Searching for $0\nu\beta\beta$ Decay with CUORE and CUPID

by

Roger Guo Huang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Yury G. Kolomensky, Chair
Professor Kam-Biu Luk
Professor Karl van Bibber

Fall 2021

Searching for $0\nu\beta\beta$ Decay with CUORE and CUPID

Abstract

Searching for $0\nu\beta\beta$ Decay with CUORE and CUPID

by

Roger Guo Huang

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Yury G. Kolomensky, Chair

Since they were first postulated, neutrinos have been one of the most mysterious fundamental particles known to us. The discovery of neutrino oscillation has shown that contrary to our original assumptions, neutrinos are not all massless. This has renewed interest in the idea of Majorana neutrinos as an explanation for the small but nonzero neutrino masses, and the search for neutrinoless double-beta ($0\nu\beta\beta$) decay is currently the most sensitive way to probe this possibility. An observation of this process would constitute the first example of violation of lepton number conservation, demonstrate that neutrinos have a Majorana nature, and help set the scale of their absolute masses. CUORE (Cryogenic Underground Observatory for Rare Events) is one of the leading experiments in the current international program looking for evidence of $0\nu\beta\beta$ decay.

In part I of this dissertation I present a $0\nu\beta\beta$ search based on analysis of CUORE data from its first tonne-year of $^{\text{nat}}TeO_2$ exposure, corresponding to 288.8 kg·yr of ^{130}Te exposure. We observe no evidence of $0\nu\beta\beta$ decay of ^{130}Te and set a Bayesian 90% C.I. lower limit on the corresponding half-life of $T_{1/2}^{0\nu} > 2.2 \times 10^{25}$ years, as well as a Frequentist 90% C.L. lower limit of $T_{1/2}^{0\nu} > 2.6 \times 10^{25}$ years.

As CUORE continues to take data, efforts are already underway to build towards its eventual successor CUPID (CUORE Upgrade with Particle ID). In part II of this dissertation I present the work I have contributed towards the realization of CUPID, including light yield characterization and simulation for TeO_2, analysis efforts for the CUPID-Mo demonstrator, and the development of cryogenic front-end electronics for CUPID.

To my family

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank first and foremost my advisor, Professor Yury Kolomensky, who has taught me along every step of my journey as a graduate student and shown me the kind of physicist I aspire to be. It is with the inspiration of my time at Berkeley that I have chosen to continue pursuing a career in experimental physics.

I would like to acknowledge the other members of my committee, Professors Kam-Biu Luk and Karl van Bibber, for their efforts and valuable feedback on my work, both during the writing of this dissertation and before it.

I would like to acknowledge Professor Gabriel Orebi Gann and, from her group at Berkeley, Javier Caravaca and Benjamin Land, with whom the CHESS measurements were a collaborative effort.

I would like to thank Yuan Mei, who has guided me through my first foray into electronics research, and from whom I have learned many intricacies of electronics development and testing.

I would also like to thank the past and present members of the Kolomensky group at Berkeley for many interesting and useful discussions over lunch, coffee, and elsewhere. The group's people have made it an enjoyable and supportive environment to be part of, for which I am very grateful. In particular, I wish to thank Vivek Singh and Brad Welliver for teaching me much about the black magic of cryogenics, and Benjamin Schmidt and Giovanni Benato for mentoring me through my first research project as a graduate student.

And I would like to thank the other members of the CUORE collaboration, whose efforts have made this work possible and who I have greatly enjoyed working with and talking to over the years, and who I look forward to seeing again when our meetings and conferences are no longer so restricted by the conditions of a pandemic.

The work in this dissertation was done with funding support provided by the UC Berkeley Physics Department, by the U.S. Department of Energy Office of Science, Office of Nuclear Physics under Contract Nos. DE-FG02-08ER4155 and DE-SC0017617, and by the Laboratory Directed Research and Development (LDRD) funding from Lawrence Berkeley National Lab, provided by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Part I

CUORE

Chapter 1

Introduction

How and why does matter in the universe exist? The question is simple to state and is a natural thing for anyone to wonder by just looking around at our world. A common response for laymen is to say the universe originated with the Big Bang, but this doesn't quite answer the question. The high density of energy in the early universe could produce matter and antimatter, but how did we end up in a universe that appears to be made practically entirely of matter, with no large-scale antimatter structures anywhere? The name "antimatter" often evokes a sense of the exotic or abnormal, but as far as we know there's no reason the Earth, our solar system, and our galaxy couldn't have been originally been constructed out of antimatter instead of matter. It somehow just seems to be the case that very slightly more matter was produced than antimatter in the early universe, and as the universe cooled this excess was frozen in while the rest of the matter and antimatter mutually annihilated away. This process is known as baryogenesis, wherein some interactions create a net positive number of baryons, which are the particles that constitute most of the normal matter we see, including protons and neutrons.

Formally speaking, this is the matter-antimatter asymmetry problem, closely tied to the question of whether there is a way to distinguish between matter and antimatter other than just saying they're opposites of each other. The answer to this latter question turns out to be yes, because something known as CP symmetry is violated in Nature's laws, causing matter and antimatter to be treated differently. If CP symmetry were a good symmetry of nature, all particles would behave the same under CP transformation, which is the combined application of the charge conjugation (C) transformation and the parity (P) transformation. Charge conjugation flips all internal quantum numbers (such as positive and negative electric charge), while parity flips the sign of spatial coordinates. Said in plainer terms, CP symmetry is the idea that the laws of Nature are the same if we swap all particles and antiparticles and swap the directions of left and right. In our modern understanding of particle physics under the Standard Model, the vast majority of interactions are CP-symmetric, but there do exist rare processes that violate CP symmetry [1]. This at least assures us that there are indeed nonzero fundamental differences between matter and antimatter, but the known sources of CP violation are far too small to explain the predominance of matter that we see in the universe.

The nature of the neutrino, which is currently not fully understood, may provide a possible answer. A full discussion of the questions posited in this section is beyond the scope of my work, but this dissertation will focus specifically on the phenomenon of double beta decay: why it's interesting, what it can tell us about the neutrino, and how the CUORE and CUPID experiments are studying it. In particular, I will present my contributions to the CUORE experiment and its most recent results, as well as the work I have done working towards the future CUPID experiment, which has not yet finalized its design. In this chapter I will start with an overview of the neutrino's properties and how it may play a role in answering the big questions about the universe.

1.1 The Origins of the Universe

Back in 1967, Andrei Sakharov asked the question: if it's true that our expanding universe first originated from a super-hot super-dense state as we believe, then what conditions must have been satisfied for the universe we now observe to have been the result? He could not think of a mechanism whereby matter and antimatter could have been created in equal amounts and then somehow separated on a macroscopic scale, so he postulated that the modern universe contains only matter. Modern cosmological observations confirm this, as we have mostly excluded the possibility of appreciable amounts of baryonic antimatter by looking for the gamma ray signatures we would expect from their annihilation with typical matter at any boundaries [2]. From this, Sakharov hypothesized that in order to generate our matter-dominated universe, there must have been processes that violated baryon number (B) conservation, charge conjugation (C) symmetry, and CP symmetry, and that these processes must have existed outside of thermal equilibrium as well [3]. These are now known as the Sakharov conditions for matter production.

The first condition, a need for a baryon number violating process, is somewhat obvious - if all processes produce baryons and antibaryons in equal amounts, then there's clearly no way to end up with a net positive amount of matter. The need for violation of C-symmetry and CP-symmetry essentially means that the baryon number violating process should be asymmetric in its treatment of matter and antimatter. This way, it will create matter more often than it will create antimatter, instead of having the matter-creating processes be balanced out by the complementary antimatter-creating processes occurring at the same rate. The last condition, that the process must occur out of thermal equilibrium, is the most subtle of them. It stems from the CPT theorem [4], which states that any Lorentz-invariant, local, unitary quantum field theory must obey CPT symmetry, the combined application of a CP transformation along with a time reversal (T) transformation. CPT has so far been observed to be an exact symmetry of nature, and we mostly expect any theories we develop beyond the Standard Model to still obey the assumptions laid out in the CPT theorem. If we suppose that a baryon-generating process satisfying the other Sakharov conditions also satisfies CPT symmetry, then even though CP symmetry is violated, if the process occurs only in thermal equilibrium then the time-reversed process will annihilate baryons at the same rate that they're being created. Requiring the process to occur out of thermal equilibrium allows the products of the process to get "frozen in" as the universe expanded. It is notable that all three Sakharov conditions are actually satisfied in the Standard Model: baryon number is an anomalous symmetry, C and CP violation exist due to having three generations of quarks and leptons, and out-of-equilibrium interactions occurred as the universe cooled past the electroweak phase transition. However, all three are satisfied far too weakly in the current Standard Model to explain the amount of matter we have in the universe[1].

One interesting possible mechanism for baryogenesis is actually indirect, through leptogenesis first [5]. Leptogenesis is an analogous procedure that creates leptons, such as electrons and neutrinos, instead of baryons. This excess of leptons can then be converted into an excess of baryons through sphaleron processes, which preserve the number of baryons minus leptons (B-L) but can change the total number of baryons and leptons individually [6]. Natural extensions of the Standard Model that still preserve (B-L) number require the addition of right-handed neutrinos, which also provide the mechanism for leptogenesis. Heavy right-handed neutrinos N would be produced in the early universe, but their decays would be out of equilibrium once the universe cooled to the point where there was no longer sufficient energy to create them again. These decays can follow a number of possible schemes but will generally be along the lines of $N \to X\nu$, where X could be something like a Higgs boson or W boson and we get a light neutrino ν in the final state. These decays can be CP violating through the contributions of various higher-order diagrams, allowing for leptogenesis via the creation of more neutrinos than antineutrinos as all of the relic heavy neutrinos decayed away. This chain of heavy right-handed neutrino decays yielding leptogenesis and then in turn yielding baryogenesis through sphaleron processes features the neutrino quite prominently, and so we now turn to discussion of what we know of the neutrino so far and how it is described in the Standard Model.

1.2 Neutrinos and the Standard Model

The existence of the neutrino was first postulated back in 1930 by Wolfgang Pauli to solve a dilemma observed with nuclear β decays. β decay was thought to be a fairly straightforward process, wherein a nucleus was transformed into an isobar by converting a proton into a neutron or vice versa along with the emission of a β particle (an electron or positron). In this scheme, one body (the parent nucleus) decays into two bodies (the daughter nucleus and the β particle), and so basic kinematics tells us that in order to conserve both energy and momentum, both the daughter nucleus and the released β particle should have fixed energies given the total energy of the decay. However, experiments showed that when we measure the

[1]The known sources of CP violation in the quark sector are insufficient, but our current experimental results leave room for additional sources of CP violation in the lepton sector that we may have simply not observed yet. Searching for these additional sources is the subject of a number of ongoing and future experiments.

β energy of many β decays of the same isotope, instead of seeing a peak around the expected β energy from a two-body decay we see that the β energies form a continuous spectrum from near 0 up to the expected energy. This result was mysterious enough that some physicists considered the possibility that energy and momentum weren't actually conserved quantities in this interaction.

To resolve this problem, Pauli hypothesized that there's actually a third particle in the β decay that we simply couldn't detect. If this is the case, then the decay has 3 particles in the final state and the distribution of energy among the decay products is no longer deterministic, which would explain the energy spectrum observed with the β particles. He called this hypothetical particle the *neutrino*, named for the idea that it must be a small neutral particle for it to avoid triggering any of the particle detectors people used. However, while Pauli's hypothesis would allow us to salvage the concepts of energy and momentum conservation, he did wonder if he had committed a sin by making an untestable prediction - if the neutrino existed but was undetectable as he said, then how could we ever test whether his theory was correct? Luckily, the neutrino turned out to actually be detectable, but just with extraordinarily low interaction rates compared to all other fundamental particles we know of. The first direct detection of the neutrino was made in 1956 using a nearby nuclear reactor as the neutrino source [7], marking a success for Pauli's original prediction.

Neutrino interactions were eventually incorporated into the theory of the electroweak interaction, which has become part of the Standard Model. In the Standard Model, neutrinos are massless, electrically neutral leptons coming in 3 flavors paired with the 3 charged leptons: the electron (e), muon (μ), and tau (τ) particles. They are unique among the fundamental particles in the Standard Model in that they interact only via the weak force, as they possess no charge under either the electromagnetic or strong forces. For this reason, they have an exceptionally small but still nonzero interaction rate with normal atoms that we might use in a detector.

The beginning of the end of this picture of the neutrino came late in the 1960s, when Ray Davis put into practice a neutrino detection method that Bruno Pontecorvo and Luis Alvarez had thought of [8], and used it to count electron neutrinos of solar origin, produced as a result of the fusion cycles that power the sun. His experiment used 520 tons of chlorine (contained in the form of C_2Cl_4) placed in an underground tank at the Homestake mine, and it counted neutrino interactions by looking for the conversion of ^{37}Cl into ^{37}Ar induced by the absorption of an incoming neutrino, $^{37}\mathrm{Cl} + \nu_e \rightarrow {}^{37}\mathrm{Ar} + e^-$. The neutrino interactions were counted by periodically flushing all of the argon out of the tank and counting the number of ^{37}Ar decays. Using the neutrino interaction cross section calculations and solar neutrino production rate calculations available at the time, he predicted an average of 2 to 7 solar neutrino captures on the chlorine in his detector per day. This approach was remarkably successful, but with the puzzling result that the total neutrino count was about a third of the expected number from contemporary solar models [9]. Given the potential uncertainties associated with both the theoretical models and the experimental method, this discrepancy was not immediately universally considered a matter of great concern.

If one were to accept the Homestake experiment's results as correct, one possible explanation was the idea of neutrino oscillation, wherein the mass eigenstates in which neutrinos evolve through time are not the same as the flavor eigenstates through which they participate in the weak interaction. Formally speaking, this would mean that the flavor states $\alpha = e, \mu, \tau$ could be written as superpositions of the mass states $i = 1, 2, 3$

$$|\nu_\alpha\rangle = \sum_{i=1}^{3} U^*_{\alpha i} |\nu_i\rangle$$

where U is what is now called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. From a theoretical perspective, a similar phenomenon had already been observed by that time in neutral kaon oscillation, where the mass eigenstates were not the same as the CP eigenstates, and where the phenomenon of weak interactions being able to change quark flavors was eventually described by the analagous Cabibbo-Kobayashi-Maskawa (CKM) matrix. If the process of neutrino oscillation existed, electron neutrinos produced in the sun would oscillate between the three flavors as they made their way out of the sun and approached the Earth. Since the Homestake experiment's detection method was only sensitive to electron-type neutrinos, it would end up detecting only some fraction of the total expected solar neutrino flux.

A combination of results from other large experiments over the next few decades ultimately ended up vindicating Davis and the Homestake detector, showing that their measurement was not in error and was indeed the first observation of the effects of neutrino oscillation. By 1998 the Super-Kamiokande experiment in Japan, measuring Cherenkov light with a total detector volume of 50000 tons of pure water, similarly reported a large deficiency in the total solar neutrino flux compared to the prediction by the standard solar model while also determining that there was no distortion of the expected energy spectrum [10], ruling out many alternative explanations. At the same time, Super-Kamiokande also reported a zenith-angle dependent deficiency of muon-type neutrinos of atmospheric origin [11], which could likewise be explained as a result of neutrino oscillation. On the other side of the world in Canada, the SNO experiment was able to finally resolve the solar neutrino problem by deploying 1000 tons of heavy water (D_2O). The deuterons in the experiment were sensitive to both charged-current ($\nu_e + d \to e^- + p + p$) and neutral-current ($\nu + d \to \nu + p + n$) interactions, but only electron-type solar neutrinos would be able to participate in the charged-current interactions while any flavor of neutrino could induce a neutral-current interaction. The SNO detector was also sensitive to elastic scattering interactions ($\nu + e \to \nu + e$), which any flavor of neutrino can participate in but which has a greatly enhanced rate for ν_e, since they can participate in this interaction with either a Z or W^- boson as the mediator[2]. This allowed them to measure the flux of all different flavors of neutrinos of solar origin at the same time. They ultimately determined that electron-type neutrinos do indeed make up about a third of the total solar neutrino flux, while adding in the observed muon-type and tau-type neutrino fluxes makes up the difference [12], finally putting to rest the question of where the

[2]With only the Z-mediated diagram available to them, the elastic scattering cross-sections for ν_μ and ν_τ are both about 0.15 times the cross-section for ν_e.

missing Homestake neutrinos had gone. Ray Davis won the Nobel Prize in 2002 for his first detection of extraterrestrial neutrinos, and the work done by the Super-Kamiokande and SNO experiments eventually won their leaders the Nobel Prize in 2015 for the discovery of neutrino oscillation[3].

The Modern Understanding of Neutrino Mixing

Our present understanding of neutrinos now accepts that the 3 Standard Model neutrino flavors are in fact superpositions of at least 3 neutrino mass states and can oscillate between flavors as they propagate through space. We parametrize this behavior with 3 angles θ_{ij} which denote the amount of mixing between the mass states and with a CP-violating phase factor δ_{CP}. There are additionally two more CP-violating Majorana phase factors α_1, α_2 which are only relevant if neutrinos possess a Majorana nature. Using the shorthand $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$, the PMNS matrix is commonly expanded as:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \cdot \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \cdot \begin{pmatrix} c_{21} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} e^{i\alpha_1} & 0 & 0 \\ 0 & e^{i\alpha_2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Thanks to a combination of results from neutrino oscillation experiments looking at solar neutrinos, atmospheric neutrinos, nuclear reactor neutrinos, and particle accelerator neutrinos, we have now been able to measure most of the mixing angles θ_{ij} (other than the octant of θ_{23}), as well as the absolute value of the differences between two pairs of mass states. One of the major questions that remains unresolved is the value of δ_{CP}, which indicates the degree of CP violation that occurs in neutrino interactions. The other is the question of the neutrino mass hierarchy, which refers to the problem that we do not yet know the ordering of all 3 neutrino mass states. This is because neutrino oscillations in vacuum are only sensitive to the square of the mass differences. From matter effects in solar neutrino oscillations we have been able to determine that $m_2 > m_1$, but we have not been able to determine the ordering of m_3 relative to them. We refer to the two possibilities as the normal hierarchy ($m_1 < m_2 < m_3$) and the inverted hierarchy ($m_3 < m_1 < m_2$), where the inverted case is so named because the two larger mass states would be almost degenerate[4]. A schematic of these two mass hierarchies along with what we know so far about the neutrino states is

[3]Technically, the SNO result was not proof of neutrino oscillation by itself - they showed that solar neutrinos were somehow changing flavors before they reached the Earth, but other possible explanations still existed. However, when the SNO results were combined with evidence for nuclear reactor antineutrino oscillation from the KamLAND experiment [13], our modern picture of 3 neutrino states oscillating between each other became clear.

[4]There's another possibility referred to as the quasi-degenerate scheme, in which even the lightest neutrino masses is significantly larger than the sizes of the mass splittings. In this case, the 3 neutrino masses would be almost degenerate relative to the magnitudes of their differences, so regardless of the mass ordering it would no longer be appropriate to call them hierarchical. The quasi-degenerate scheme is not yet completely ruled out by experimental limits on the neutrino masses, but we still tend to use the terms "normal hierarchy" and "inverted hierarchy" to refer to the possible mass orderings.

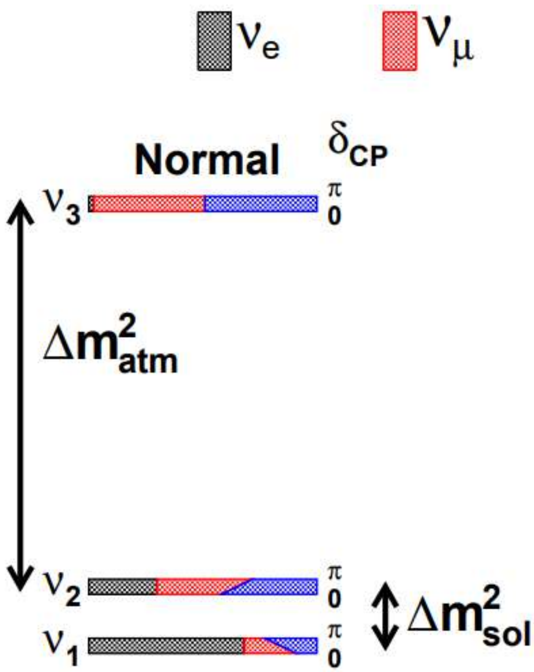


Figure 1.1: Diagram of the information we currently know about neutrino mass splittings and the PMNS matrix parameters under both the normal and inverted mass hierarchy possibilities, which are both compatible with our current measurements of $\Delta m^2_{atm} \approx 2.5\times10^{-3}\text{eV}^2$ and $\Delta m^2_{sol} \approx 7.5 \times 10^{-5}\text{eV}^2$. The colored shadings indicate the flavor composition of each mass eigenstate, with the diagonal borders indicating the dependence on the still-unknown true value of δ_{CP}. Reprinted from [16].

shown in Fig. 1.1. We expect both the δ_{CP} and mass hierarchy questions to be answered within the next 10-20 years, as next-generation experiments like DUNE, Hyper-Kamiokande, and JUNO come online and begin acquiring data [14, 15].

There is also a question of how many light neutrinos there are. We have fairly conclusively determined the number of active light neutrinos to be 3, obtained by precise measurements of Z boson decays assuming lepton universality [17]. However, additional neutrinos can still be squeezed into this picture if they're either too heavy to be a permitted decay channel for the Z boson or don't couple to it at all. The latter case is referred to as the possibility of light sterile neutrinos, and in this scenario the sterile neutrinos could still mix with the 3 active neutrinos, expanding the PMNS matrix beyond its known 3×3 structure. Measurements of the PMNS matrix elements are not yet precise enough to constrain it to 3×3 based on unitarity alone, and a number of anomalies in short-baseline neutrino experiments have hinted at the possibility of a 4th light neutrino mixing with the 3 known ones [18]. However, these anomalies are generally not simultaneously solvable with just the addition of a single

4th neutrino into the oscillation framework, and the evidence for this 4th neutrino from each of them individually is not yet strong given the systematic uncertainties they each face.

So far we have discussed the history of how neutrino oscillation was discovered, as well as the phenomonology of how they are modeled. However, this still leaves open the question of how neutrino oscillation is even possible in the first place. Namely, neutrino oscillation requires that at least 2 of the neutrino mass states be non-zero, or else there couldn't be 3 distinct mass states, but the Standard Model assumed completely massless neutrinos. To account for neutrino masses we must therefore extend the Standard Model.

1.3 Neutrino Masses and the Seesaw Mechanism

All other fundamental fermions in the Standard Model obtain their masses through the Higgs mechanism, whereby the Lagrangian first contains terms respecting SU(2) invariance in the form:

$$-y\bar{L}He_R + h.c.$$

Here, y is the Yukawa coupling serving as a free parameter, H is the Higgs doublet, L is the left-handed SU(2) doublet for the electron generation of leptons, e_R is the electron right-handed SU(2) singlet, and $h.c.$ is the Hermitian conjugate. After spontaneous electroweak symmetry breaking with the Higgs picking up a vacuum expectation value (vev), this turns into mass terms of the form:

$$-m_e(\bar{e}_L e_R + e_L \bar{e}_R)$$

where the mass m_e is determined by the values of the Yukawa coupling and the Higgs vev. Analagous terms exist for the quarks and other charged leptons as well to give them their masses. With the discovery of the Higgs boson at the LHC [19, 20], we now have experimental confirmation that this mechanism is indeed the source of fermion masses in the Standard Model. Neutrinos ended up as massless in this model because this mechanism pairs the left-handed SU(2) doublets with the right-handed SU(2) singlets through the Higgs doublet in order to obtain the mass terms, but right-handed neutrinos do not exist in the Standard Model. However, this was not for any fundamental reason. Since right-handed particles do not participate in the weak interaction and neutrinos do not participate in the electromagnetic and strong interactions, right-handed neutrinos would not interact with any of the forces in the Standard Model at all. For this reason, right-handed neutrinos are also often called sterile neutrinos. Prior to the first hints of neutrino oscillation there was also no evidence that neutrinos had mass, so right-handed neutrinos would have served no purpose in the Standard Model and might as well be left out.

With the present knowledge that nonzero neutrino masses exist, one could add right-handed neutrinos into the Standard Model and allow neutrinos to gain their masses by the same Higgs mechanism that all the other fermions use. This would be a straightforward addition, but it would raise another question: if neutrinos obtain their mass through the same mechanism as the other fermions, then why are neutrino masses so much smaller than

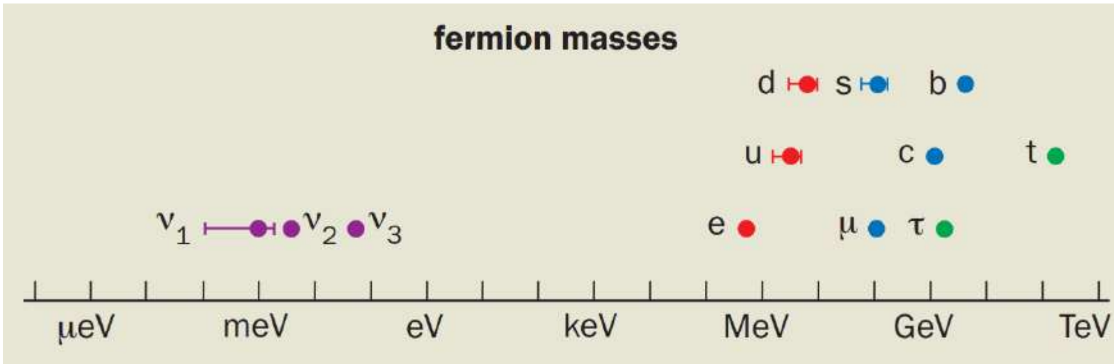


Figure 1.2: Plot of the known fermion masses in the Standard Model. The neutrino masses listed are the values under the normal hierarchy if we assume the masses are of similar to the values of the mass differences obtained from neutrino oscillation data, as the actual values of the neutrino masses have not yet been measured. It can be seen that the scale of the neutrino masses is far below the scale of the rest of the fermions. Reprinted from [21].

the other fermions' masses? While we have not yet measured the actual value of the neutrino masses, if we consider some current reasonable upper limits as shown in Fig. 1.2, we can see that neutrino masses are not merely smaller than the masses of other fermions – the neutrino masses are vastly smaller by several orders of magnitude. Now, there is also no fundamental reason this couldn't still be the case. The masses of the fermions in the Standard Model are determined by the strength of their couplings to the Higgs boson, but these couplings are free parameters entirely determined by experimental results. Correspondingly, there aren't any actual restrictions saying the strength of the neutrino couplings to the Higgs couldn't just happen to be many orders of magnitude smaller than the other fermion couplings[5]. But physicists generally agree that this would be, to use a technical term, very weird. Neutrino masses are on an entirely different scale from the other fermion masses, and so it seems reasonable to think that there is something fundamentally different about the origin of neutrino masses.

