

2. **Trap assisted excitation:** charge carriers can be thermally excited into energy levels lying

mid-band gap due to available states from charge traps. From these mid-gap states, the carriers can then be excited fully and become a free charge carrier. This process is also known as Shockley-Read-Hall (SRH) generation, and is the dominant transition for semiconductors with indirect band gaps [58]. For the case of one trap level of energy E_t, the rate of carrier generation per volume is given by

$$R_{SRH} = \frac{pn - n_i^2}{\tau_p \left[n + n_i \exp\left(\frac{E_t - E_i}{k_B T}\right)\right] + \tau_n \left[p + n_i \exp\left(\frac{E_i - E_t}{k_B T}\right)\right]}, \tag{5.11}$$

where E_i is the Fermi energy of the intrinsic semiconductor, n_i is the intrinsic carrier concentration, and τ_p, τ_n are carrier lifetimes. The lifetimes are inversely proportional to the capture cross section, the number of traps, and the thermal velocity of the carriers. It's crucial to note that this process is heavily temperature dependent.

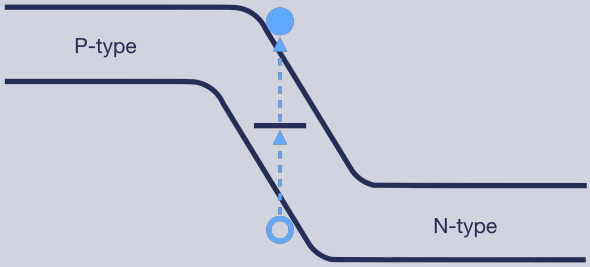

Figure 5.9: Diagram showing charge carrier generation through the Shockley-Read-Hall mechanism.

3. **Band-to-band tunneling:** due to the quantum mechanical nature of the electron wavefunctions in the crystal, there is some probability that charge carriers can tunnel across the potential barrier imposed by the band gap. The tunneling probability can be determined using the WKB approximation assuming a uniform electric field $\vec{E}$ across a triangular barrier [58]:

$$R_{BTBT} \approx \exp\left(-\frac{4\sqrt{2m^*}E_g^{3/2}}{3q\hbar|\vec{E}|}\right). \tag{5.12}$$

A strong electric field makes the potential barriers more narrow and thus increases the rate. The rate is only loosely a function of the temperature due to the band gap's temperature dependence.

This is sometimes called Zener tunneling.

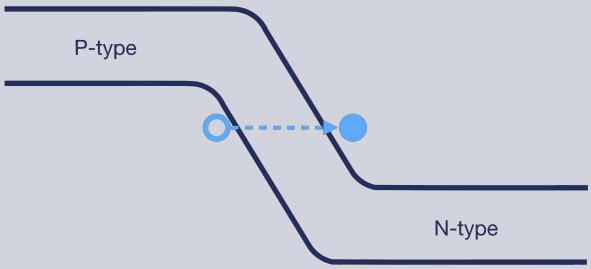

Figure 5.10: Diagram showing charge carrier generation through band to band tunneling.

4. **Trap assisted tunneling:** the presence of available energy states due to traps within the band gap allows for a charge carriers to tunnel to these states and then either tunnel again fully across the gap or to thermally excite across the gap. These mid-gap states both lower the potential barrier and decrease the barrier width that the carriers need to tunnel across making tunneling more probable. It is common to model this process as a SRH generation process modified by terms of the form $1+\Gamma$ that encode the increased population of trap states due to the tunneling:

$$R_{TAT} = \frac{pn - n_i^2}{\frac{\tau_p}{1+\Gamma_p}\left[n + n_i \exp\left(\frac{E_t - E_i}{k_B T}\right)\right] + \frac{\tau_n}{1+\Gamma_n}\left[p + n_i \exp\left(\frac{E_i - E_t}{k_B T}\right)\right]}. \tag{5.13}$$

This is often referred to as the Hurkx model [190]. The electric field dependent term is given by

$$\Gamma_{n,p} = \frac{\Delta E_{n,p}}{k_B T} \int_0^1 du \, \exp\left[\frac{\Delta E_{n,p}}{k_B T} u - \frac{4}{3} \frac{\sqrt{2m^*}(\Delta E_{n,p})^{3/2}}{q\hbar|\vec{E}|} u^{3/2}\right], \tag{5.14}$$

and is an integral across the depletion region, where $\Delta_{n,p}$ depends on the relative position of the trap level and the conduction-band minimum (or valence band maximum). Thus this process is both strongly temperature and field dependent.

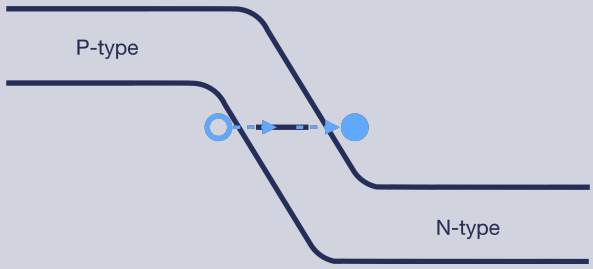

Figure 5.11: Diagram showing charge carrier generation through trap assisted tunneling.

The DCR is a function of both the bias voltage and the temperature of the SiPM (it is also a function of the electric field and doping profile, but these are not controlled by the user of a SiPM). Functionally, it can be decomposed into the rate of carrier generation multiplied by the probability that the generated carriers will trigger a charge avalanche. At lower bias voltages, the electric field is weaker and thus generated carriers have a lower probability to be accelerated enough to create an avalanche (this is encoded in the lower impact ionization coefficient at lower fields). A semi-empirical relationship for the voltage dependence of room temperature SiPMs was used in Ref. [188] which uses a simple exponential model for the carrier generation rate (where the exponential dependence of the types of rates is assumed to dominate the relationship) [58]:

$$DCR(V_{OV}) = \exp\left(a + b \cdot V_{OV}\right)\left[1 - \exp\left(-\alpha \cdot V_{OV}\right)\right], \tag{5.15}$$

where the term in square brackets represents the Geiger triggering probability and approaches 1 as the bias increases. The behavior of the DCR as a function of voltage is shown in Fig. 5.12. In general the DCR is a function of the total applied voltage, but the presence of the extra fit factor a allows for parameterization in terms of the overvoltage.

The temperature dependence of the DCR is more complicated. A good model above 200 K suggests that a combination of SRH generation of carriers and trap-assisted tunneling is responsible for the DCR [187]. This rate of generation is well-described by the semi-empirical modified Arrhenius relation:

$$DCR_{SRH}(T) = AT^{3/2} \exp\left(-\frac{E_g}{k_B T}\right). \tag{5.16}$$

Below 200 K the change in DCR as a function of temperature is much less pronounced — this suggests

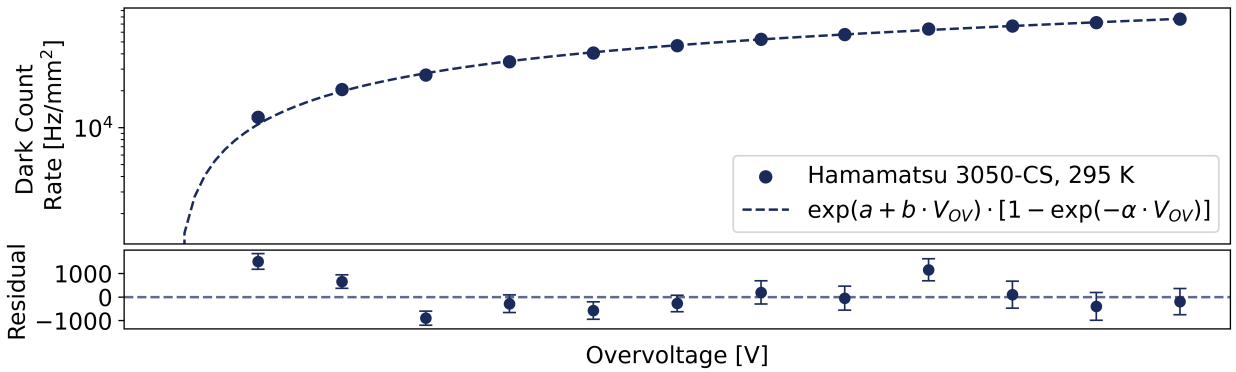

Figure 5.12: Voltage dependence of the dark count rate for a Hamamatsu 3050-CS SiPM at room temperature. The effective Geiger triggering probability shape parameter for dark counts is $\alpha = 0.5 \pm 0.08$ V^{-1}

that the mostly temperature-independent band-to-band tunneling dominates [171]. The tunneling rate can be lowered by reducing the electric field through inventive doping profiles and the cryogenic DCR can be reduced by ingenious device manufacturing [171, 178]. Below roughly 150 K the DCR drops very suddenly and could potentially be explained by a decrease in the avalanche triggering probability. This decrease is caused by what Refs. [186, 187] call carrier freezeout; however, it would be more preferential to refer to this effect as the activation of new charge traps. At low temperatures (< 100 K), ionized donor and acceptor impurities can act as traps for excited charge carriers [191–193]. Carrier freezeout (sometimes called hard carrier freezeout) is a more specific term — for when donor and acceptor impurities are unable to ionize and thus impede establishment of a depletion region — and should be reserved for cases where the breakdown voltage increases as temperature decreases, which occurs in silicon $p - n$ junctions at 30 K [194], or when avalanche multiplication ceases [193, 195].

5.2.8 Correlated noise

SiPMs have two further sources of noise that are correlated with primary avalanches: cross talk and afterpulsing. Correlated noise is present for both real photon-induced signals and dark counts.

Optical crosstalk

Electrons accelerated during a Geiger avalanche can emit photons in the near infrared, generating on the order of 10^{-5} photons per electron at the peak emissivity of 1.14 eV photons [188, 196]. The mechanism for

the generation of these photons is poorly understood, and models range from hot-intraband luminescence [171] to bremsstrahlung [197]. These optical photons are emitted isotopically and can induce signals in other SPADs in the following ways:

Cross talk mechanisms

1. **Direct/Prompt cross talk**: optical photons can be directly absorbed in the depleted region or very near the depleted region in a neighboring SPAD. For typical inter-SPAD distances, this leads to a delay between pulses on the order of hundreds of femtoseconds, well below the sampling resolution of most digitizers. Thus, direct cross talk looks like two SPADs firing at the exact same time.

2. **Delayed cross talk**: optical photons can be absorbed in the undepleted region beneath neighboring cells. The produced charge carriers can then drift into the depleted region and trigger an avalanche. The time scale for diffusion is on the order of a few to hundreds of nanoseconds [172].

3. **External cross talk:** for SiPMs that have an entrance window or are directly coupled to an external optical medium such as a scintillator, the optical photons can be reflected back into a SPAD. This type of cross talk is difficult to characterize.

The functional dependence of cross talk can be modeled by the following [188]:

$$P_{CT} = f(T)V_{OV}C_{eff} \cdot PDE_{CT}(V_{OV}, \lambda_{NIR}, T), \tag{5.17}$$

where f is the average number of optical photons produced per electron in the avalanche, $V_{OV}C_{eff}$ is the gain in number of electrons, and $PDE_{CT}(V_{OV}, \lambda_{NIR}, T)$ is the average photon detection efficiency of these near-infrared (NIR) optical photons. The cross talk PDE depends strongly on the geometry of the device, as trenches built in-between SPADs can decrease this parameter strongly. Following Ref. [188] this can be expressed in terms of an average geometric factor γ that encodes the absorption probability, and the

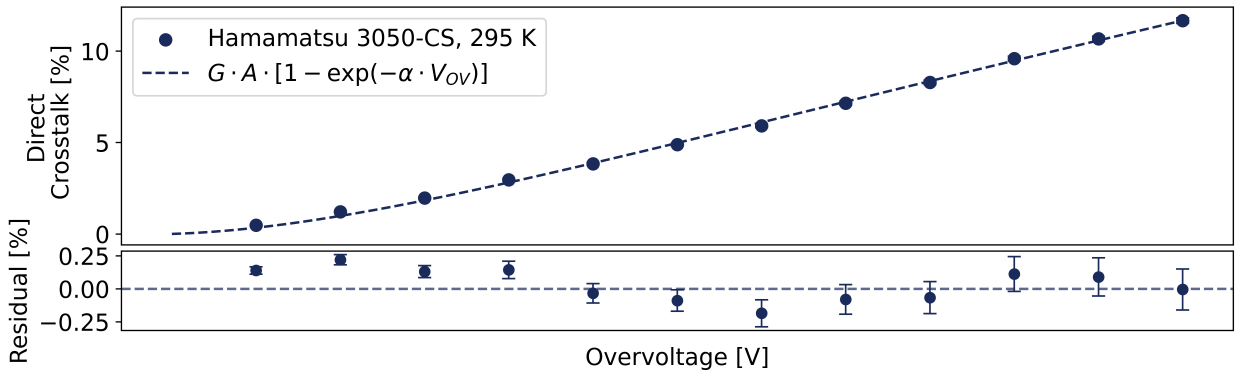

Figure 5.13: Voltage dependence of the direct cross talk probability for a Hamamatsu 3050-CS SiPM at room temperature. The effective Geiger triggering probability shape parameter for direct cross talk optical photons is $\alpha = 0.5 \pm 0.03$ V^{-1}. The effective number of absorbed optical photons produced per avalanche is $(4 \pm 0.1) \times 10^{-8}$

avalanche triggering probability:

$$P_{CT} = f(T) V_{OV} C_{eff} \cdot \gamma \cdot (1 - \exp(-\alpha V_{OV})). \tag{5.18}$$

The voltage dependence of the cross talk on voltage is shown in Fig. 5.13. The temperature dependence of the cross talk is not well understood, Ref. [187, 198] measured no temperature dependence while Ref. [188] saw a weak decrease at lower temperatures. This is somewhat surprising as the PDE could be a weak function of temperature, although it is possible that the PDE for cross talk photon wavelengths does not change very much.

Afterpulsing

Afterpulsing occurs when a delayed release of charge triggers a secondary avalanche in a SPAD that was recharging from a primary avalanche. Afterpulsing dominantly occurs due to primary avalanche charge carriers getting stuck in charge traps in the depletion region — these are typically impurities in the lattice. Once the charge traps release the charges, they can go on to trigger another avalanche within the same cell. Afterpulses are characterized by pulses that are reduced in height compared to the primary avalanche due to the lower electric field in the recharging cell. Afterpulsing can also occur when optical cross talk photons are absorbed into the undepleted region below the recharging cell and diffuse into the depleted region [172].

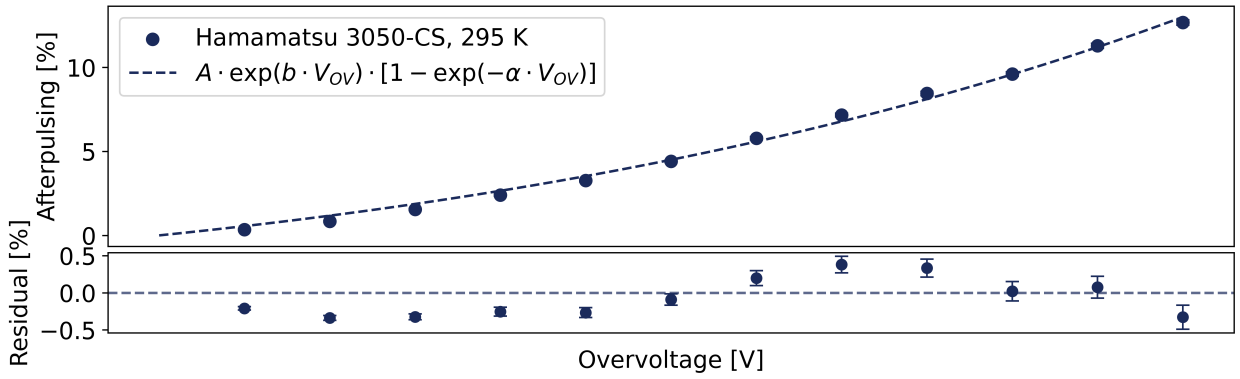

Figure 5.14: Voltage dependence of the afterpulsing probability for a Hamamatsu 3050-CS SiPM at room temperature. The effective Geiger triggering probability shape parameter for afterpulsing is $\alpha = (5 \pm 1) \times 10^{-5}$ V^{-1}.

Although, some prefer to classify this as a type of delayed optical cross talk.

Like cross talk, afterpulsing increases as a function of overvoltage due to the creation of more primary charge carriers. This increases the likelihood of encountering a trap. An empirical model is complicated to produce, as the trapped charges need to be released on a time scale shorter than the cell recovery time. One could attempt to parametrize the afterpulsing probability as a product of the number of charge carriers per avalanche times the probability $\epsilon_{\text{trap/release}}$ they are trapped and re-released on the appropriate time scale times the avalanche triggering probability:

$$P_{AP}(V) = V_{OV} C_{eff} \cdot \epsilon_{\text{trap/release}} \cdot (1 - \exp(-\alpha V_{OV})). \tag{5.19}$$

The release of charge traps is voltage dependent, as the field strength can increase the tunneling probability out of traps. Following Ref. [188] this term is mostly dominated by an exponential dependence and thus can be recast as

$$P_{AP}(V) = A \cdot \exp(-\beta V_{OV}) \cdot (1 - \exp(-\alpha V_{OV})). \tag{5.20}$$

The voltage dependence of the afterpulsing is shown in Fig. 5.14.

The temperature dependence of afterpulsing relates to the lifetimes of charge traps: at higher temperatures, trap lifetimes are shorter and thus trapped charges are released fast enough to not constitute an

afterpulse. At lower temperatures trap lifetimes increase and the afterpulsing probability does too. At temperatures below $\sim$ 100 K it is possible that the ionized impurities in the depletion region act as additional charge traps increasing the afterpulsing rate in this temperature regime. Because the afterpulsing rate is proportional to the avalanche probability, reducing the electric field strength via doping profile selection reduces the afterpulsing rate. This comes at a tradeoff for a reduced PDE at lower overvoltages.

5.3 Measurement principles

5.3.1 Gain

The gain in principle can be determined by measuring the mean charge produced during one-photon-induced avalanches. However, a more robust estimate of the gain can be produced by measuring the distances between adjacent peaks in the charge spectrum and averaging them together. One way to do this is to fit a line to the location of the mean of each photoelectron peak in the charge spectrum.

Gain measurement procedure

1. Waveforms are collected, triggering on the rising edge in dark conditions or synchronous with trigger from a light pulse.
2. Digitized waveforms are converted from ADC units to voltage, and then from voltage to current using the gain of any amplifiers present.
3. Current waveforms are integrated to get the charge spectrum. The waveforms must be integrated over a time range several times as long as the cell recovery time to make sure that all charge is integrated.
4. Distances between peaks in the charge spectrum are used to determine the gain. Determination of the gain allows for the charge spectrum to be normalized into number of photons.

5.3.2 Breakdown voltage

Two common methods to determine the breakdown voltage are as follows:

Breakdown voltage measurement procedures

- **Reverse I-V measurement:** the breakdown voltage can be determined by identifying the voltage at which the reverse current suddenly rapidly increases due to the onset of avalanche multiplication. Due to noise fluctuations during measurements of the reverse I-V curve, it is recommended that data be obtained in steps of 0.1 V with sufficient dwell time at each point. There are numerous methods to extract the location of the breakdown voltage, as illustrated in detail in Ref. [174]. Of these numerous methods, the maximum of the derivative of the logarithm of the data yielded the most consistent results. The logarithm of the data smooths any noise present pre-breakdown.

- **Gain measurement:** the breakdown voltage can be determined by identifying the voltage at which the gain is equal to zero, following Eqn. 5.4. The gain is typically measured over a range of overvoltages and then a linear fit is performed to extrapolate the gain to zero. This method was found to give more robust and repeatable measurements of the breakdown voltage, and was the method of choice for the following work.