As a matter of fact, a possible answer comes from another unique aspect of neutrinos. As the only known fermions neutral to both electric charge and color in the Standard Model, they are the only ones that could be Majorana particles, meaning that they are their own antiparticles. This is as opposed to being Dirac particles like the other fermions, where the particle and antiparticle states are distinct from each other. Majorana right-handed neutrinos would be permitted to have a Majorana mass term of the form $-iM\nu_R^c\nu_R$, which are normally forbidden for particles that carry any kind of conserved quantum number. If

[5]This possibility could be directly tested by measuring Higgs decays to find the strength of its coupling to neutrinos, but given how weak this coupling would have to be, we are currently nowhere near having the ability to do this measurement in a particle accelerator.

we consider the case of just one neutrino flavor for simplicity and add this mass term to the Dirac mass terms obtained through the Higgs mechanism, rewriting ν_L, ν_R in terms of the Weyl spinors ψ_L, ψ_R, we can write down the general neutrino mass terms in the Lagrangian as:

$$\mathcal{L}_{\nu,mass} = -m_D\bar{\psi}_L\psi_R - m_D\bar{\psi}_R\psi_L - M_R\bar{\psi}_R\psi_R = \begin{pmatrix}\bar{\psi}_L & \bar{\psi}_R\end{pmatrix}\begin{pmatrix}0 & m_D \\ m_D & M_R\end{pmatrix}\begin{pmatrix}\psi_L \\ \psi_R\end{pmatrix}$$

This gives a contribution from both the Dirac mass term m_D and the Majorana mass term M_R, and the mass eigenstates are the combinations of ψ_L, ψ_R that diagonalize the matrix $\begin{pmatrix}0 & m_D \\ m_D & M_R\end{pmatrix}$. The straightforward solution then gives that the mass eigenvalues are $\sqrt{m_D^2 + \frac{1}{4}M_R^2} \pm \frac{1}{2}M_R$. In the case that $M_R >> m_D$, the solutions approximate to m_D^2/M_R and M_R. If we suppose the Dirac mass m_D is O(100 GeV), at the electroweak scale, then to obtain a neutrino mass of O(10 meV) we just need the Majorana mass M_R to be O(10^{15} GeV), somewhere around the energy scale of a grand unified theory. Of course, these numbers for m_D and M_R are completely arbitrary, but the point is to show that through the seesaw mechanism we can obtain appropriately small neutrino masses without scale mismatches. This is called the minimal type-I seesaw mechanism, requiring only the addition of right-handed neutrinos that transform as singlets in the Standard Model gauge groups. There exist proposals of other types of seesaw mechanisms to explain neutrino masses as well, for which a discussion can be found in [22].

The addition of these Majorana neutrino mass terms would result in the possibility of lepton number violation, which so far has never been observed[6]. The total number of leptons minus antileptons in the universe would then no longer be preserved, permitting some process of leptogenesis. This makes the type-I seesaw mechanism particularly interesting for its connections to baryogenesis discussed earlier in this chapter. Adding massive right-handed Majorana neutrinos to the Standard Model has the potential to simultaneously explain why the 3 neutrinos we're already familiar with have such small masses and provide a mechanism satisfying the Sakharov conditions to allow for baryogenesis through leptogenesis [23]. From a theoretical standpoint, this is quite appealing, as physicists tend to like ideas where one idea or principle can answer multiple questions at once. The natural next question is then: how can we test it? The masses generally expected of M_R for the type-I seesaw are far beyond the reach of any particle accelerators we have now, though there do exist models with more reachable energies and we can expect other tests to be possible [22, 24]. One of the most promising avenues for an experimental tests, however, turns out to come from nuclear physics instead of high-energy physics. This is the search for neutrinoless double beta decay.

[6]Like baryon number, lepton number is an anomalous symmetry in the Standard Model.

Chapter 2

Double-Beta Decay

The fact that one of the best avenues for investigating the Majorana nature of neutrinos comes from nuclear physics is basically a consequence of the size of Avogadro's number. In the modern era where many searches for new physics involve very rare interactions or require very precise measurements, statistics is important not only to distinguish signal from noise, but to allow for the possibility of having any signal in the experiment in the first place. High-energy experiments using particle accelerators to study rare processes are often limited by their statistics, and even detectors using the abundant natural neutrino flux on Earth are statistically limited by the neutrino's naturally low interaction rate. In comparison, when we study nuclear decays the number of nuclei in even a small amount of macroscopic material will dwarf the number of collisions we can reach with the highest luminosity particle accelerators right now. The comparison of number of nuclei to number of collisions is not exactly apples-to-apples, but it holds true that through studying nuclear decays we are able to probe the Majorana nature of the neutrino much more effectively than any high-energy physics experiment is currently capable of doing.

This chapter will begin with a discussion of the theory behind neutrinoless double beta decay ($0\nu\beta\beta$) and how it is related to the possibility of a Majorana neutrino. I will then discuss the general considerations that go into any experimental search for $0\nu\beta\beta$ decay and summarize the current experimental landscape, along with a discussion of complementary efforts to study neutrino masses and how they relate to the $0\nu\beta\beta$ search effort.

2.1 Theory of $2\nu\beta\beta$ and $0\nu\beta\beta$

Double-beta decay was first postulated by Maria Goeppert-Mayer back in 1935, building upon Fermi's theory of β decay with a modification to simultaneously emit two electrons and two neutrinos [25]. This process, often abbreviated as $2\nu\beta\beta$ decay, corresponds to the nuclear transition:

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

This is a second-order weak process and is thus extremely rare. Goeppert-Mayer's original calculations estimated $2\nu\beta\beta$ half lives in excess of 10^{17} years, a rate that's slow enough that natural radioactive backgrounds will obscure its signal unless an experiment takes special measures to remove or reject these background events. In addition, the rarity of this process means it's only practically observable in even-even nuclei where the single β decay is energetically forbidden or otherwise highly suppressed[1], so that the nucleus can only decay through the much rarer double β decay. This is illustrated in Fig. 2.1, where one can see that for some isobars the nuclear mass does not monotonically converge towards the most stable configuration. A simplified explanation of this phenomenon is that nuclear structures tend to prefer even numbers of protons and neutrons so that they can pair up among themselves, and so on isobars with even atomic number A, some odd-odd configurations will be less stable than the even-even configuration that's further away from the globally favored state. Many decades after it was theorized, $2\nu\beta\beta$ decay was first directly observed in 1987, with a measurement of a half life of O(10^{20} years) in ^{82}Se [26].

Interestingly, the idea of neutrinoless double beta decay ($0\nu\beta\beta$) had already been proposed not long after Goeppert-Mayer's original paper about $2\nu\beta\beta$ decay. After Majorana suggested the possibility of the neutrino being its own antiparticle, people realized that if this were true, then the $2\nu\beta\beta$ decay could be modified to include only the exchange of a virtual neutrino and have no neutrinos in the final state. This would correspond to a violation of lepton number by 2, creating 2 electrons by the process:

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

Under the modern picture, these two versions of double beta decay are represented by the Feynman diagrams in Fig. 2.2. Early proposals of $0\nu\beta\beta$ decay actually suggested it could have a relatively high rate, as the momentum space accessible by the virtual neutrino enhances the decay rate significantly [28]. This possibility was discarded after the discovery that the weak interaction was maximally parity violating, meaning that the $0\nu\beta\beta$ diagram in Fig. 2.2 was impossible since the neutrino emitted from one W vertex would have the wrong helicity to be captured by the other W vertex.

With the discovery that neutrinos have nonzero mass, we now know that helicity and chirality are not equivalent for neutrinos, and so $0\nu\beta\beta$ decay can indeed be facilitated with the exchange of a virtual light Majorana neutrino as depicted in Fig. 2.2. However, instead of being enhanced by the virtual exchange, it is instead suppressed by the requirement that the virtual neutrino undergoes a helicity flip in order to interact through a left-handed current at both W^- vertices. For neutrino masses that are very small compared to typical energies involved in the decay, the resulting amplitude is then proportional to this small m. Since the nuclear decays under consideration have energies on the MeV scale, this condition is easily

[1] A notable example of the second option is ^{48}Ca, for which single β decay is energetically permitted but is a fourth to sixth order forbidden transition. This means that the single β decay requires the neutrino-electron pair to carry away 4 to 6 units of orbital angular momentum, suppressing its rate so heavily that $2\nu\beta\beta$ decay ends up being faster.

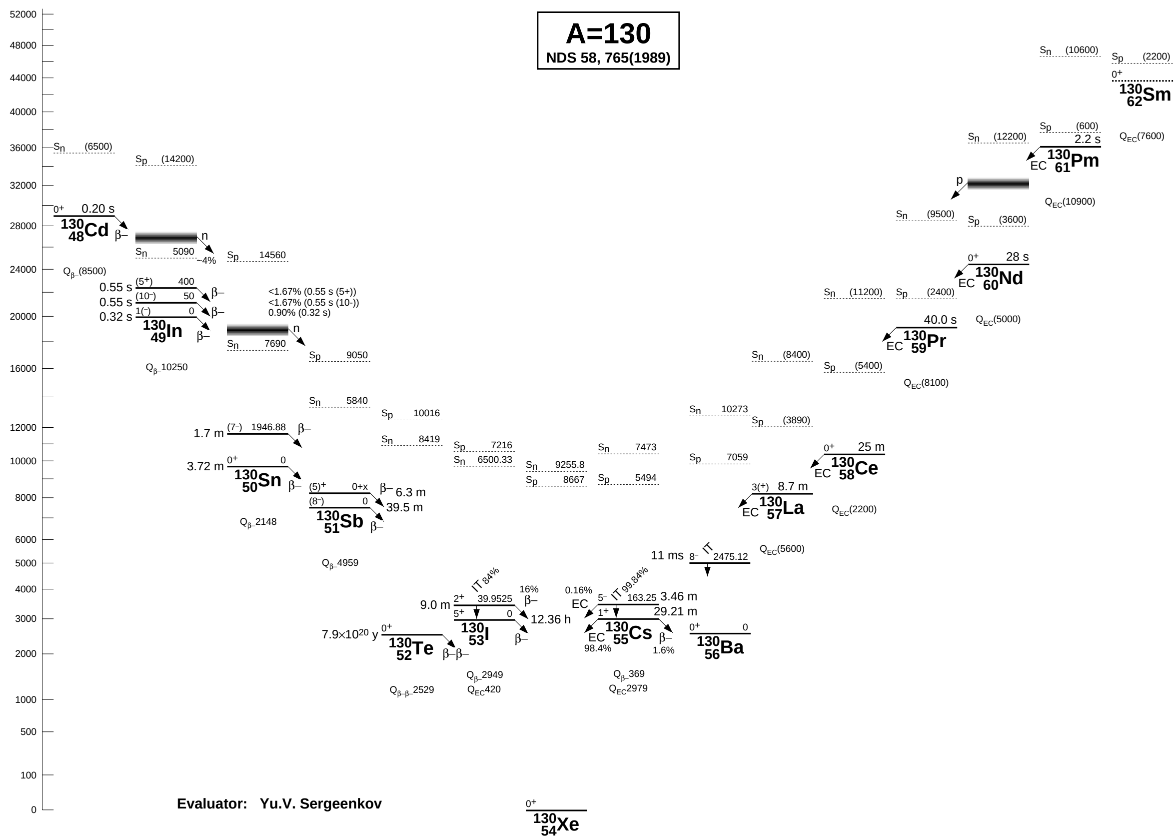


Figure 2.1: Stability curve for the A=130 isobar, with the nuclear mass excess relative to the most stable configuration on the y-axis and the atomic number on the x-axis. One can see that the even-even nuclei are relatively more stable and odd-odd nuclei are relatively less stable. At the points closest to the stable nucleus ^{130}Xe for this isobar, this difference becomes extreme enough that ^{130}Te is energetically forbidden from single β decaying to ^{130}I, but it is energetically allowed to skip it and double β decay directly to ^{130}Xe. Figure from [27].

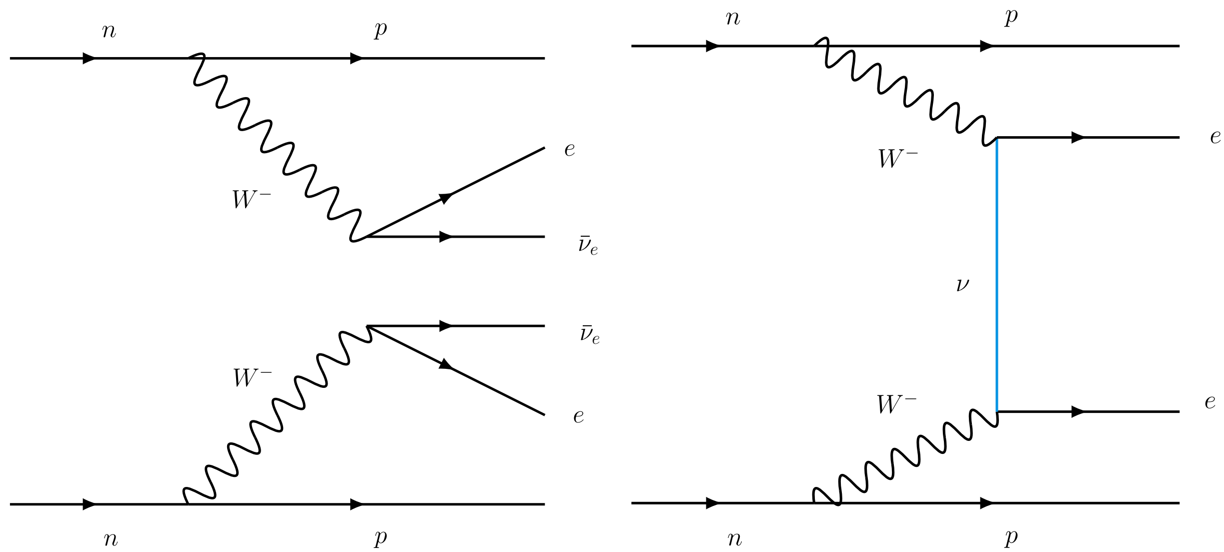


Figure 2.2: Feynman diagrams for $2\nu\beta\beta$ decay (left) and $0\nu\beta\beta$ decay (right). The $2\nu\beta\beta$ decay is a Standard Model process, and the diagram shown here for $0\nu\beta\beta$ decay is assuming it is mediated by the exchange of a virtual light Majorana neutrino.

satisfied by current limits on the light neutrino masses. Now we have to note that the m of the propagator could actually be any of the 3 light neutrino masses, and so the decay rate is actually proportional to the effective Majorana mass $m_{\beta\beta}$, defined as:

$$m_{\beta\beta} = \left|\sum_i U_{ei}^2 m_i\right| = \left|m_1|U_{e1}|^2 + m_2|U_{e2}|^2 e^{i(\alpha_2-\alpha_1)} + m_3|U_{e3}|^2 e^{i(-\alpha_1-2\delta_{CP})}\right|$$

This effective mass sums over the light neutrino mass states m_i because our knowledge of neutrino mixing tells us that even though the virtual propagator interacts with the W^- vertices in the electron flavor, it can propagate as any of the mass states. We thus pick up factors of the PMNS matrix elements U_{ei}^2 from the mixing of each mass state with the electron flavor eigenstate, with the squared factor arising from the fact that the virtual neutrino interacts with two W^-, e^- vertices. It is notable that this results in a sum over U_{ei}^2 instead of something of the form $|U_{ij}|^2$, which is what shows up in neutrino oscillation probability calculations. This means that unlike oscillation experiments, $m_{\beta\beta}$ is sensitive to the Majorana phases α_1, α_2 in the PMNS matrix, and there are effectively no experimental constraints on possible values of these phases at the moment. Curiously, given what we currently know of the neutrino masses and PMNS parameters, if neutrinos follow the normal mass hierarchy there exist values of α_1, α_2 that could cause $m_{\beta\beta}$ to vanish. If this turns out to be the case, there would need to be some other new physics mediating $0\nu\beta\beta$ decay for it to occur. However, there's a very small range of values that would result in this scenario.

In the absence of any reason to believe that the phases should conspire to suppress $m_{\beta\beta}$ so heavily, we proceed for now with the assumption that nature is not so randomly cruel.

The diagram in Fig. 2.2 also hides much of the complexity involved in the nuclear decay. Since we're dealing with an entire nucleus and not just lone nucleons, there are various effects from nuclear physics that must be accounted for as well. These effects are all captured in what we call the nuclear matrix elements of the decay, denoted as:

$$M_{0\nu} = M_{0\nu}^{GT} - \frac{g_V^2}{g_A^2} M_{0\nu}^{F} + M_{0\nu}^{T}$$

Listed from largest to smallest contributions, the 3 components are from the possibilities of Gamow-Teller (GT) decay, Fermi (F) decay, and tensor (T) decay, where the last one is often left out due to its relative insignificance. Due to the complexity of nuclear many-body systems, there are significant theoretical uncertainties associated with the calculations of these nuclear matrix elements, a summary of which can be found in [29]. Different models make different sorts of approximations to perform this calculation; some prominent models at the moment include the shell model (SM), quasiparticle random phase approximation (QRPA) approach, energy-density functional theory (EDF), and the interacting boson model (IBM). In the absence of a strong reason to believe any particular model will be more accurate than the others in calculating $M_{0\nu}$ at the moment, this results in a spread of a factor of 2-3 in the resulting nuclear matrix element calculations, as can be seen in Fig. 2.3. There is also uncertainty associated with "quenching" of the axial vector coupling g_A. This refers to the fact that predictions of Gamow-Teller transition strength for β and $2\nu\beta\beta$ decays tend to systematically overshoot the experimentally measured values for heavier nuclei. The predictions line up better with the measured values for these heavier nuclei if the Gamow-Teller operator is multiplied by a coefficient < 1, or if equivalently the g_A that shows up in these rates is replaced with a smaller g_A^{eff}. This ad hoc adjustment of the value of g_A is what we call quenching, but we currently have no accepted theoretical explanation for this phenomenon. If it turns out that g_A for the Gamow-Teller contribution to $0\nu\beta\beta$ decay rates is strongly quenched too, the rates would be even lower than expected and would impact the projected sensitivity of $0\nu\beta\beta$ experiments. Work is ongoing in the nuclear theory community to resolve all these questions, but no clear answers lurk on the horizon yet.

Lastly, we have phase space factors $G_{0\nu}(E_0, Z)$ that must be included on top of the squared amplitudes, obtained by integrating over the permitted energy-momentums for the two electrons in the decay. The phase space factors are relatively straightforward to calculate, simply a function of the total energy E_0 available to the electrons and the atomic number Z providing the nuclear Coulomb potential. E_0 is known as long as we know the Q-value of the decay, which is the total energy released in the process and can be found by measuring the masses of the parent and daughter isotopes. Generally speaking, $G_{0\nu}$ increases as the Q-value of the decay increases, due to the larger range of energies in the integral [30]. This means that candidate $0\nu\beta\beta$ isotopes with larger Q-values will also have their $0\nu\beta\beta$ rates enhanced by $G_{0\nu}$, but practically speaking other considerations tend to be much more important than

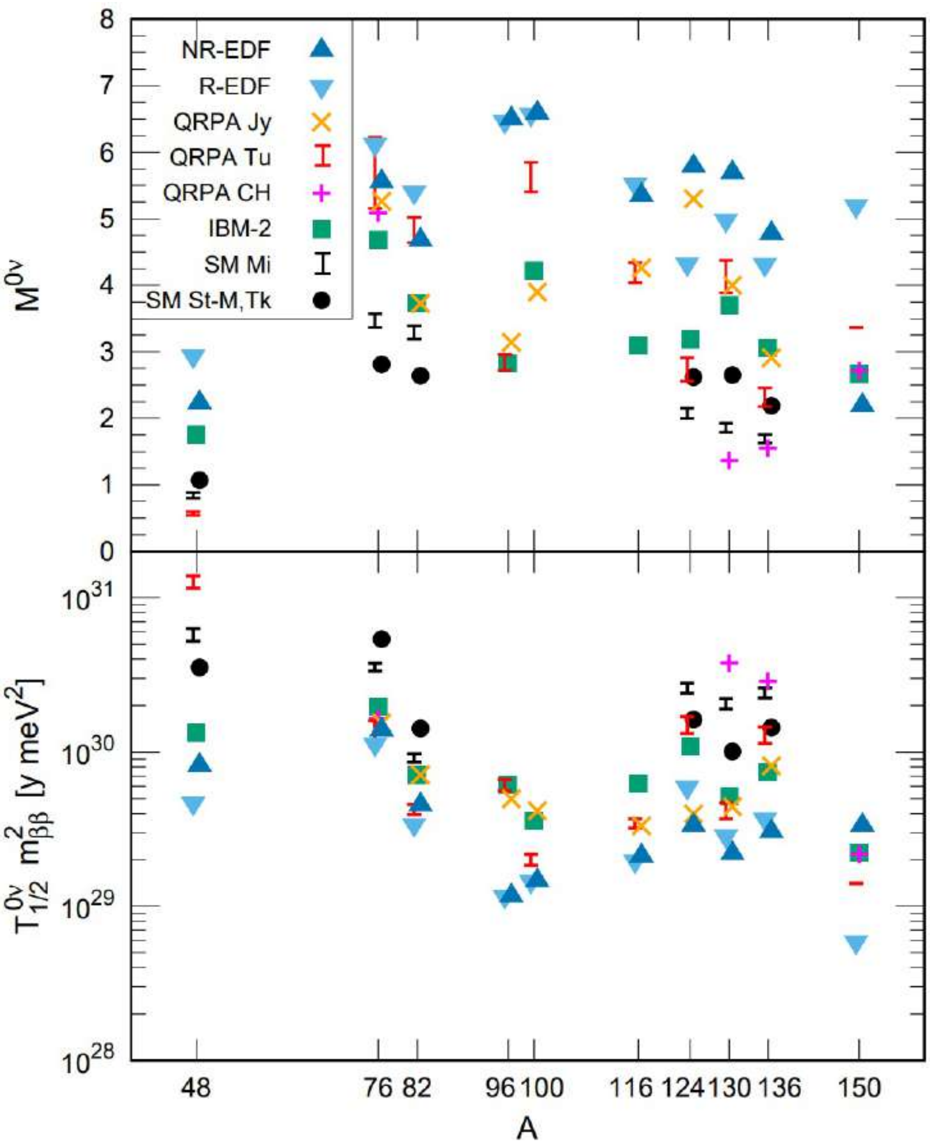


Figure 2.3: Nuclear matrix element calculations for a number of candidate double beta decay isotopes, showing the results from a selection of prominent nuclear models. The calculated values of $M_{0\nu}$ are shown in the top half. By folding in the phase space factors, one can obtain the predicted $0\nu\beta\beta$ half-lives scaled by the unknown $m^2_{\beta\beta}$, which are shown on the bottom half. Reprinted from [29].

this one. Putting everything together, the full expression for the $0\nu\beta\beta$ decay rate $\Gamma_{0\nu}$ from ground state to ground state takes the form:

$$\Gamma_{0\nu} = \ln(2) G_{0\nu}(E_0, Z) |M_{0\nu}|^2 m_{\beta\beta}^2$$

Since $\Gamma_{0\nu}$ is the physical quantity that experiments actually measure or set limits on, this expression allows us to translate those results into a limit on the more fundamental quantity of interest $m_{\beta\beta}$.

It must be noted that so far we have discussed only the prospect of $0\nu\beta\beta$ decay mediated by the exchange of a light Majorana neutrino, but there are in fact many alternative models that could result in $0\nu\beta\beta$ decay. The light neutrino mechanism is just the simplest idea, requiring minimal additions to the Standard Model and being well-motivated by the idea of the type-I seesaw. By itself, simply observing $0\nu\beta\beta$ decay does not distinguish between the different possible mechanisms. However, the Schechter-Valle theorem, also known as the black-box theorem, shows that existence of $0\nu\beta\beta$ necessarily means that neutrinos have a nonzero Majorana mass even if the decay is primarily mediated by some other mechanism [31]. The idea behind this theorem is basically shown in Fig. 2.4: even if we don't know how $0\nu\beta\beta$ decay occurs, if we know it exists then we can treat whatever process mediates it as a "black box" reaction of the form $dd \to uue^-e^-$. By crossing symmetry this black box process can be rearranged to mediate a $\bar{\nu}_e\nu_e$ term, granting an effective Majorana mass. The effective Majorana mass that arises only from this black-box operator would be far too small to explain the known light neutrino mass states [32], but the black-box theorem at least would allow us to make a qualitative statement about the Majorana nature of neutrinos in the event of a $0\nu\beta\beta$ discovery.