5.3.3 Photon detection efficiency

The photon detection efficiency can be measured by illuminating the SiPM and computing the mean number of detected photons divided by the total number of photons incident on the SiPM surface:

$$\mathrm{PDE} = \frac{N_{SiPM}}{\Phi_{SiPM} A_{SiPM}}. \tag{5.21}$$

One way to determine the incident flux of photons at the SiPM surface (Φ_{SiPM}) is to use an integrating sphere to uniformly distribute light onto the surface of the SiPM and the surface of a photodiode with known responsivity [188, 199, 200]; this photodiode is called the reference diode as it is used as a reference to determine the flux of photons. Integrating spheres uniformly distribute light due to high-reflectivity diffuse coatings. They preserve the input power of the incident light, but distribute it spatially. If the reference diode is also a SiPM with a known $PDE_{\mathrm{ref.}}$, and the mean number of photons incident on the reference diode $N_{\mathrm{ref.}}$ is also computed (to be discussed in detail later on, see Eqn. 5.26), then the PDE can be computed by

PDE measurement

$$\mathrm{PDE} = \frac{N_{SiPM}}{\frac{N_{\mathrm{ref.}}}{\mathrm{PDE}_{\mathrm{ref.}}\mathrm{A}_{\mathrm{ref.}}} A_{SiPM}}. \tag{5.22}$$

The mean number of photons detected at the SiPM can be computed in three ways. The first method, often called the "current method", relies on using a picoammeter to measure the sustained current I in a SiPM under continuous low-light illumination [4][199, 202, 203]:

$$N_{SiPM} = \frac{I}{G \cdot e \cdot ECF}, \tag{5.23}$$

where ECF is the excess charge factor that describes the fractional amount of charge produced due to correlated noise [174]. This method enables use of a continuous light source. In the second method, the mean number of detected photons is determined by taking the mean of the gain-corrected charge spectrum giving $\bar{N}_{det.}$ [174, 202]. This number is then corrected by the ECF, which can be approximated by $(1 + P_{AP})(1 + P_{CT})$ [113, 201]:

$$N_{SiPM} = \frac{\bar{N}_{det.}}{(1 + P_{AP})(1 + P_{CT})}. \tag{5.24}$$

This measurement can be taken with either a continuous light source, or with a pulsed light source. The two methods described both require excellent determination of the ECF, which is difficult to do reliably. The third method is a counting method that is unbiased by the presence of correlated noise such as crosstalk. This method is often called the "pedestal method" of measuring the PDE, and is described in Ref. [188]. To summarize, waveforms must be acquired synchronously with a pulsed light source so that baseline events, where the SiPM did not detect photons, are recorded. If the number of detected photons is Poisson distributed, then the probability of not detecting any photoelectrons, $P(N_{\mathrm{detected}} = 0)$, gives an unbiased estimate of the Poissonian mean ($\overline{N_\gamma}$) because the baselines are unbiased by any correlated noise [188]:

$$P(N_{\mathrm{detected}} = 0) = \exp\left(-\overline{N_\gamma}\right). \tag{5.25}$$

The probability of not detecting any photons is equal to the number of baseline events (N_0) divided by the

[4]Linear response of a SiPM is possible under continuous low light illumination [201].

total number of LED pulses (N_{total}). In this way, the mean number of detected photons becomes

Mean number of photons

$$\overline{N_\gamma} = -\ln\left(\frac{N_0}{N_{\text{total}}}\right). \tag{5.26}$$

5.3.4 Dark count rate

Dark counts are described by Poisson statistics because the creation of any given dark count is independent of previous dark counts. Namely, the probability of a given number of dark counts to occur in a given interval is given by a Poisson distribution. The dark count rate can be determined by looking at the distribution of the time between events for data taken in dark conditions, under no illumination. For a Poisson process the time to the next event is given by an exponential distribution [57]. Analysis routines for computing the times to next event are given below in Sec. 5.4.1. A histogram of the time to next event is shown in Fig. 5.15 where it is clear to see a population of events due to delayed cross talk and aferpulsing at short time scales. It is therefore derisible to estimate the DCR by fitting only large inter-event time events from the histogram with the function [172, 174, 188]:

Dark count rate measurement

$$\frac{dN}{dt} = A\exp(-DCR\cdot dt). \tag{5.27}$$

5.3.5 Correlated noise

The cross talk fraction is easy and robust to calculate. For data taken under dark conditions, cross talk always engenders a pulse that is at least 2 p.e. high, the cross talk fraction is simply the number of pulses with height greater than 1.5 p.e. divided by the number of pulses with height greater than 0.5 p.e.:

$$P_{CT} = \frac{f_{N>1.5}}{f_{N>0.5}}. \tag{5.28}$$

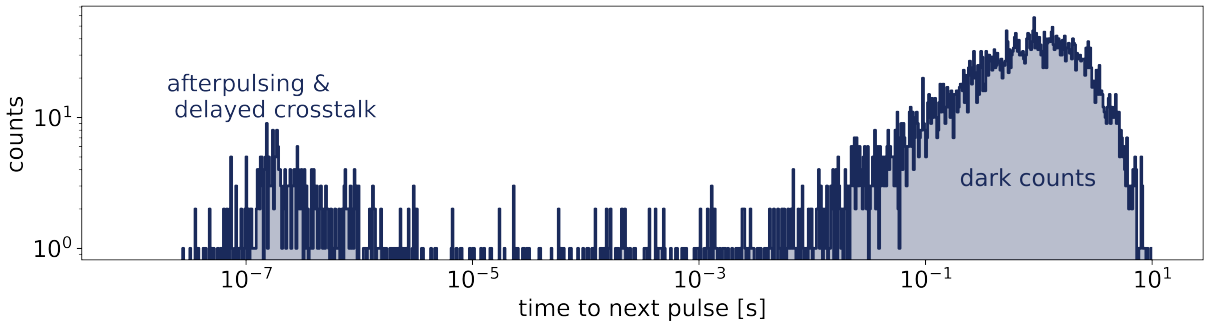

Figure 5.15: Histogram of the time to next event after triggering on a one photoelectron primary pulse with data taken in dark conditions with no illumination. The amplitude of the next event does not inform the creation of this histogram; as such, there are numerous direct cross talk events within the long inter-event time population.

This cross talk fraction sums the contribution from direct cross talk and delayed cross talk coincident with dark counts. There is also the potential for contamination from two dark counts within the same digitization sample, but the probability of this is negligible. Likewise the probability that a delayed cross talk event simultaneously with a dark count within a 1 ns digitization window is also negligible. To good approximation then, the direct optical cross talk probability is

Direct cross talk

$$P_{DiCT} = \frac{f_{N>1.5}}{f_{N>0.5}}. \tag{5.29}$$

The afterpulsing fraction and delayed optical cross talk fraction both rely on the inter-event arrival time distribution for their measurement. Afterpulses can be separated from delayed cross talk due to the fact that latter always has a pulse height equal to 1 p.e. while the former is strictly less than 1 p.e. The most correct computation of the afterpulsing rate then leverages both amplitude and arrival time information for its calculation. To start, all events with amplitude less than 1 p.e. are selected and the afterpulsing probability at time dt is modeled as the product of a constant that depends on the Geiger triggering probability, the probability that the trapped charge is released relative to the mean trap lifetime τ_{trap}, and the relative amplitude due to the cell recovering on a scale $\tau_{rec.}$ [204, 205]:

$$\frac{dN_{AP}}{dt} = A \exp\left(-\frac{dt}{\tau_{trap}}\right)\left[1 - \exp\left(-\frac{dt}{\tau_{rec.}}\right)\right]. \tag{5.30}$$

The cell recovery time is determined by fitting the selected events in dt, A space with

$$A = A_{1p.e.}\left[1 - \exp\left(-\frac{dt}{\tau_{rec.}}\right)\right]. \tag{5.31}$$

The event selection consists of both afterpulses and dark counts, so the arrival time information can be histogrammed and fit using the following:

$$\frac{dN}{dt} = A \exp\left(-DCR \cdot dt\right) + B \exp\left(-\frac{dt}{\tau_{trap}}\right)\left[1 - \exp\left(-\frac{dt}{\tau_{rec.}}\right)\right]. \tag{5.32}$$

Here DCR and $\tau_{rec.}$ are fixed parameters determined using methods described above. The afterpulsing probability is then computed by integrating Eqn. 5.30 to five times the trap lifetime [204]:

Afterpulsing probability

$$P_{AP} = \frac{\int_0^{5\tau_{trap}} \frac{dN_{AP}}{dt} dt}{N_{tot.}} \tag{5.33}$$

The delayed cross talk probability can be estimated similarly by selecting all events with an amplitude of 1 p.e. or above and then fitting the resulting inter-event histogram with Eqn. 5.30 without the recovery time term [205].

This complex method can be approximated by-eye by placing a cut on both arrival times and pulse heights and labeling everything within this bounding box as an afterpulse. The afterpulse probability is then the ratio of this number to the total number of counts. It is required to note the cut values used for computation for this method [206].

A final method for measuring the afterpulsing and delayed cross talk utilizes only the inter-event arrival time histogram and no pulse amplitude information. As such, it is also an approximation to these rates. After subtracting off the per-bin contribution of the dark count rate determined from a fit to large inter-event times, the population of residuals at short inter-event times comes from both afterpulses and delayed

cross talk [172, 188]. The residual populations are split by-eye with a time cut; typically events with less than 10 ns inter-event arrival times are classified as delayed cross talk, and the remaining population from $10 - 1000$ ns is treated as afterpulsing. The total number of counts in each division is determined and from these the afterpulsing and delayed cross talk probabilities are computed.

5.4 Analysis techniques

5.4.1 `sipm_studio`: waveform-level SiPM characterization software

In order to quickly perform device characterizations, we developed the Python-based software package `sipm_studio`. Characterization of the noise parameters DCR, afterpulsing, and cross talk require that waveforms are acquired under dark conditions, and that the digitizer triggers on the rising edge. Care must be taken so that the signal-to-noise ratio (SNR) is large enough that the threshold can be placed low enough to trigger on all 1 p.e. events. The mean number of detected photons is typically measured when the device is illuminated; it is common to pulse a light source and trigger the DAQ system synchronous with the pulse for this type of characterization. The gain in principle can be determined under dark conditions; however, its analysis is similar to the determination of the mean number of detected photons. Presently, there is no main-line analysis routine for forced trigger SiPM waveforms; although, the processors and routines described below can be readily applied to this scenario.

Light measurements using `sipm_studio`

The crux of the analysis of illuminated SiPM data is the creation of the charge spectrum and locating the precise positions of the photoelectron peaks. To that end, `sipm_studio` runs the following algorithm

Light analysis workflow with `sipm_studio`

1. If the user doesn't specify, the polarity of the pulses is determined. This is done by taking the average of all waveforms and checking if the absolute value of the minimum value is greater than or less than the maximum value, see Fig. 5.16.

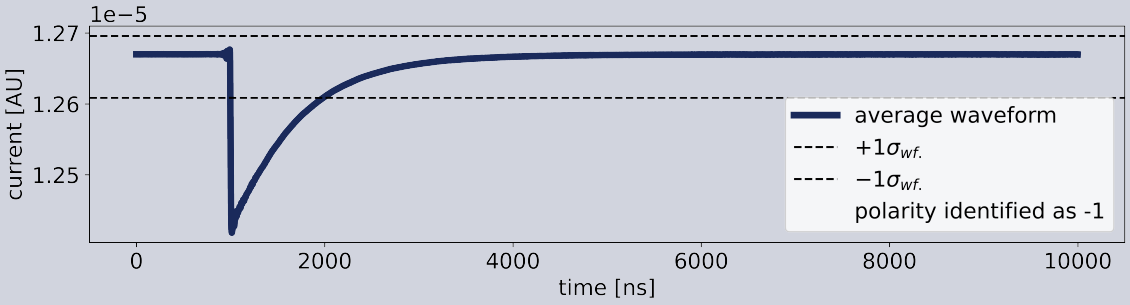

Figure 5.16: Demonstration of polarity determination for SiPM analysis

2. If the user doesn't specify, the baseline length is found. The samples of the average waveform are histogrammed. The median of the histogram corresponds to the average baseline, and the RMS of the baseline can also be determined from this histogram. From the start of the waveform, the time point at which the next two samples are both above the $5\sigma_{bl.}$ threshold is identified. The last sample of the baseline is then chosen as the sample 20 ns before the threshold trigger. See Fig. 5.17

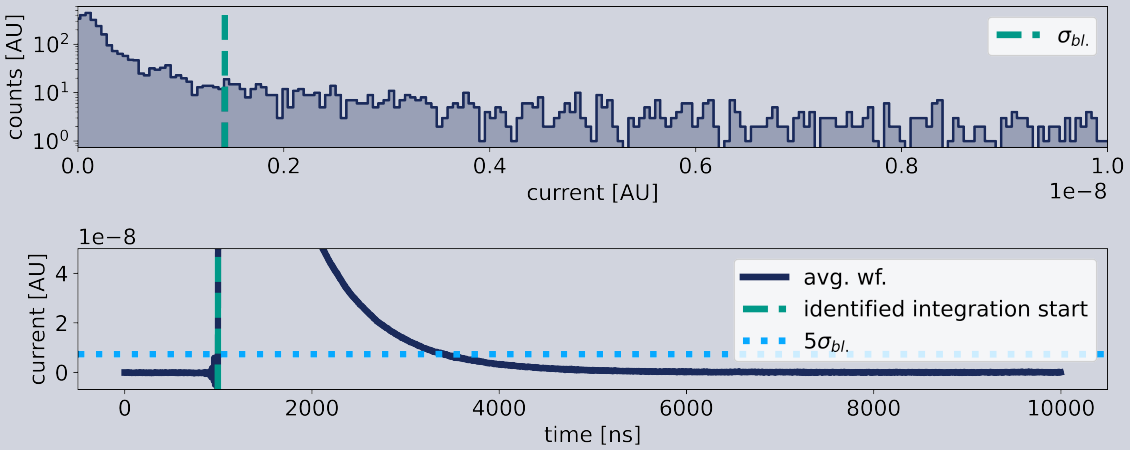

Figure 5.17: Finding the trigger time of an average SiPM waveform

3. If the user doesn't specify, the integration time is set to 10τ, where τ is the time constant determined by finding the time point where the maximum of the average waveform has fallen to $1/e$ of its value, as shown in Fig. 5.18. This integration time represents a first guess of the

integration time, and is optimized later on.

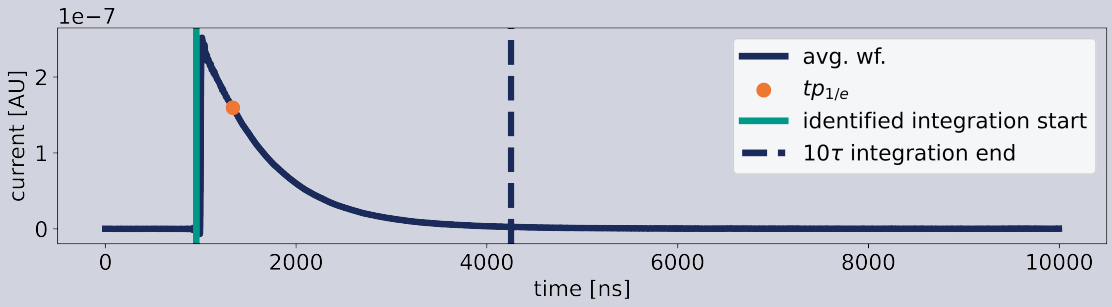

Figure 5.18: Finding the integration time of an average SiPM waveform by estimating the RC time constant

4. A baseline correction is applied to center all the waveforms' baselines to zero. This is achieved by fitting the first n samples of the waveform, as determined in step 2, with a line (fitting ADC. vs. samples) and then subtracting the intercept from every value of the waveform.

5. The charge spectrum is generated by integrating the waveforms and making a histogram

6. A rough estimate of the gain is determined by taking the fast Fourier transform of the charge spectrum, dividing out $1/f$ noise, and finding the tallest peak in charge-frequency space. This peak's location represents the gain, and its value is inverse Fourier-transformed and recorded, see Fig. 5.19. If requested, the routine stops here.

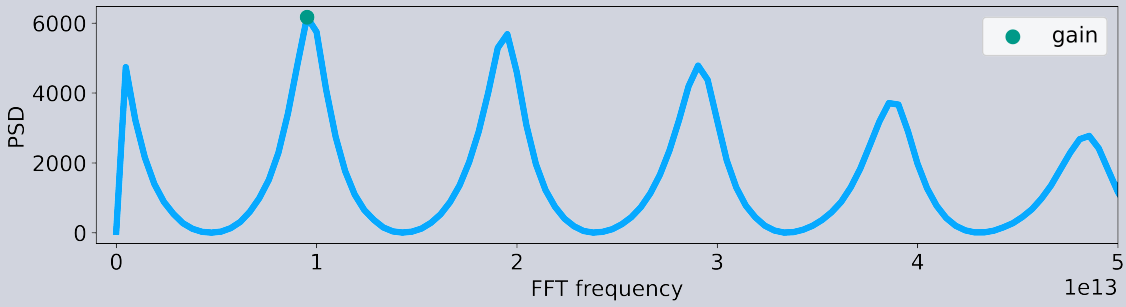

Figure 5.19: Fourier transform of a SiPM charge spectrum. The largest peak here corresponds to the gain of the SiPM due to the periodic structure of the photoelectron peaks in the charge spectrum.

7. A guess for the distances between peaks in the charge spectrum is produced by smoothing the charge spectrum with a Gaussian filter and then taking a derivative. The zero-crossing of the resulting curve coincide with the locations of the peaks and is shown in Fig. 5.20.

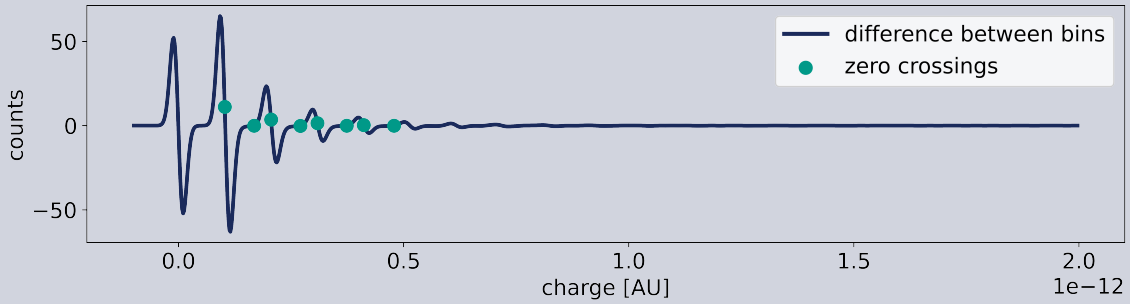

Figure 5.20: Smoothed derivative of a SiPM charge spectrum. The distances between every second zero crossing approximate the distances between peaks. These distances are then used to estimate the gain.

8. The results of the previous steps are used to seed a peak finding routine `scipy.signals.find_peaks`.

9. A guess for the width of a photoelectron peak is produced by finding the tallest peak and walking forward until the inflection point between peaks is found. The results of this procedure are shown in Fig. 5.21

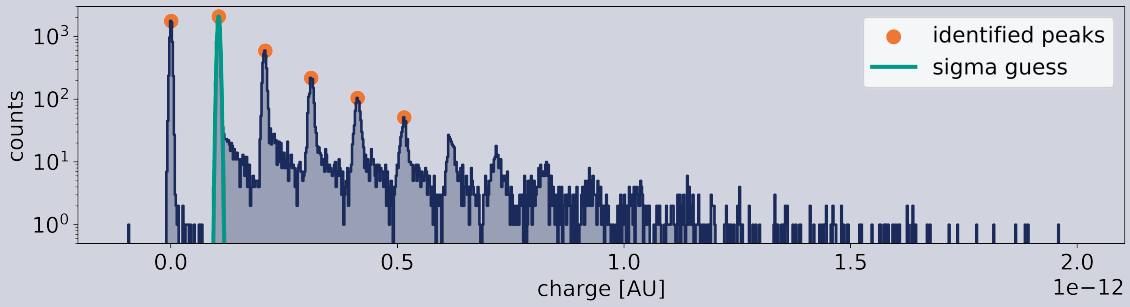

Figure 5.21: SiPM charge spectrum with a guess for the photoelectron peak width shown. The width of the peak is determined by walking off the tallest peak until the inflection point is found. The results of peak finding are also illustrated.