2.2 Experimental Considerations

From an experimental point of view, searching for $0\nu\beta\beta$ decay has the particularly notable advantage that its experimental signature is very clean. In the Standard Model $2\nu\beta\beta$ process, the total energy of the decay is split among the two neutrinos and two electrons. Since we have no way to reliably detect neutrinos, particle detectors will only measure the energy of the two electrons, and measurements of $2\nu\beta\beta$ decay show up as the measurement of a continuous energy spectrum from near 0 to near the Q-value of the decay. On the other hand, $0\nu\beta\beta$ naturally has no neutrinos to invisibly take away energy, and so the two electrons carry the full energy of the decay. A search for $0\nu\beta\beta$ is thus a search for a distinctive cluster of events forming a peak around the $\beta\beta$ Q-value (sometimes abbreviated $Q_{\beta\beta}$) at the end of the $2\nu\beta\beta$ spectrum as shown in Fig. 2.5. Since the Q-value can be determined quite precisely by separate measurements of the nuclear masses of both the parent and daughter nuclei, $0\nu\beta\beta$ experiments know exactly where their signal should show up and what it should look like. We then only need to worry specifically about the backgrounds around the appropriate $Q_{\beta\beta}$, and detectors can be optimized for performance in this specific energy region as well. This can be contrasted with experiments searching for WIMP dark matter, for example, where

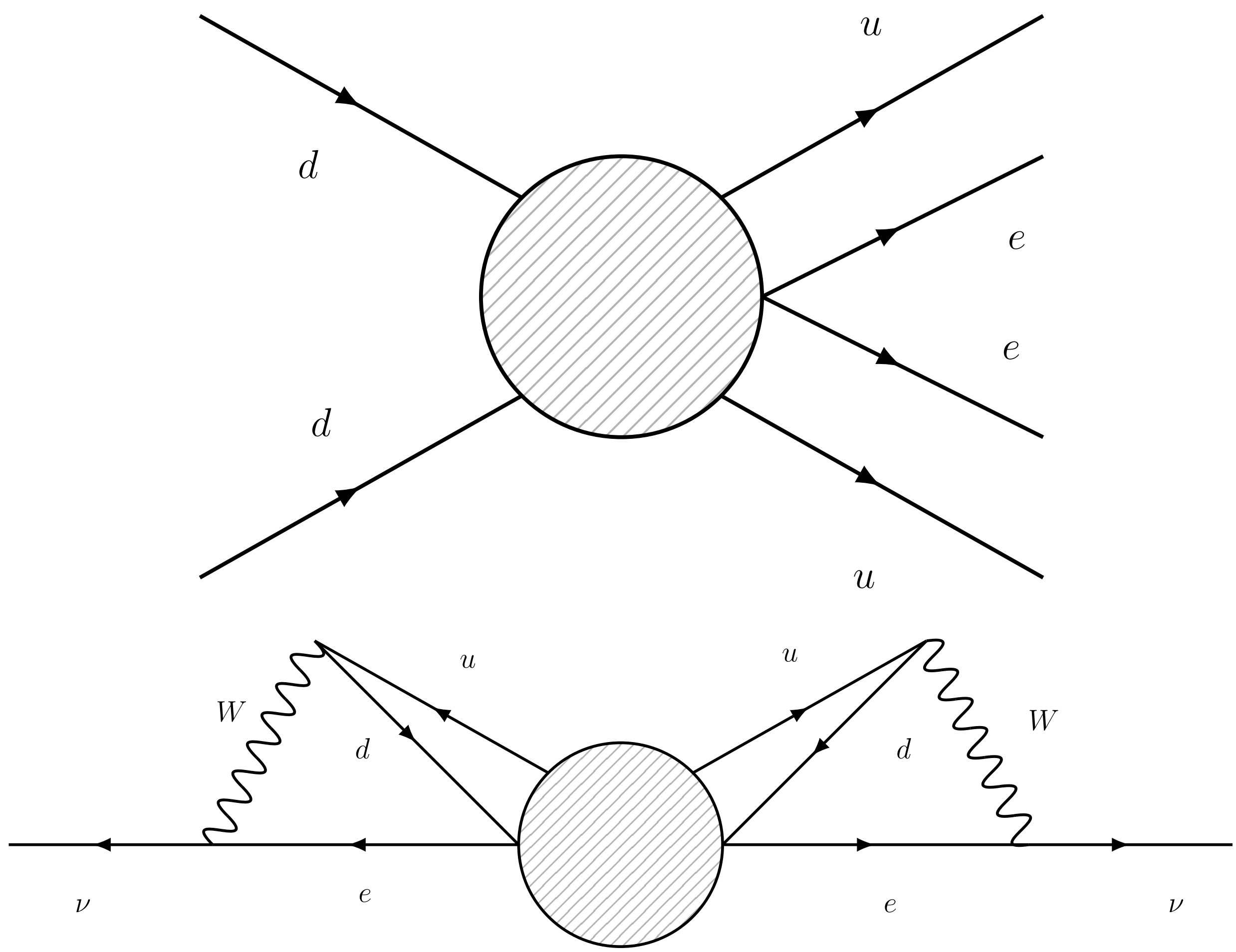


Figure 2.4: **Top**: a diagram that corresponds to the $dd \to uue^-e^-$ result of $0\nu\beta\beta$ decay, where the shaded circle represents whatever mechanism mediates the process. Regardless of whatever happens in the black box allowing $0\nu\beta\beta$ decay, we can use crossing symmetry to rearrange the inputs and outputs of the black box. **Bottom**: an effective ν_e Majorana mass operator mediated by the $0\nu\beta\beta$ black box process.

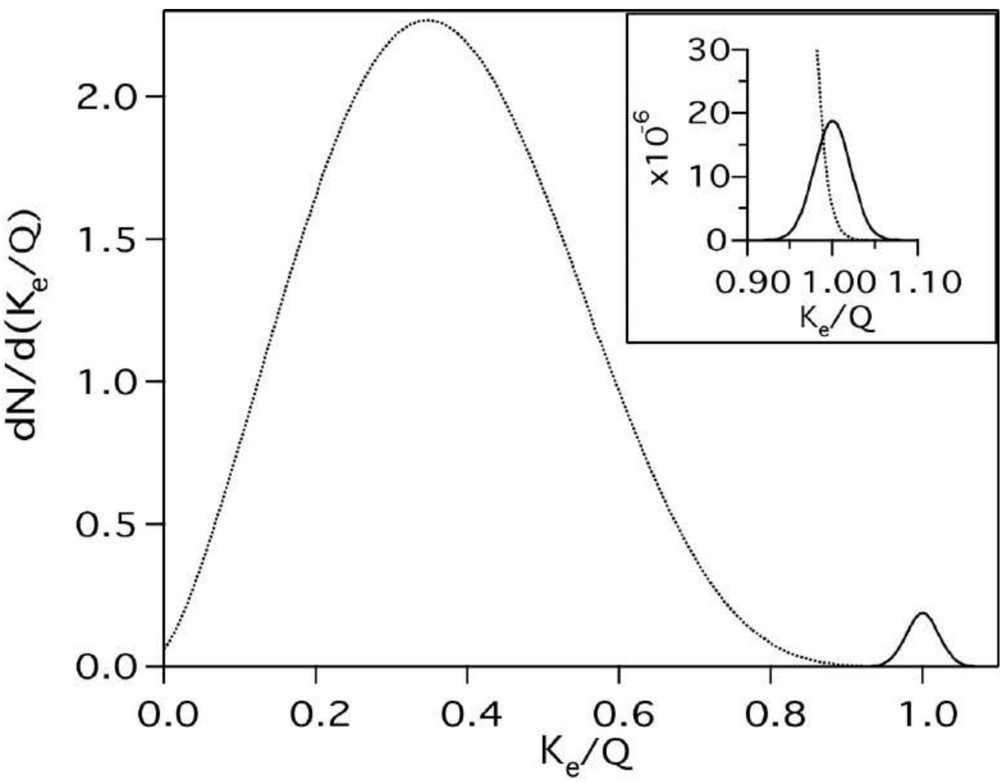


Figure 2.5: A sample normalized $2\nu\beta\beta$ spectrum along with what a theoretical $0\nu\beta\beta$ signal would look like, forming a peak at the end of the spectrum. A 5% energy resolution is assumed here, which gives the $0\nu\beta\beta$ peak its width. The zoom in on the Q-value in the top right shows how the 2ν spectrum spills into the potential 0ν signal, but the overlap is smaller if the energy resolution is improved. Reprinted from [33].

the theoretical mass of a dark matter WIMP could be anything in a large range of values, forcing experiments to worry about a whole range of possible backgrounds and requiring searches over a large range of energies for potential WIMP signals. The disadvantage is that $2\nu\beta\beta$ events constitute an irreducible background, since any $0\nu\beta\beta$ candidate will necessarily be capable of $2\nu\beta\beta$ decay as well. This background has the potential to contaminate the $0\nu\beta\beta$ search region, depending on the experiment's energy resolution and pileup rejection capabilities.

At this point we can see that searching for $0\nu\beta\beta$ decay is a question of looking for a cluster of events at the right energy for the $\beta\beta$ isotope in question. In the case of an experiment that is still background limited, the $0\nu\beta\beta$ peak must be large enough that it is clearly not due to random fluctuations in the number of background events expected in the region. We can then define a few experimental parameters of interest:

$$\begin{aligned}
a &= \text{fraction of the experimental mass that is the } 0\nu\beta\beta \text{ isotope} \\
\epsilon &= \text{detection efficiency for } 0\nu\beta\beta \text{ signal} \\
M &= \text{experimental mass} \\
t &= \text{active runtime of the experiment} \\
B &= \text{background index at } Q_{\beta\beta} \\
\Delta E &= \text{energy resolution at } Q_{\beta\beta}
\end{aligned}$$

Clearly the number of $0\nu\beta\beta$ events we expect to see is proportional to $a\epsilon Mt$, scaling with the amount of $\beta\beta$ isotope we have and how long we wait for a decay. The expression Mt is often referred to as the exposure of an experiment, with greater exposure corresponding to more available data. The background index B requires a bit more explanation - it is often expressed in terms of counts / (keV·kg·yr), interpreted as the number of background events one expects in a 1 keV energy range for each kilogram of detector mass and each year the detector is operated for. Following this definition, the number of background events we expect to see is $MtB\Delta E$, where the energy resolution ΔE shows up because this is roughly how wide of an energy range we will be looking in for a 0ν signal. Treating the occurrence of background events as a Poisson process, this gives us a statistical uncertainty of $\sqrt{MtB\Delta E}$ on the number of observed background events. Since the uncertainty on the number of background events obscures our ability to determine whether we've seen $0\nu\beta\beta$ events or not, we find that the sensitivity of such an experiment is given by:

$$T^{0\nu}_{1/2} \text{ sensitivity} \propto a\epsilon\sqrt{\frac{Mt}{B\Delta E}}$$

On the other hand, in a background-free experiment, we would be certain that any observed events at $Q_{\beta\beta}$ are from $0\nu\beta\beta$ decay, and so the sensitivity scales as:

$$T^{0\nu}_{1/2} \text{ sensitivity} \propto a\epsilon Mt$$

Of course, no experiment can actually be 100% background free, as there will always be at least some small uncertainty associated with its background modeling and background rejection capabilities. But the point holds that as an experiment reaches the point where the number of expected background events becomes close to 0, its sensitivity scales linearly with the exposure Mt. Experiments with non-negligible backgrounds can only scale their sensitivity with $\sqrt{Mt}$ since increased exposure will increase the number of both background and possible signal events. Background-free experiments can thus obtain much more favorable scaling of their sensitivity with increased exposure, justifying the difficulty of reaching the background-free benchmark. This in combination with the other experimental parameters that appear in the sensitivity expressions tells us what to consider when designing a $0\nu\beta\beta$ experiment.

Isotope Choice

The choice of $\beta\beta$ isotope for a $0\nu\beta\beta$ search naturally has important ramifications for many of the experimental parameters that determine the experimental sensitivity. An obvious factor is the rarity of the element and the natural abundance of the isotope of interest, which will determine the cost of enriching the element to obtain the $\beta\beta$ isotope in high enough concentrations for an experiment. This is not an insigificant consideration, as these enrichment costs are often the biggest expense for $0\nu\beta\beta$ experiments. Another factor is the Q-value of the $\beta\beta$ decay, which as mentioned before affects the phase space factor going into the predicted $0\nu\beta\beta$ rate. More importantly, the Q-value determines which natural radioactive backgrounds could contaminate the $0\nu\beta\beta$ search region. Primordial isotopes like ^{232}Th and ^{238}U have half lives on the same scale as the age of the Earth and are thus omnipresent, meaning that the various daughter isotopes in their decay chains are present as well. As a result, backgrounds like the 2615 keV γ from ^{208}Tl decay in the ^{232}Th chain and the 3272 keV β from ^{214}Bi decay in the ^{238}U chain cannot be easily eliminated from the experimental environment. The few $\beta\beta$ isotopes with Q-values above these energies do not have to worry about background contributions from those radioactive contaminants; the rest have to contend with them either through passive shielding or active forms of background rejection. The Q-values and natural abundances of some notable $\beta\beta$ isotopes are shown in Fig. 2.6.

A perhaps less obvious consideration is how the choice of isotope impacts the detector technologies that can be used. There exist detection methods that are completely agnostic to the $\beta\beta$ isotope being used, the most prominent example being the SuperNEMO/NEMO3 experiments, which deploy their $\beta\beta$ isotope external to a tracking calorimeter. Despite the various advantages of such an approach, the fact that the $\beta\beta$ isotope is outside the detector volume causes the detection efficiency to be relatively low, estimated at 18% for NEMO3 and 30% for SuperNEMO [34]. Due to the importance of the detection efficiency ϵ for an experiment's sensitivity, this has caused such approaches to generally fall out of favor compared to what are known as source=detector methods. In source=detector arrangements, the $\beta\beta$ isotope is incorporated into the detector itself in some fashion, naturally granting very high detection efficiency. Not all isotopes are suited to such an approach, but notable examples perfectly suited for this include ^{76}Ge and ^{136}Xe. High-purity germanium detectors are widely used as particle detectors for their extremely precise energy resolution, and the ionization and scintillation properties of liquid xenon are already exploited in dark matter experiments. This makes ^{76}Ge and ^{136}Xe appealing choices, despite the fact that one can see in Fig. 2.6 that neither their isotopic abundance nor Q-value are anything special. A number of other $\beta\beta$ isotopes can be grown into stable crystals or dissolved into a liquid detector, allowing them to used in source=detector approaches as well.

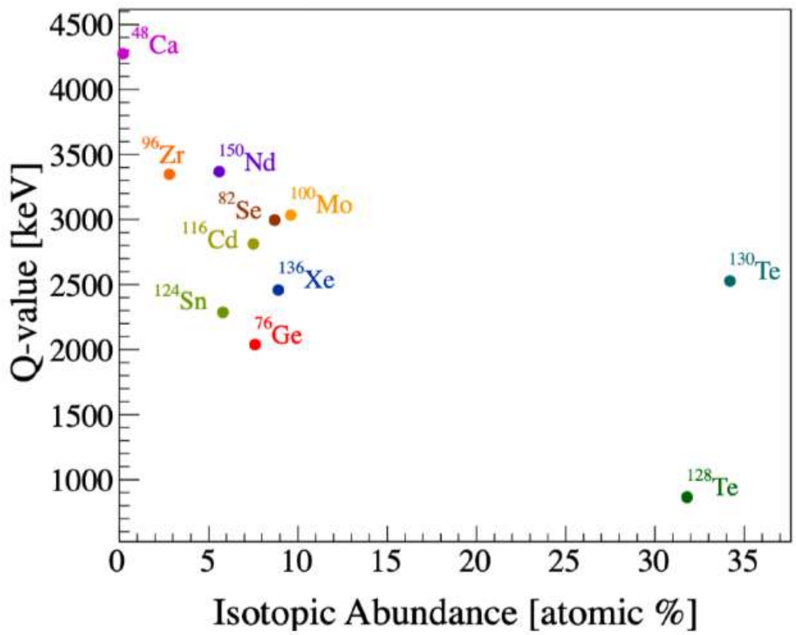


Figure 2.6: Natural isotopic abundance and Q-value for some $\beta\beta$ isotopes most frequently considered for use in $0\nu\beta\beta$ experiments. Being further to the right and further up on this plot are both advantageous, but for most isotopes there is a tradeoff, to say nothing of considerations regarding their chemical properties.

2.3 Current Experimental Landscape

The current international landscape of $0\nu\beta\beta$ decay research is both large and varied, with many different approaches attaining competitive results. To date no one has yet observed $0\nu\beta\beta$ decay[2], and so investments and collaboration sizes have grown corresponding to the increasing size of the experiments necessary to continue probing higher and higher $0\nu\beta\beta$ half-lives. A summary of the current leading limits is shown in Fig. 2.7, sometimes colloquially called a "lobster plot" for its vague resemblance to the shape of a two-clawed lobster. As mentioned before, the quantity that experiments actually set a limit on is the half-life of the decay, but to compare results between different isotopes these are converted into limits on $m_{\beta\beta}$, introducing uncertainties from the nuclear matrix element calculations.

It is also worth noting how results from other neutrino experiments complement the limits

[2] A subset of collaborators working on the Heidelberg-Moscow experiment declared that they had observed $0\nu\beta\beta$ decay in ^{76}Ge back in 2001 [35], but this was not generally believed due to their controversial analysis methods for event selection. More recent ^{76}Ge experiments have now fairly conclusively excluded the $0\nu\beta\beta$ rate that they had claimed to observe.

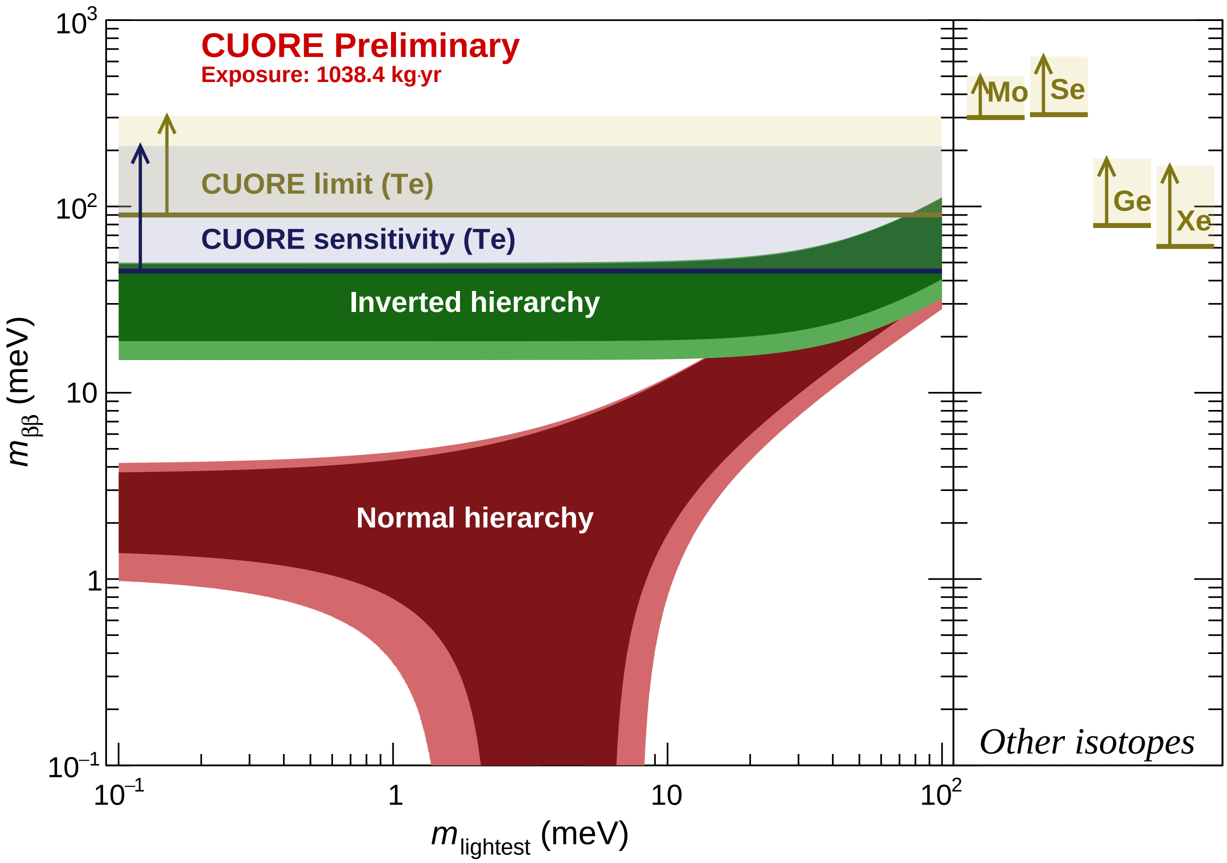


Figure 2.7: Plot of current leading limits on $m_{\beta\beta}$ for various isotopes. The shaded regions for each element are the upper limits on $m_{\beta\beta}$ from the corresponding $\beta\beta$ isotope, with the width of the regions corresponding to uncertainties from the nuclear matrix elements. Permitted $m_{\beta\beta}$ values for the inverted and normal mass hierarchy scenarios are obtained from current neutrino oscillation measurements [36]. Results for Ge are from GERDA [37], for Xe from KamLAND-Zen [38], for Se from CUPID-0 [39], and for Mo from CUPID-Mo [40]. The CUORE limit is presented in this dissertation, and the CUORE sensitivity is the expected sensitivity from CUORE's full lifetime.

from $0\nu\beta\beta$ experiments. The red and green shaded bands in Fig. 2.7 show the possible values of $m_{\beta\beta}$ in the cases of the normal and inverted mass hierarchies, assuming the 3-neutrino mixing paradigm and using the current measured values in the PMNS matrix from neutrino oscillation experiments. Here we can see a visualization of the fact that it is possible for $m_{\beta\beta}$ to vanish in the normal hierarchy if the Majorana phases take the right values. In addition, while the $0\nu\beta\beta$ experiments constrain space along the y-axis in Fig. 2.7 as they set tighter upper limits on $m_{\beta\beta}$, other experiments that attempt to set limits on the neutrino masses will constrain space along the x-axis. Two classes of experiments in particular are relevant along the x-axis. One is direct limits on the neutrino masses by measuring the endpoint of β decay spectra, where the current strongest limit from KATRIN is $m_\nu < 0.8$ eV [41]. This limit depends only on the assumption of conservation of energy and is thus widely accepted as valid, but it is weak enough that it is not in range of the x-axis of this plot. The other is limits on the sum of the 3 neutrino masses from cosmological measurements, for which the current strongest limit is $\sum m_\nu < 0.12$ eV from the Planck satellite [42]. These limits are obtained by measuring the effects of neutrinos on the cosmic microwave background and matter power spectrum of the universe. This present cosmological limit is strong enough to suggest that current $0\nu\beta\beta$ experiments will probably not observe $0\nu\beta\beta$ decay. However, the cosmological limits are heavily dependent on our models of early cosmological evolution, and so we generally do not think they should be treated as an immutable statement. Of course, all limits on $m_{\beta\beta}$ from $0\nu\beta\beta$ experiments also completely depend on the assumption that neutrinos are Majorana particles and that $0\nu\beta\beta$ decay is mediated by a light Majorana neutrino, but that is why it is useful to have different classes of experiments studying neutrino masses in different ways with different model assumptions.

A more detailed overview of the many $0\nu\beta\beta$ experiments that either have recent results or will have results in the near future can be found in [43], but I will make explicit mention of the current leading limits. The overall leading limit of $m_{\beta\beta} < 61 - 165$ meV comes from a half-life limit of $T^{0\nu}_{1/2} > 1.07 \times 10^{26}$ yr in ^{136}Xe from KamLAND-Zen [38], which uses a xenon-loaded liquid scintillator approach[3]. It has a relatively poor energy resolution of $\sigma \sim 7.3\%/\sqrt{\text{Energy (MeV)}}$, which is wide enough that backgrounds from $2\nu\beta\beta$ decays are a concern, but compensates with its ability to use its large detector volume to provide active shielding against backgrounds and load a large mass of ^{136}Xe, attaining an exposure of 504 kg·yr of ^{136}Xe. The next leading limit of $m_{\beta\beta} < 79 - 180$ meV comes from a half-life limit of $T^{0\nu}_{1/2} > 1.8 \times 10^{26}$ yr in ^{76}Ge from GERDA [37], which operates high-purity germanium detectors with a liquid argon shield. They were able to achieve energy resolutions as good as $\sigma/Q_{\beta\beta} \sim 0.05\%$, and this in combination with their active background rejection methods resulted in an essentially background free $0\nu\beta\beta$ search, with an estimated background index of 5.2×10^{-4} counts / (keV·kg·yr). This allowed them to achieve a stronger half-life limit with a comparatively smaller exposure of 127 kg·yr of ^{76}Ge, although this still corresponds to

[3]KamLAND-Zen benefited from a statistical underfluctuation in the number of events around $Q_{\beta\beta}$ for this result. There's nothing unusual about this, but it does mean they obtained a stronger limit than would have been expected on average given their experimental parameters.

a weaker limit on $m_{\beta\beta}$ after accounting for phase space factors and nuclear matrix elements.

Finally, the last leading limit on this plot comes from CUORE's result with ^{130}Te, which shall be the focus of the rest of part 1 of this dissertation.