10. Each identified peak is then fit with a Gaussian, where the initial guess for the width comes from the previous step. The results of this fitting procedure are shown in Fig. 5.22

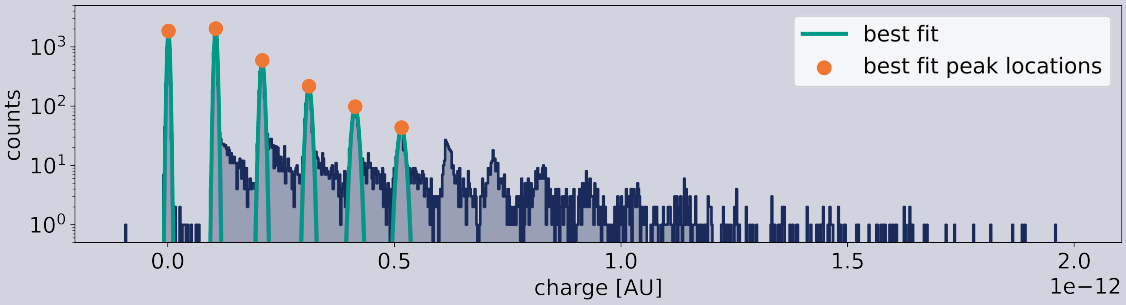

Figure 5.22: SiPM charge spectrum with Gaussian fits to photoelectron peaks

11. If the user doesn't specify, the presence or absence of the pedestal is determined.

12. Each peak is labeled by the number of photoelectrons it corresponds to. The gain is then computed by fitting a line to the best-fit location of each peak and taking its slope.

13. The normalized charge spectrum is returned, see Fig. 5.23.

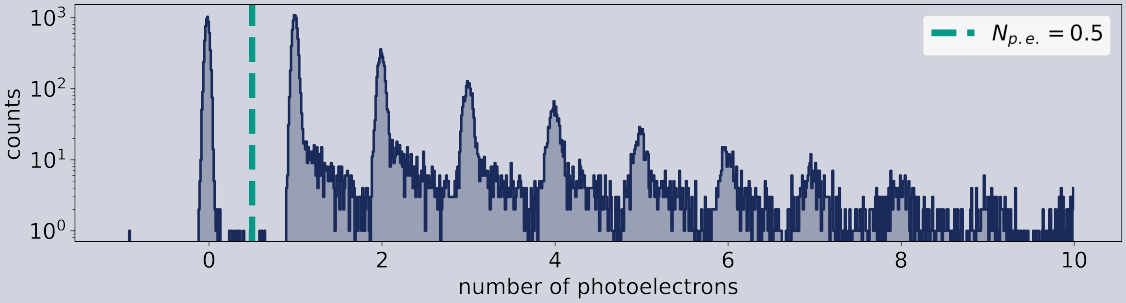

Figure 5.23: SiPM charge spectrum normalized to the number of detected photoelectrons.

14. The integration time, which was initially set to 10τ in Step 3, is optimized by maximizing Ashman's D which is computed as

$$D = \frac{\mu_{1p.e.} - \mu_{ped.}}{\sqrt{2(\sigma^2_{1p.e.} + \sigma^2_{ped.})}}. \tag{5.34}$$

This is done by changing the integration time on an integer grid from $1\tau - 20\tau$ and computing the resulting figure of merit.

The mean number of photons is computed using Eqn. 5.26 and approximating the number of baseline events N_0 as the number of waveforms with $N < 0.5$ within the integration gate.

Dark characterization using `sipm_studio`

In order to characterize the noise parameters of the SiPMs under dark conditions, the heights and times of each pulse within a waveform need to be identified. A simple peak finding routine does not work because it cannot accurately determine the heights of afterpulses: afterpulses occurring immediately after a primary discharge would be assigned an amplitude near unity, when these should be assigned the smallest amplitude. The hypercurrent method — pulse heights of the derivative of the waveform are used for analysis — employed by LEGEND-200 (see Sec. 3.2.3) fixes this issue, but fails to identify closely-spaced peaks and tends to underestimate amplitudes. The approach `sipm_studio` takes employs the optimal filter — the Wiener filter — to deconvolve the expected response of the one photoelectron avalanche. A non-negative least-squares template fitting approach would also work, but these two procedures are degenerate and should perform identically. The Wiener filter template and the identification of pulses is performed as follows:

Dark analysis with `sipm_studio`

- The charge spectrum is generated as above in Sec. 5.4.1
- Waveforms from within $\pm 1\sigma$ of the 1 p.e. peak are selected by applying a cut on the integrated charge
- The 1 p.e. waveforms are averaged to create a template pulse, see Fig. 5.24.

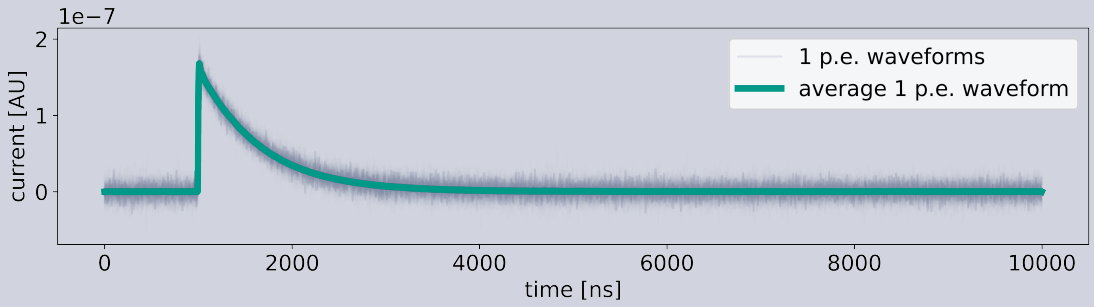

Figure 5.24: SiPM superpulse created from averaging waveforms corresponding to one photoelectron.

- Waveforms from the pedestal are used to create a representative noise pulse
- The Wiener filter is constructed following

$$G(\omega) = \frac{1}{H(\omega)} \left[\frac{1}{1 + \frac{1}{|H(\omega)|^2 SNR^2}} \right] \tag{5.35}$$

, where $H(\omega)$ is the Fourier transform of the impulse response, and SNR is the signal-to-noise ratio. We can approximate the signal to noise ratio by taking $\frac{Y(\omega)}{N(\omega)}$, where $Y(\omega)$ is the power spectral density of the signal, and $N(\omega)$ is the power spectral density of the noise. An example of applying the Wiener filter to a SiPM waveform is shown in Fig. 5.25

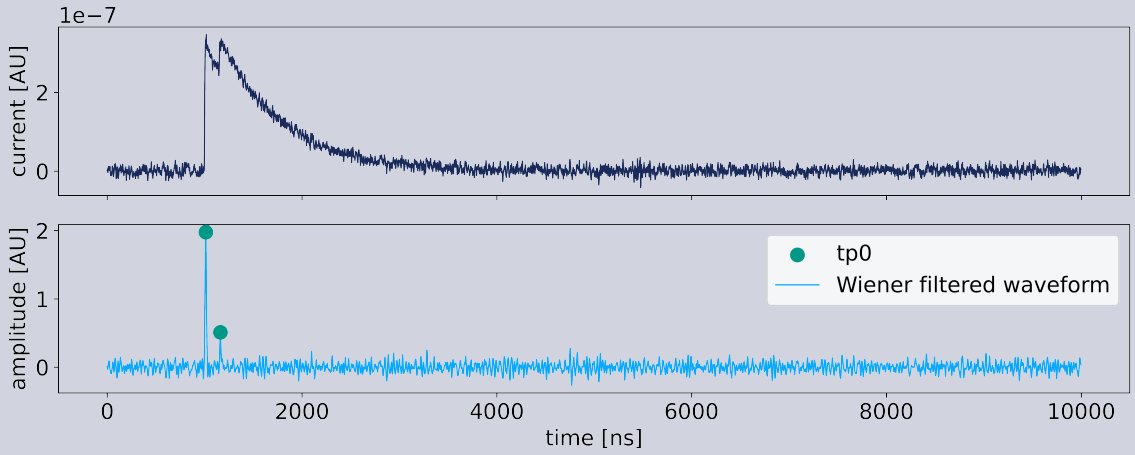

Figure 5.25: Example of Wiener filter deconvolution on a SiPM waveform. The amplitude of the afterpulse is easily determined.

- The Wiener filter is then convolved with all waveforms.
- The peaks above threshold in the Wiener convoluted waveforms are identified; their locations and heights are saved.
- The heights of the identified pulses are calibrated to number of photoelectrons, see Fig. 5.26

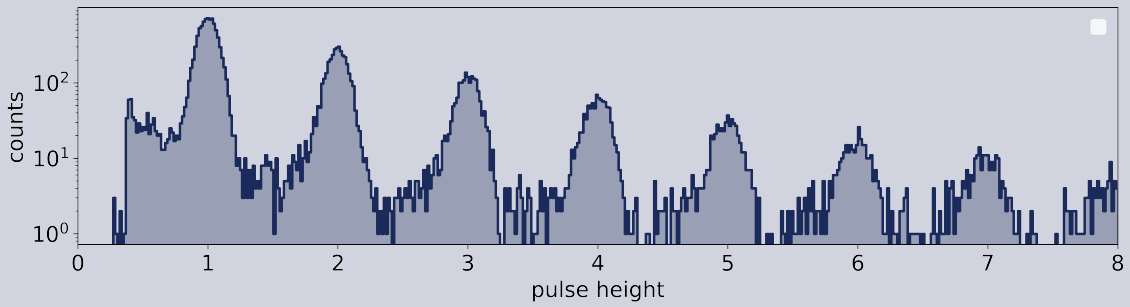

Figure 5.26: Pulse height spectrum created after applying a Wiener filter to SiPM waveforms and reading off the identified peaks in the Wiener deconvolution.

- The inter-event time plot is made by triggering on all 1 p.e. events and finding the time to the next event. An example of this plot made by the processing is shown in Fig. 5.27

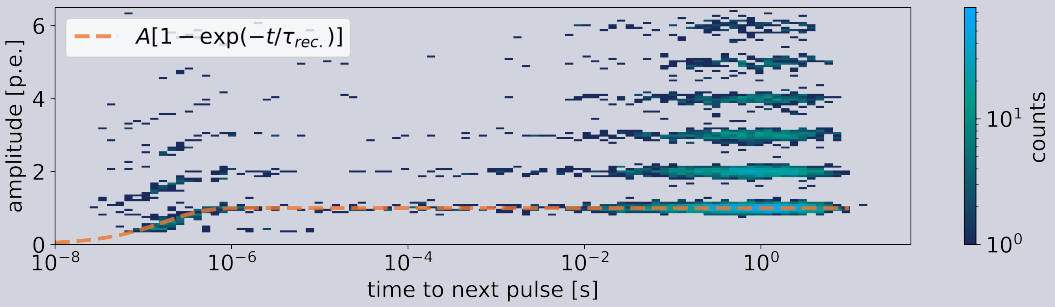

Figure 5.27: Two dimensional histogram of the time to next event and the next event's amplitude for a SiPM under dark conditions. The different noise contributions show up as different distributions in this representation of the data. The events with amplitudes below 1 photoelectron and with delay times less than 10^{-4} s are afterpulses. The events with one photoelectron amplitude clustered around 1 s delay times are primary dark counts. The clusters of events with delay times around 1 s but with amplitudes of $2, 3, \ldots$ are dark counts with accompanying cross talk events — again, because the probability of two dark counts occurring close in time is negligible.

The analysis of the inter-event time plot is usually done outside of the automatic processing, as the optimal window for determining the DCR requires manual tuning.

5.5 Cryogenic characterization of silicon photomultipliers

Several rare-event search experiments have operated SiPMs at cryogenic temperatures in liquid argon [161, 207–210] and liquid xenon [211], or are planning on operating with SiPMs optimized to detect the scintillation light from liquid argon [212–214] and liquid xenon [215–217]. The discovery sensitivity of these experiments is directly tied to the photon detection efficiency of the SiPMs used for light readout, and the PDE of SiPMs at cryogenic temperatures is not well studied. The few available measurements of the cryogenic PDE of SiPMs tend to focus on SiPMs that are optimized for cryogenic operation [218–220], although a small number of measurements of the cryogenic PDE of non-cryogenically-optimized SiPMs have been made and are available [186, 187]. Several of these measurements have shown that the PDE consistently drops by 20% to 30% at LAr cryogenic temperatures compared to room temperature values across a range of wavelengths [186, 218]. The physical mechanism responsible for this decrease is not well understood, but is hypothesized to originate from a decrease in either the quantum efficiency, the avalanche triggering probability, or a combination of the two [186, 187]. A drop in quantum efficiency can be explained by an increase

in absorption length due to the temperature dependence of the bandgap [58], and a drop in avalanche triggering probability could be explained by increased charge trapping lower temperatures. The characterization of SiPMs at cryogenic temperatures is important for understanding these effects, but is quite challenging as it requires a dedicated setup that holds only the SiPM at cryogenic temperature while ensuring that other optical properties of the setup do not change. The design of our apparatus cools and maintains the SiPM, at ambient pressure, at 77 K by keeping the SiPM in thermal contact with a simple liquid nitrogen reservoir in a dewar. Other experimental setups rely on vacuum systems to guarantee optical stability, while our approach uses liquid nitrogen boil off to similar effect, as systematic studies in section 5.5.4 show. The liquid nitrogen bath only allows us to probe one temperature — unlike other experiments that use closed-cycle cryostats or heat-exchangers — close to LEGEND's 87 K operating temperature where the results are the most critical to understanding the impact of a changing PDE on the overall light collection efficiency of LEGEND's liquid argon detector.

5.5.1 Experimental apparatus and method

A SiPM characterization test stand has been built at the University of Washington to measure the PDE and other properties, such as the dark count rate and breakdown voltage, at both room temperature and cryogenic temperature. This experimental apparatus was first introduced in Ref. [113], but several modifications have taken place since its inception. The setup consists of a ThorLabs IS200-4 integrating sphere suspended above a liquid nitrogen bath inside of a dewar contained inside a dark box as shown in Fig. 5.28. At each port of the integrating sphere, with perpendicular lines of sight, are the SiPM under test, a BROADCOM AFBR-S4K33C0135L SiPM (the reference SiPM), and a pulsed light emitting diode (LED) light source peaked at 562 nm with a full width at half maximum of 11 nm [221]. This wavelength was chosen because it is within LEGEND's WLS fibers' emission spectrum, approximately 460 nm to 580 nm [165], and was readily at hand. One NanOptics^{TM}(030UC-00S) clear, double-cladded optical fiber connects the LED light source —located outside of the dewar — to the integrating sphere. The SiPM under test sits on top of a copper cold finger, which allows it to be coupled thermally to a liquid nitrogen bath. The cold finger also ensures that the integrating sphere and the surface of the SiPM under test are not submerged in liquid nitrogen, which would alter the transmissivity and prohibit a clear measurement of the intrinsic PDE; however, future

measurements should be done to measure the transmissivity in liquid argon to determine its impact on the overall liquid argon detector efficiency. Components internal to the dewar, such as the integrating sphere and amplifier board for the SiPM under test, are colder than room temperature due to convection from liquid nitrogen; the precise temperature these components reach was not studied, but it was determined that the operating temperature of these components did not systematically impact the study, see Sec. 5.5.4 for more details.

The uncalibrated BROADCOM SiPM is used as a reference for the relative light level in the integrating sphere because an estimate of its room temperature PDE is available from the manufacturer and is constant across measurements. For room temperature PDE measurements, the reference diode is mounted directly to the integrating sphere; for cryogenic PDE measurements, the reference SiPM is placed at the lid of the dewar, thermally isolated from the LN, and optically coupled to the integrating sphere by a bundle of clear fibers. The SiPM signals are read out when triggered by an Agilent 33521 pulse generator that pulses the LED. The SiPM signals are pre-processed first with Texas Instruments LMH6629 transimpedance amplifiers near the integrating sphere and then with pre-amps consisting of two Analog Devices AD8014 amplifiers in series outside the dewar, based on the amplifier found in [222]. The Texas Instruments LMH6629 amplifier near the integrating sphere increases the signal-to-noise ratio so that single-photon resolution can be achieved during room temperature measurements. The layout of the amplifier is given in Ref. [113].

The PDE at liquid nitrogen temperature is calculated by taking the ratio of the number of detected photons at the SiPM (N_{SiPM}) and the incoming flux of photons at the SiPM port (Φ_{SiPM}) multiplied by the SiPM area (A_{SiPM}). Calculation of N_{SiPM} is done using Eqn. 5.40 and is detailed in section 5.5.2. We are using the reference SiPM to calculate Φ_{SiPM}, and because the reference SiPM has not been absolutely calibrated (we are taking its value from the manufacturer's datasheet), the PDE we measure is relative to the reference and is given by

$$\mathrm{PDE}^{\mathrm{rel.}}(V_{OV}, \lambda, 77\ \mathrm{K}) = \frac{N_{SiPM}(V_{OV}, \lambda, 77\ \mathrm{K})}{\Phi_{SiPM} A_{SiPM}}. \tag{5.36}$$

The incident flux on the test SiPM port is assumed to be equal to the incident flux on the reference SiPM port. During liquid nitrogen PDE measurements, the reference SiPM is placed at the lid of the dewar and is optically coupled to the integrating sphere by a bundle of the clear, double-cladded optical fibers, in order

to preserve the known response ($\mathrm{PDE}^{\mathrm{ref.}}$) of the reference SiPM. Because the LED's emission spectrum, $\rho(\lambda)$, spans a range of wavelengths, the reference SiPM's PDE should be a weighted average of the PDE across the LED's emission spectrum,

$$\mathrm{PDE}^{\mathrm{ref.}} = \int \frac{\lambda}{hc} \left[\frac{\rho(\lambda)}{\eta(\lambda)}\right] d\lambda \,. \tag{5.37}$$

$\eta(\lambda)$ is the spectral response of the reference SiPM; both $\eta(\lambda)$ and $\rho(\lambda)$ are available from the manufacturer. The reference SiPM is operated at a constant applied overvoltage of 5 V. The incident flux on the test SiPM is therefore expressed as

$$\Phi_{SiPM} = \frac{N_{\mathrm{ref.}}}{A_{\mathrm{ref.}}\mathrm{PDE}^{\mathrm{ref.}} f}, \tag{5.38}$$

where $N_{\mathrm{ref.}}$ is the number of photons detected by the reference SiPM during the liquid nitrogen temperature measurement, $A_{\mathrm{ref.}}$ is the area of the reference SiPM, and f is the flux ratio that corrects for the transmission coefficient to the reference SiPM (including, but not limited to, optical fiber transmission and coupling to the fibers, which are measured in Section. 5.5.4) and any increased reflectivity in the sphere due to the fibers. The equality in Eq. 5.38 assumes that light is distributed equally to both integrating sphere ports, that the transmission coefficient does not vary with temperature, and that the SiPMs are aligned perfectly normal to the port. These assumptions are verified in Section. 5.5.4. The flux ratio can be measured by first comparing the number of photons measured by the reference SiPM at room temperature in this lid PDE configuration to the number of photons measured at room temperature when the reference SiPM is placed directly on the sphere. This is then normalized by the number of photons detected by the SiPM under test in order to correct for any fluctuations in the LED's brightness or increased reflectivity from the fibers,

$$f = \frac{N^{\mathrm{lid}}_{\mathrm{ref.}}}{N^{\mathrm{dir.}}_{\mathrm{ref.}}} \frac{N^{\mathrm{dir.}}_{\mathrm{SiPM}}}{N^{\mathrm{lid}}_{\mathrm{SiPM}}}. \tag{5.39}$$

Here $N^{\mathrm{dir.}}_{\mathrm{ref.}}$ and $N^{\mathrm{dir.}}_{\mathrm{SiPM}}$ are measured when both SiPMs are mounted directly on the integrating sphere, while $N^{\mathrm{lid}}_{\mathrm{ref.}}$ and $N^{\mathrm{lid}}_{\mathrm{SiPM}}$ are measured with the test SiPM on the integrating sphere and the reference SiPM at the lid. These four quantities are measured at room temperature and the value of f is of order ~ 1.