Chapter 3

The CUORE Experiment

CUORE (Cryogenic Underground Observatory for Rare Events), an experiment designed to search for the $0\nu\beta\beta$ decay of ^{130}Te, serves as the latest and largest development in a long history of using the cryogenic calorimetric technique for particle detection. It is located at LNGS (Laboratori Nazionali del Gran Sasso) in Italy, an underground research facility sheltered beneath the Gran Sasso mountains, which also houses other neutrino, dark matter, and $0\nu\beta\beta$ experiments. Like these other rare event searches, CUORE must be sheltered from cosmic muons, which are muons resulting from the decay of pions created when energetic cosmic rays interact with nuclei in the atmosphere. These cosmic muons have a flux of about 1 / cm^2 / minute at sea level, which could overwhelm a large detector looking for very rare events. The rock overburden provided by the Gran Sasso mountains is the shielding equivalent of about 3800 meters of water, reducing this muon flux by about 6 orders of magnitude to a much more manageable 2×10^{-6} counts / cm^2 / minute in the underground halls of LNGS where experiments are located [44]. From a human perspective, the LNGS experiment halls also have the advantage of being not too far from civilization and being easily accessible by car, unlike many other underground labs around the world located in former mines.

^{130}Te has a number of favorable properties for a $0\nu\beta\beta$ search. Looking again at Fig. 2.6, ^{130}Te is notable for having the highest natural isotopic abundance of all $\beta\beta$ isotopes at $\sim$34%. By using this isotope, CUORE is thus able to use unenriched tellurium and still have a competitive amount of $\beta\beta$ isotope, saving on the normal costs of isotope enrichment. ^{130}Te also has a relatively high Q-value of 2527.5 keV [45], above most natural β/γ backgrounds. While this is still below the 2615 keV γ emitted by ^{208}Tl, it does fall above the Compton edge for this γ ray. This means that the number of ^{208}Tl γ events that can pollute the $0\nu\beta\beta$ search region is significantly reduced, since they must be multi-Compton scattered to deposit an amount of energy near $Q_{\beta\beta}$.

From an elemental perspective, tellurium also has the benefit that it can be grown into high-purity TeO_2 crystals [46]. CUORE loads its ^{130}Te payload in the form of 988 $5\times5\times5$ cm^3 crystals of TeO_2, for a total mass of 742 kg. Using unenriched tellurium, this corresponds to a total of 206 kg of ^{130}Te. The crystals are then cooled to a base temperature of $\sim$10 mK and operated as cryogenic calorimeters, detecting particles by the rise in temperature they induce when they deposit energy in a crystal. This has the advantage of being a source=detector approach, since the TeO_2 crystals both contain the ^{130}Te and serve as the particle detectors. It is also worth noting that this method is actually not limited to tellurium; there are several $\beta\beta$ isotopes that can be grown into crystals suitable for this approach, which grants it some flexibility in isotope choice. Some examples of experiments exploring the cryogenic calorimetric method with other isotopes include CUPID-Mo using ^{100}Mo in Li_2MoO_4 [40], AMoRE using ^{100}Mo in $CaMoO_4$ [47], CUPID-0 using ^{82}Se in ZnSe [39], and CANDLES using ^{48}Ca in CaF_2 [48].

This chapter will focus on the specific design of the CUORE experiment and how it searches for $0\nu\beta\beta$ decay, discussing the general principles behind the cryogenic calorimetric technique and its advantages and then describing the one-of-a-kind cryostat that has allowed CUORE to employ this technique on a larger scale than ever before.

3.1 The Cryogenic Calorimetric Technique

In general, a calorimeter is just any detector that measures energy deposits. Cryogenic calorimeters have the particular approach of performing this task by measuring a temperature increase ΔT in an absorber and corresponding the ΔT to some amount of energy, using a basic combination of some absorber and some thermometer. For a $0\nu\beta\beta$ search, we're interested in resolving energies of O(1 MeV) with O(keV) resolution, which on the scale of nuclear physics are not unusual energies. However, if we remember that our calorimeters are macroscopic objects (in the case of a CUORE crystal, weighing 750 g), we can restate this as wanting to detect an O(0.1 pJ) energy deposit with O(0.1 fJ) resolution. Clearly these kinds of energy deposits would not create measurable changes in temperature under normal conditions. Debye's Law tells us that the phonon contribution to the heat capacity $C(T)$ of a solid scales with temperature T as:

$$C(T) \propto \left(\frac{T}{T_D}\right)^3$$

where T_D is the Debye temperature of the material, which can in principle be calculated from the speed of sound in the material but is generally experimentally determined instead. From this expression we see that the heat capacity scales as T^3, so as we operate the calorimeters at colder temperatures we also get larger relative changes in temperature from any particular energy deposit[1]. A larger relative change in temperature means a better resolution when we estimate the magnitude of the energy deposit using the temperature change as a proxy. We also see that the colder we can go the better, as the heat capacity continues decreasing as

[1]For conductors, the electron contribution to heat capacity, which scales linearly with T, is significant at low temperatures. The crystals we use as calorimeters are in general insulating materials, so their heat conduction is dominated by the phonon modes and their heat capacity is accurately modeled by the T^3 scaling.

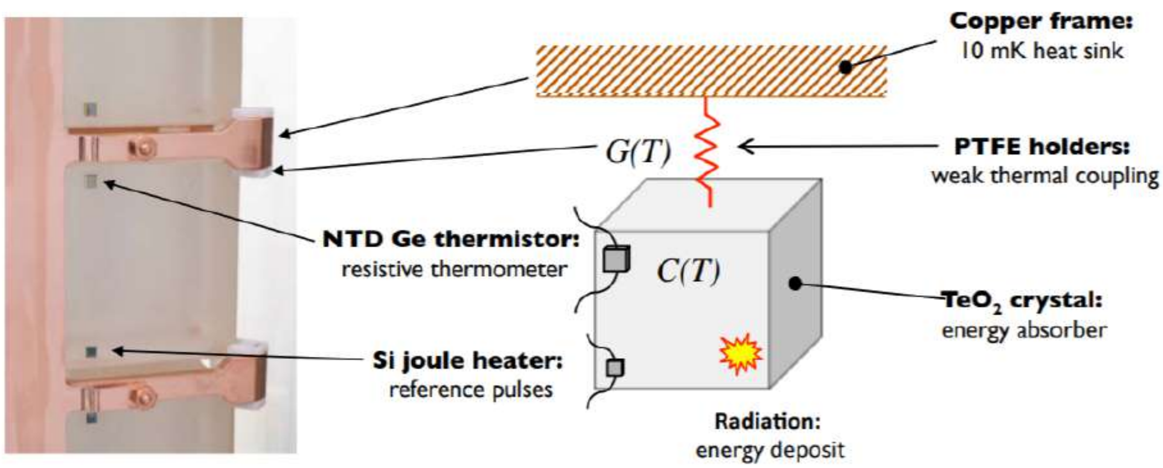


Figure 3.1: Schematic of how a CUORE crystal is operated as a calorimeteric detector. The TeO_2 crystal serves as an energy absorber with some heat capacity $C(T)$ and is linked to a thermal bath by a weak thermal link with thermal resistance $G(T)$. In CUORE's case, the thermal bath is provided by the copper support structures holding the detectors in the cryostat and is operated somewhere near ~10 mK, and the weak thermal links are provided by teflon (PTFE) holders that are the only link between the detectors and their copper supports. The crystal experiences a rise in temperature when radiation deposits energy in it and then slowly releases the energy back to the thermal bath through the weak thermal link, allowing it to return to the base temperature.

a cubic function of temperature all the way down to absolute zero[2]. This contrasts with, for example, germanium detectors operated at cryogenic temperatures, where they must be cold enough that thermal noise is negligible, but going any colder doesn't yield further improvements since it doesn't improve the nature of their signal. For TeO_2 with a Debye temperature of $T_D = 232$ K [49], we can reach a heat capacity of roughly $C \approx 1$ MeV / 100 μK at T=10 mK.

In CUORE this cryogenic calorimetric principle is employed through a setup shown in Fig. 3.1. The cryostat provides a thermal bath that dictates the base temperature of the crystals serving as calorimeters. In order for the temperature rise induced by an energy deposit in a crystal to be measurable, the crystal cannot return to the equilibrium temperature too quickly. This is arranged by having the crystals be attached to the copper support structures serving as the thermal bath with only small teflon (PTFE) holders, which act as

[2]In practice, the energy resolution of a cryogenic calorimeter doesn't actually continue improving as the operating temperature gets arbitrarily low. Factors such as the stability of the cryogenic system and properties of the thermometer attached to the absorber also come into play, so there is a balance to find the optimal operating temperature.

weak thermal links. Each crystal is additionally instrumented with a silicon-based heater [50], which can receive a current to manually inject power into the crystal for purposes of response stabilization, and a thermistor to measure the temperature of the crystal.

In CUORE, the thermistors are made of neutron-transmutation doped (NTD) germanium. These are created by subjecting pure germanium to a controlled neutron flux from a nuclear reactor, transmuting ^{70}Ge into ^{71}Ga, ^{74}Ge into ^{75}As, and ^{76}Ge into ^{77}Se, thereby doping the material[3]. Since the neutron capture cross section on Ge is low, with a sufficiently thin Ge wafer this process results in a uniformly random distribution of charge carrier sites throughout the material. At low temperatures, the resistance of NTD Ge wafers acquire an exponential dependence on temperature [51], as very few charge carriers have sufficient thermal energy to exist in the conduction band. Instead, they effectively tunnel between charge carrier sites in order to carry current. The CUORE NTD Ge thermistors are thus operated by applying a bias current across them and reading the output voltage; the sharp change in temperature from a particle event causes the NTD thermistor's resistance to drop correspondingly, which the electronics will then see as a change in voltage. This signal is brought out to a front-end electronics setup at room temperature outside of the fridge, passing through a differential pre-amplifier, second-stage programmable amplifier, and an anti-aliasing active filter [52]. The analog signal is finally passed through a $\Sigma\Delta$ ADC (AD7732), with each detector's voltage reading digitized as an 18-bit signal at a rate of 1000 samples per second. This ADC resolution is chosen to be negligible compared to the typical resolution of normal noise in the detectors, and the sampling rate provides sufficient timing resolution for pulse shape studies of our slow signals. All data from the ADCs are stored with no hardware trigger cut so that the full data stream is always available for every detector, and all triggering is done in software. Fig. 3.2 shows a characteristic pulse for one detector after the signal has been digitized.

The fundamental limitation of the energy resolution obtainable by this method comes from statistical fluctuations in the internal energy of the crystal serving as the absorber. Since the crystal is connected to the cryostat's thermal bath, it can be treated as a canonical ensemble with average energy $\langle E\rangle = -\frac{\partial \ln Z}{\partial \beta}$, where $\beta = 1/(k_B T)$. Then the heat capacity is given by:

$$C = \frac{\partial \langle E\rangle}{\partial T} = \frac{1}{k_B T^2}(\langle E^2\rangle - \langle E\rangle^2)$$

From this we can derive the typical size of the fluctuation in energy:

$$\langle \Delta E^2\rangle \equiv \langle E^2\rangle - \langle E\rangle^2 = k_B T^2 C$$

$$\sqrt{\langle \Delta E^2\rangle} \propto T\left(\frac{T}{T_D}\right)^{3/2}$$

This can also be understood as being due to the statistical fluctuation in the number of phonons in the system. Phonons have a mean energy of $k_B T \sim 1\mu$eV at 10 mK, and there

[3]Naturally occurring isotopes of Ge also include ^{72}Ge and ^{73}Ge, but when these isotopes capture a neutron they simply turn into another stable isotope of Ge.

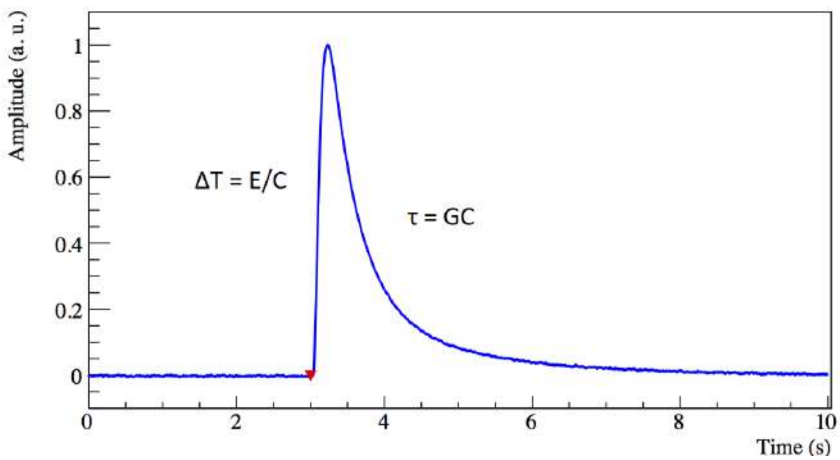


Figure 3.2: A typical example of a CUORE pulse with energy near $Q_{\beta\beta}$ of ^{130}Te. The pulse height is determined by the change in temperature $\Delta T = E/C$, and the decay time $\tau = GC$ is given by how long it takes the absorber to release the heat back into the thermal bath through the weak thermal link, where E is the energy of the deposit, C is the heat capacity of the detector, and G is the strength of the thermal link between the detector and the thermal bath. For CUORE pulses, it takes on the order of seconds to return to the baseline state.

are a total of $N = C/k_B$ effective phonon modes. The size of the fluctuation in energy is then given by $\sqrt{N}\langle E_{phonon}\rangle = \sqrt{k_B T^2 C}$, the same expression derived above. For a TeO_2 crystal at 10 mK, this corresponds to a resolution limit of less than 10 eV. In practice, CUORE and other cryogenic calorimetric $0\nu\beta\beta$ experiments are nowhere near this fundamental limit due to a number of reasons. For one, the experimental setup is not actually an idealized system, as each component has its own heat capacities and the thermal boundaries are imperfect. Other factors such as the stability of the cryogenic environment and electronic noise also play a large role in the resolution we can actually attain. CUORE has typically measured resolutions of $\sigma \sim 2.5 - 3.5$ keV at 2615 keV, which is still excellent in the current $0\nu\beta\beta$ landscape. In particular, this is good enough that CUORE does not need to worry about $2\nu\beta\beta$ events or natural radioactive peaks spilling into the $Q_{\beta\beta}$ region where they could be confused for a ^{130}Te $0\nu\beta\beta$ decay.

We have seen that advantages of the cryogenic calorimetric technique include flexibility in isotope choice, a source=detector arrangement, and a very good energy resolution. One of the disadvantages is that thermal signals are very slow; looking at Fig. 3.2 again, one can see that it takes seconds for a detector to return to its baseline state and be ready for another event. However, since CUORE is a rare-event search with very controlled backgrounds, the event rates are low enough that this actually doesn't matter. A more relevant concern is the lack of particle and event-type discrimination capability. Since the only signal being detected is the total heat deposit in an entire crystal, α, β, and γ events all look the same, as do events occurring near the surface and events occurring in the center of the detector[4], even though in principle these kinds of events all deposit energy in different ways. This actually does limit $0\nu\beta\beta$ sensitivity, since ideally we should be able to take advantage of the fact that $0\nu\beta\beta$ events are β events mostly taking place in the bulk of the crystals in order to reject other types of events as backgrounds. Solving this problem requires an upgrade to the entire experimental setup of CUORE, which we will return to in the second half of this dissertation when we discuss CUPID.

3.2 Cryogenic Infrastructure

Although the cryogenic calorimetric method and the use of ^{3}He-^{4}He dilution fridges both have long histories, and although the basic physics behind them is well-understood, CUORE represents a major technical advancement by virtue of its sheer size. CUORE uses a custom-built dilution fridge able to cool 15 tonnes of material to below 4 K, including 3 tonnes of material that are cooled to below 50 mK. The sensitive detector mass of 742 kg of TeO_2 kept near 10 mK has demonstrated high-quality stable performance over the course of 4 years of operation at this point, without once re-opening the fridge after CUORE began its first dataset. The search for $0\nu\beta\beta$ with CUORE thus also serves as a demonstration of the feasibility of tonne-scale cryogenics in general, which potentially have more general applications in fields such as quantum computing, dark matter searches, and gravitational wave detectors, all of which could benefit from large-scale stable cryogenic environments in the sub-100 mK range.

^{3}He-^{4}He Dilution Refrigerators

The ^{3}He-^{4}He dilution fridge is the basic technology used to cool and maintain macroscopic amounts of material at the temperatures needed for cryogenic calorimeters. This method is based on exploiting properties of ^{3}He and ^{4}He as quantum liquids, allowing for large amounts of cooling power even at temperatures of a few mK. In particular, ^{4}He is a boson and thus obeys Bose-Einstein statistics, beginning its transition into a superfluid state at 2.17 K as every particle can condense into the same ground state. By contrast, ^{3}He is a fermion and must obey Fermi-Dirac statistics following the Pauli exclusion principle[5]. The result is that for a mixture of ^{3}He and ^{4}He, the phase diagram takes the form shown in Fig. 3.3. As the temperature of the mixture goes to 0 K it will separate into a ^{3}He-rich phase and a ^{4}He-rich

[4]There has been work showing it is possible to deposit Al films on crystal surfaces so that surface events and bulk events end up having different pulse shapes, allowing for this kind of discrimination [53], but CUORE is not equipped to do this.

[5]Eventually ^{3}He does undergo a superfluid transition as well, but at a much lower temperature, below 3 mK. This happens when the fermionic nuclei form Cooper pairs in a way analagous to how electrons pair up in superconductors.

phase, but contrary to the classical expectation, the ^{4}He-rich phase never fully expels the ^{3}He. This is because ^{3}He has a smaller mass and therefore a larger zero-point motion than ^{4}He, so given the "choice" of settling into a pure ^{3}He or pure ^{4}He liquid, both types of atoms will prefer to join the ^{4}He solution where they are more tightly bound. However, since ^{3}He is also a fermion, it has to occupy progressively higher energy states as more of it is piled into a ^{4}He solution, and eventually it is no longer entropically favorable for it to join the ^{4}He-rich mixture instead of existing on its own as pure ^{3}He. For T=0 K, this saturation point occurs at the 6.6% ^{3}He that we see in the phase diagram. Meanwhile the ^{3}He-rich side of the phase diagram settles into a 100% ^{3}He state at T=0 K, as ^{4}He has no problem stacking with itself.

These two possible phases for a ^{3}He-^{4}He mixture at low temperatures are sometimes called the *dilute* phase and *concentrated* phase, referring to the concentration of ^{3}He in each mixture (up to 6.6% in the dilute phase and near 100% in the concentrated phase). Measurements of the specific heat of these mixtures tell us that the enthalpy of ^{3}He is higher in the dilute phase than in the concentrated phase, so cooling occurs every time a ^{3}He atom is transferred from the concentrated phase to the dilute phase. This forms the basic principle of a dilution fridge, sketched in Fig. 3.3. A *mixing chamber* at the lowest temperature stage contains liquid helium in the concentrated ^{3}He phase floating on top of liquid helium in the dilute ^{3}He phase, kept separated by their natural difference in density. A *still* operated at around 700 mK is connected to the dilute phase of the mixing chamber and is continuously pumped on.

At the temperature that the still operates, the vapor pressure of ^{3}He is quite high but the vapor pressure of ^{4}He, which is in its superfluid state, is near 0. Pumping the still thus removes mostly ^{3}He, resulting in a helium mixture that is in the dilute ^{3}He phase but that is unsaturated with ^{3}He as it is continuously being pumped out. Since the still is connected to the dilute phase in the mixing chamber, osmotic pressure pushes ^{3}He from the mixing chamber's dilute phase up to the still to replace the ^{3}He being lost. ^{3}He is then pulled from the concentrated phase in the mixing chamber to replace this lost ^{3}He in the dilute phase, and this endothermic process provides the cooling power of the fridge. The ^{3}He pumped out from the still is then cycled back down to the concentrated phase in the mixing chamber, recooled by the ^{3}He being pulled up from the mixing chamber through some heat exchanger, and this completes the cycle.

In a fridge that simply pumps on pure liquid ^{4}He, obtaining its cooling power through the heat of vaporization, the cooling power scales as $e^{-1/T}$, as the vapor pressure drops exponentially with temperature but the heat of vaporization remains mostly constant. By contrast, the dilution fridge's cooling method uses the heat of mixing between the concentrated and dilute ^{3}He phases in the mixing chamber, making it dependent only on the difference in specific heat of the two phases, since both of these phases are stable all the way down to absolute 0 K. The specific heat of both the concentrated and dilute ^{3}He mixtures are roughly proportional to the temperature T for low T, so the cooling power in the dilution fridge scales as T^2 instead. This allows dilution fridges to have the cooling power to sustain temperatures of O(10 mK) at the mixing chamber stage. Of course, since this process described here is already being fed with liquid helium, some other process must be used to precool the ^{3}He-^{4}He

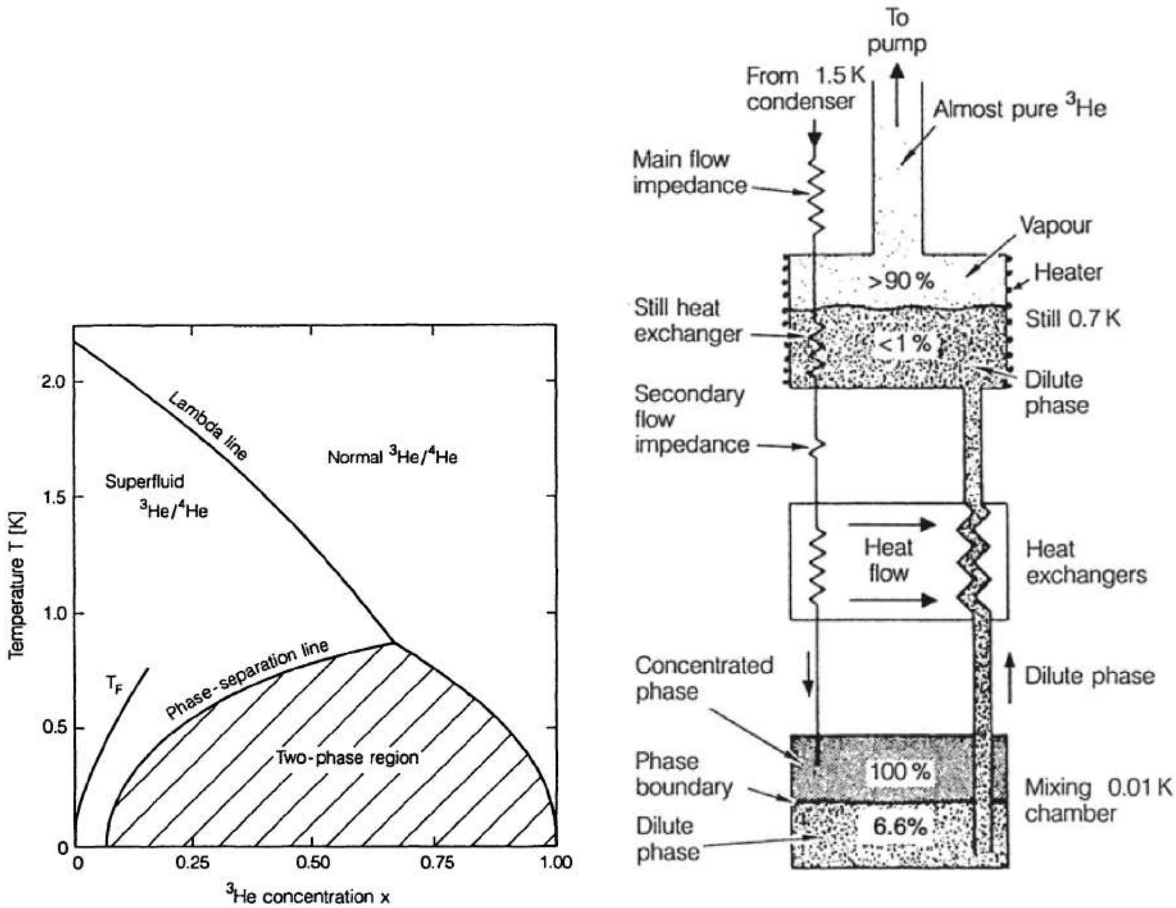


Figure 3.3: **Left**: phase diagram for liquid ^{3}He-^{4}He mixtures at saturated vapor pressure. The shaded region is inaccessible, showing that as temperature drops below about 0.867 K, a ^{3}He-^{4}He mixture must split into two phases falling on either side of the shaded region: one rich in ^{3}He and one in rich in ^{4}He. **Right**: schematic of a typical ^{3}He-^{4}He dilution fridge. The still and mixing chamber are shown, as well as the ^{3}He abundance in the mixture in each part of the cycle. Reprinted from [54].

mixture used in the dilution unit to ∼4 K, whereupon it can then take over the rest of the cooling process. This is where the structure of the rest of the cryostat comes into play.

The CUORE Cryostat

A schematic of the custom-built CUORE cryostat is shown in Fig. 3.4, mostly following the typical structure of a ^{3}He-^{4}He dilution fridge but at a much larger scale. The cryostat is separated into different temperature stages, with the temperature and size of each stage decreasing as one moves further into the cryostat. When pumped out to near-vacuum, the layered stages provide shielding for the inner colder stages against blackbody radiation from the outer warmer stages. The cooling power above 4 K is supplied by 5 Cryomech pulse tube coolers, which can each provide 1.2 W of cooling power at the 4 K stage. Only 4 pulse tubes are kept active during normal operation, and only 3 are actually required to maintain the base state, with the extras built in for redundancy. The use of these pulse tube coolers makes CUORE's cryostat a "dry" fridge, in contrast with a more traditional "wet" fridge. Wet fridges use a liquid helium bath to bring the temperature down to 4 K, with a bath at 1 K provided by pumping on a ^{4}He bath[6]. This 1 K bath is used for precooling the ^{3}He in the dilution unit. Wet fridges require periodic downtime to replenish the bath, limiting the livetime of the experiment and therefore the rate at which it can accumulate exposure, in addition to imposing additional maintenance costs from having to constantly acquire more liquid helium. A dry fridge avoids all these problems but has the downside that the moving parts of the pulse tubes can contribute mechanical vibrations that affect the stability of detectors at the coldest stage. CUORE suppresses this noise by burying the flexlines in a sandbox filled with quartz powder, in addition to mechanically decoupling the moving parts from the cryostat [55]. CUORE also utilizes a novel technique that tunes the phases of the pulse tubes' periodic pressure waves to force them to cancel each other out as much as possible, minimizing any remaining vibrational noise [56].