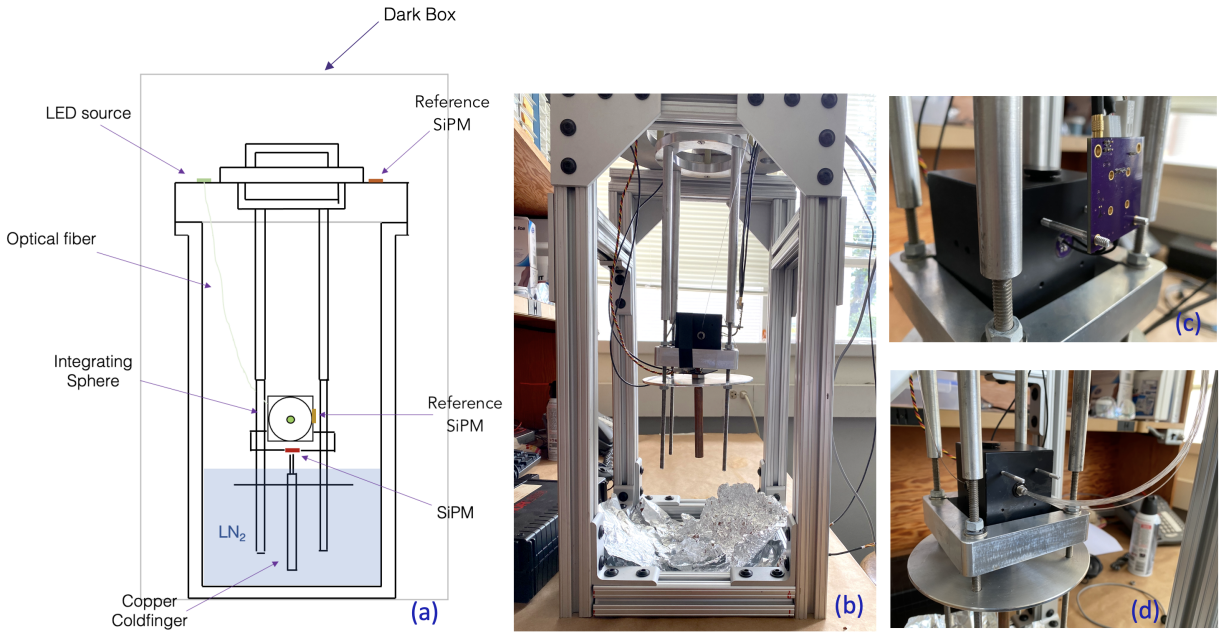

Figure 5.28: (a) A cross-sectional view of the apparatus used to perform direct and lid cryogenic PDE measurements. The integrating sphere is suspended above a dewar filled with liquid nitrogen (LN_2) and a copper cold finger thermally coupled to the SiPM extends into the LN_2 volume. (b) A photo of the integrating sphere suspension set-up that is inserted into the dewar. (c) A photo of the direct PDE reference SiPM configuration with the first stage amplification board at the integrating sphere. (d) A photo of the lid PDE measurement configuration with the optical fiber bundle connected to the integrating sphere.

5.5.2 Data acquisition and processing

The signal output of the SiPMs were recorded by a CAEN DT5730 digitizer with a 500 MS/s sampling rate, 14-bit resolution, and dynamic range of 2 Vpp. The data were acquired using CAEN CoMPASS software [223]. The LED was pulsed with a 16 ns pulse width and a 1 kHz frequency from the function generator. Waveforms of length 10 μs, with the trigger at 1.5μs, were acquired. These waveforms were long enough to ensure that the last microsecond of data — which is several microseconds after the LED has turned off and the SiPM has recharged — is taken in dark conditions so that a dark count rate could be measured. Three sets of waveform data were taken with 90 s acquisition time at each overvoltage in order to improve statistics. This process of collecting waveform data at each overvoltage was done with the direct PDE configuration at room temperature, and with the lid PDE configuration at room and liquid nitrogen temperatures. Further dark data were taken to enable a full characterization of the two SiPMs, see appendix 5.5.3.

The mean number of detected photons was computed using the pedestal method described in Eqn. 5.26 because it is unbiased by correlated noise. To get the number of baseline events N_0 the SiPM signals are first integrated in a 1 μs window around the LED trigger time as well as the last 1 μs of the waveform to get the dark count rate. Examples of the synchronously triggered waveforms are shown in Fig. 5.29a. When the histograms of these integrated signals are made, the peaks that correspond to the integral of the baseline events are called the pedestal peaks, as can be seen in Fig. 5.29b. The light level of the pulsed LED was set low enough so that an appreciable number of baseline events were acquired by making sure the pedestal peak was resolvable by eye when using the CoMPASS online energy spectrum. This corresponds to roughly 5.73 ± 0.08 photons per pulse incident on the SiPM face, as measured by the reference diode in the direct PDE configuration of the test stand. A Gaussian is then fit to the pedestal peaks in the histograms to determine the pedestal peak locations μ and standard deviations σ. The number of baseline events N_0 in light and dark conditions are then determined by the number of counts below a $\mu + 3\sigma$ threshold. This routine is necessary because differences in the electronics noise between room and liquid nitrogen temperatures preclude the use of a single threshold. To get the mean number of photons detected by either the test SiPM or the reference SiPM (N_γ) requires subtraction of the dark count rate,

$$N_\gamma = \left[-\ln\left(\frac{N_{0,\mathrm{light}}}{N_{\mathrm{total}}}\right) + \ln\left(\frac{N_{0,\mathrm{dark}}}{N_{\mathrm{total}}}\right) \right] . \tag{5.40}$$

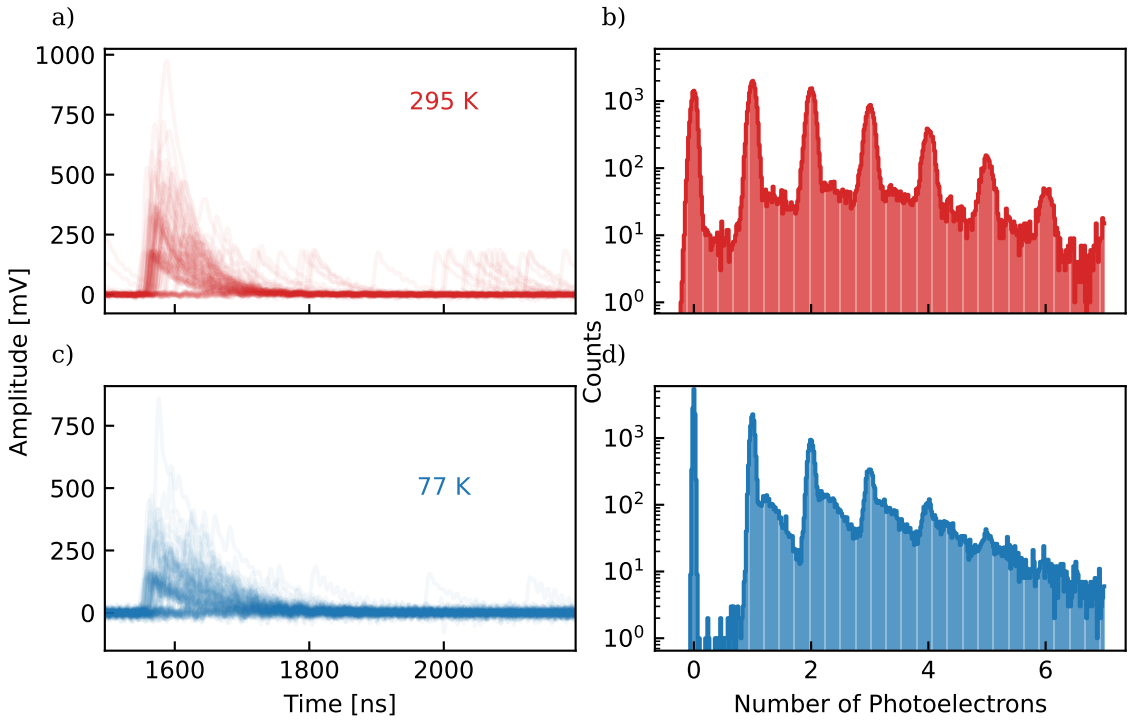

Figure 5.29: Hamamatsu SiPM waveforms at 2.5 V overvoltage, triggered synchronously with the LED pulse, for room temperature (a) and liquid nitrogen temperature (c). Normalized charge spectra are used to calculate the mean number of photons detected and are shown at room temperature (b) and liquid nitrogen temperature (d), where the peak shape for $N_{>0}$ of Hamamatsu 77 K data was found to be distorted by after pulsing, see Appendix.

$N_{0,\mathrm{dark}}$ is the number of baseline events in dark conditions, $N_{0,\mathrm{light}}$ is the number of baseline events synchronous with the LED pulse. The total number of pulses is the same for both light and dark because the dark counts come from the portion of the waveform long after the LED turns off.

Post-processing of the data from CAEN CoMPASS is done with custom open source software, `SiPM Studio` (see Sec. 5.4.1) [5]. This software parallel processes PDE data and automatically performs the Gaussian fits to the pedestal peak.

5.5.3 Cryogenic dark characterization

To estimate the SiPM dark count behavior, we took additional data under no LED illumination while the digitizer triggered on waveforms close to the 0.5 photoelectron amplitude. Waveforms with 10 μs long

[5] https://github.com/SamuelBorden/sipm_studio

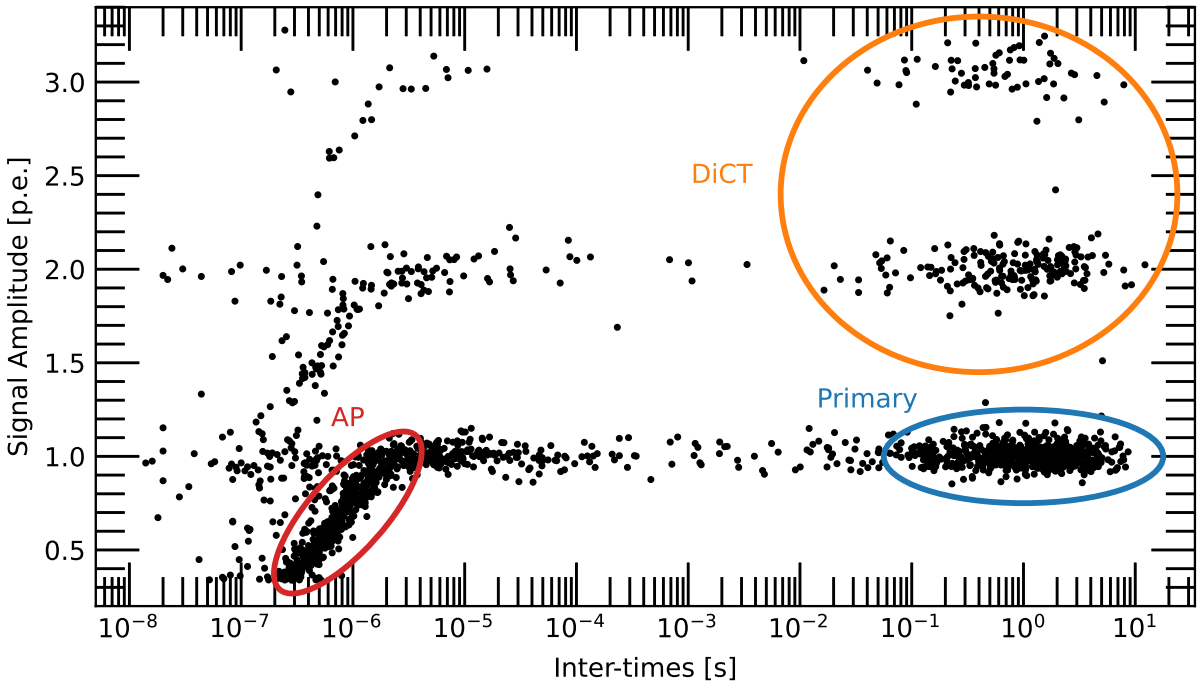

Figure 5.30: The different dark count populations shown as waveform amplitude vs waveform inter-time spacing for a 4 V overvoltage KETEK at liquid nitrogen temperature. Inter-time is defined as the time to the next waveform the triggered waveform. Primary dark counts are produced from thermal excitations, DiCT is produced from a SPAD triggering one or more neighboring SPADs simultaneously, and AP is from the trapping and re-releasing of charge carriers that generate a secondary avalanche while the SPAD is not fully recovered.

traces were acquired for 60 seconds at each reverse bias at room temperature; however, at liquid nitrogen temperatures, an hour of data per reverse bias was taken to ensure high enough statistics. The custom software processed dark data by Wiener filtering the waveforms to get a more precise determination of dark pulses' amplitude and arrival time. Wiener filtering allows for an unbiased estimate of pulse height even when a pulse occurs on the decay tail of a previous pulse.

Under no illumination, SPADs can trigger an avalanche from thermal excitations due to the environment's ambient temperature along with other correlated noise: direct crosstalk (DiCT), delayed crosstalk (DeCT), and afterpulsing (AP). The DCR was computed using Eqn. 5.27, the DiCT was computed using Eqn. 5.29, and the AP was approximated by taking the ratio of counts with amplitude less than 0.8 p.e. (and within a window of 500 ns for room temperature data and 1000 ns for liquid nitrogen data) to the total number of counts. The dark count populations for liquid nitrogen temperature are shown in Fig. 5.30.

The dark count rate was measured for several overvoltages in both devices at both room temperature and liquid nitrogen temperature. The dark count rate is suppressed by several orders of magnitude at cryogenic temperatures, as seen in Fig. 5.31(a, b). This is to be expected as the thermal generation of carriers follows a modified Arrhenius relation — thermal generation is suppressed so much that, at liquid nitrogen temperatures, band-to-band tunneling dominates [187].

The direct crosstalk was also measured at a variety of overvoltages for both devices. The crosstalk probability is larger at cryogenic temperatures for both devices, but there is a cross-over point where it becomes equal to the room temperature crosstalk probability, as seen in Fig. 5.31(c) and Fig. 5.31(d). This is somewhat surprising as the crosstalk probability depends on the gain (via the number of secondary photons produced) and the avalanche triggering probability. The gain does not change significantly at cryogenic temperatures, but the avalanche triggering probability might, as was discussed in Section 5.5.4 [187]. This could be evidence for an increased avalanche triggering probability arising from the larger mobility of charge carriers at liquid nitrogen temperatures [186]. The cross-over could indicate that the avalanche triggering probability saturates faster as a function of overvoltage in the liquid nitrogen case. The slight temperature dependence of the crosstalk probability agrees with the elevated crosstalk probability at liquid nitrogen temperatures reported in [198]. The difference between room and liquid nitrogen temperature crosstalk probabilities is also not as substantial as it is for the afterpulsing.

The Hamamatsu and KETEK afterpulsing rates are significantly higher at liquid nitrogen temperatures, as shown in Fig. 5.31(e,f). A higher afterpulsing rate is potentially explained by new charge traps or by the increased charge trapping lifetime at cryogenic temperatures: trapped carriers are more likely to be re-released while the cell is recovering and are thus more likely to trigger an afterpulse [224]. Carrier freeze-out could lead to the activation of new shallow charge traps that could explain both the high afterpulsing rate as well as the decrease in PDE via a decrease in the avalanche triggering probability due to carrier loss [187]. In addition to charge trapping re-releases, afterpulsing can also occur due to secondary carriers drifting from the substrate into the recharging SPAD [225]. This type of afterpulsing could also increase due to increased mobility of charge carriers at cryogenic temperatures and longer SPAD recovery times: carriers in the bulk silicon are more easily drifted into the active region and trigger an afterpulse [226]. Fig. 5.29d shows that the higher afterpulsing rate degrades the Hamamatsu charge spectrum at liquid nitrogen temperatures. A

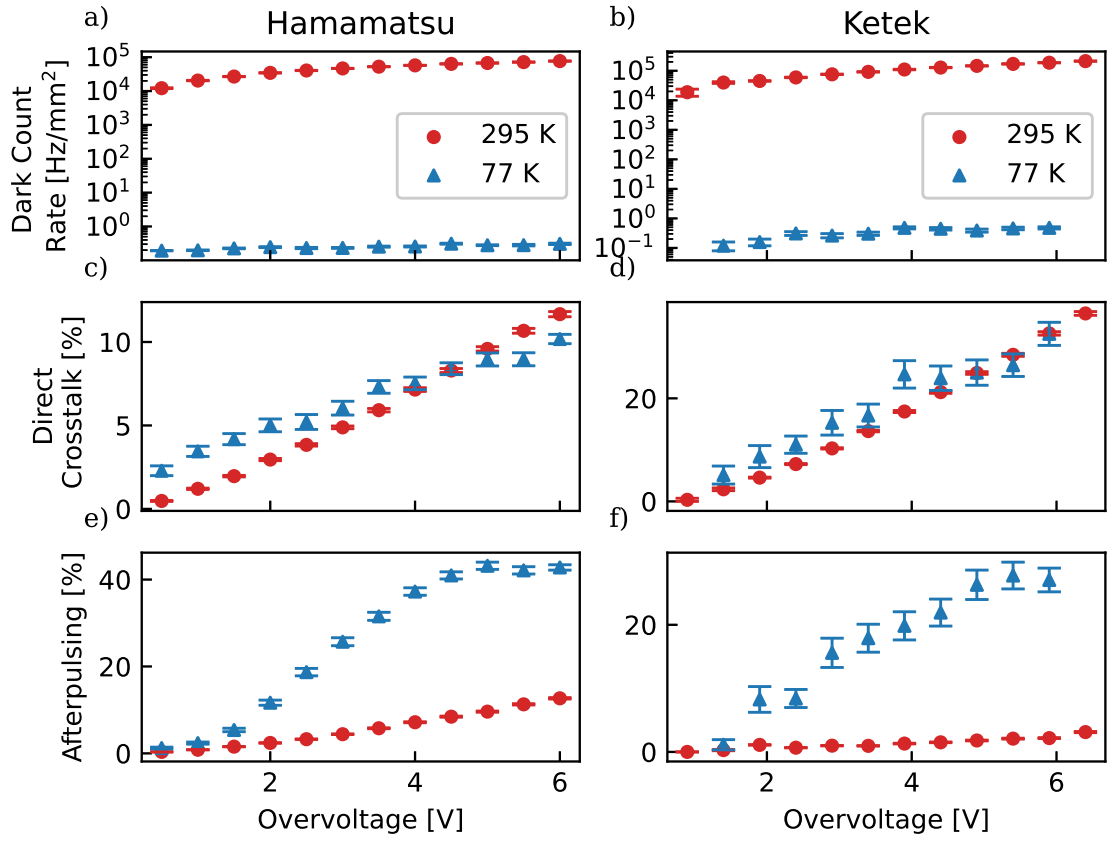

Figure 5.31: Dark characterization data for the KETEK PM3325-WB-D0 and Hamamatsu S13360-3050CS SiPMs at room and liquid nitrogen temperatures as a function of overvoltage. Room temperature data for the Ketek were taken with a 0.1 V offset from the breakdown voltage, so liquid nitrogen data were acquired with the same offset. In addition, the number of events for the Ketek SiPM at 0.4 V overvoltage at liquid nitrogen temperature was too low to analyze due to the extremely low dark count rate, so the data point was omitted.

similarly high cryogenic afterpulsing rate for a Hamamatsu device was recently reported in [227].