While the pulse tube systems are sufficient to sustain low-temperature operations, their overall cooling power is low enough that it would take months to reach 4 K starting from a room temperature of 300 K using them alone. To supplement them, CUORE also has a fast cooling system powered by a separate external cryostat [55]. This system cools helium gas and pushes it through the inner vacuum chamber (IVC) of the CUORE cryostat to speed up the initial cooldown, with special controls on the temperature and pressure of this circuit so that air and radon cannot leak in. The fast cooling system is used to facilitate the cooldown from roughly 200 K to 50 K, after which it no longer provides significant additional cooling power over the normal pulse tube system. At that point, it is turned off and the helium from its circuit is pumped out of the IVC, letting the pulse tube coolers do the rest of the cooling work before the dilution unit can become active. CUORE uses a custom-built ^{3}He-^{4}He dilution unit from Leiden cryogenics to bring the final temperature of its coldest stages

[6]The cooling from room temperature down to 77 K is generally done with a liquid nitrogen bath first, before switching to liquid helium for the rest of the cooling.

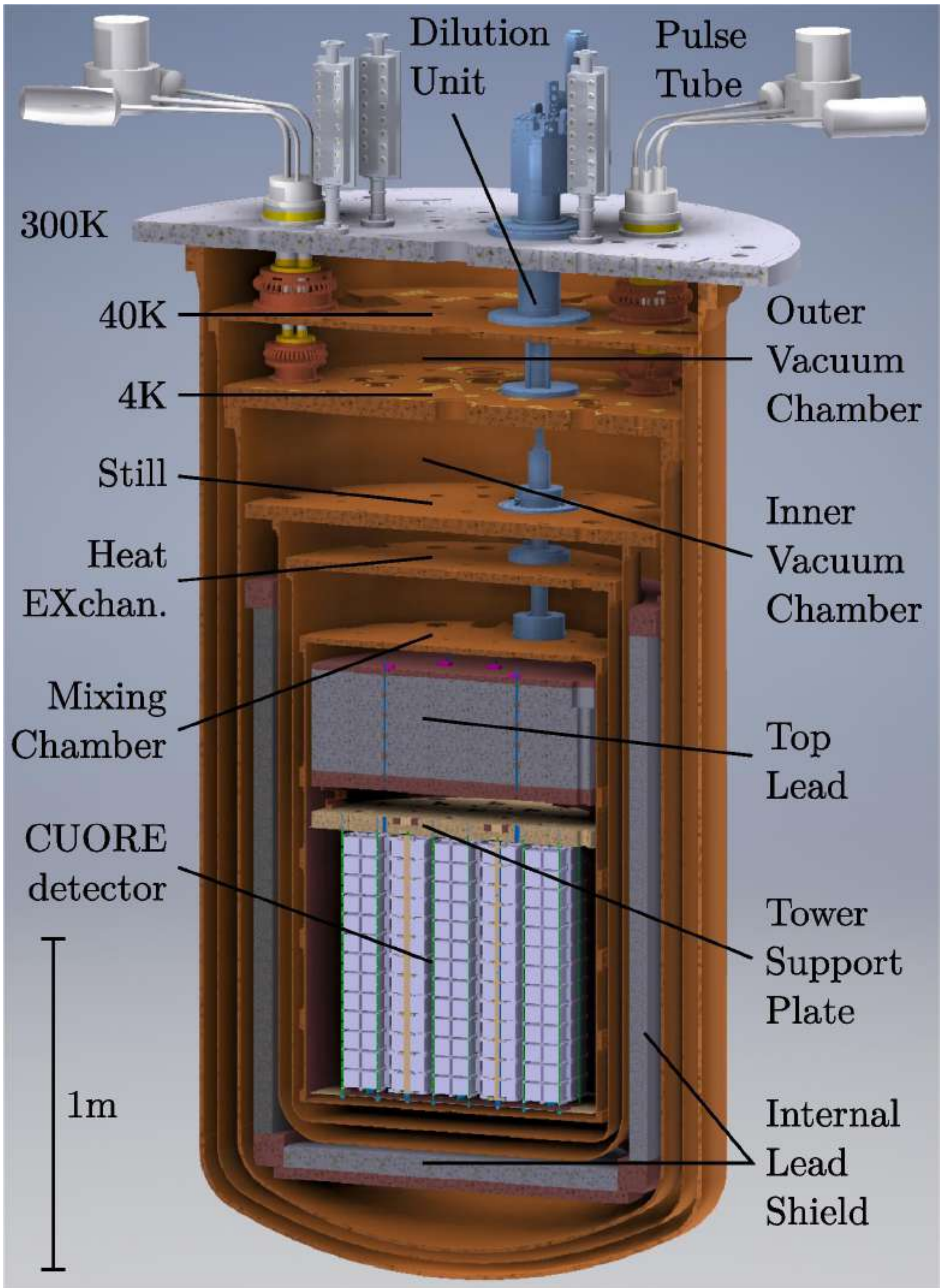


Figure 3.4: Rendering of the CUORE cryostat and its different stages and chambers. The different stages are labeled by their nominal temperatures when the cryostat is in operation. In addition, the still operates at about 800 mK, the heat exchanger at 50 mK, and the mixing chamber at 10 mK. The tower support plate holds the CUORE detectors below it and is mechanically decoupled from the rest of the cryostat. The internal lead shields are made of Roman lead due to their proximity to the detectors, while the top lead shield is made of modern lead since it lies further away. Reprinted from [55].

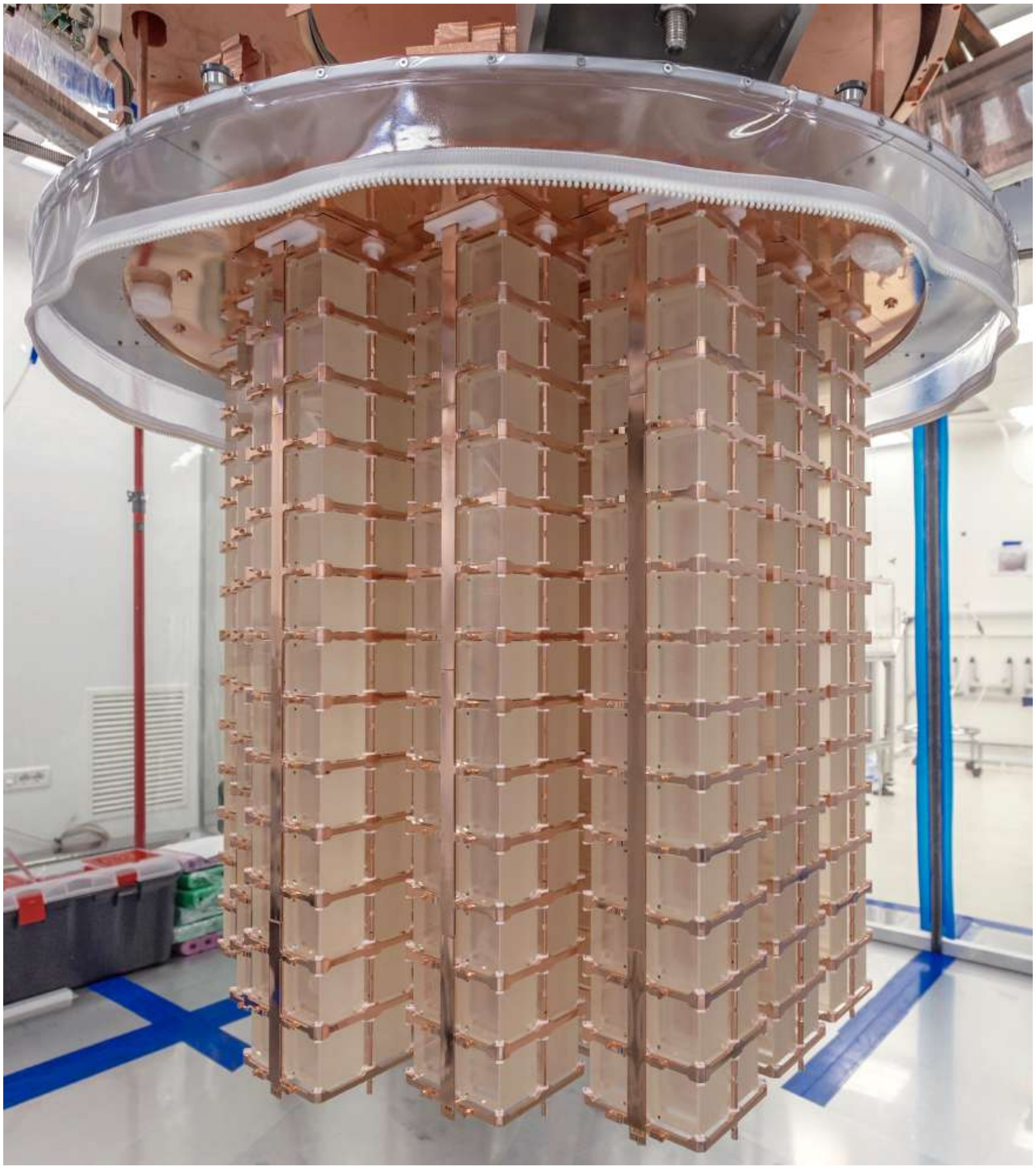


Figure 3.5: The CUORE towers after installation, before the cryostat was closed up. These towers occupy the innermost chamber of the cryostat, which is lowered to ∼10 mK during normal operations.

down to ∼10 mK, with an estimated cooling power of 4 μW at 10 mK [55]. With the aid of the fast cooling system, CUORE is able to complete a cooldown from room temperature to base temperature in ∼20 days.

Since CUORE is a rare event search, radioactive backgrounds are top concerns. Potential radioactive contaminants in all materials used in the construction of the cryostat must be accounted for, with the materials closest to the actual CUORE detectors requiring the strictest controls. Materials are screened for low concentrations of long-lived contaminants like ^{232}Th and ^{238}U, and the materials are stored underground after procurement so as to avoid cosmogenic activation of more short-lived radioactive isotopes. The copper cans which form the temperature stages of the cryostat serve as adequate shielding against α and β backgrounds from the underground lab environment, but lead shielding is required to further mitigate external γ backgrounds, which have greater penetrative power. However, lead shielding itself can carry additional radioactive contaminants, mostly in the form of ^{210}Pb, which shows up in the decay chain of ^{238}U. Even when uranium is removed during the lead ore purification process, any ^{210}Pb that appeared in the ore from uranium decays prior to processing will still be in the final product. With a half-life of 22 years, it would take too long to wait for the ^{210}Pb to decay away, so this contaminant is quite difficult to remove in any modern lead. To solve this issue, CUORE uses ancient lead salvaged from Roman shipwrecks for the inner shielding nearest the detectors. This lead is old enough that almost all of the ^{210}Pb has already decayed away, and since it was at the bottom of the sea during that entire time it was also protected from cosmic rays that could have activated new short-lived isotopes. Assays of Roman lead show its ^{210}Pb activity is extremely low at less than 715 μBq/kg [57], and even its ^{238}U/^{232}Th levels are less than half those of modern lead [58]. The top lead shield above the detectors is made of modern lead, since it is more separated from the detectors and therefore has less strict radiopurity requirements.

The TeO_2 crystals that serve as the CUORE detectors are placed in the innermost stage of the cryostat, which reaches a base temperature of ∼10 mK provided by cooling from the mixing chamber. The 988 crystals are arranged into 19 towers, with each tower containing 13 vertical layers of 4 crystals each. One of these towers was first assembled and tested in the CUORE-0 experiment, which was operated in a smaller cryostat as the predecessor to CUORE [59]. After demonstrating that this CUORE-style tower met the radiopurity and detector performance standards needed for CUORE, the rest of the towers were assembled and installed in the CUORE cryostat. An image of the CUORE towers shortly after installation but before the cryostat was closed up for the cooldown is shown in Fig. 3.5.

Besides all of these components needed for actual cryogenic operation, CUORE was originally also equipped with an internal calibration system [60]. This system featured the ability to deploy ^{232}Th calibration strings directly into the innermost stage of the cryostat, circumventing the problem of trying to calibrate the detectors with a source outside the many layers of shielding that are situated to protect the detectors from environmental radiation. In principle, this internal calibration system would allow a more even distribution of calibration statistics on the CUORE detectors and the ability to resolve calibration peaks from low-energy γ particles that would otherwise be mostly blocked by CUORE's shielding, all without disturbing the cryostat's operational abilities. Since the calibration strings were in close proximity to the detectors, we were able to obtain a $\sim$100 mHz calibration event rate in each detector with relatively low ^{228}Th activities of 3.6 Bq in each of the 6 innermost calibration strings and 19.4 Bq in each of the 6 outermost strings. However, this system was retired after CUORE's third dataset due to the fact that it did sometimes interfere with the cryogenic system. Instead, it was found that mixed ^{232}Th-^{60}Co sources could be deployed outside the cryostat as 8 evenly spaced calibration strings, each with an activity of 5.3 kBq of ^{232}Th and 5.0 kBq of ^{60}Co, and still provide enough calibration lines with sufficient statistics for a satisfactory calibration of the detector response. This comes at the cost that the calibration event rates are no longer evenly spread among the detectors, with the detectors closest to the outer shields receiving much higher rates than the detectors in the very middle of the cryostat, but this has minimal effect on the resulting detector performance and is well-worth the ability to perform calibrations with a system entirely external to the cryogenic apparatus.

CUORE began its first dataset in 2017, and the rate at which it has been accumulating exposure since then is shown in Fig. 3.6. There were two periods of extended maintenance early on, during which the stability and performance of the cryogenic apparatus were improved. The current external calibration system was introduced during the second maintenance period. Since the end of this second maintenance in early 2019, CUORE has been running stably and accumulating data at an almost continuous rate, with only brief pauses in between datasets for calibration and routine chores. Most recently, data corresponding to 1120 kg$\cdot$yr of raw TeO_2 exposure has been unblinded and used for a $0\nu\beta\beta$ analysis. The next chapters discuss the data collection and analysis procedure for this time period, as well as the final results of this analysis yielding CUORE's most recent $0\nu\beta\beta$ limit.

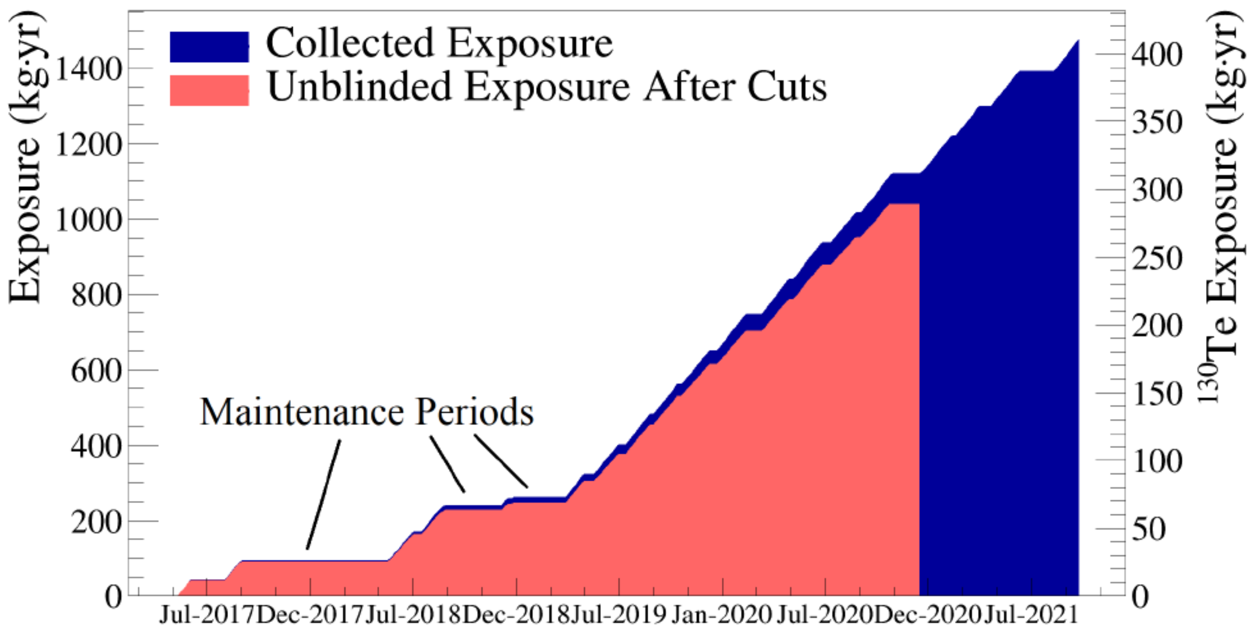


Figure 3.6: Exposure accumulation since CUORE first began taking data. The cumulative TeO_2 exposure is labeled on the left y-axis, and the corresponding ^{130}Te exposure relevant for a $0\nu\beta\beta$ analysis is labeled on the right y-axis. The long pauses correspond to extended periods of maintenance that were imposed to improve the performance and stability of the cryogenic apparatus. Since the beginning of 2019, CUORE has not required such long downtimes and has been accumulating exposure at an almost continuous rate, with only brief pauses between datasets for short routine maintenance and for calibration.

Chapter 4

CUORE Data Analysis

Starting from the raw data of pulses in voltage obtained from reading the current-biased NTDs on the CUORE crystals, there is a good deal of processing needed to clean it and properly understand each event. This chapter details the process by which CUORE collects its data, structures it, and finally processes it all into an energy spectrum of good events suitable for a physics analysis.

4.1 Data Collection Procedure

CUORE data are organized into *datasets*, which are further divided into *runs*. Each run in a dataset consists of roughly 24 hours of data and is classified as either a *calibration run* or a *physics run*. Calibration runs involve the deployment of radioactive sources, whose known decay lines are used to calibrate the response of our detectors. In the first three datasets of CUORE, this was done with internal ^{232}Th sources through the previously mentioned internal calibration system. For the fourth dataset this was done with an external ^{232}Th source, and for all datasets since then this has been done with an external mixed ^{232}Th-^{60}Co source. Physics runs collect data that are used for actual physics analyses (referred to as *physics data*), where the events in the detectors are primarily from natural radioactive decays, whether they are environmental backgrounds or $\beta\beta$ decays. Each dataset generally lasts for one to two months, beginning and ending with a few days of calibration runs that bookend a long set of physics runs. Between datasets, other types of optimization runs are sometimes performed as well. These can include load curve characterization for the NTD sensors on each crystal to find the optimal bias points, pulse tube phase scans for noise minimization, and tests of detector response to different amounts of injected energy using the heaters. Any particular dataset will be completed using roughly the same operating conditions so that the overall performance is consistent across the dataset, but changes in the operating conditions can (but do not necessarily) occur between datasets. The room temperature front-end electronics are all designed to have small thermal drift and high gain stability over the timescale of many datasets, so that any variations in detector performance between datasets primarily do not come from the electronics [52].

During data collection, the previously described digitized data streams of each detector are continuously read out into shared memory on the DAQ computers. Because our signals are slow, we can use a relatively low sampling rate of 1000 samples per second for each of the 1024 channels that are digitized[1], which makes it possible to perform all triggering in software. This also allows us to copy all raw data off the DAQ computers to be permanently stored in remote server farms for later reanalysis using more complex trigger algorithms, with the data accumulating at a manageable rate of ∼3.5 terabytes per month. For live data-quality monitoring purposes, the DAQ computers run a derivative trigger (DT) algorithm on the data as it comes in and saves the timestamps of each trigger [61]. This algorithm checks for a rise in voltage over some period of time that greatly exceeds what could be expected from normal baseline noise. The DT computes a moving average dV/dt over the most recent 40 ms of data and triggers once it exceeds some tunable threshold. Combined with a deadtime of 1 second after each trigger, the threshold is tuned depending on the noise levels of each dataset so that the trigger rate is not above a few mHz per detector. This triggering serves as a way to flag possible events of interest so that they can be subjected to basic monitoring procedures. For the purposes of analysis, an *event* in CUORE is defined as a 10 second time window consisting of 3 seconds before the trigger time and 7 seconds after. In addition to the derivative trigger to identify physical events, random noise triggers are also injected every ∼80 seconds on average so that the baseline conditions can be analyzed. The silicon heaters are activated at regular intervals of about once per 10 minutes for each detector to check for variations in detector response, and these are explicitly flagged in the trigger system as heater events as well.

The derivative-triggered events and noise events are used to construct the average pulse (AP) and average noise power spectrum (ANPS) for each detector, which are used in the optimum filter. Besides this, the live data monitoring is used to flag time periods of excessive noise or otherwise unstable conditions[2], which are then excluded from the rest of the analysis. This live data processing also allows us to check that calibrations are proceeding with sufficient statistics and to check for variations in calibration response between the beginning and end of a dataset. In total, live processing goes through amplitude evaluation, stabilization, and calibration in order to allow us to perform these checks, which shall be covered in the following section detailing the entirety of the offline analysis chain.

[1]Besides the 988 channels from the calorimeters, the extra channels are used for various diagnostic and auxiliary devices that help monitor the cryostat conditions, such as microphones and accelerometers.

[2]We contend with the fact that Italy frequently experiences small earthquakes, which the CUORE detectors are sensitive enough to be affected by even when the epicenters are fairly far away. Luckily, it's easy to identify and eliminate time periods where the cryostat was disturbed by an earthquake, and the overall resulting loss of livetime is low.

4.2 The Analysis Chain

Optimum Filter and Triggering

After collection of a dataset is completed, all of the calibration and physics runs of that dataset are subjected to an offline reprocessing, which involves a fuller and more sophisticated suite of techniques compared to the live data monitoring. The first step of this is using an optimum trigger to re-identify events in the data stream, replacing the job done by the derivative trigger. This uses the optimum filter, which is also used for amplitude evaluation. The optimum filter optimizes the signal-to-noise ratio when there is a known signal template and known noise power spectrum, using the transfer function in frequency space for each component frequency ω_k:

$$H(\omega_k) \propto e^{i\omega_k t_{max}} \frac{s^*(\omega_k)}{N(\omega_k)} h(\omega_k)$$

Here, $H(\omega_k)$ is the resulting filtered waveform, $s(\omega_k)$ is the Fourier transform of the signal template, $N(\omega_k)$ is the spectral power density of the noise, $h(\omega_k)$ is the actual waveform being filtered, t_{max} is the position of the maximum point in the time window being filtered, and the proportionality is only up to a normalization factor so that the filter has unity gain. This takes advantage of the difference in the power spectra of typical signal-like pulses and the ambient noise, shown in Fig. 4.1. The signal templates are obtained by taking the average pulse of derivative triggered events that exceed some tunable amplitude threshold in each detector, and the noise templates are obtained from the random noise triggers. Evaluating events by the amplitude of their optimum-filtered pulses suppresses the contribution of noise and improves the energy resolution compared to using the amplitude of the unfiltered pulses. The optimum trigger builds on this by also triggering on the optimum-filtered pulses instead of the raw pulses like the derivative trigger does [62]. Since this filter more readily distinguishes low-energy pulses from noise, the primary gain here is a drastic lowering of the energy threshold, which on average drops from tens of keV to < 10 keV. This doesn't affect our ability to trigger on events with MeVs of energy like $0\nu\beta\beta$ candidate events, but it provides an auxiliary benefit by improving multiplicity analyses (see later in this section). The gains offered by the optimum trigger are of particular interest for future low-energy analyses, as well as for improving the background model of CUORE.

Stabilization

After the optimum-filtered amplitudes of every triggered event are calculated, we must account for possible variations in detector response due to slow drifts in their baseline operating temperatures. We calculate the *baseline* for each event by looking at the 2.25 seconds spanning the beginning of the event's time window to 75% of the time elapsed before the trigger, which is defined to be 3 seconds into the time window. A linear fit is performed on this subset of the pretrigger time period, and the mean readout voltage is what we define as the event's baseline before the physical pulse occurred. This pretrigger voltage essentially serves

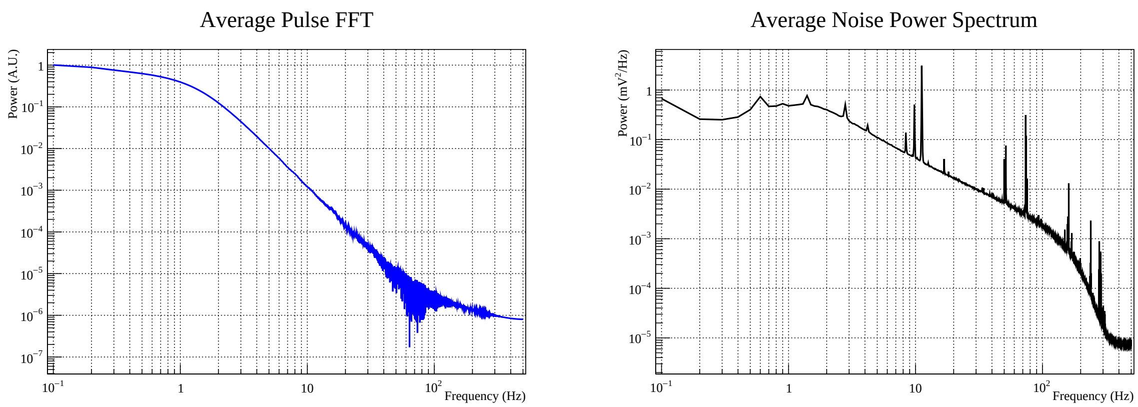


Figure 4.1: **Left**: a normalized power spectrum of the average pulse of a typical CUORE detector, which is constructed from the data to approximate its "good" response to a physical event. **Right**: the average noise power spectrum of the same detector from the same dataset, obtained from sampling the noise during periods of time without real triggers. One can see the noise features a number of characteristic peaks absent in the signal's power spectrum, which the optimum filter helps suppress.

as a measure of the detector's NTD resistance before the energy deposit, which in turn serves as a proxy for the detector's pre-event temperature. Events with a baseline slope that is too large are rejected as bad events, since this indicates they were not in a stable state and we cannot properly model their expected response to a physical event of any given energy.