5.5.4 Cryogenic characterization of the photon detection efficiency

The relative PDE was measured at room temperature and at liquid nitrogen temperature for the KETEK SiPM and Hamamatsu SiPM, see Fig. 5.32. Errors on all three quantities in Eq. 5.36 were propagated during error calculations. In order to account for a ± 0.1 V uncertainty on the breakdown voltage of the reference SiPM, a systematic uncertainty on $\mathrm{PDE}^{\mathrm{ref.}}$ of 1% was found from the manufacturer's datasheet and was used in the analysis. The room temperature values of the PDE match the manufacturer's data sheets. The Hamamatsu SiPM has a much larger PDE than the KETEK SiPM. Although the doping profiles are proprietary, it is likely that the Hamamatsu device has an electric field structure that leads to both a larger PDE as well as a larger afterpulsing rate, as seen in Fig. 5.31. The room temperature was 72 °F (295 K) held by an air conditioning unit and the liquid nitrogen temperature was measured to be 77 K with a calibrated PT1000 sensor placed in between the SiPM board and copper cold finger. These temperatures were stable within 1 K during data taking. The breakdown voltage of the KETEK SiPM was measured at room temperature to be 24.1 V; the breakdown voltage at 77 K was 20.6 V. The breakdown voltages were obtained by extrapolating the SiPM gain, G, as a function of reverse bias to $G = 1$ and were confirmed with a reverse IV curve from a Keithley 2450 Sourcemeter®. Likewise, the breakdown voltage of the Hamamatsu SiPM was measured at room temperature to be 51.5 V; the breakdown voltage at 77 K was 42.2 V. Because the breakdown voltage depends on the temperature, the breakdown voltage was also used as a cross-check on the temperature. The breakdown voltages at liquid nitrogen temperature were determined by submerging the devices completely in liquid nitrogen in a separate experiment prior to the PDE measurement. While the reference SiPM was mounted on the lid, the breakdown voltage of the reference SiPM was found to not change after an LN fill, confirming that it was kept at room temperature.

The PDE was measured for a range of overvoltages and results show a significant decrease in PDE at liquid nitrogen temperature for all overvoltages. This decrease in PDE is present even when accounting for possible changes in the optical properties of the setup at different temperatures or between configurations. Any systematic impact of the light uniformity in the integrating sphere or SiPM alignment would be seen during room temperature PDE measurements. Our room temperature PDE measurements are in

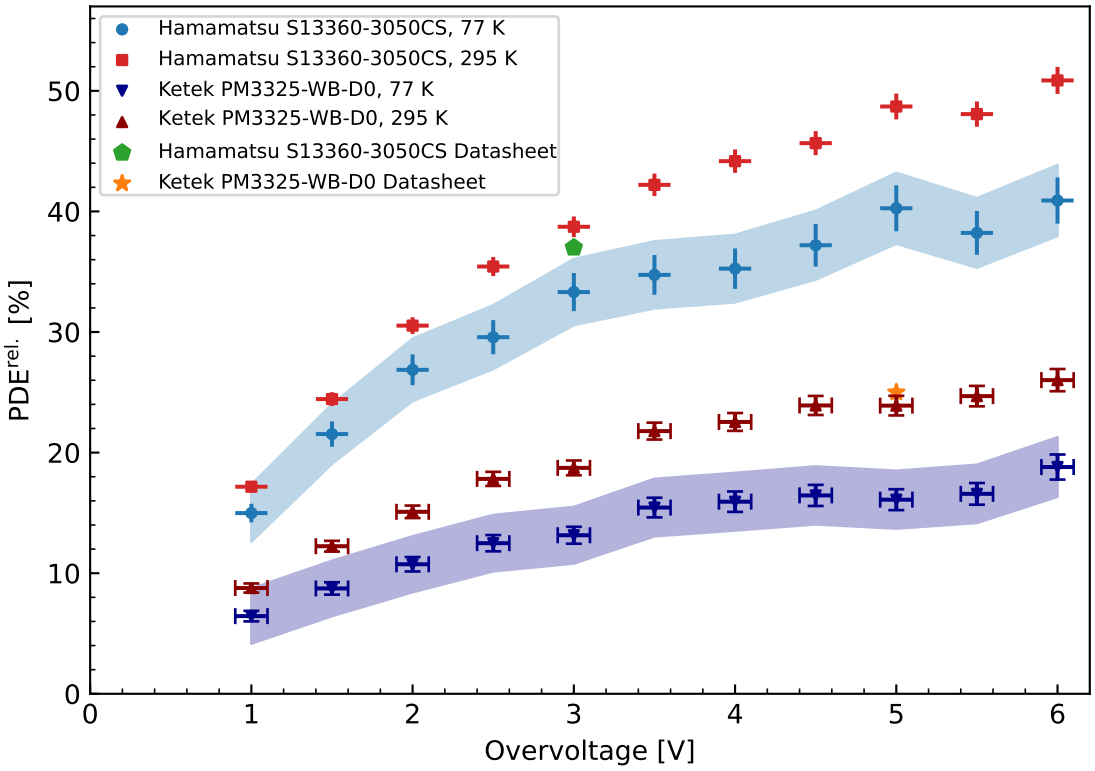

Figure 5.32: Measurement of the relative PDE for a KETEK PM3325-WB-D0 SiPM and a Hamamatsu S13360-3050CS SiPM for a range of overvoltages. Measurements are done at room temperature and liquid nitrogen temperature using 562 nm light. The datasheet values for the PDEs at the peak LED wavelength used are provided for reference. The error bands correspond to combined statistical and systematic (2% from the flux ratio f and 1% from $\mathrm{PDE}^{\mathrm{ref.}}$) uncertainties.

good agreement with the manufacturer's datasheets, as seen in Fig. 5.32. Moreover, the agreement between manufacturer and measured room temperature PDE values indicates that all three SiPMs are consistent with their manufacturer reported values, within our experimental uncertainties; although, a common systematic offset between all three cannot be excluded. Any differences in SiPM alignment when the reference diode is moved to the lid for lid configuration PDE measurements are accounted for in measurements of the flux ratio, Eq. 5.39. Systematic studies were performed to determine if there was any temperature dependence in the transmission coefficient of the optical fibers connecting to the reference SiPM, the emission spectrum of the LED, or the reflectivity of the integrating sphere.

We characterized the temperature dependence of the optical fibers' transmission coefficient (and the flux ratio f) by taking the dewar insert as shown in Fig. 5.28(d) and running 25 cm of the 30 cm long fibers through a Styrofoam cooler that was then filled with liquid nitrogen. The fibers were thus the only part of the test stand that was cold. One end of the fibers was connected to the integrating sphere, with the Hamamatsu SiPM placed at the opposite port, while the other end was coupled to the reference SiPM. The integrating sphere, the Hamamatsu SiPM, and the reference SiPM were all at room temperature. We then took PDE data with and without liquid nitrogen on the fibers while triggering on LED pulses sent into the integrating sphere. Fig. 5.33 shows the effective PDE — the number of photons detected by the Hamamatsu on the integrating sphere divided by the number of photons detected by the reference SiPM at the end of the fiber — when the fibers were at room and liquid nitrogen temperatures. The difference between the mean effective PDEs measured over several trials was found to be 2%, which places a limit on the systematic error in our cryogenic PDE measurements.

We also investigated the systematics arising from the finite width of the LED's emission spectrum. Using Eq. 5.37 to calculate the reference SiPM PDE instead of using the reference SiPM's PDE at the peak emission wavelength of the LED introduces only an average 0.02% difference in our final cryogenic PDE results, which is below our statistical sensitivity. We therefore compute the cryogenic PDE using only the peak emission wavelength of the LED.

We checked that $\rho(\lambda)$ (the LED's emission spectrum) and the integrating sphere's optical properties were not functions of temperature. While the LED emission spectrum should not change because the LED is placed outside of the dewar, the optical fiber that connects the LED to the integrating sphere could have

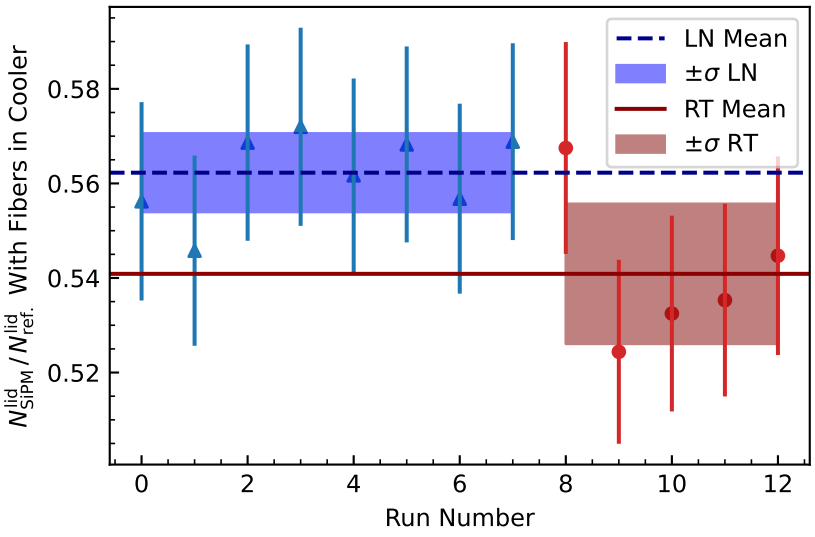

Figure 5.33: The fluctuation in the effective PDE, $\frac{N^{\text{lid}}_{\text{SiPM}}}{N^{\text{lid}}_{\text{ref.}}}$, when only the optical fibers were cooled down sets a limit on the systematic error at 2%.

a changing transmission coefficient as a function of temperature. To constrain this possibility, we compared the number of photons detected at the reference SiPM at 77 K and 295 K while the test stand was in the lid PDE measurement configuration. The reference SiPM has a comparable spectral response to the SiPMs under test. The results, shown in Fig. 5.34, indicate that the reference SiPM detects roughly the same number of photons when the test stand is at both room and cryogenic temperatures. Thus there is neither a shift in the emission spectrum of the LED nor a shift in the reflectivity of the integrating sphere. This measurement also serves as an *in situ* cross-check that the optical fibers' transmission probability is not temperature dependent.

We ensured that condensation was not forming inside the integrating sphere or on the surfaces of the SiPMs. First, prior to data taking, a liquid nitrogen fill was performed and the dewar insert was lifted to visually check that no condensation was present. Second, the number of photons detected by the reference SiPM at the lid does not change significantly between room temperature and cryogenic temperature measurements in Fig. 5.34: any change in reflectivity engendered by condensation inside the integrating sphere would cause a systematic change in the number of photons detected by the reference SiPM. The nitrogen gas boil-off from the liquid nitrogen ensures that condensation cannot form. Therefore, any changes in the optical properties of the test stand between the two temperatures do not account for the observed decrease in PDE.

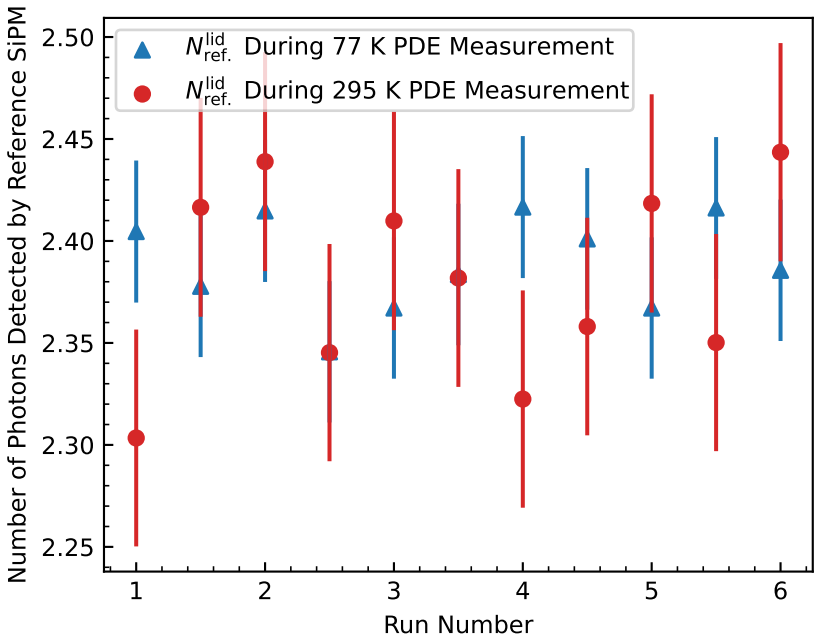

Figure 5.34: Comparing the number of photons detected by the reference SiPM located at the lid during room and cryogenic temperature PDE measurements shows that there are no systematic shifts in the LED's emission spectrum or the transmission coefficient of the optical fibers.

The decrease in PDE at 77 K indicates that some combination of the quantum efficiency, avalanche triggering probability, and effective fill factor may be changing with temperature. Similar reductions of the PDE at cryogenic temperatures have been found in other SiPMs at 10 K using a calibrated system [228], and at 77 K using uncalibrated systems [186, 187, 219, 229] and calibrated systems [215, 218]. These systems show similar decreases in PDE of roughly 30% at liquid nitrogen temperatures for wavelengths close to 560 nm. An FBK NUV-HD-cryo SiPM operated at 3 V overvoltage was measured to have a PDE drop from ∼24% at room temperature to ∼18% at 75 K for 525 nm light; however, the decrease in PDE at 5 V overvoltage is much smaller, dipping from ∼34% down to only ∼29% [186]. A VUV-sensitive Hamamatsu SiPM at 4 V overvoltage displayed a similar decrease in PDE from 28.4% at room temperature to 20.1% at 77 K for 570 nm light as reported in [218]. Because silicon is an indirect band-gap semiconductor and fewer phonons are present in the crystal lattice at liquid nitrogen temperature, the probability of charge carriers moving to the conduction band from the absorption of a photon will be lower. This is reflected in the increase of the absorption depth of photons at a given energy at liquid nitrogen temperatures; this effect could reduce the quantum efficiency [186, 187]. The band-gap of silicon also increases with decreasing temperature,

but this may produce negligible effects on the quantum efficiency [230]. In addition, the recombination probability for electron-hole pairs before reaching the avalanche region may be higher at liquid nitrogen temperature. Similar findings were observed when researching the quantum efficiency of silicon thin film solar cells [231]. The avalanche triggering probability could also decrease due to carrier loss from carrier freeze-out [187]; however, the triggering probability may also increase slightly due to larger carrier mobility at liquid nitrogen temperatures. More work needs to be done to disentangle the many effects that might be causing the decrease in the PDE. It is not expected that thermal expansion significantly impacts the effective fill factor because the thermal expansion coefficient for silicon at liquid nitrogen temperatures amounts to only a 0.02% change in area, which is too small to explain the observed decrease in PDE.

5.6 Future work

The mechanism behind the observed reduction of the PDE at cryogenic temperatures is still not well understood. There is a complex interplay between the absorption depth of a photon in silicon and the avalanche triggering probability as a function of absorption depth, as hypothesized in Eqn. 5.6 that makes the temperature dependence hard to determine. For UV light, however, because the short absorption length ensures that all avalanches are electron-driven, the PDE can be well-modeled by Eqn. 5.10 which contains only two fit parameters [179]: PDE_{max}, which is comparable to the quantum efficiency (and contains information on the absorption length), and k_e, which is a semi-empirical parameter that encodes the strength of the voltage dependence of the avalanche triggering probability. Studying the temperature dependence of these two fit parameters for UV wavelength PDE data would shed some light on the physics responsible for the observed PDE decrease.

Finding a model-independent way to disentangle the origin of the PDE decrease could also be fruitful and could validate the model used in Eqn. 5.10. Operating the SiPM below its breakdown voltage in linear mode would not disentangle the effects of the quantum efficiency and avalanche triggering probabilities; even in the linear mode the SiPM acts as an avalanche photodiode (APD) with avalanche multiplication. Operating the SiPM below the linear mode, where there is no internal gain mechanism and the SiPM behaves more like a p-i-n photodiode, would enable measurements of the isolated quantum efficiency of the device. However, the operating voltage in such a regime would mean that the depletion region is different in its

physical extent compared to when the SiPM is operated above breakdown. More work should be performed on simulations of the depletion region to see if useful information could be extracted for UV wavelength data.

The parametrization of the PDE in Eqn. 5.7 can in principle allow for a detailed study of the temperature dependence of the fit parameters of this model — this could also help to disentangle the origin of the measured PDE decrease. It could also prove fruitful to study the functional dependence of Eqn. 5.7 on absolute voltage instead of overvoltage, because the geometry of the depletion region should depend on the absolute value of the voltage. It could also be interesting to look at the dependence on the relative overvoltage (defined as the overvoltage divided by the breakdown voltage), as was done in Ref. [188]. A full dedicated campaign to measure the cryogenic PDE at multiple wavelengths is thus recommended.

5.7 Conclusions

An experimental setup has been built at the University of Washington to characterize SiPMs at liquid nitrogen temperature to assess the efficiency of the liquid argon veto readout system to be deployed in LEGEND-1000. The SiPM test stand consists of an integrating sphere suspended above a liquid nitrogen bath in a dewar housed inside a dark box. The SiPM is in thermal contact with a copper cold finger that is partially submerged in liquid nitrogen and the integrating sphere is in a room-temperature thermal bath. Numerous systematics studies were performed to ensure that no other optical properties of the experimental apparatus changed at liquid nitrogen temperatures. Measurement of both a KETEK PM3325-WB-D0 SiPM and Hamamatsu S13360-3050CS SiPM PDE at liquid nitrogen temperature suggests the liquid argon veto readout system has a lower efficiency compared to the expected PDE for room temperature operation. We expect the same behavior to persist at liquid argon temperatures because they are only 10 K away from the measured liquid nitrogen temperatures; furthermore, past studies have shown that the PDE at 87 K for similar wavelengths is only 1.1 times larger than the PDE at 77 K [187]. The reduction in PDE seems to be caused by temperature dependencies to the intrinsic quantum efficiency and avalanche triggering probability of the SiPM. Changes to the effective fill factor may be present but are most likely negligible in comparison to the changes in quantum efficiency and avalanche triggering probability. Additional measurements of the PDE at different wavelengths at cryogenic temperatures should be pursued.

The measurement of a decrease in PDE greater than 20% for all overvoltages at cryogenic temperatures for a SiPM sensitive to the visible spectrum is significant information for assessing and improving the background suppression capability of LEGEND-1000. In LEGEND-1000 background components are strongly vetoed by either the LAr anti-coincidence cut or by the PSD cut, but not both [89]. Conservatively, reducing the projected LAr anti-coincidence efficiencies for all background components in the LEGEND-1000 background model (presented in Ref. [89]) by 20 % yields a background index of 1.1×10^{-5} counts/keV/kg/yr. This is only a 10 % increase in the projected background index. A dedicated Monte Carlo of the background model for LEGEND-1000 should be performed with a decreased SiPM PDE to accurately assess the impact of the PDE decrease on the projected background index of the experiment.

Based on the cryogenic characterization work presented in this chapter, the Hamamatsu SiPM presents a compelling choice for use in the LEGEND experiment because its PDE is roughly two times larger than the current KETEK device used in LEGEND-200, as shown in Fig. 5.32. The Hamamatsu SiPM does have a higher afterpulsing probability, but the analysis routines used in tagging LAr anti-coincidence can use optimal filtering to easily distinguish afterpulsing events from primary physics signals. Future work should be done to characterize the Fondazione Bruno Kessler (FBK) SiPMs that LEGEND-1000 plans to use for its LAr detector light readout system — these SiPMs are the same FBK NUV-HD-Cryo SiPM technology being used for DarkSide-20K [213]. Compared to the Hamamatsu device, these FBK SiPMs have a comparable PDE of 40% at the green wavelengths LEGEND uses for its optical fibers, a lower dark count rate on the order of 1×10^{-2} counts per square millimeter, and a photosensitive area that is roughly ten times larger than that of the Hamamatsu SiPMs [186]. These excellent device parameters make the FBK SiPM the baseline design choice for the LEGEND-1000 LAr light readout system.