Most detectors use the regular heater pulses to stabilize their response across drifting baselines. An example of one such pulse is shown in Fig. 4.2, where we can see the pulse baseline and amplitude in voltage space. Since the heater events are always injected at the same energy for each dataset, usually in the 2-3 MeV range, we can use them to track how a calorimeter's amplitude response depends on its baseline. A linear fit is performed for the amplitudes of these heater events against their baselines for each detector in each run. This fit is used to stabilize the amplitude of all events for the detector as a function of their baselines, implicitly assuming that the amplitude response only varies linearly with the baseline and that this linear variation is the same for all energies. This additionally assumes that the change in heater response due to different baselines is the same as the change in response to radiative energy deposits. Experience has shown these assumptions to be generally true[3], so this method of using the heater pulses provides a mostly reliable way to track the detectors' change in response in real time as the baseline undergoes small drifts. This process is referred to as heater-based thermal gain stabilization (heaterTGS).

[3]We have seen that the energy resolution of heater events tends to be different from those of physics events, but tracking the variation in amplitude over different baselines using the heaters and applying the result to physics event stabilization still seems valid.

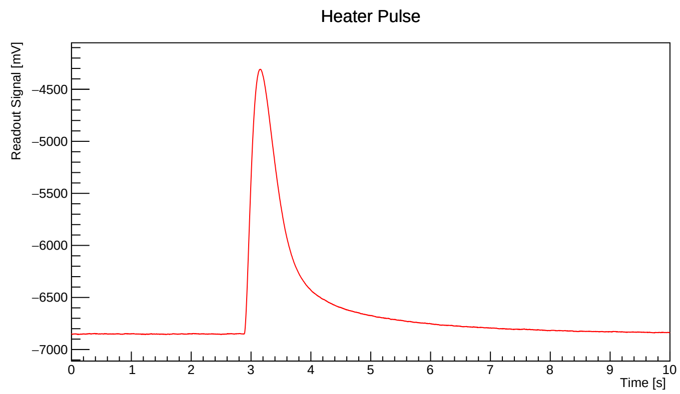


Figure 4.2: A typical pulse caused by the activation of the heater in one of the CUORE calorimeters. From pulses like this, we can track how the heater pulse amplitude depends on the baseline value at which the heater was activated, yielding trends such as that shown in Fig. 4.3.

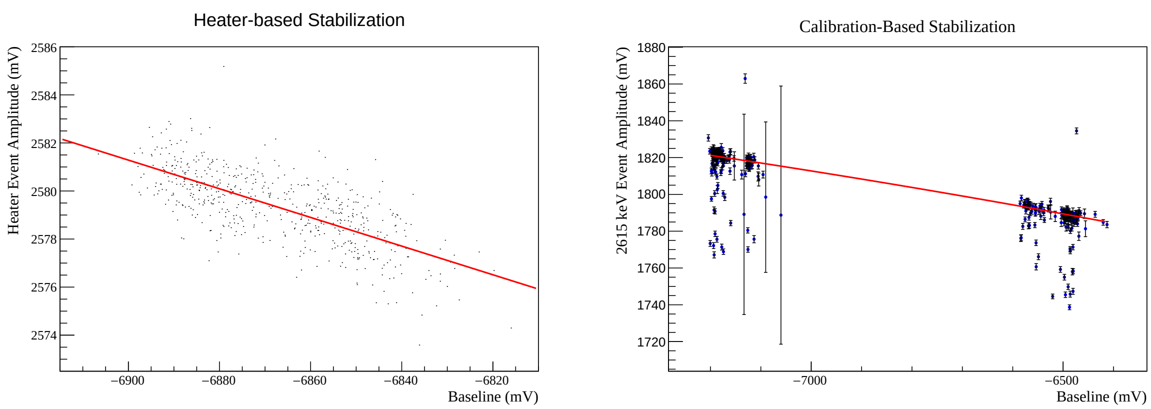


Figure 4.3: Amplitude (in arbitrary uncalibrated units) versus baseline (in mV) for heater events and 2615 keV calibration events in the same detector from the same dataset, showing how the heaterTGS and calibrationTGS algorithms work and showing the general linear dependence of the amplitude response on the baseline. **Left**: amplitudes of heater events versus baseline for one detector over 5 consecutive runs, with the linear fit that is used for the heaterTGS correction drawn in red. **Right**: amplitudes of 2615 keV calibration events versus baseline, also showing the linear fit function used for the calibrationTGS correction drawn in red.

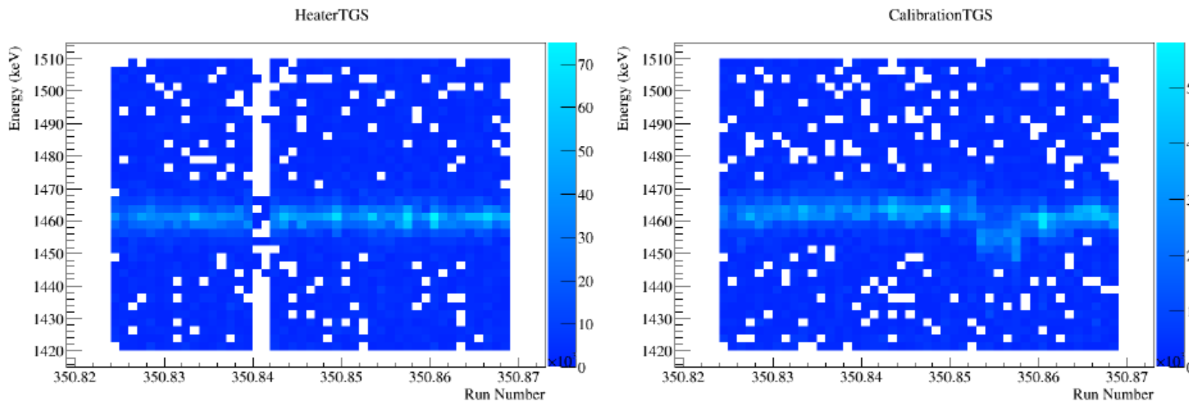


Figure 4.4: Heat map of events near the 1460 keV γ peak from natural ^{40}K decays in physics data over the course of the many runs comprising one dataset. **Left**: the energies obtained using heaterTGS, where the peak around 1460 keV is clearly visible on top of the roughly flat background, and where one can see the peak's position is stable across all runs. Events for some runs are missing because many heaters failed during those runs, making application of the heaterTGS algorithm impossible. **Right**: the energies obtained using calibrationTGS, where one can see the peak position drifts for a few runs. This indicates that the calibrationTGS procedure does not properly stabilize the amplitudes in those runs, a result of a failed interpolation from the calibration runs at the beginning and end of this dataset to the operating conditions in the middle of the dataset.

Out of the 984 functional detectors in CUORE, 29 of them have nonfunctional heaters due to issues during installation. In addition, some detectors' heaters sporadically have stability issues for particular runs or datasets that make them unreliable for a heater-based stabilization. To handle these detectors, we employ another algorithm called calibration-based thermal gain stabilization (calibrationTGS). This relies on our knowledge that the highest energy line from our calibration will always be the 2615 keV ^{208}Tl γ peak and that the variation in amplitude response with different baselines is small enough that this peak will still be clearly visible in unstabilized data. These physical events from the calibration can then basically be treated the same way as the manually injected heaters; they should mostly correspond to actual deposits of 2615 keV of energy, so we can perform a linear fit of the actual amplitude responses versus the baselines they occurred at to stabilize the response. Both the heaterTGS and calibrationTGS methods are applied to each detector where possible, with examples of their application to the same detector from one dataset shown in Fig. 4.3. This results in two amplitude estimators for every event, from which we choose one for each detector after calibrating.

The disadvantage of the calibrationTGS method is that we only have events from the opening and closing calibrations of the dataset, so we have to interpolate to any other baselines at which the intervening physics runs may have operated. This interpolation can sometimes miss the mark, resulting in improperly stabilized amplitudes for some of the operating conditions in the middle of a dataset. On the other hand, the heaterTGS method uses heater events that are present in every run, so it uses information being obtained in real time alongside the actual physics data. This can result in stark differences in performance like that shown in Fig. 4.4, where the calibrationTGS method fails to properly stabilize the amplitudes of events collected under specific operating conditions in the dataset but the heaterTGS method is fine, even though both had reasonable performance in calibration data. Due to the fact that datasets are up to two months long, some of the physics runs in this case had notably different conditions than the calibration runs, so that the calibrationTGS interpolation fails. For this reason, we default to using heaterTGS where possible and only use the calibrationTGS method when the heaterTGS is either unavailable or has clearly failed in some way.

Calibration

After we obtain stabilized amplitudes for all events in a dataset, we can apply the calibration procedure to determine the physical energy deposits that they correspond to. This is done by looking at the stabilized amplitude spectra for each detector in the calibration data and identifying the physical peaks that we expect to see from our calibration sources. With the known energies of the peaks, this gives us a collection of points to match stabilized amplitudes with energies, which we then fit with a calibration function to convert the entirety of the stabilized amplitude spectrum into an energy spectrum. For each detector, we use an empirically determined calibration fit function of form:

$$E = a_1 x + a_2 x^2$$

where x is the stabilized amplitude of an event, E is the corresponding energy, and a_1, a_2 are the free parameters determined by the calibration procedure. This imposes a requirement that 0 amplitude is 0 energy and allows us to include nonlinear effects in the responses of our detectors. As mentioned before, CUORE has tried an internal calibration system using a ^{232}Th source string before moving to an external calibration system, which initially used only ^{232}Th strings but later had ^{60}Co added into them as well. A comparison of the resulting calibration spectra from each of these methods is shown in Fig. 4.5, where one can see that the internal calibration system did indeed provide more peaks to calibrate on, even if it was dropped for its technical complexity costs discussed in Chapter 3.

For the internal calibration system, the calibration function is determined with fits to the positions of the peaks at 239, 338, 583, 911, 969, and 2615 keV. These come from ^{212}Pb, ^{228}Ac, ^{208}Tl, ^{228}Ac, ^{228}Ac, and ^{208}Tl decays respectively, which are all part of the ^{232}Th decay chain. They were picked as the most prominent peaks in the calibration spectrum, distributed over a healthy range of energies to properly constrain the calibration function. The external ^{232}Th calibration uses only the peaks at 511 and 2615 keV, which are the only

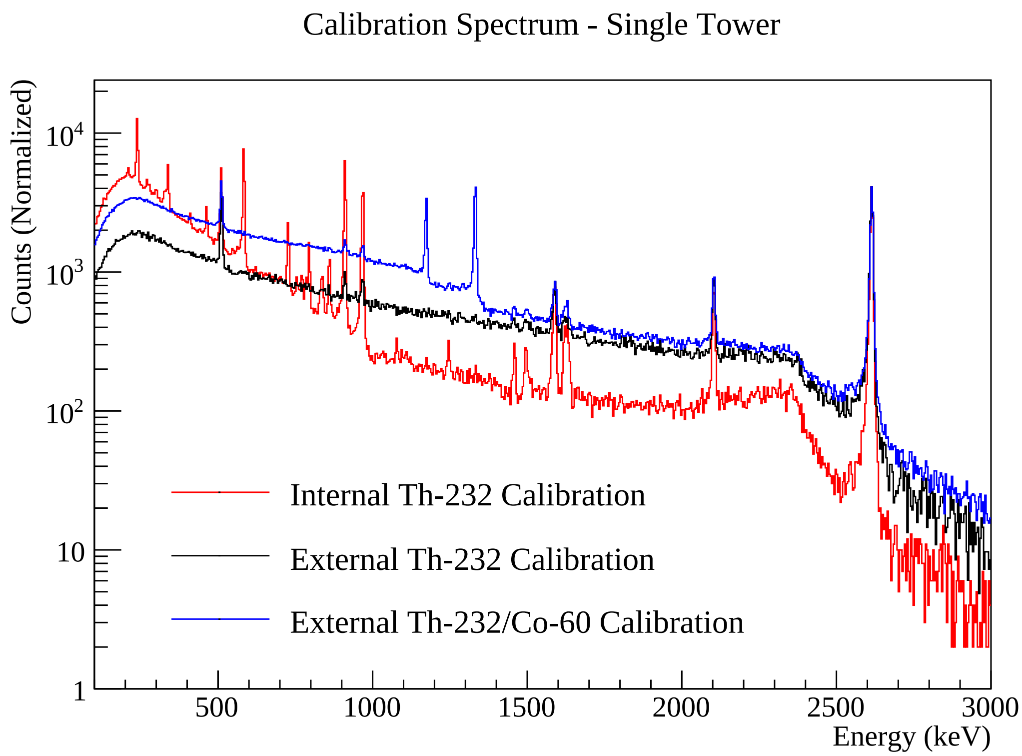


Figure 4.5: Overlaid calibration spectra for a single tower in 3 different datasets, which respectively used the internal ^{232}Th calibration system, external ^{232}Th calibration strings, and external mixed ^{232}Th-^{60}Co calibration strings. The spectra are normalized to have the same peak height at 2615 keV. One can see that the internal calibration system allows us to see many more low-energy peaks that can help constrain the calibration function, while these low-energy peaks are heavily suppressed or nonexistent when using the external calibration system.

peaks with sufficient statistics to calibrate on, as the other low-energy γ rays are mostly blocked by the cryostat shielding[4]. This has the notable flaw that it leaves only 2 points to fit a calibration function that has 2 free parameters. It also has no constraining peaks for the entire 2 MeV of energies between the 2 peaks, which turns out to be an energy range of great interest to a $0\nu\beta\beta$ search looking at the ^{130}Te Q-value of 2528 keV. This motivated the

[4]Although there is a 510.8 keV γ ray from ^{208}Tl decays, this 511 keV peak is actually primarily not the result of that decay, but is a γ ray coming from e^+e^- annihilation. This pair annihilation is possible after e^+e^- pair production first occurs through the interaction of a high-energy γ ray in one of the crystals. This allows the 511 keV peak to still be prominent when using an external calibration source. The overlap with the natural γ decay causes the peak to be broadened, which is why it wasn't used in the internal calibration, but we have little choice for external calibrations, and its peak position is expected to be good enough for calibration purposes.

addition of the ^{60}Co sources to the external calibration strings, which brings in two more γ peaks at 1173 and 1333 keV.

Calibration is performed for both of the stabilized amplitude estimators, from heaterTGS and from calibrationTGS. This results in two corresponding energy estimators. At this point, one of the energy estimators is chosen for each detector in a dataset, defaulting to the heaterTGS-derived energy value but using the calibrationTGS-derived energy when it is either necessary or has obviously better performance. After this, we finally have an energy spectrum for the physics and calibration data of each detector, and from here we can perform a few more advanced analyses to further clean the data.

Coincidence Analysis

While the cryogenic calorimetric detection method does not allow us to pinpoint the location of an interaction within any particular crystal, we do have 988 crystals spread throughout the cryogenic volume of CUORE. By checking for whether multiple crystals register energy deposits at the same time, we can gather additional information about the nature of the event or events that caused the deposits. Given the low event rates in the physics data, simultaneous energy deposits in two nearby crystals are more likely to be caused by some kind of shared source than to be completely uncorrelated events that just randomly happened at the same time. To give a couple of examples, this could be a γ ray from a natural background scattering in one crystal and then depositing the rest of its energy in another crystal, or it could be an α decay near the surface of one crystal depositing part of its energy in its source crystal and the rest in an adjacent crystal. This ties into another advantage of CUORE's source=detector setup. Monte Carlo simulations tell us that slightly over 88% of $0\nu\beta\beta$ events should have their energy fully contained in the crystal that they originated in [63]. We can then limit ourselves to considering events where only one crystal registers an energy deposit, rejecting events with energy deposits in multiple crystals at once. This allows us to capture most of the potential $0\nu\beta\beta$ signal while rejecting a significant number of potential background events. This is one of the primary motivations for the coincidence analysis of CUORE events.

We define the multiplicity of an event as the number of detectors that register energy deposits in coincidence with each other. This requires the specification of 3 factors that determine what counts as a coincidence:

1. **Energy threshold**: the minimum energy of a triggered event to count towards a coincidence. Lower energy thresholds let us capture more true coincidences but also increase the chance of including random noise that got triggered as a physics-like event.

2. **Distance**: the maximum distance between two crystals that allows them to count as being in coicidence. Physically correlated events will generally occur in nearby crystals, so simultaneous energy deposits in two crystals on opposite sides of the cryostat are more likely to be due to random chance.

3. **Time window**: the maximum time between two crystals' events that allow them to count as being in coincidence. Even α particles, the slowest particles that can still travel between crystals to deposit energy in more than one, are fast enough that the two energy deposits should be less than 1 ms apart (which is the rate at which we sample our data). However, both the timing resolution and timing accuracy of our events are imperfect, so we must pick a time window wide enough to capture coincidences of interest while not being so wide that we capture uncorrelated events that are just happening close in time. A number of algorithms are used to try to correct for "jitter" between the detectors that cause systematic errors in the calculation of their timing differences. One tries to synchronize different detectors by comparing their delays between the nominal trigger times of events and their pulse peaks, and another tries to synchronize detectors in the same tower by looking at coincidences in calibration data, which have much higher statistics than physics data.

For the $0\nu\beta\beta$ analysis, we choose an energy threshold of 40 keV, no distance limit, and a time window of 5 ms. The energy threshold of 40 keV is comfortably above the trigger threshold for most detectors and is low enough that it can identify 2615 keV γ rays that could enter our $0\nu\beta\beta$ analysis region of 2490 to 2575 keV by Compton scattering in one crystal before deposting the rest of their energy in another (e.g. by splitting its energy as 40 keV in one detector and 2575 keV in another). This maximizes our ability to reject potential backgrounds to the $0\nu\beta\beta$ analysis through the coincidence analysis. The relatively narrow time window of 5 ms leaves a small chance of getting accidental coincidences with the unlimited distance window, and the absence of a limit on distance allows us to capture correlated noise that may not be in adjacent detectors or even just γ scattering events where the γ travels to a distant detector after scattering in a different one. The low trigger threshold provided by the optimum trigger aids in this, as it allows us to reliably trigger on > 40 keV events that factor into this multiplicity analysis.

Other choices for these values to calculate coincidences are possible for other types of analyses. For instance, we also calculate coincidences using an energy threshold of 70 keV, distance radius of 15 cm, and time window of 150 ms, which is more useful for background model studies. A 15-cm distance cut corresponds to only allowing coincidences between a detector and any of its immediate neighbors, which are the other crystals on the same or adjacent floors in the same tower. This permits the use of a larger time window to better study α events involving a nuclear recoil in one crystal with the α being deposited in an adjacent crystal. Understanding these types of events helps us better characterize our natural α-emitting backgrounds.

An example of the energy distribution of multiplicity-2 events (events with energy deposits in 2 crystals in coincidence) for one dataset is shown in Fig. 4.6, using the 40-keV threshold and 5-ms time window for the $0\nu\beta\beta$ analysis. One can see distinct diagonal bands of events corresponding to natural monoenergetic γ decays that split their energy between two detectors. There are also vertical and horizontal bands corresponding to correlated decays, such as the ^{208}Tl γ rays at 2615 keV and 583 keV, or the ^{60}Co decays at 1173 and 1333

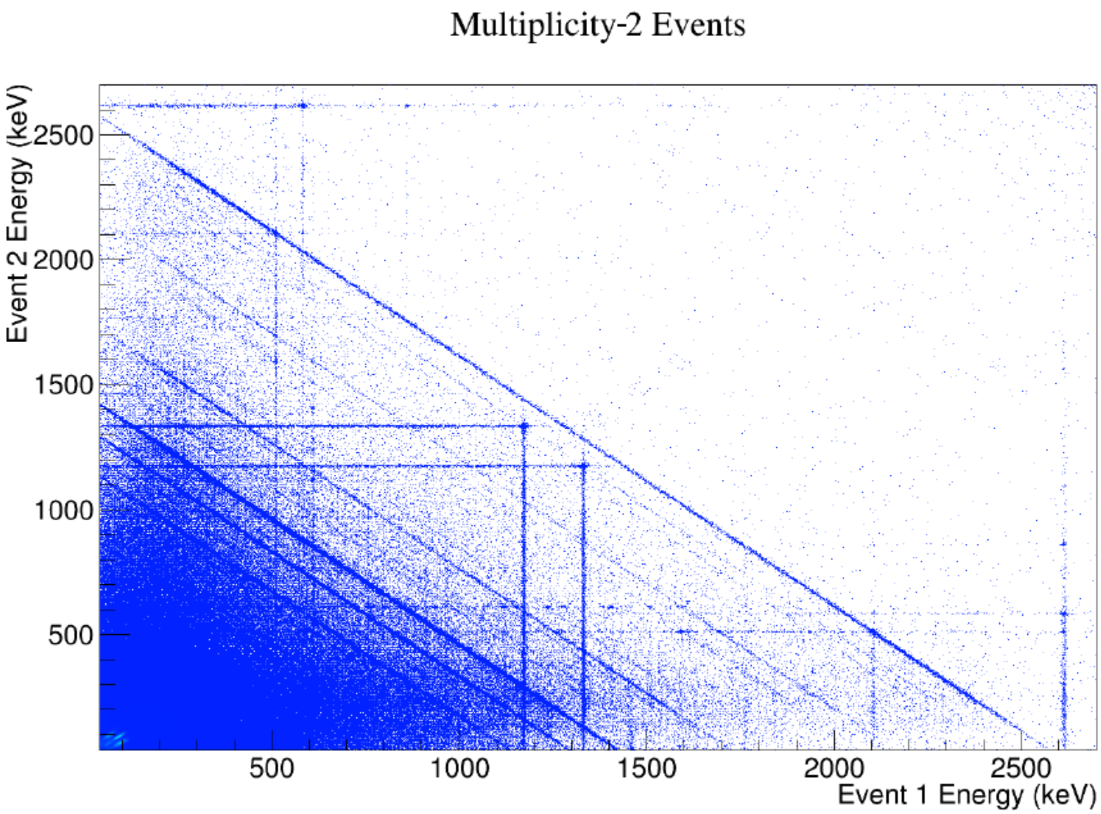


Figure 4.6: Energy spectrum of multiplicity-2 events for one dataset, using the 40-keV threshold and 5-ms window described in the text. For each pair, the energy of the "first" event is on the x-axis, and the energy of the "second" event is on the y-axis (but which is first and which is second is arbitrary). There are distinct diagonal lines corresponding to natural γ peaks at 1173, 1333, 1460, and 2615 keV, where they scatter in one detector and deposit part of their energy there first before eventually reaching another detector and depositing the remainder of their energy. As a result, we can see these populations where the energies of event 1 and 2 sum up to the γ-peak energy.

keV. All events falling in this multiplicity-2 spectrum end up being cut for a $0\nu\beta\beta$ analysis, where for now we can consider only multiplicity-1 events (events with just one detector registering an energy deposit).

Pulse Shape Analysis

Lastly, we perform a pulse shape analysis to remove any events that are due to pileup or are otherwise nonphysical in some way. This can be due to mechanical or electrical noise that cause transient instabilities in the detector operating conditions that pass our trigger algorithms and are able to emulate the characteristics of normal events caused by physical radiation. We include basic data quality cuts at the beginning of the processing to eliminate obvious cases of pileup or noise, but these are often insufficient to catch all undesirable events. This procedure was previously done with six pulse-shape parameters calculated for every event:

- Rise time: time for the unfiltered pulse to rise from 10% of its maximum value to 90% of the maximum value
- Decay time: time for the unfiltered pulse to decay from 90% of its maximum value to 10% of that value
- Optimum filter delay: the time between the trigger time and the pulse maximum in the optimum filtered waveform
- Optimum filter test value (left): a χ^2 measure of the similarity of the leftside part of the optimum filtered waveform to the optimum filtered average pulse
- Optimum filter test value (right): a χ^2 measure of the similarity of the rightside part of the optimum filtered waveform to the optimum filtered average pulse
- Baseline slope: the slope of the baseline in the unfiltered waveform before the trigger

These 6 pulse-shape variables were normalized as a function of energy for each detector in a dataset, using empirically chosen fit functions to match their dependence on energy. From this, the 6 normalized variables could be combined into one normalized measure using a Mahalanobis distance MD. For each event with its normalized pulse-shape parameters contained in a 6-dimensional vector $\boldsymbol{x}$, this distance is defined as:

$$MD(\boldsymbol{x}) = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^{\intercal}\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

Here, $\boldsymbol{\mu}$ is the vector containing the mean values of the 6 normalized pulse-shape parameters over all events for a detector in a dataset, and $\boldsymbol{S}$ is the covariance matrix of these normalized parameters. We could then cut events that had too large of a Mahalanobis distance. This would correspond to events that had one of their pulse-shape variables be unusually high or low, or just had a combination of several of the pulse-shape variables being slightly unusual. While this technique was generally successful in cutting bad events while maintaining decent efficiency on good events, we sometimes experienced issues with pathological normalizations causing entire chunks of events from certain energy ranges to be cut. This could generally be solved by eliminating the corresponding detectors from the analysis after identifying them through extensive manual inspection, but this loss in efficiency led us to drop this approach. Instead, we have adopted a procedure based on principal component analysis, which shall be the topic of the next chapter.

Chapter 5

Pulse Shape Discrimination

Pulse shape discrimination (PSD) is the process by which we distinguish "good" pulses from "bad" pulses in CUORE, as determined by considering their shapes. This occurs after all other processing, which has already calculated traits of events like the energy and multiplicity. From a basic understanding of the physics of a cryogenic calorimeter, we know a detector's response to an energy deposit should look something like that shown in Fig. 3.2, characterized by a flat baseline, a sharp rise near the trigger time, and a gradual smooth decay back to the baseline afterwards. Loosely defined, a good pulse is an event caused by a single physical energy deposit from some kind of radiation, where the detector has responded the way we expect it to. These events include natural radioactive backgrounds and $\beta\beta$ decay candidate events. A bad pulse is an event that either has multiple energy deposits at once or is not caused by radiation at all, instead being some kind of noise or change in detector conditions that has caused a response similar to what we expect from a radiative event. This is a very broad level of PSD, not even attempting to distinguish between different types of "good" events, which is rather difficult with CUORE's detectors.

This chapter discusses the principles of principal component analysis, as well as how we have incorporated it into CUORE to perform pulse shape discrimination at a better level than we were previously able to achieve using basic pulse-shape variables.

5.1 Principal Component Analysis

The basic idea of principal component analysis is to transform data into a new basis ordered by how much of the data's variance lies along each coordinate [64]. Studying the data through this transformed basis better reflects its interesting aspects, allowing for dimensionality reduction and feature extraction. For instance, consider studying a collection of photos of the sky. The data have a naturally high dimensionality given by the number of pixels in the photos, but most of the pixels may be of just the blank blue sky, which is probably not interesting. A principal component analysis (PCA) would allow one to determine where the variation between the photos is, which is probably more interesting for analysis. PCA is thus frequently used as a transformation before feeding data into a machine-learning algorithm, as this does some of the preliminary work of reducing high-dimensional data into a few components which are likely to be important for whatever the algorithm may be trying to do.

This sort of transformation naturally has an application to studying the form of the pulses that constitute CUORE events as well. A CUORE pulse is a 10000-dimensional piece of data (from a 10 second data stream sampled at a rate of 1 kHz), but many of the samples are often uninteresting; for instance, the first 2000 samples will normally be mostly flat for an event and can be summarized in far fewer numbers, but we collect it all because sometimes there will be an unusual feature that needs to be accounted for. PCA provides a method to reduce the 10000-dimensional pulses to just a few most interesting features, which we can then use to perform additional pulse shape analysis beyond what has already been done in the processing chain.

From a mathematical perspective, one in general considers a set of N pieces of data, each being D-dimensional. This can be assembled as a $N \times D$ data matrix $\boldsymbol{X}$, where each row is a vector $\boldsymbol{x}_n = (x_1, x_2, ..., x_D)$ containing the data of one observation. We want to project this data into a number of dimensions $M < D$, with the direction of each of the M dimensions determined by a D-dimensional unit vector $\boldsymbol{u}_i$ satisfying $\boldsymbol{u}_i^\intercal \boldsymbol{u}_i = 1$. If we consider the first coordinate $\boldsymbol{u}_1$ of this new M-dimensional basis, each data vector $\boldsymbol{x}_n$ can be projected onto it as a scalar $\boldsymbol{u}_1^\intercal \boldsymbol{x}_n$. Our goal is to maximize the variance of the projections onto this first coordinate vector, given by:

$$\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{u}_1^\intercal \boldsymbol{x}_n - \boldsymbol{u}_1^\intercal \bar{\boldsymbol{x}})^2 = \boldsymbol{u}_1^\intercal \boldsymbol{S} \boldsymbol{u}_1$$

where we introduce the terms $\bar{\boldsymbol{x}}$ as the mean of the data vectors and $\boldsymbol{S}$ as the data's covariance matrix, defined by:

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n$$

$$\boldsymbol{S} = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_n)^\intercal$$

This now leaves us with the task of finding the $\boldsymbol{u}_1$ that maximizes $\boldsymbol{u}_1^\intercal \boldsymbol{S} \boldsymbol{u}_1$ while obeying the constraint that $\boldsymbol{u}_i^\intercal \boldsymbol{u}_i = 1$. This can be done by introducing a Lagrange multiplier λ_1 and maximizing with respect to $\boldsymbol{u}_1$ the expression:

$$\boldsymbol{u}_1^\intercal \boldsymbol{S} \boldsymbol{u}_1 + \lambda_1(1 - \boldsymbol{u}_1^\intercal \boldsymbol{u}_1)$$

Taking $\partial/\partial \boldsymbol{u}_1$ tells us that the stationary points satisfy:

$$\boldsymbol{S}\boldsymbol{u}_1 = \lambda_1 \boldsymbol{u}_1$$

This is a statement that $\boldsymbol{u}_1$ must be an eigenvector of $\boldsymbol{S}$. A simple rearrangement tells us that $\boldsymbol{u}_1^\intercal \boldsymbol{S} \boldsymbol{u}_1 = \lambda_1$, so the variance of the projections onto $\boldsymbol{u}_1$ is maximized when λ_1 is maximized. Our solution for $\boldsymbol{u}_1$ is thus the eigenvector of $\boldsymbol{S}$ with the largest eigenvalue. We call $\boldsymbol{u}_1$ the leading principal component of the data. To obtain the subsequent vectors $\boldsymbol{u}_i$, we simply continue maximizing $\boldsymbol{u}_i^\intercal \boldsymbol{S} \boldsymbol{u}_i$ while imposing an orthogonality requirement that $\boldsymbol{u}_i^\intercal \boldsymbol{u}_j = 0 \; \forall \; i \neq j$. The general solution for the leading M principal components turns out to be the eigenvectors of $\boldsymbol{S}$ with the M largest eigenvalues $\lambda_1, \lambda_2, ..., \lambda_M$. The proof follows by straightforward induction: since we have already shown this is the case for $M = 1$, assume it is true for some general M and consider how to obtain the $(M+1)$th component. In order to maximize the variance along this new component while requiring orthogonality to the existing principal components, introduce $(M+1)$ Lagrange multipliers denoted y_i and maximize the expression:

$$\boldsymbol{u}_{M+1}^\intercal \boldsymbol{S} \boldsymbol{u}_{M+1} + y_{M+1}(1 - \boldsymbol{u}_{M+1}^\intercal \boldsymbol{u}_{M+1}) + \sum_{i=1}^{M} y_i \boldsymbol{u}_{M+1}^\intercal \boldsymbol{u}_i$$

Taking the derivative with respect to the desired new vector $\boldsymbol{u}_{M+1}$ and setting it to 0 gives the stationary points defined by:

$$2\boldsymbol{S}\boldsymbol{u}_{M+1} = 2y_{M+1}\boldsymbol{u}_{M+1} - \sum_{i=1}^{M} y_i \boldsymbol{u}_i$$

Multiplying both sides by $\boldsymbol{u}_{M+1}^\intercal$ and recalling the orthonormality conditions gives:

$$\boldsymbol{S}\boldsymbol{u}_{M+1} = y_{M+1}\boldsymbol{u}_{M+1}$$

This is the same form we saw for the case of $M = 1$, telling us that $\boldsymbol{u}_{M+1}$ is the eigenvector of $\boldsymbol{S}$ with the largest eigenvalue λ_{M+1} not already taken by one of the other components.

The overall outcome of the principal component analysis procedure is we obtain M principal components, where M is chosen by us, and the corresponding projections of the data onto each of these components. The forms of the principal components $\boldsymbol{u}_1, ..., \boldsymbol{u}_M$ reduce the data into the M most "important" parts, and the projection $\boldsymbol{u}_i^\intercal \boldsymbol{x}$ of a data vector $\boldsymbol{x}$ onto the component $\boldsymbol{u}_i$ provides information about how that principal component features in $\boldsymbol{x}$. Deciding how to extract and make use of this information is central issue when applying PCA to a specific set of data.

5.2 PCA in CUORE

Using PCA to analyze data requires that we first have a way of extracting the principal components that best describe the data. By its nature, PCA is very sensitive to outliers in data, since these will tend to dominate the variance that we are trying to maximize. However, these outliers tend not to be what we most care about, since we are usually trying to obtain components that provide information about the bulk of events. The process of extracting the principal components is what we call training, and this requires us to first obtain a mostly clean sample of events. Once reasonable principal components are calculated using this training sample, the rest of the data can be projected onto these components for study as well.

For CUORE, individual detectors can have different characteristic responses to energy deposits, so each detector should be assigned its own principal components. The training can be attempted using events from the calibration γ peaks, which will both provide much greater statistics compared to using physics data and consist of mostly "good" events[1]. The resulting PCA components for a typical detector using this method are shown in Fig. 5.1, where the baselines of each event have first been offset to 0. We see that the leading principal component is similar to an average pulse for the detector, approximating the shape of a good pulse. This tells us that most of the variance between events lies in their different pulse heights, corresponding to the different energies of the events. The subleading components do not have obvious characteristic features, but they can be understood as capturing the effects of pileup in calibration data. Since events are much more frequent in calibration data, events that have more than one pulse in their time window are not unusual, but since the additional pulses from pileup can show up at any point in the time window, we see their effects all over the place in the subleading principal components.

While a qualitative analysis of the principal components in Fig. 5.1 already suggests that the first component is far more important than the others, we can quantify this by considering each component's explained variance as well. As previously described, each component $\boldsymbol{u}_i$ is chosen to maximize the value of $\boldsymbol{u}_i^\intercal \boldsymbol{S} \boldsymbol{u}_i$, which is a measure of the variance of the data projected onto $\boldsymbol{u}_i$. We can thus compare the importance of different components by looking at the relative magnitudes of their explained variance $\boldsymbol{u}_i^\intercal \boldsymbol{S} \boldsymbol{u}_i$. This value for each of the components is shown in Fig. 5.2, normalized so that the explained variances sum to 1 for the components we have considered. We see that the vast majority ($> 90\%$) of the variance is captured with just the leading component. In combination with the observation that the leading component is approximately what we want a good pulse to look like, this inspires an approach of using the leading principal component to perform pulse shape discrimination.

PSD with PCA Reconstruction Errors

One of the ways that principal components are often used is by feeding the projections of the data onto the leading few principal components as input to a machine-learning algorithm, which might be a classifier of some sort. The projections serve as a proxy for the "important" information about the data as represented by the principal components. When performing pulse shape discrimination in CUORE, we are essentially trying to create a classifier as well,

[1] "Bad" events generally do not get reconstructed at any particular energy, so they tend to be spread out over the energy spectrum. As a result, if we look specifically around the energies of a known physical peak, where we know any actual events from that natural radioactive decay should reconstruct, the proportion of "good" to "bad" events is much higher.

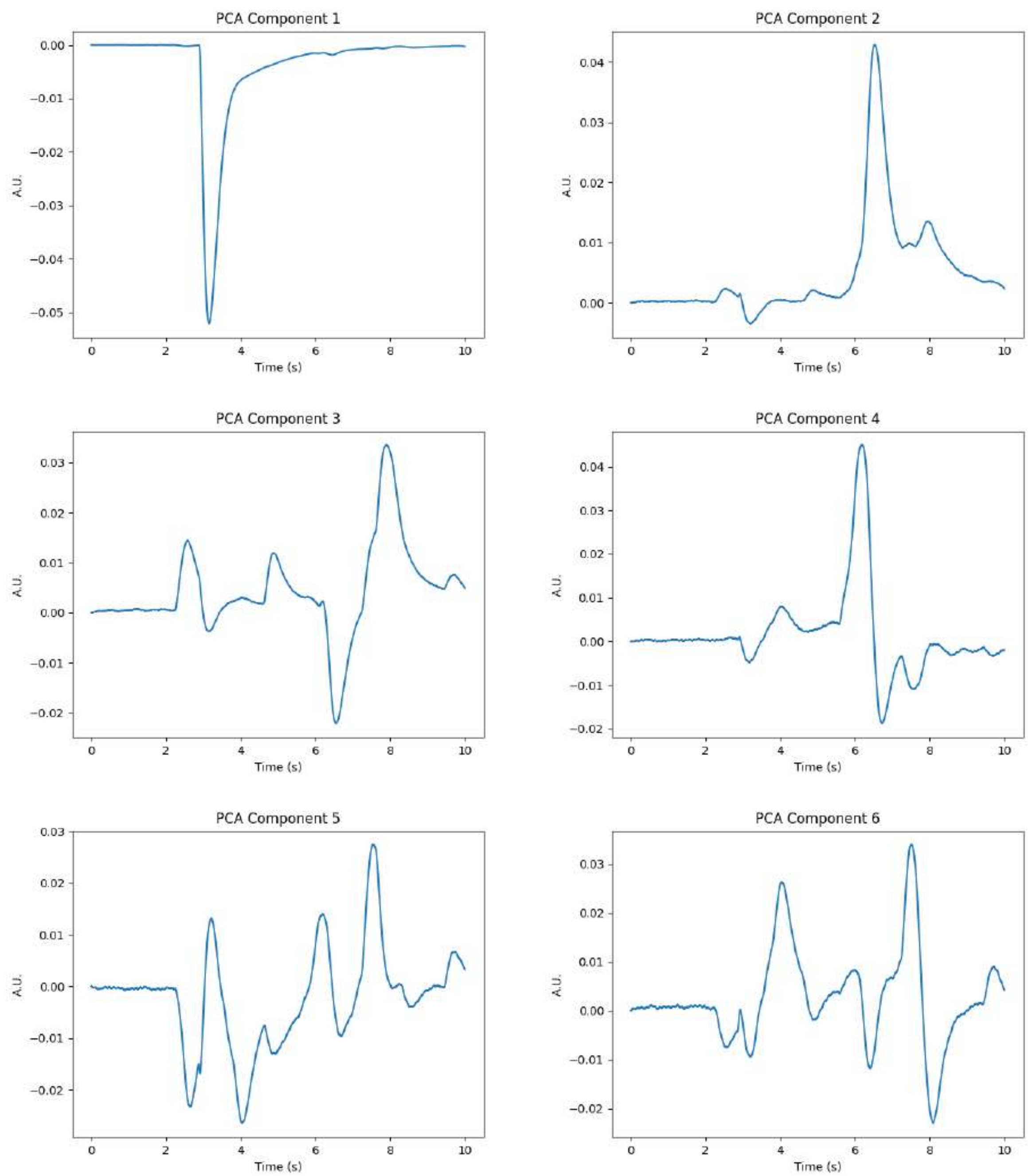


Figure 5.1: Leading 6 principal components for a single CUORE detector, obtained by training on events from the γ peaks in the calibration data. The leading component is similar to an average pulse for the channel, mostly capturing the expected shape of a normal pulse. This can be understood as a statement that most of the variance between events is in their different energies, which result in different pulse heights. The rest of the components capture various anomalous features that constitute the rest of the variance in the data, likely corresponding to different types of pileup, which is not uncommon in the calibration data.

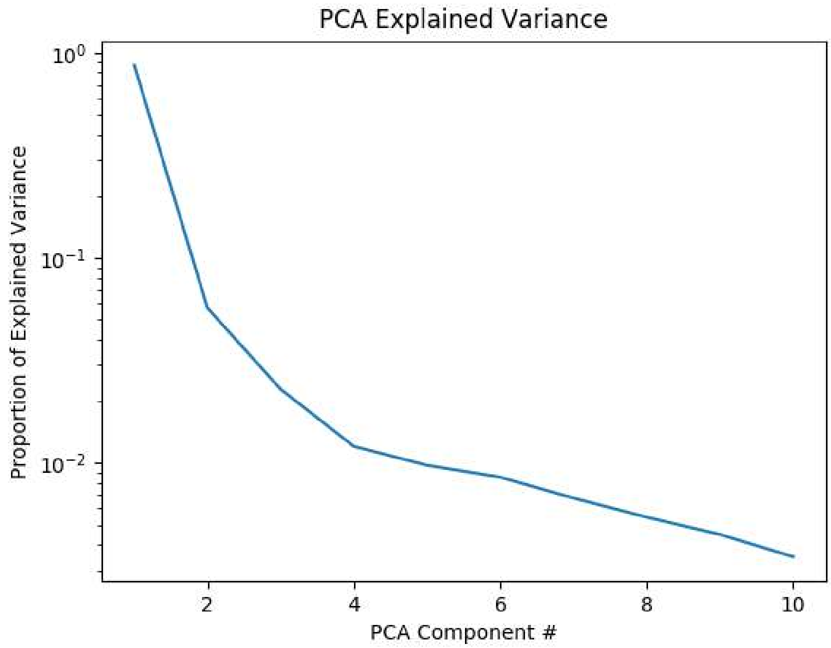


Figure 5.2: Proportion of the variance in the data explained by each of the principal components in Fig. 5.1, normalized so that they sum to 1. The vast majority of the variance is captured with just the first component, telling us that the subleading components may not be important to understanding the behavior of most events.

separating good from bad events. However, using the principal-component projections in this manner for PSD in CUORE has a number of problems. There are many ways for an event's pulse shape to be anomalous, so the manner in which a bad event gets projected onto the leading components that we have trained is not easily predictable. Typical classifiers solve this issue by simply training on samples of bad and good events, so that they can learn how their projections onto the principal components differ. But we do not have *a priori* labels of the events like this in CUORE, or we would not need to perform an additional PSD procedure in the first place. This makes it nonobvious how we could algorithmically identify a bad event by just looking at the values of the PCA projections.

Instead, we use the idea of a "reconstruction error". This builds on the idea that while we don't know what a bad pulse looks like, we do know that a good pulse should look quite similar to the leading principal component that we have obtained. Using M principal components, we can attempt to reconstruct a data vector $\boldsymbol{x}$ by calculating:

$$\sum_{i=1}^{M} (\boldsymbol{x} \cdot \boldsymbol{u}_i)\boldsymbol{u}_i$$

Here, $\boldsymbol{x} \cdot \boldsymbol{u}_i$ is the projection of $\boldsymbol{x}$ onto $\boldsymbol{u}_i$. In the limit of $M = D$ (using the same number of principal components as the number of dimensions in the data), this expression should perfectly reconstruct each data vector $\boldsymbol{x}$. When we attempt to reconstruct the data using only the first few principal components, we omit the ability to reconstruct features that would have been described by the trailing principal components. In particular, if we attempt to reconstruct each pulse $\boldsymbol{x}$ using only the leading component $\boldsymbol{u}_1$, then any pulses that deviate from just a linear scaling of $\boldsymbol{u}_1$ will be poorly reconstructed. In CUORE we idealize the leading component by using the average pulse instead, and we calculate the reconstruction error (RE) as:

$$RE = \sqrt{\sum_{d=1}^{D} \left[x_d - (\boldsymbol{x} \cdot \boldsymbol{u}_{AP}) u_d\right]^2}$$

where x_d are the values of the D-dimensional data vector $\boldsymbol{x}$, $\boldsymbol{u}_{AP}$ is the leading principal component role served by the detector's average pulse, and u_d are the values of this principal component vector $\boldsymbol{u}_{AP} = (u_1, u_2, ...u_D)$. Pulses with shapes very similar to that of the average pulse will have a low reconstruction error, and pulses with anomalous features will have a high reconstruction error. The formula uses the squared difference between the data samples and the reconstructed pulse samples to penalize short but large deviations from the expected shape (like pileup) instead of small deviations over the whole window (like normal noise).

The reconstruction error of events from a detector has a dependence on energy, since the expected RE of good events will rise as the pulses get larger. This relation is shown in Fig. 5.3, which also shows examples of the reconstruction errors of good and bad events. We normalize away this energy-dependence with a quadratic fit, which was empirically determined to be a reasonable fit function. This is also justified by our prior knowledge that the detectors have a nonlinear response with energy, as seen in the form of the calibration functions. This normalization tells us what the expected reconstruction error of an event is given its energy, with some natural variation in the expected values. Here we take advantage of our knowledge that bad events are the ones with the larger reconstruction errors[2], since we know that $\boldsymbol{u}_{AP}$ should capture most of the behavior of a good pulse.

The RE normalization function is iteratively fitted for each detector, using events from the physics data that have already passed basic quality cuts from earlier in the processing and that have multiplicity 1. This leaves us with a sample of mostly good events to find the "expected" variation of the reconstruction error with energy. Each iteration of the fit calculates the median absolute deviation (MAD) of the data points from the fit function and then trims 5% of the data that have the largest positive residuals from the fit, corresponding to the highest reconstruction errors. This removes the events that are most likely to be bad

[2]Notably, this is something we cannot do with basic pulse-shape variables or with the bare values of the projections onto the principal components. For instance, in the absence of labeled training samples, we in general do not know whether bad pulses are the ones with smaller or higher decay times, and we do not know whether bad pulses should have a small or large projection onto the second principal component.

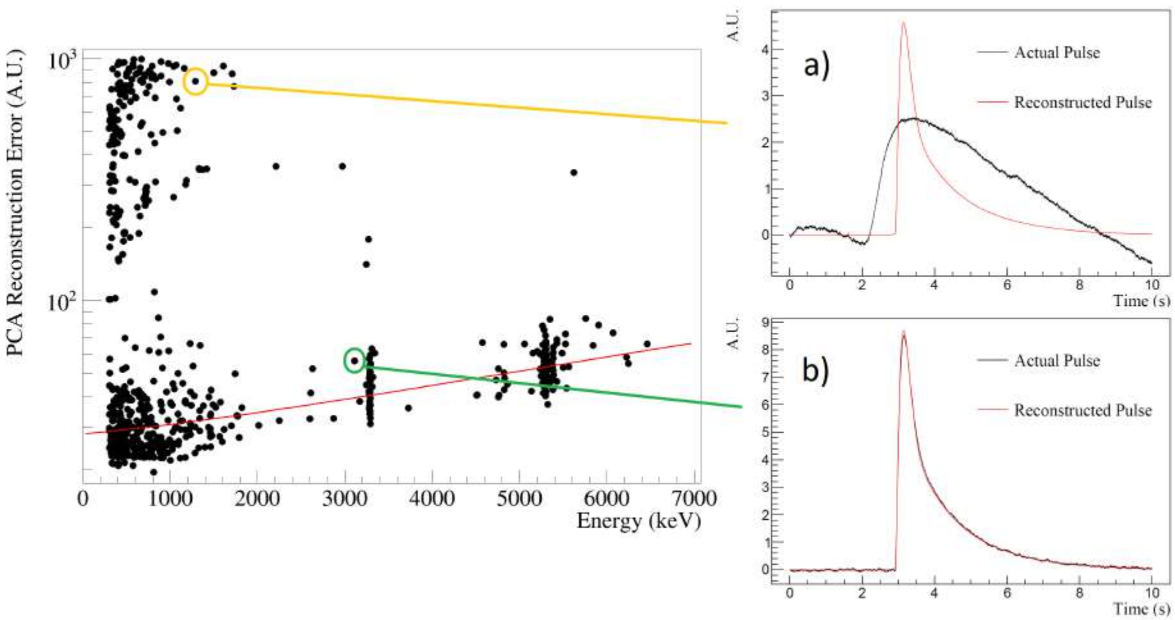


Figure 5.3: Plot of the reconstruction error versus energy for events from one dataset's physics data in one detector. The red line is a quadratic fit to normalize the expected RE as a function of energy. Two examples of pulses are highlighted, with the actual pulse drawn in black and the reconstructed pulse using $\boldsymbol{u}_{AP}$ drawn in red. One can see that the pulse with a low reconstruction error is a normal-looking pulse while the pulse with a large reconstruction error is an abnormal response, likely due to some kind of detector instability.

from the fit procedure, so that we eventually end up with a normalization function that is fit to the good population of events. This iterative procedure is continued until there are no remaining points that are several MADs away from the fit, or until 50% of the data has been trimmed from the fit. The value of this threshold is tuned by hand to obtain acceptable results, but it is justified by our expectation that we can safely assume that bad events do not comprise more than half of the total events from physics data in a detector, as determined by manual inspection of random samples of events. An example of the result of this procedure is shown in Fig. 5.4.