Chapter 6

Conclusion

Current generation neutrinoless double-beta decay experiments are probing an exciting new chunk of parameter space, and the bottom of the inverted ordering parameter space is in reach for the sensitivities of next-generation experiments, and hopefully a discovery is soon to follow. Experiments such as LEGEND will require ultra-low background indices and large sensitive masses in order to reach the target discovery sensitivities. The LEGEND-200 experiment is exploring new parameter space and testing the detector technology that will enable LEGEND-1000 to fully cover the inverted ordering. The first results from the LEGEND-200 detector are promising, and the experiment will continue accumulating exposure until LEGEND-1000 comes on-line. The path to LEGEND-1000 requires that every key analysis and hardware parameter be highly optimized so that the lowest background and largest sensitive exposure are achieved. From there, the quest for neutrinoless double-beta decay is then a matter of statistical inference.

6.1 Frequentist statistical inference for quasi-background free experiments

This work presents the frequentist statistical inference for the first dataset from the LEGEND-200 detector. No evidence for a signal was observed ($p = 0.1$) and a lower limit on the half-life of neutrinoless double-beta decay for ^{76}Ge was set at $T^{0\nu}_{1/2} > 5 \times 10^{25}$ years (90% confidence level). A combined analysis of data from MAJORANA DEMONSTRATOR (MJD), Germanium Detector Array (GERDA), and LEGEND-200 was also presented, again yielding no evidence for a signal ($p = 0.29$). The combined limit on the half-life was set at $T^{0\nu}_{1/2} > 1.9 \times 10^{26}$ years (90% confidence level). This work was enabled by the development of a

Python software package `freqfit` that performs frequentist statistical inference on unbinned data.

During the frequentist analysis of LEGEND-200, it was noticed that the behavior of the test statistic distributions deviated from asymptotic predictions by Wilks' theorem. The behavior of the frequentist statistical treatment was then investigated for generic rare-event peak searches in the quasi-background-free regime. In this regime the departure from asymptotic behavior persists as the total count rate is so low. It is thus recommended that the full frequentist profile likelihood ratio treatment be implemented for analyses of quasi-background-free experiments: toys are necessary. A discussion of one-sided and two-sided hypothesis tests was also presented. The use of a one-sided hypothesis test for upper-limit setting does not constitute flip-flopping, as the statistical hypothesis being tested is different from the two-sided test and is not conditioned on the results of the two-sided test — as long as the two-sided test and/or the likelihood to generate either is *also* provided. Results of using the one-sided $\tilde{q}_\mu$ test statistic for exclusion limits were presented for the LEGEND-200 dataset; an almost 20% difference in the limit was computed. It is thus recommended that quasi-background-free experiments generate two sets of expected-and-observed limit plots: one using the two-sided statistic $\tilde{t}_\mu$ for discovery and two-sided confidence interval construction, and one using the one-sided $\tilde{q}_\mu$ for reporting an upper limit.

6.2 Cryogenic SiPM characterization for LEGEND

In order to reach the desired background index for the LEGEND-1000 experiment, it is necessary to hyper-optimize the performance of every single detector in the experiment. The measured reduction in the photon detection efficiency at cryogenic temperatures thus represents an important input for the background modeling of the LEGEND-1000 experiment. However, the SiPMs tested in this work are not from the manufacturer used in the baseline design of LEGEND-1000, which plans on using custom-designed SiPMs from Fondazione Bruno Kessler (FBK). It is thus highly recommended that the cryogenic PDE of these devices be measured as soon as possible to see if there is a similar drop in the PDE. The mechanism for the observed reduction in the PDE is not well understood. Follow-up measurements of the cryogenic PDE for numerous wavelengths of light from the UV to the IR could help disentangle the origins of the drop. The SiPM test stand at the University of Washington could be upgraded to a vacuum cryostat system with liquid nitrogen feed-thru so that systematic effects from the optical transmission fibers can be completely removed from the

analysis. Furthermore, if a pulse tube cryocooler is used, a wider range of temperatures could be scanned in order to observe a broader picture of the temperature dependence of the PDE.

6.3 Future ideas for background rejection R&D in LEGEND-1000

The projected background index of the LEGEND-1000 experiment is predicated on the acquisition of underground source argon for the liquid argon detector. Underground argon is depleted in concentration of ^{42}Ar whose progeny ^{42}K is one of the largest backgrounds for LEGEND-200. If underground argon is not acquired, the current PSD performance and liquid argon anti-coincidence cut performance would not enable LEGEND-1000 to reach its background index goal. As such, the acquisition of underground argon is the largest single technical risk to the LEGEND program, and should be mitigated. Currently, there is R&D work on encapsulating the germanium detectors entirely in scintillating material to reduce the direct exposure to liquid argon, as well as the development of new PSD cuts targeting ^{42}K surface beta-decays. This R&D work requires a large activity of ^{42}Ar to enable testing, on the order of hundreds of becquerels in order to distinguish the decay from background. Procurement of ^{42}Ar is no easy feat, as concentrations in the atmosphere are on the order of microbecquerel per kilogram. One possible way to create a sample is via high energy α bombardment on ^{40}Ar at a cyclotron facility, such as at the 88-inch cyclotron at Lawrence Berkeley National Laboratory. The production of a high activity ^{42}Ar is urgently needed. Work, which is not presented in this thesis, is currently being performed at the University of Washington to create a target ^{40}Ar sample for irradiation at one of these facilities.

For further R&D work on ^{42}Ar background mitigation, it could be beneficial to understand the mechanism by which ^{42}K arrives on the surfaces of the germanium detectors. It is unknown whether the ionized potassium ions have their drift dominated by convection currents in the liquid argon cryostat, or by electric forces due to fringe fields from the germanium detectors. Furthermore, the relative abundance of charge states of the ionized potassium are also unknown. All ^{42}K ions are expected to be initially positively charged from the beta-decay of ^{42}Ar; however, GERDA saw evidence of both positive and negatively charged potassium ions adhering to the detector surface. One easy way to measure the relative fraction of positively to negatively charged states produced in the decay of ^{42}Ar in a liquid argon environment is to perform the following experiment. Two copper rods are held at high voltage and are inserted into a U-shaped liquid

argon cryostat, with one rod being positively biased and the other being negatively biased. A lead brick is inserted in the middle of the U, and then the argon cryostat is spiked with a known activity of ^{42}Ar. Outside of the cryostat are two germanium detectors that are used to measure rate of counts in the 80% intensity 1565.86 keV gamma line from the decay of ^{42}K. If a calibration source of known activity is placed coaxial to the copper rods, then the efficiency of the germanium detectors can be calibrated. Thus the relative rates of ^{42}K counts in each detector can be used to determine the relative fraction of the charge states populated during the decay of ^{42}Ar in the argon cryostat. This information could be useful if only one charge state is dominantly populated: cuts could be optimized for the corresponding detector surface to which it would adhere.

6.4 The road ahead

The discovery sensitivity of future neutrinoless double-beta decay experiments rests on increasing the sensitive exposure and decreasing the expected background index. The sensitive exposure is primarily a function of isotopic mass and thus is cost-limited. Reduction of the background index, however, is a function of human ingenuity. Exciting hardware R&D work is being done on the LEGEND-1000 experiment to improve the liquid argon detector, and alternate ^{42}Ar mitigation strategies are being pursued from different directions. On the analysis front, a delicate balance rests on decreasing the background while keeping sensitive mass as high as possible; future work on multivariate background cuts and machine learning-based approaches will be invaluable in these endeavors. With the discovery sensitivity optimized through both hardware and analysis advancements, these next generation neutrinoless double-beta decay experiments will completely cover the allowed inverted ordering parameter space — hopefully a 3σ discovery is soon to come.

References

1. KamLAND-Zen Collaboration *et al.*, *Phys. Rev. Lett.* **130**, Publisher: American Physical Society, 051801 (Jan. 2023).
2. H. Acharya *et al.*, *First Results on the Search for Lepton Number Violating Neutrinoless Double Beta Decay with the LEGEND-200 Experiment*, Version Number: 4, 2025.
3. W. Pauli, *Pauli letter collection*.
4. C. D. Ellis, W. A. Wooster, *Proc. Roy. Soc. Lond. A* **117**, 109–123 (1927).
5. C. D. Ellis, W. A. Wooster, en, *Mathematical Proceedings of the Cambridge Philosophical Society* **22**, 849–860, ISSN: 1469-8064, 0305-0041 (Nov. 1925).
6. W. Pauli, *On the earlier and more recent history of the neutrino*, INIS Reference Number: 23084481 (Cambridge University Press, United States, 1991), ISBN: 978-0-521-36452-2.
7. B. Pontecorvo, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **1**, Originally published 1946, 25–31 (1991).
8. F. Reines, C. L. Cowan, *Phys. Rev.* **92**, Publisher: American Physical Society, 830–831 (Nov. 1953).
9. A. D. Franklin, A. D. Marino, *Are There Really Neutrinos?: An Evidential History*, en, Google-Books-ID: QbXjDwAAQBAJ (CRC Press, May 2020), ISBN: 978-0-429-57843-4.
10. F. E. Wietfeldt, E. B. Norman, *Physics Reports* **273**, 149–197, ISSN: 0370-1573 (1996).
11. B. Pontecorvo, *Sov. Phys. JETP* **6**, 429–431 (1958).
12. B. Pontecorvo, *Sov. Phys. JETP* **7**, 172–173 (1958).

13. A. B. Balantekin, B. Kayser, *Annual Review of Nuclear and Particle Science* **68**, _eprint: https://doi.org/10.1146/annurc nucl-101916-123044, 313–338 (2018).

14. L. Stanco, *EPJ Web Conf.* **164**, ed. by L. Bravina, Y. Foka, S. Kabana, 01031, ISSN: 2100-014X (2017).

15. M. Aker *et al.*, en, *Nat. Phys.* **18**, Number: 2 Publisher: Nature Publishing Group, 160–166, ISSN: 1745-2481 (Feb. 2022).

16. M. Thomson, *Modern Particle Physics*, en, Google-Books-ID: BV1sAAAAQBAJ (Cambridge University Press, Sept. 2013), ISBN: 978-1-107-03426-6.

17. E. Majorana, it, *Nuovo Cim* **14**, 171–184, ISSN: 1827-6121 (Apr. 1937).

18. M. Fukugita, T. Yanagida, *Physics of Neutrinos* (Springer, Berlin, Heidelberg, 2003), ISBN: 978-3-642-07851-4 978-3-662-05119-1.

19. E. Akhmedov, in *The Physics of Ettore Majorana: Theoretical, Mathematical, and Phenomenological*, ed. by S. Esposito (Cambridge University Press, Cambridge, 2014), pp. 303–353, ISBN: 978-1-107-04402-9.

20. J. J. Gomez-Cadenas, J. Martin-Albo, M. Mezzetto, F. Monrabal, M. Sorel, *La Rivista del Nuovo Cimento* **35**, arXiv:1109.5515 [hep-ex, physics:hep-ph, physics:nucl-ex], 29–98, ISSN: 0393697X, 0393697X (Jan. 2012).

21. S. M. Bilenky, en, *Universe* **6**, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, 134, ISSN: 2218-1997 (Sept. 2020).

22. M. Agostini, G. Benato, J. A. Detwiler, J. Menéndez, F. Vissani, *Rev. Mod. Phys.* **95**, Publisher: American Physical Society, 025002 (May 2023).

23. H. Murayama, *Phys. World* **15**, 35–39, ISSN: 0953-8585, 2058-7058 (May 2002).

24. B. J. P. Jones, *The Physics of Neutrinoless Double Beta Decay: A Primer*, arXiv:2108.09364 [hep-ph, physics:nucl-ex, physics:nucl-th], Feb. 2022.

25. F. T. Avignone, S. R. Elliott, J. Engel, *Rev. Mod. Phys.* **80**, Publisher: American Physical Society, 481–516 (Apr. 2008).

26. G. Altarelli, D. Meloni, en, *J. High Energ. Phys.* **2013**, 21, ISSN: 1029-8479 (Aug. 2013).

27. A. D. Sakharov, en, *Sov. Phys. Usp.* **34**, Publisher: IOP Publishing, 392, ISSN: 0038-5670 (May 1991).

28. S. Dell'Oro, S. Marcocci, M. Viel, F. Vissani, en, *Advances in High Energy Physics* **2016**, Publisher: Hindawi, e2162659, ISSN: 1687-7357 (Apr. 2016).

29. W. H. Furry, *Phys. Rev.* **56**, Publisher: American Physical Society, 1184–1193 (Dec. 1939).

30. W. H. Furry, *Phys. Rev.* **54**, Publisher: American Physical Society, 56–67 (July 1938).

31. M. Goeppert-Mayer, *Phys. Rev.* **48**, Publisher: American Physical Society, 512–516 (Sept. 1935).

32. S. R. Elliott, A. A. Hahn, M. K. Moe, *Phys. Rev. Lett.* **59**, Publisher: American Physical Society, 2020–2023 (Nov. 1987).

33. J. Schechter, J. W. F. Valle, *Phys. Rev. D* **22**, Publisher: American Physical Society, 2227–2235 (Nov. 1980).

34. M. J. Dolinski, A. W. Poon, W. Rodejohann, *Annual Review of Nuclear and Particle Science* **69**, _eprint: https://doi.org/10.1146/annurev-nucl-101918-023407, 219–251 (2019).

35. J. Engel, J. Menéndez, *Rep. Prog. Phys.* **80**, 046301, ISSN: 0034-4885, 1361-6633 (Apr. 2017).

36. V. Cirigliano *et al.*, *Phys. Rev. Lett.* **120**, Publisher: American Physical Society, 202001 (May 2018).

37. A. Belley *et al.*, *Phys. Rev. Lett.* **132**, Publisher: American Physical Society, 182502 (Apr. 2024).

38. J. Torres, *LobsterPlot*, original-date: 2021-10-14T19:33:41Z, Aug. 2025.

39. K. J. Kelly, P. A. N. Machado, I. Martinez-Soler, S. J. Parke, Y. F. Perez-Gonzalez, *Phys. Rev. Lett.* **123**, Publisher: American Physical Society, 081801 (Aug. 2019).

40. A. Ashtari Esfahani *et al.*, en, *Phys. Rev. Lett.* **131**, 102502, ISSN: 0031-9007, 1079-7114 (Sept. 2023).

41. W. Elbers *et al.*, en, *Phys. Rev. D* **112**, 083513, ISSN: 2470-0010, 2470-0029 (Oct. 2025).

42. The NOvA Collaboration *et al.*, en, *Nature* **646**, 818–824, ISSN: 0028-0836, 1476-4687 (Oct. 2025).

43. A. Abusleme *et al.*, *Chinese Phys. C* **49**, 033104, ISSN: 1674-1137, 2058-6132 (Mar. 2025).

44. CUORE Collaboration *et al.*, *Phys. Rev. Lett.* **129**, Publisher: American Physical Society, 222501 (Nov. 2022).

45. G. Anton *et al.*, en, *Phys. Rev. Lett.* **123**, 161802, ISSN: 0031-9007, 1079-7114 (Oct. 2019).

46. X. Ni, presented at the 2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD), pp. 1–1, ISBN: 9798350338669.

47. O. Azzolini *et al.*, *Phys. Rev. Lett.* **129**, Publisher: American Physical Society, 111801 (Sept. 2022).

48. AMoRE Collaboration *et al.*, *Phys. Rev. Lett.* **134**, Publisher: American Physical Society, 082501 (Feb. 2025).

49. J. H. Fremlin, M. C. Walters, en, *Proc. Phys. Soc. A* **65**, 911, ISSN: 0370-1298 (Nov. 1952).

50. E. Fiorini, A. Pullia, G. Bertolini, F. Cappellani, G. Restelli, *Physics Letters B* **25**, 602–603, ISSN: 0370-2693 (Nov. 1967).

51. F. T. Avignone *et al.*, *Physics Letters B* **256**, 559–561, ISSN: 0370-2693 (Mar. 1991).

52. A. A. VASENKO *et al.*, en, *Modern Physics Letters A*, Publisher: World Scientific Publishing Company (Nov. 2011).

53. H. S. Miley, *Phys. Rev. Lett.* **65**, 3092–3095 (1990).

54. M. Berglund, M. E. Wieser, de, *Pure and Applied Chemistry* **83**, Publisher: De Gruyter, 397–410, ISSN: 1365-3075 (Jan. 2011).

55. Majorana Collaboration *et al.*, *Phys. Rev. Lett.* **130**, Publisher: American Physical Society, 062501 (Feb. 2023).

56. GERDA Collaboration *et al.*, *Phys. Rev. Lett.* **125**, Publisher: American Physical Society, 252502 (Dec. 2020).

57. G. F. Knoll, *Radiation Detection and Measurement*, en, Google-Books-ID: HKBVAAAAMAAJ (Wiley, Jan. 2000), ISBN: 978-0-471-07338-3.

58. S. M. Sze, Y. Li, K. K. Ng, *Physics of Semiconductor Devices*, en, Google-Books-ID: svYkEAAAQBAJ (John Wiley & Sons, Mar. 2021), ISBN: 978-1-119-61800-3.

59. I. J. Arnquist *et al.*, en, *Phys. Rev. C* **107**, 045503, ISSN: 2469-9985, 2469-9993 (Apr. 2023).

60. C. Kittel, P. McEuen, *Introduction to Solid State Physics*, en, Google-Books-ID: rAMujwEACAAJ (John Wiley & Sons, 2005), ISBN: 978-81-265-3518-7.

61. S. Richard, F. Aniel, G. Fishman, en, *Phys. Rev. B* **70**, 235204, ISSN: 1098-0121, 1550-235X (Dec. 2004).

62. S. H. Simon, *The Oxford Solid State Basics*, en, Google-Books-ID: lSNoAgAAQBAJ (OUP Oxford, June 2013), ISBN: 978-0-19-150210-1.

63. G. Wang *et al.*, en, *Materials Science in Semiconductor Processing* **39**, 54–60, ISSN: 13698001 (Nov. 2015).

64. P. S. Barbeau, J. I. Collar, O. Tench, *J. Cosmol. Astropart. Phys.* **2007**, 009–009, ISSN: 1475-7516 (Sept. 2007).

65. H. Spieler, *Semiconductor Detector Systems*, en, Google-Books-ID: yy_ymbccPL8C (OUP Oxford, Aug. 2005), ISBN: 978-0-19-152365-6.

66. W. Shockley, *J. Appl. Phys.* **9**, 635–636, ISSN: 0021-8979 (Oct. 1938).

67. D. C. Radford, *siggen*.

68. I. Abt *et al.*, en, *J. Inst.* **16**, Publisher: IOP Publishing, P08007, ISSN: 1748-0221 (Aug. 2021).

69. D. Butta *et al.*, *IEEE Trans. Nucl. Sci.* **71**, 2171–2179, ISSN: 0018-9499, 1558-1578 (Sept. 2024).

70. V. D'Andrea *et al.*, "LEGEND Note 24-004 Analysis Methods and Procedures for Energy Scale", LEGEND internal document, Apr. 2025.

71. I. Arnquist *et al.*, *J. Inst.* **18**, P09023, ISSN: 1748-0221 (Sept. 2023).

72. U. Fano, en, *Phys. Rev.* **72**, 26–29, ISSN: 0031-899X (July 1947).

73. R. C. Alig, S. Bloom, en, *Phys. Rev. Lett.* **35**, 1522–1525, ISSN: 0031-9007 (Dec. 1975).

74. F. Goulding, D. Landis, *IEEE Transactions on Nuclear Science* **41**, 1145–1149, ISSN: 1558-1578 (Aug. 1994).

75. A. Owens, A. Peacock, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Proceedings of the 5th International Workshop on Radiation Imaging Detectors **531**, 18–37, ISSN: 0168-9002 (Sept. 2004).

76. F. S. Goulding *et al.*, *IEEE Transactions on Nuclear Science* **31**, 285–299, ISSN: 1558-1578 (Feb. 1984).

77. J. Roth, J. H. Primbsch, R. P. Lin, *IEEE Transactions on Nuclear Science* **31**, 367–371, ISSN: 1558-1578 (Feb. 1984).

78. F. Petry *et al.*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **332**, 107–112, ISSN: 0168-9002 (July 1993).

79. C. E. Aalseth, F. T. Avignone, R. L. Brodzinski, H. S. Miley, J. H. Reeves, en, *J Radioanal Nucl Chem* **233**, 119–125, ISSN: 1588-2780 (July 1998).

80. J. Hellmig, H. V. Klapdor-Kleingrothaus, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **455**, 638–644, ISSN: 0168-9002 (Dec. 2000).

81. D. González *et al.*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **515**, 634–643, ISSN: 0168-9002 (Dec. 2003).

82. T. Comellato, M. Agostini, S. Schönert, en, *Eur. Phys. J. C* **83**, 236, ISSN: 1434-6052 (Mar. 2023).

83. P. Luke, F. Goulding, N. Madden, R. Pehl, *IEEE Transactions on Nuclear Science* **36**, 926–930, ISSN: 1558-1578 (Feb. 1989).

84. S. I. Alvis *et al.*, en, *Phys. Rev. C* **99**, 065501, ISSN: 2469-9985, 2469-9993 (June 2019).

85. M. Agostini *et al.*, en, *Eur. Phys. J. C* **79**, 978, ISSN: 1434-6052 (Nov. 2019).

86. I. J. Arnquist *et al.*, en, *Eur. Phys. J. C* **82**, 226, ISSN: 1434-6052 (Mar. 2022).

87. R. J. Cooper, D. C. Radford, P. A. Hausladen, K. Lagergren, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **665**, 25–32, ISSN: 0168-9002 (Feb. 2011).

88. N. Abgrall *et al.*, presented at the, p. 020027.

89. LEGEND Collaboration *et al.*, *LEGEND-1000 Preconceptual Design Report*, Version Number: 1, 2021.

90. D. Budjáš, M. B. Heider, O. Chkvorets, N. Khanbekov, S. Schönert, en, *J. Inst.* **4**, P10007, ISSN: 1748-0221 (Oct. 2009).

91. M. Agostini *et al.*, en, *Eur. Phys. J. C* **73**, 2583, ISSN: 1434-6052 (Oct. 2013).

92. M. Agostini *et al.*, en, *J. Inst.* **6**, P03005, ISSN: 1748-0221 (Mar. 2011).

93. J. Gruszko, o. b. o. t. M. Collaboration, en, *J. Phys.: Conf. Ser.* **888**, Publisher: IOP Publishing, 012079, ISSN: 1742-6596 (Sept. 2017).

94. E. R. Engelhardt, en, PhD thesis, The University of North Carolina at Chapel Hill University Libraries.

95. K. Freund *et al.*, en, *Eur. Phys. J. C* **76**, 298, ISSN: 1434-6052 (May 2016).

96. E. Segreto, en, *Phys. Rev. D* **103**, 043001, ISSN: 2470-0010, 2470-0029 (Feb. 2021).

97. T. Heindl *et al.*, *EPL* **91**, 62002, ISSN: 0295-5075, 1286-4854 (Sept. 2010).

98. C. Vogl *et al.*, *J. Inst.* **17**, C01031, ISSN: 1748-0221 (Jan. 2022).

99. I. Abritta Costa *et al.*, *J. Inst.* **18**, P09007, ISSN: 1748-0221 (Sept. 2023).

100. G. R. Araujo *et al.*, en, *Eur. Phys. J. C* **82**, 442, ISSN: 1434-6052 (May 2022).

101. S. Calgaro *et al.*, "LEGEND Note 24-010 Data Partitioning and Exposures for the First LEGEND-200 Unblinding", LEGEND internal document, Feb. 2025.

102. P. Agnes *et al.*, en, *Eur. Phys. J. C* **81**, 359, ISSN: 1434-6044, 1434-6052 (Apr. 2021).

103. B. Bos, en, PhD thesis, The University of North Carolina at Chapel Hill University Libraries, 2025.

104. G. G. Marshall, eng, Publication Title: Doctoral thesis, UCL (University College London)., Doctoral, UCL (University College London), Jan. 2025.

105. S. J. Borden *et al.*, "LEGEND Note 24-005 Data Cleaning and Quality Procedures for Ge Detectors", LEGEND internal document, Jan. 2025.

106. S. Diebold *et al.*, presented at the Optics for EUV, X-Ray, and Gamma-Ray Astronomy VIII, ed. by G. Pareschi, S. L. O'Dell, p. 63, ISBN: 978-1-5106-1255-6 978-1-5106-1256-3.

107. F. Mölder *et al.*, eng, *F1000Res* **10**, 33, ISSN: 2046-1402 (2021).

108. F. Henkes *et al.*, "LEGEND Note 25-002 Juleana Methods and Procedures for the First LEGEND-200 Unblinding", LEGEND internal document, July 2025.

109. L. Pertoldi *et al.*, *legend-daq2lh5*, July 2025.

110. J. Detwiler *et al.*, *legend-pydataobj*, Aug. 2025.

111. M. Agostini *et al.*, *pygama*, Sept. 2025.

112. I. Guinn *et al.*, *dspeed*, Sept. 2025.

113. N. W. Ruof, English, ISBN: 9798379410407, Ph.D. University of Washington, 2023.

114. G. Grünauer, LEGEND Collaboration, en, presented at the Proceedings of XVIII International Conference on Topics in Astroparticle and Underground Physics — PoS(TAUP2023), p. 261.

115. M. Agostini *et al.*, en, *Eur. Phys. J. C* **81**, 682, ISSN: 1434-6044, 1434-6052 (Aug. 2021).

116. W. S. Quinn *et al.*, "LEGEND Note 24-011 Data Cleaning and Quality: HPGe Quality Cuts and Efficiencies", LEGEND internal document, Jan. 2025.

117. M. Agostini *et al.*, en, *Eur. Phys. J. C* **82**, 284, ISSN: 1434-6052 (Apr. 2022).

118. A. Hostiuc, en (Aug. 2023).

119. I. Guinn *et al.*, "L-Note 24-013, Pulse Shape Discrimination for the 2024 Unblinded Analysis of LEGEND-200", LEGEND internal document, Feb. 2025.

120. T. Comellato, M. Agostini, S. Schönert, en, *Eur. Phys. J. C* **81**, 76, ISSN: 1434-6044, 1434-6052 (Jan. 2021).

121. T. Dixon, R. Gala, M. P. Green, I. S. Guinn, L. Pertoldi, "LEGEND Note 24-008 Active mass and $0\nu\beta\beta$ decay containment efficiency", LEGEND internal document, Feb. 2025.

122. S. Calgaro *et al.*, "L-Note 24-007 Background analysis for LEGEND-200 data release at Neutrino 2024", LEGEND internal document, Feb. 2025.

123. K. von Sturm, *LEGEND-200: First glance at the background in physics data*, University of Vienna, Aug. 2023.

124. J. A. Detwiler, M. Agostini, "L-Note 24-006, Blinding Scheme and Unblinding Procedure for the First $0\nu\beta\beta$ Decay Search with LEGEND-200", LEGEND internal document, Feb. 2025.

125. GERDA Collaboration *et al.*, *Phys. Rev. Lett.* **120**, Publisher: American Physical Society, 132503 (Mar. 2018).

126. P. H. Garthwaite, I. T. Jolliffe, B. Jones, *Statistical Inference*, en (Oxford University Press, 2002), ISBN: 978-0-19-857226-8.

127. D. R. Cox, *Principles of Statistical Inference*, en, Google-Books-ID: EMYWpoVn7vcC (Cambridge University Press, Aug. 2006), ISBN: 978-0-521-68567-2.

128. S. Calgaro, eng, Accepted: 2025-01-11T18:58:24Z Publisher: Università degli studi di Padova (Dec. 2024).

129. R. D. Cousins, en, *Am. J. Phys.* **63**, Publisher: AIP Publishing, 398–410, ISSN: 0002-9505 (May 1995).

130. M. G. Kendall, J. K. Ord, S. F. Arnold, *Kendall's Advanced Theory of Statistics*, en, Google-Books-ID: XG5tjwEACAAJ (Edward Arnold, 1994), ISBN: 978-0-470-23412-9.

131. F. James, M. Roos, en, *Computer Physics Communications* **10**, 343–367, ISSN: 00104655 (Dec. 1975).

132. J. D. Perezgonzalez, *Front Psychol* **6**, 223, ISSN: 1664-1078 (Mar. 2015).

133. O. Behnke, K. Kröninger, G. Schott, T. Schörner-Sadenius, *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*, en, Google-Books-ID: U77IAgAAQBAJ (John Wiley & Sons, Aug. 2013), ISBN: 978-3-527-41058-3.

134. G. Cowan, K. Cranmer, E. Gross, O. Vitells, en, *Eur. Phys. J. C* **71**, 1554, ISSN: 1434-6052 (Feb. 2011).

135. S. D. Biller, S. M. Oser, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **774**, 103–119, ISSN: 0168-9002 (Feb. 2015).

136. G. J. Feldman, R. D. Cousins, *Phys. Rev. D* **57**, Publisher: American Physical Society, 3873–3889 (Apr. 1998).

137. S. S. Wilks, *The Annals of Mathematical Statistics* **9**, Publisher: Institute of Mathematical Statistics, 60–62, ISSN: 0003-4851, 2168-8990 (Mar. 1938).

138. P. R. Bevington, D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, en, Google-Books-ID: JYaZPwAACAAJ (McGraw-Hill, 2003), ISBN: 978-0-07-119926-1.

139. L. Heinrich, M. Feickert, G. Stark, *scikit-hep/pyhf: v0.7.6*, Jan. 2024.

140. L. Varriano *et al.*, *cenpa/freqfit: v0.3*, Apr. 2025.

141. H. Dembinski *et al.*, *scikit-hep/iminuit*, Apr. 2025.

142. S. K. Lam *et al.*, *numba/numba: Numba 0.61.2*, Apr. 2025.

143. A. statistics forum, "Frequentist Limit Recommendation", ATLAS unpublished, Feb. 2011.

144. R. D. Cousins, presented at the, ADS Bibcode: 2006sppp.conf...75C, p. 75.

145. R. D. Cousins, *Phys. Rev. D* **62**, Publisher: American Physical Society, 098301 (Sept. 2000).

146. R. D. Cousins, L. Wasserman, *PHYSTAT Informal Review: Marginalizing versus Profiling of Nuisance Parameters*, arXiv:2404.17180 [physics], Apr. 2024.

147. L. Demortier, *arXiv: Data Analysis, Statistics and Probability* (Dec. 2003).

148. Majorana Collaboration *et al.*, *Phys. Rev. C* **100**, Publisher: American Physical Society, 025501 (Aug. 2019).

149. W. Xu, L. Hehn, X. Zhu, S. R. Elliott, "Statistical Analysis for Setting Limit on the $\beta\beta(0\nu)$ Decay for the Majorana Demonstrator, Unidoc #M-TECHDOCUNI-2017-024", June 2018.

150. DOE/NSF Nuclear Science Advisory Committee & Long Range Plan Working Group, N. None, "A New Era of Discovery: The 2023 Long Range Plan for Nuclear Science", tech. rep. None, 2280968, None, 2280968.

151. C. F. Jiao, J. Engel, J. D. Holt, en, *Phys. Rev. C* **96**, 054310, ISSN: 2469-9985, 2469-9993 (Nov. 2017).

152. F. F. Deppisch, L. Graf, F. Iachello, J. Kotila, en, *Phys. Rev. D* **102**, 095016, ISSN: 2470-0010, 2470-0029 (Nov. 2020).

153. D. Baxter *et al.*, en, *Eur. Phys. J. C* **81**, 907, ISSN: 1434-6052 (Oct. 2021).

154. Majorana Collaboration *et al.*, *Phys. Rev. C* **99**, Publisher: American Physical Society, 065501 (June 2019).

155. S. J. Borden *et al.*, "L-Note 25-001 $0\nu\beta\beta$ statistical analysis of the first year of LEGEND-200 data", LEGEND internal document, Feb. 2025.

156. K. Alfonso *et al.*, *Sensitivity of the CUPID experiment to $0\nu\beta\beta$ decay of ^{100}Mo*, Version Number: 1, 2025.

157. S. Algeri, J. Aalbers, K. D. Morå, J. Conrad, en, *Nat Rev Phys* **2**, Publisher: Nature Publishing Group, 245–252, ISSN: 2522-5820 (May 2020).

158. E. Aprile *et al.*, en, *Phys. Rev. D* **84**, 052003, ISSN: 1550-7998, 1550-2368 (Sept. 2011).

159. L. Lista, en, *EPJ Web of Conferences* **55**, ed. by T. Delemontex, A. Lucotte, 03003, ISSN: 2100-014X (2013).

160. S. Borden, J. Detwiler, W. Pettus, N. Ruof, *J. Inst.* **19**, P12014, ISSN: 1748-0221 (Dec. 2024).

161. M. Schwarz *et al.*, *EPJ Web of Conferences* **253**, 11014 (Dec. 2021).

162. J. Janicskó Csáthy, H. Aghaei Khozani, A. Caldwell, X. Liu, B. Majorovits, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654**, 225–232 (Oct. 2011).

163. P. Krause, PhD Thesis, Munich, Tech. U., Munich, Tech. U.

164. M. Kuźniak, A. M. Szelc, *Instruments* **5**, 4 (Dec. 2020).

165. C. Wiesinger, PhD Thesis, Munich, Tech. U., Munich, Tech. U., 2020.

166. M. Agostini *et al.*, *The European Physical Journal C* **83** (Apr. 2023).

167. C. Piemonte, A. Gola, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **926**, 2–15 (May 2019).

168. C. Betancourt, A. Dätwyler, P. Owen, A. Puig, N. Serra, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **958**, 162851 (Apr. 2020).

169. S. Gundacker *et al.*, *Physics in Medicine and Biology* **65**, 025001 (Jan. 2020).

170. M. Nemallapudi, S. Gundacker, P. Lecoq, E. Auffray, *Journal of Instrumentation* **11** (Oct. 2016).

171. S. Gundacker, A. Heering, *Physics in Medicine and Biology* **65** (Aug. 2020).

172. F. Acerbi, S. Gundacker, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Silicon Photomultipliers: Technology, Characterisation and Applications **926**, 16–35, ISSN: 0168-9002 (May 2019).

173. D. Renker, E. Lorenz, *Journal of Instrumentation* **4** (Apr. 2009).

174. R. Klanner, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **926**, 36–56 (May 2019).

175. P. Buzhan *et al.*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **504**, 48–52 (May 2003).

176. K. G. McKay, *Phys. Rev.* **94**, Publisher: American Physical Society, 877–884 (May 1954).

177. W. Oldham, R. Samuelson, P. Antognetti, *IEEE Trans. Electron Devices* **19**, 1056–1060, ISSN: 0018-9383, 1557-9646 (Sept. 1972).

178. M. Capasso *et al.*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **982**, 164478, ISSN: 0168-9002 (Dec. 2020).

179. G. Gallina *et al.*, *IEEE Transactions on Electron Devices* **66**, 4228–4234, ISSN: 1557-9646 (Oct. 2019).

180. K. Morimoto, E. Charbon, *Sensors (Basel)* **21**, 3447, ISSN: 1424-8220 (May 2021).

181. P. Eraerds, M. Legré, A. Rochas, H. Zbinden, N. Gisin, EN, *Opt. Express, OE* **15**, Publisher: Optica Publishing Group, 14539–14549, ISSN: 1094-4087 (Oct. 2007).

182. F. Corsi, M. Foresta, C. Marzocca, G. Matarrese, A. D. Guerra, *J. Inst.* **4**, P03004–P03004, ISSN: 1748-0221 (Mar. 2009).

183. A. Duara *et al.*, en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **979**, 164483, ISSN: 01689002 (Nov. 2020).

184. E. Currás-Rivera *et al.*, *J. Inst.* **20**, C06076, ISSN: 1748-0221 (June 2025).

185. C. R. Crowell, S. M. Sze, *Appl. Phys. Lett.* **9**, 242–244, ISSN: 0003-6951 (Sept. 1966).

186. F. Acerbi, G. Paternoster, S. Merzi, N. Zorzi, A. Gola, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1046**, 167683 (Jan. 2023).

187. G. Collazuol, M. G. Bisogni, S. Marcatili, C. Piemonte, A. Del Guerra, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, VCI 2010 **628**, 389–392, ISSN: 0168-9002 (Feb. 2011).

188. A. Otte *et al.*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **567**, 360–363 (Nov. 2006).

189. A. d. S. Croix *et al.*, *Mapping the Photon Detection Efficiency of VUV Sensitive SiPMs from the Ultra-Violet to the Near Infra-Red*, arXiv:2508.16005 [physics], Aug. 2025.

190. G. Hurkx, D. Klaassen, M. Knuvers, *IEEE Trans. Electron Devices* **39**, 331–338, ISSN: 00189383 (Feb. 1992).

191. R. Pires, R. Dickstein, S. Titcomb, R. Anderson, en, *Cryogenics* **30**, 1064–1068, ISSN: 00112275 (Dec. 1990).

192. M. J. Penn, B. L. Dougherty, B. Cabrera, R. M. Clarke, B. A. Young, en, *Journal of Applied Physics* **79**, 8179–8186, ISSN: 0021-8979, 1089-7550 (June 1996).

193. E. A. Gutierrez-D, M. J. Deen, C. L. Claeys, *Low Temperature Electronics: Physics, Devices, Circuits, and Applications*, en, Google-Books-ID: e677kp7OhzAC (Academic Press, 2001), ISBN: 978-0-12-310675-9.

194. K. I. Nuttall, M. W. Nield, en, *International Journal of Electronics* **24**, 69–78, ISSN: 0020-7217, 1362-3060 (Jan. 1968).

195. J. Ren, G. Nimtz, J. Jakumeit, R. Wollrab, en, *Appl Phys A* **65**, 325–328, ISSN: 1432-0630 (Sept. 1997).

196. M. Hampel *et al.*, en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **976**, 164262, ISSN: 01689002 (Oct. 2020).

197. N. Akil, S. Kerns, D. Kerns, A. Hoffmann, J.-P. Charles, *IEEE Trans. Electron Devices* **46**, 1022–1028, ISSN: 00189383 (May 1999).

198. F. Acerbi *et al.*, *IEEE Transactions on Electron Devices* **64**, 521–526 (Feb. 2017).

199. G. Zappalà *et al.*, *J. Inst.* **11**, P08014–P08014, ISSN: 1748-0221 (Aug. 2016).

200. P. Eckert, H.-C. Schultz-Coulon, W. Shen, R. Stamen, A. Tadday, en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **620**, 217–226, ISSN: 01689002 (Aug. 2010).

201. A. Nagai *et al.*, *J. Inst.* **14**, P12016–P12016, ISSN: 1748-0221 (Dec. 2019).

202. G. Bonanno *et al.*, en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **610**, 93–97, ISSN: 01689002 (Oct. 2009).

203. K. Yamamoto *et al.*, presented at the 2006 IEEE Nuclear Science Symposium Conference Record, pp. 1094–1097, ISBN: 978-1-4244-0560-2.

204. A. Nagai *et al.*, en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **948**, 162796, ISSN: 01689002 (Dec. 2019).

205. F. Nagy *et al.*, en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **759**, 44–49, ISSN: 01689002 (Sept. 2014).

206. A. Ghassemi, K. Sato, K. Kobayashi, *MPPC -Technical guide to silicon photomultipliers (MPPC)*.

207. M. Agostini *et al.*, *The European Physical Journal C* **78** (May 2018).

208. B. Abi *et al.*, *Journal of Instrumentation* **15** (Dec. 2020).

209. P. Agnes *et al.*, *The European Physical Journal C* **81** (Nov. 2021).

210. C. E. Aalseth *et al.*, *The European Physical Journal C* **81** (Feb. 2021).

211. M. Chiappini, *Journal of Instrumentation* **18** (Oct. 2023).

212. A. Falcone *et al.*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **985**, 164648 (Jan. 2021).

213. C. E. Aalseth *et al.*, *The European Physical Journal Plus* **133** (Mar. 2018).

214. I. Kochanek, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **980**, 164487 (Nov. 2020).

215. A. Jamil *et al.*, *IEEE Transactions on Nuclear Science* **65**, 2823–2833 (Oct. 2018).

216. R. Peres, *Journal of Instrumentation* **18** (Mar. 2023).

217. P. Ferrario, *Journal of Instrumentation* **17** (Jan. 2022).

218. R. Álvarez-Garrote *et al.*, *Measurement of the photon detection efficiency of Hamamatsu VUV4 sipms at cryogenic temperature*, Publication Title: arXiv.org, Apr. 2024.

219. P. Nakarmi *et al.*, *Journal of Instrumentation* **15** (Jan. 2020).

220. T. Pershing *et al.*, *Journal of Instrumentation* **17** (Apr. 2022).

221. Thorlabs, *LED560L* (May 2021).

222. M. D'Incecco *et al.*, *IEEE Transactions on Nuclear Science* **65**, 1005–1011 (Apr. 2018).

223. *CAEN CoMPASS Multiparametric DAQ Software for Physics Applications*.

224. A. Gola *et al.*, *Sensors* **19**, 308 (Jan. 2019).

225. F. Acerbi *et al.*, *IEEE Transactions on Nuclear Science* **62**, 1318–1325 (June 2015).

226. I. Tobehn-Steinhäuser *et al.*, *The 8th International Symposium on Sensor Science* (May 2021).

227. M. Guarise *et al.*, *Journal of Instrumentation* **16** (2021).

228. M. Ma *et al.*, *Photo-detection efficiency measurement for FBK HD near-UV sensitive sipms at 10 K temperature*, Publication Title: arXiv.org, Dec. 2023.

229. M. Biroth, P. Achenbach, W. Lauth, A. Thomas, *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)* (Oct. 2016).

230. W. Bludau, A. Onton, W. Heinke, *Journal of Applied Physics* **45**, Publisher: American Institute of Physics, 1846–1848, ISSN: 0021-8979 (Apr. 1974).

231. T. Wagner, U. Rau, presented at the Proceedings of 3rd World Conference onPhotovoltaic Energy Conversion, 2003, vol. 2, 1162–1165 Vol.2.

Chapter A

Appendix One

A.1 Test statistic for toys with no signal-like events

We derive the expected test statistic for a special case, where we have M partitions each with n_i events far from $Q_{\beta\beta}$ and 1 partition with no events. We include nuisance parameters for the energy bias correction, energy resolution, and efficiency scale α. Noting that any number of empty partitions may be combined into a single empty partition, this "special case" therefore describes the vast majority of toys that are performed in the frequentist Monte Carlo simulation.

Let us assume that we have M partitions, each with n_i events that are far enough away from $Q_{\beta\beta}$ in energy so that the Gaussian component of the PDF is 0. Furthermore, let the total number of events in the whole experiment be N so that $N = \sum_i n_i$. We combine all partitions without events into a single empty partition. We will also let the likelihood have Gaussian constraints on σ, Δ, α and that these have nominal values of $\tilde{\sigma}, \tilde{\Delta}, \tilde{\alpha}$. The likelihood is therefore

$$\mathcal{L} = e^{-(s_0+b_0)} \prod_{i=1}^{M} \frac{e^{-(s_i+b_i)}}{n_i!} \frac{(b_i)^{n_i}}{(\Delta E)^{n_i}} e^{-\frac{(\alpha-\tilde{\alpha})^2}{2\delta_\alpha^2}} e^{-\frac{(\sigma_i-\tilde{\sigma}_i)^2}{2\delta_{\sigma_i}^2}} e^{-\frac{(\Delta_i-\tilde{\Delta}_i)^2}{2\delta_{\Delta_i}^2}} \tag{A.1}$$

The negative two log-likelihood (-2LL) is therefore

$$-2\ln(\mathcal{L}) = \frac{(\alpha-\tilde{\alpha})^2}{\delta_\alpha^2} + 2\sum_{i=0}^{M}(s_i+b_i) + 2\sum_{i=1}^{M}\ln(n_i!) - 2\ln\left[\frac{(b_i)^{n_i}}{(\Delta E)^{n_i}}\right] + \frac{(\sigma_i-\tilde{\sigma}_i)^2}{\delta_{\sigma_i}^2} + \frac{(\Delta_i-\tilde{\Delta}_i)^2}{\delta_{\Delta_i}^2} \tag{A.2}$$

If every partition has the same background index, then we can factorize $\mathcal{B}$ out and use the fact that $\sum_{i=1}^{M} \ln(\mathcal{B}^{n_i}) = \ln(\prod_{i=1}^{M} \mathcal{B}^{n_i}) = \ln(\mathcal{B}^{\sum n_i}) = \ln(\mathcal{B}^N)$

$$-2\ln(\mathcal{L}) = \frac{(\alpha-\tilde{\alpha})^2}{\delta_\alpha^2} - 2\ln(\mathcal{B}^N) + 2\sum_{i=0}^{M}(s_i+b_i) + 2\sum_{i=1}^{M}\ln(n_i!) - 2\ln[(\mathcal{E}_i)^{n_i}] + \frac{(\sigma_i-\tilde{\sigma}_i)^2}{\delta_{\sigma_i}^2} + \frac{(\Delta_i-\tilde{\Delta}_i)^2}{\delta_{\Delta_i}^2} \tag{A.3}$$

A.1.1 Best-fit

Because all of the events fall far from $Q_{\beta\beta}$ I leave it without proof that Γ must be 0 in the best-fit. The other parameter values we can find from maximizing the NLL.

$$-2\frac{\partial\ln(\mathcal{L})}{\partial\mathcal{B}} = -2\frac{N}{\mathcal{B}} + 2\Delta E\sum_{i=0}^{M}\mathcal{E}_i = 0 \quad \Longrightarrow \quad \hat{\mathcal{B}} = \frac{N}{\Delta E\sum_{i=0}^{M}\mathcal{E}_i} \tag{A.4}$$

The nuisance parameters σ_i, Δ_i get set to their central value because there is no data to constrain them:

$$-2\frac{\partial\ln(\mathcal{L})}{\partial\Delta_i} = 2\frac{\Delta_i-\tilde{\Delta}_i}{\delta_{\Delta_i}^2} = 0, \qquad -2\frac{\partial\ln(\mathcal{L})}{\partial\sigma_i} = 2\frac{\sigma_i-\tilde{\sigma}_i}{\sigma_{\sigma_i}^2} = 0 \tag{A.5}$$

The best-fit value of the global efficiency scaling is different however

$$-2\frac{\partial\ln(\mathcal{L})}{\partial\alpha} = 2\Gamma\sum_{i=0}^{M}\delta_{\epsilon_i}\mathcal{E}_i + 2\frac{\alpha-\tilde{\alpha}}{\delta_\alpha^2} = 0 \tag{A.6}$$

But, because we have argued that for these events the best-fit for Γ is $\Gamma = 0$, we get that

$$\frac{\alpha-\tilde{\alpha}}{\delta_\alpha^2} = 0 \tag{A.7}$$

The NLL at the global best-fit is therefore

$$-2\ln(\mathcal{L}(\hat{\Gamma},\hat{\mathcal{B}},\hat{\alpha},\hat{\sigma}_i,\hat{\Delta}_i)) = -2\ln(\hat{\mathcal{B}}^N) + 2\sum_{i=0}^{M}\hat{b}_i + 2\sum_{i=1}^{M}\ln(n_i!) - 2\ln[(\mathcal{E}_i)^{n_i}] \tag{A.8}$$

where $\hat{b}_i = \Delta E\mathcal{E}_i\hat{B}$.

A.1.2 Profile at $\Gamma \geq 0$

Nothing in the likelihood changes in this scenario for the background index when Γ is fixed – Eq. A.4 still holds. Similarly, the nuisance parameters σ_i, Δ_i are again fixed to their central value. The critical value of α when Γ is fixed and non-zero is

$$-2\frac{\partial \ln(\mathcal{L})}{\partial \alpha} = 2\Gamma \sum_{i=0}^{M} \delta_{\epsilon_i}\mathcal{E}_i + 2\frac{\alpha - \tilde{\alpha}}{\delta_\alpha^2} = 0 \quad \Longrightarrow \quad \alpha = \tilde{\alpha} - \Gamma\delta_\alpha^2 \sum_{i=0}^{M} \delta_{\epsilon_i}\mathcal{E}_i \tag{A.9}$$

The -2LL is therefore:

$$-2\ln(\mathcal{L}(\Gamma, \hat{\mathcal{B}}, \hat{\alpha})) = -2\ln(\hat{\mathcal{B}}^N) + 2\sum_{i=0}^{M}(\hat{s}_i + \hat{b}_i) + 2\sum_{i=1}^{M}\ln(n^i!) - 2\ln[(\mathcal{E}_i)^{n_i}] + \Gamma^2\delta_\alpha^2\left(\sum_{i=0}^{M}\delta_{\epsilon_i}\mathcal{E}_i\right)^2 \tag{A.10}$$

where $\hat{b}_i = \Delta E \mathcal{E}_i \hat{B}$, $\hat{s}_i = \Gamma(\epsilon_i + \delta_{\epsilon_i}\hat{\alpha})\mathcal{E}_i$, and $\hat{\alpha}$ is given by Eq. A.9

A.1.3 Test statistic

The test-statistic under our definition of $\tilde{t}_\Gamma$ is just the difference between Eqns. A.10 and A.8:

$$\tilde{t}_\Gamma^{\text{bkg.only}} = -2\ln(\mathcal{L}(S, \hat{\mathcal{B}}, \hat{\alpha})) + 2\ln(\mathcal{L}(\hat{S}, \hat{\mathcal{B}}, \hat{\alpha})) \tag{A.11}$$

$$\tilde{t}_\Gamma^{\text{bkg.only}} = 2\sum_{i=0}^{M}\hat{s}_i + \Gamma^2\delta_\alpha^2\left(\sum_{i=0}^{M}\delta_{\epsilon_i}\mathcal{E}_i\right)^2 \tag{A.12}$$

$$\tilde{t}_\Gamma^{\text{bkg.only}} = 2\Gamma\sum_{i=0}^{M}(\epsilon_i + \delta_{\epsilon_i}\hat{\alpha})\mathcal{E}_i + \Gamma^2\delta_\alpha^2\left(\sum_{i=0}^{M}\delta_{\epsilon_i}\mathcal{E}_i\right)^2 \tag{A.13}$$

$$\tilde{t}_\Gamma^{\text{bkg.only}} = 2\Gamma\sum_{i=0}^{M}\epsilon_i\mathcal{E}_i + \delta_{\epsilon_i}\mathcal{E}_i\left(\tilde{\alpha} - \Gamma\delta_\alpha^2\sum_{i=0}^{M}\delta_{\epsilon_i}\mathcal{E}_i\right) + \Gamma^2\delta_\alpha^2\left(\sum_{i=0}^{M}\delta_{\epsilon_i}\mathcal{E}_i\right)^2 \tag{A.14}$$

$$\tilde{t}_\Gamma^{\text{bkg.only}} = 2\Gamma\sum_{i=0}^{M}(\epsilon_i + \delta_{\epsilon_i}\tilde{\alpha})\mathcal{E}_i - \Gamma^2\delta_\alpha^2\left(\sum_{j=0}^{M}\delta_{\epsilon_j}\mathcal{E}_j\right)^2 \tag{A.15}$$

This test-statistic is a helpful diagnostic for how well our statistical tools are working. If we fix all nuisance parameters such that $\hat{\alpha} = \tilde{\alpha} = 0$ and $\delta_\alpha = 0$, we see that the test-statistic reduces to a constant irrespective of the number of background counts

$$\tilde{t}_\Gamma^{\text{bkg.only}} = 2\Gamma \sum_{i=0}^{M} \epsilon_i \mathcal{E}_i, \quad \tilde{\alpha} \to 0 \tag{A.16}$$

This means that, for fixed nuisance parameters, all toys with events far from $Q_{\beta\beta}$ should have a delta distribution in their test-statistic PDFs. The actual distribution for such toys is shown in Figure A.1, showing a narrow peak at the calculated value of the test statistic from Equation A.15. Variations in this distribution at very small values ($\sim 10^{-10} - 10^{-15}$) indicated floating point error accumulation in our toys. By rounding the test statistics at this number of decimals, we recover the expected behavior. Note that the vast majority of toys generated at Γ_0 (the blue distribution in Figure A.1) have their test statistic occur at the same value; this is the location of the median as well as upper parts of the 1 and 2 σ quantiles.

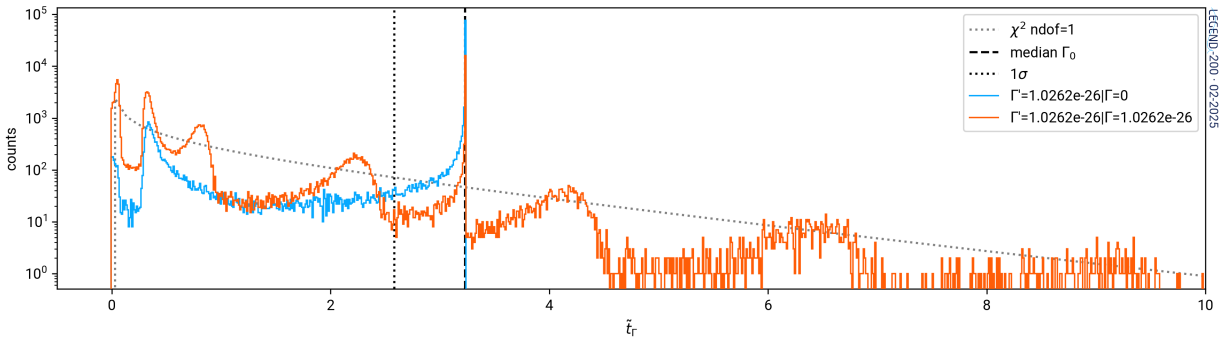

Figure A.1: PDF of the test statistic for LEGEND-200 with all nuisance parameters fixed. The large spike where the median is located is dominated by toys with only background-like events.

When we include α as a fit parameter and also vary its auxiliary measurement in toy generation, we see that it should have a spread correlated to the uncertainty δ_α and acceptance efficiency uncertainty δ_ϵ. This appears as a smearing of the test statistic distribution, as can be seen in Figure A.2. This is because different toys will draw different central values of $\tilde{\alpha}'$ during auxiliary measurement variation, leading to different values of the test-statistic. The same behavior is shown in Figure A.3, where we have indicated which parts of the $p(\tilde{t}_{\Gamma^\star|\Gamma_0})$ distribution are from toys with particular numbers of events that fall in a region $Q_{\beta\beta} \pm 3$ keV.

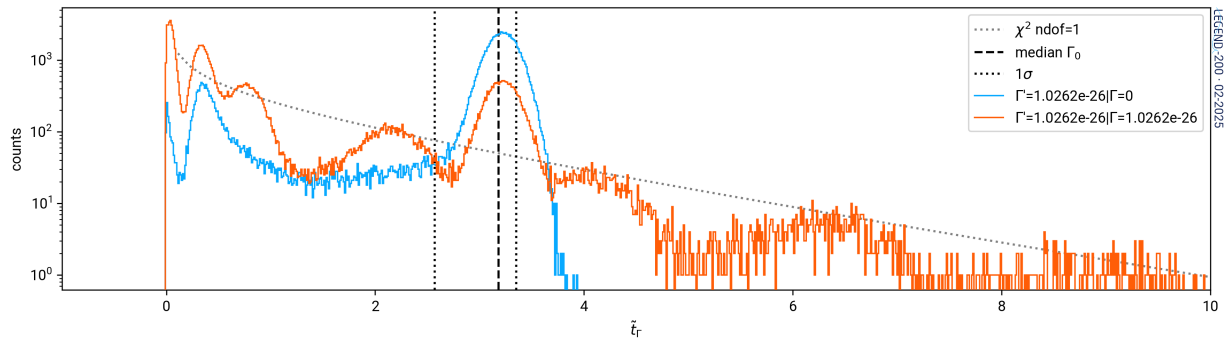

Figure A.2: PDF of the test statistic for LEGEND-200 with all nuisance parameters fixed except α. The large spike dominated by toys with only background-like events is smeared out by a finite width due to uncertainty in α.

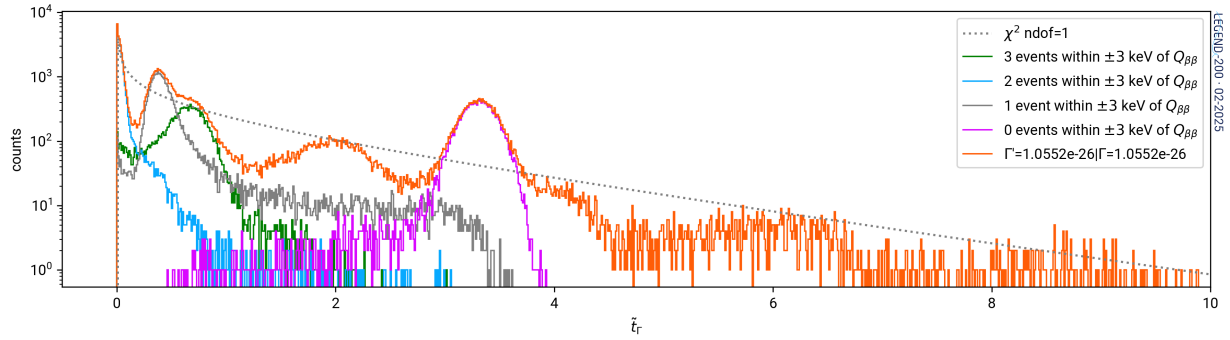

Figure A.3: PDF of the test statistic for LEGEND-200 with all nuisance parameters fixed except α with different populations of toys with events within ± 3 keV of $Q_{\beta\beta}$ marked. The large spike dominated by toys with only background-like events is smeared out by a finite width due to uncertainty in α.

Another interesting feature to fall out of this is that the delta-spike persists when testing the $\Gamma = 0$ hypothesis even when nuisance parameters are allowed to vary. We see in the limit that $\Gamma \to 0$ we have that $\tilde{t}_\Gamma \to 0$. This behavior helps explain why the median p-value on the sensitivity plots approaches 1 when $S = 0$. This behavior can be seen in Figure A.4

Finally, note that, in all cases, the test statistic distributions derived from toys are not well-approximated by the asymptotic approximation, shown in Figures A.1–A.3 as a $\chi^2_{\text{dof}=1}$ distribution. This experiment is in a low statistics regime where drawing toys is quite necessary.

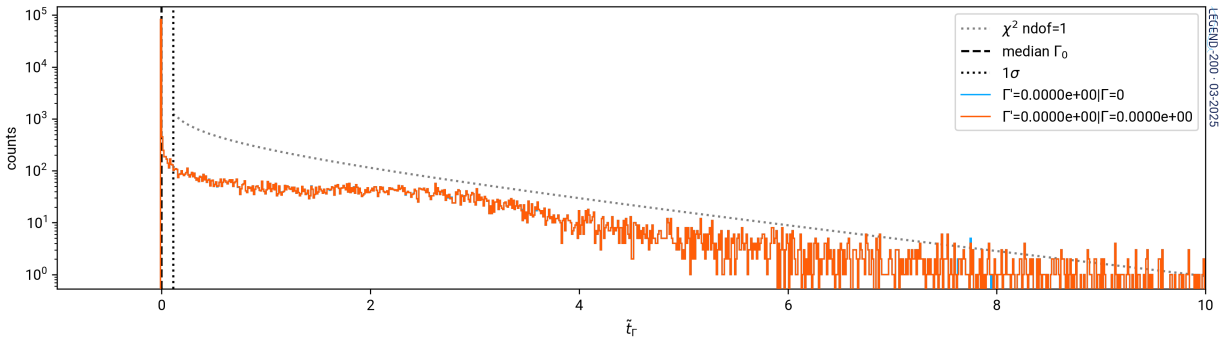

Figure A.4: PDF of the test statistic for LEGEND-200 at $\Gamma = 0$ with all nuisance parameters floating in the fit. The large spike at $\Gamma = 0$ is dominated by toys with only background-like events; however, this spike is not smeared out by the uncertainty in α due to the fact that Eqn. A.15 goes to zero as $\Gamma \to 0$.