At the end of this process we obtain a normalization function $f(E)$ for each detector. For each event with energy E, we then calculate the normalized reconstruction error (NRE) as:

$$\mathrm{NRE} = \frac{\mathrm{RE} - f(E)}{\mathrm{MAD}}$$

The NRE serves as a energy-independent measure of how well an event's pulse conforms to the expected shape given by the detector's average pulse. We then apply a cut on the NRE for the entire dataset, with the exact cut value determined by tuning a figure of merit for

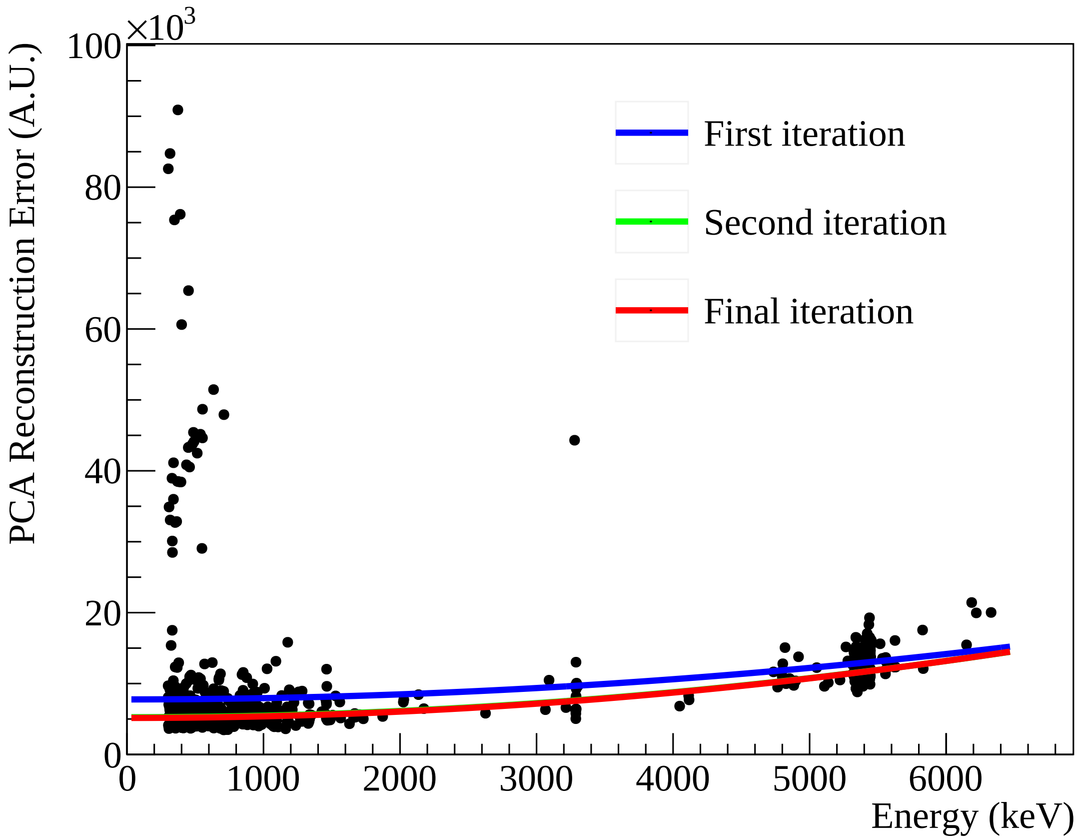


Figure 5.4: Refinement of the fitted normalization function for the PCA reconstruction errors versus energy. We start with all events from physics data in the detector passing normal data quality cuts, and we gradually cut the events that are farthest above the fit function, which is a second-order polynomial constrained to be monotonically increasing with energy. One can see the first iteration is pulled by the high outlier events, but that the iterations quickly converge to the bulk of the population at lower reconstruction errors, which are the populations of good events.

our $0\nu\beta\beta$ sensitivity.

5.3 Performance of the PCA Method

A PSD procedure is evaluated based on its ability to accept good events while rejecting bad events, for whatever definitions of good and bad the experimenter may choose. In the ideal case, we want to accept $0\nu\beta\beta$ events while rejecting everything else. Realistically, given the capabilities of our detectors, for CUORE we want to accept physical events from single radiative energy deposits, and we want to reject pileup, noise, and other nonphysical events. We optimize a simple $S/\sqrt{B}$ metric, since this is what shows up in the $0\nu\beta\beta$ sensitivity expression. We evaluate S as the efficiency of the PSD cut on the 2615 keV γ peak in physics data, which is close in energy to $Q_{\beta\beta}$, and we evaluate B as the efficiency of the PSD cut in the 2700-3100 keV region, which is a region mostly consisting of degraded alpha backgrounds and an assortment of misreconstructed events. Degraded alphas result from α decays near the surface of whatever material they originated in, so that they deposit only some fraction of their total decay energy in our detector, and these constitute the vast majority of the expected background in our $0\nu\beta\beta$ search region [58]. Although we do not expect our PSD procedure to be able to reject these degraded alpha events, this region serves as an effective proxy for testing our ability to reject the bad events scattered across our energy spectrum. The resulting figure of merit and associated efficiencies as a function of possible NRE cuts are shown in Fig. 5.5. We pick a cut of $|\mathrm{NRE}| < 8$, which can be seen against the energy spectrum in Fig. 5.6.

After determining the cut value, we must also evaluate the efficiency of this PCA-based PSD cut. This is done with events in the natural background γ peaks at 1173 (^{60}Co), 1333 (^{60}Co), 1460 (^{40}K), and 2615 (^{208}Tl) keV. We consider multiplicity-1 events at the peak energies ± 40 keV and fit them with a Gaussian signal plus flat background, calculating the number of events in the Gaussian peak before and after the PSD cut to determine the efficiency. We assume a constant efficiency as a function of energy for this energy range and average the efficiencies of all 4 peaks to obtain the overall efficiency of the PSD cut. Since single detectors do not have sufficient statistics in physics data to make this calculation, this efficiency is calculated at a dataset level. An example of this calculation is shown in Fig. 5.7, along with a comparison of the efficiency of this PCA-based approach to PSD with the efficiencies using our former PSA method described in the previous chapter for all of the datasets that have been unblinded for this $0\nu\beta\beta$ analysis. There is an overall improvement in efficiencies using the PCA method while maintaining similar background rejection capabilities, which motivated our switch to this new technique.

The ability of the PCA-based PSD cut to reject undesirable events can also be seen in its effects on calibration data, shown in Fig. 5.8. The cut mostly preserves events within the calibration peak but drastically reduces the number of events in the spectrum on either side, which are more likely to be pileup-type events. This is particularly notable in the energies above the 2615 keV peak, where we see the PSD cut reduces the number of events by roughly

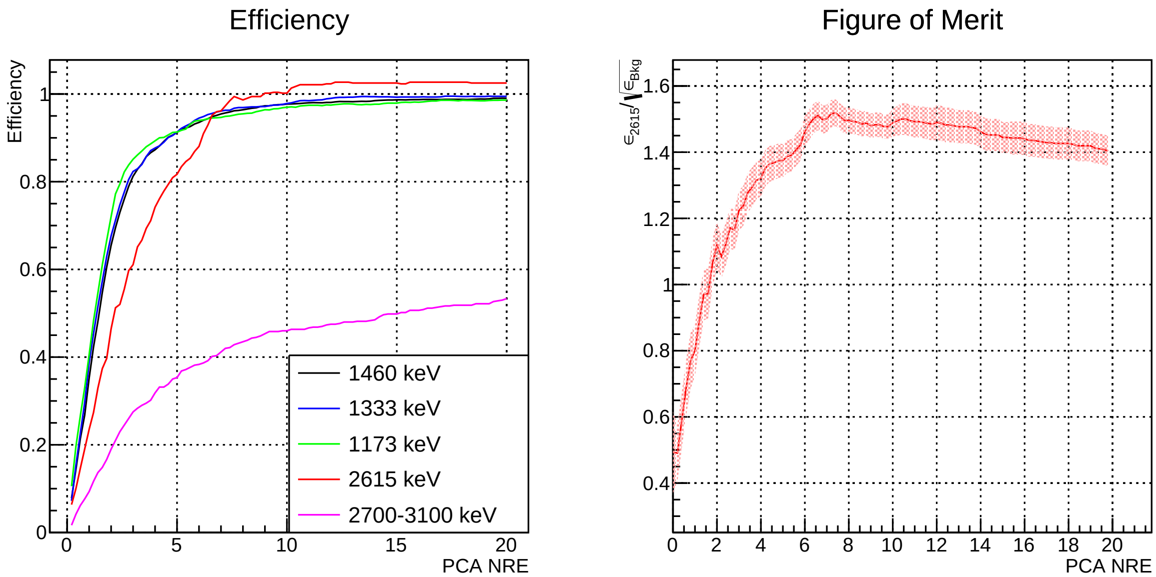


Figure 5.5: **Left**: efficiency of the PSD cut for a number of important regions as a function of the PCA NRE value that we choose as the threshold. The curves labeled with a single energy correspond to natural background γ peaks, for which we estimate the efficiency by fitting with a Gaussian signal + flat background. The efficiency can rise above 1 in this method due to uncertainties from background subtraction. **Right**: figure of merit $S/\sqrt{B}$ as a function of the PCA NRE cut value, with the shaded band corresponding to the 1σ statistical uncertainty. Signal is the efficiency on the 2615 keV γ peak, and background is the efficiency in the 2700-3100 keV region. The figure of merit reaches a maximum at some value and then starts dropping again, as cut values that are too loose allow background in while providing negligible improvements in signal efficiency.

an order of magnitude. Since the highest energy sources in our calibration strings are the 2615 keV γ rays and calibration times are too short for natural radioactive backgrounds to make significant contributions, the events above 2615 keV in calibration data are mostly the result of noise or pileup.

We use multiplicity-1 events for the final efficiency evaluation because this helps maximize the likelihood that we're considering good events and we know how to perform background subtraction for it, but using multiplicity-2 events is an interesting way to obtain an estimate of the efficiency over a continuous energy spectrum instead of just at the peak values. If we consider multiplicity-2 events where the two energies sum up to one of the γ peaks, then both events are probably good events resulting from scattering of the γ ray from the decay. However, our false coincidence rate is currently too high to reliably estimate the PSD efficiency with the multiplicity-2 events. We could perform background subtraction by, for

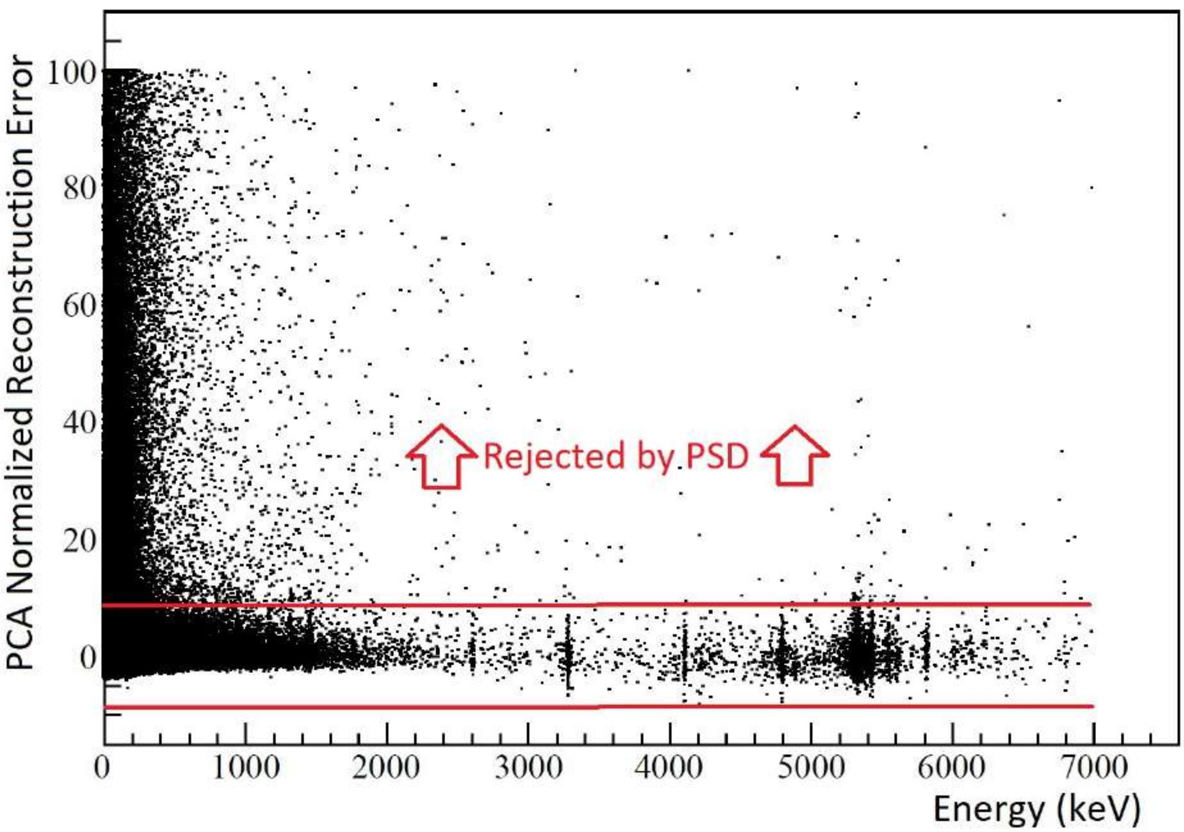


Figure 5.6: PCA normalized reconstruction error versus energy for all physics data in a single dataset, with the threshold for the final PSD cut indicated. One can see that the interesting features of the spectrum mostly lie in the permitted band of values, with the natural radioactive peaks visibly contained within the band. There is a scattering of events across the spectrum that get rejected by the PSD cut, as bad events normally do not have any characteristic energies.

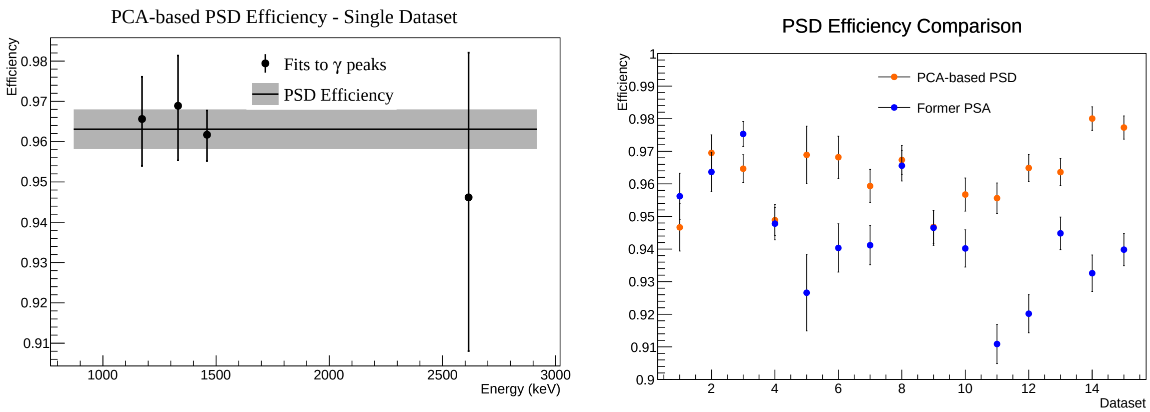


Figure 5.7: **Left**: efficiencies on the γ peaks using multiplicity-1 events for the PCA-based PSD, for a single dataset. The overall efficiency is drawn as a horizontal line, with the 1σ uncertainty shaded in gray. **Right**: PSD efficiency in each dataset using the PCA-based method and using the old PSA method with 6 pulse-shape variables for all datasets. The efficiencies for each dataset are calculated using the method shown in the left plot, and errors here are only statistical. We see a general improvement in the efficiency using the PCA method.

instance, using a time window of 1000 to 1040 ms to estimate the rate of false coincidences[3], but we have not incorporated these calculations into our analysis chain yet. Nonetheless, a qualitative comparison of the results from multiplicity-2 events should still be valid, and the multiplicity-2 event efficiency evaluations shown in Fig. 5.9 demonstrate that the old PSA cut has a clear falloff in efficiency at lower energies, while the PCA-based cut does not.

The manner in which we have used PCA for PSD here is optimized for a $0\nu\beta\beta$ analysis in CUORE, but the application of filtering away "bad" events is general-purpose and there are many ways to expand upon it. The efficiency of PCA-based PSD cuts is an improvement over our previous methods already, but the efficiency improvement is in fact even more notable at low energies, which are relatively unimportant for a $0\nu\beta\beta$ analysis. This suggests it could have applications in low energy analyses with CUORE data. The PCA components and associated projections could also help refine the energy reconstruction procedure and improve our energy resolution beyond what we currently have[4]. Lastly, in combination with a method to generate examples of pileup [65], we could train PCA components and use the projections of data onto them to perform more sophisticated PSD as well. There exist

[3]There are almost no physical processes that would cause an energy deposit in one crystal one second after a deposit in another crystal, so this would serve as a proxy for the likelihood of two unrelated events being randomly within 40 ms of each other in our multiplicity analysis.

[4]The optimum filter is optimal when the noise is stationary, but we know that this is sometimes not the case in CUORE datasets.

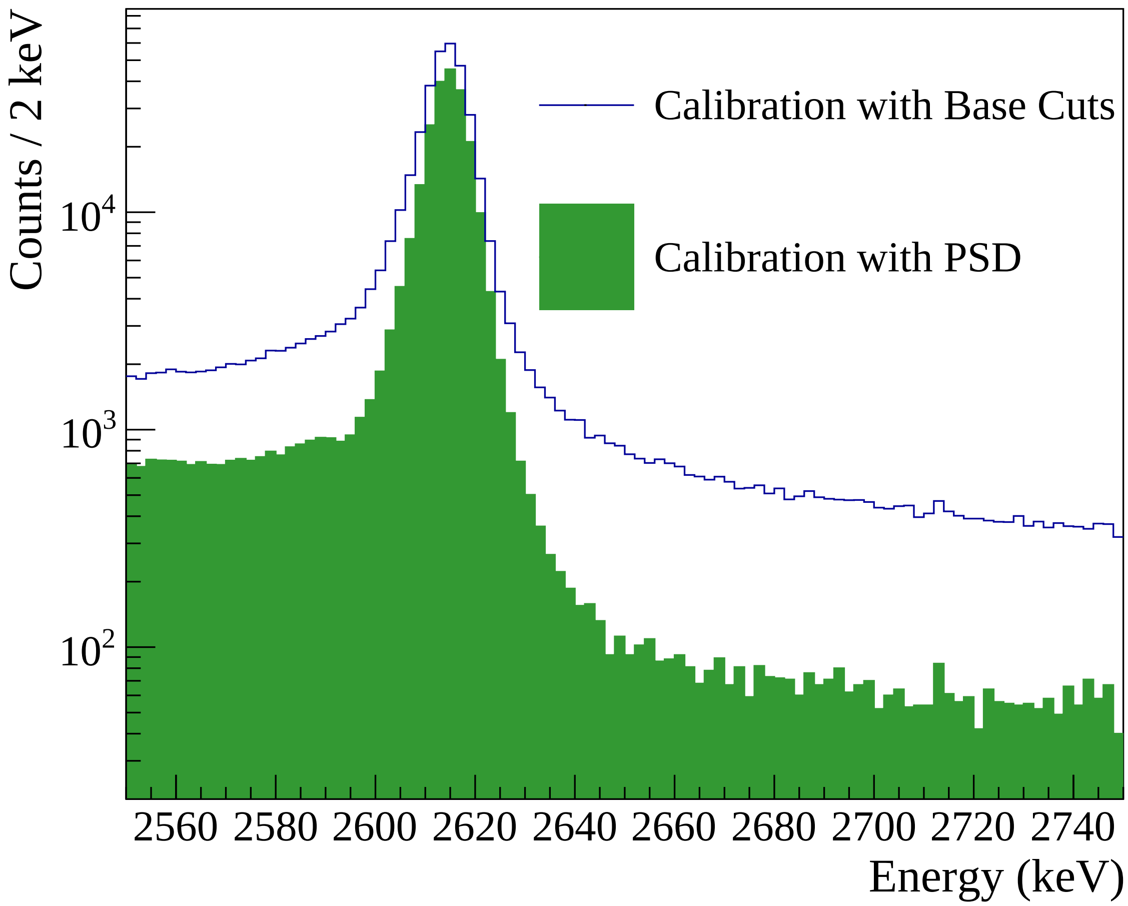


Figure 5.8: Calibration data from a single tower, before and after the PCA-based PSD cut. We see that the PSD cut is able to sharpen the resolution of the 2615 keV γ peak by eliminating pileup-induced and noisy events from around the peak. The number of events in the spectrum below and above the peak are significantly reduced as well, while a relatively small proportion of events within the peak are cut. This demonstrates that the PSD cut is keeping most of the good events, which tend to show up in the peak, and disproportionately eliminates many more bad events, which tend to show up in the continuum.

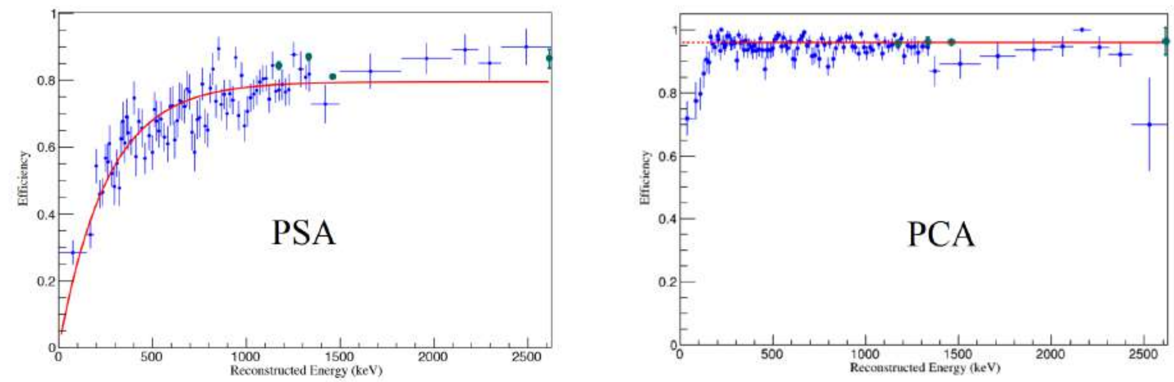


Figure 5.9: PSD efficiencies for the old PSA technique (**left**) and the PCA-based PSD (**right**) evaluated on one dataset's multiplicity-2 events whose energies sum up to one of the 1173, 1333, 1460, or 2615 keV γ peaks. The fitted efficiencies as functions of energy are drawn in red, with the fit being a constant for the PCA technique and having a rolloff for the old PSA technique (events below 300 keV are excluded in these fits as being below the PSD analysis threshold). We see that the PCA-based technique does not exhibit the same level of rolloff at lower energies, resulting in higher efficiencies for lower-energy events.

adaptions of the PCA approach that can better account for outliers in the training data, called robust PCA [66], which would be useful for improving our training procedure using actual data. These are left to future work, but they are of interest for both future CUORE analyses and for eventual CUPID analyses.

Chapter 6

$0\nu\beta\beta$ Analysis with One Tonne-Year of CUORE Data

We come now to the high-level analysis of the one tonne-year of CUORE data that gives our measurement of the $0\nu\beta\beta$ decay rate of ^{130}Te. All of the processing steps described so far do not specifically consider events in physics data near $Q_{\beta\beta} = 2527.5$ keV, but before we look there we must blind the data. Blinding stops us from tuning analysis procedures with the specific knowledge of how they would improve or weaken our result. In other words, it blocks us from massaging the data into a particular result, instead forcing us to only consider metrics like overall signal efficiency or projected backgrounds. This is done with a simple "salting" procedure, where we select a random fraction of physics events within 20 keV of the 2615 keV γ peak and shift them down by 87 keV to be around $Q_{\beta\beta}$ instead. We similarly take a random fraction of physics events within 20 keV of $Q_{\beta\beta}$ and shift them up by 87 keV to around the 2615 keV γ peak. This hides the true size of any peak that appeared near $Q_{\beta\beta}$ in the physics data and implants an artificial signal-like peak there instead, allowing us to tune our analysis on a population of events that have similar characteristics to what $0\nu\beta\beta$ would look like. Once our analysis procedure is finalized, we undo the blinding by simply shifting all events back to their true energies. I will note here that this blinding method does not truly conceal a potential $0\nu\beta\beta$ signal, since by studying specific characteristics of the pulses with salted energies one could determine that their assigned energy is off, but it is an effective blinding for anyone looking at the data in a normal way. The blinding procedure is for our benefit as the analyzers, so that there is no risk of us accidentally making biased decisions based on how they would affect our $0\nu\beta\beta$ result. It is not meant to guard against someone who is actively or maliciously trying to see through the blinding. A comparison of the unblinded and blinded spectra can be seen in Fig. 6.1.

This chapter details how we characterize our detector performance and apply cuts for a $0\nu\beta\beta$ analysis and then presents the results of CUORE's $0\nu\beta\beta$ search using 1038.4 kg·yr of TeO_2 exposure, corresponding to 288.8 kg·yr of ^{130}Te exposure. Our official results use a Bayesian framework, but we calculate a limit in the Frequentist style for comparison against other experiments as well.

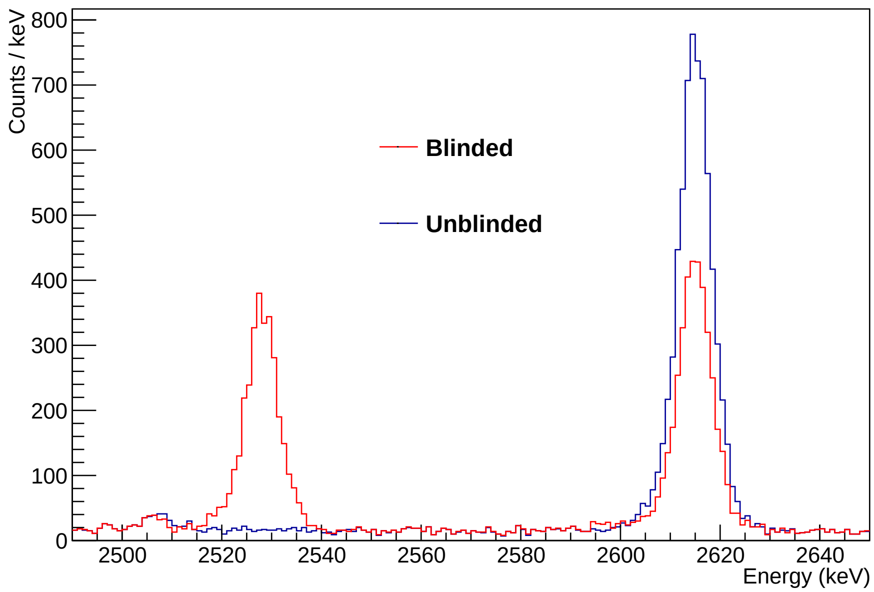


Figure 6.1: Comparison of the blinded and unblinded spectra in the physics data. Blinding is performed by simply shuffling events near $Q_{\beta\beta}$ and events near the 2615 keV γ peak, moving them up and down 87 keV respectively.

6.1 Inputs to the $0\nu\beta\beta$ Analysis

Detector Performance

We first determine the detector response functions $f(E)$, which in general vary among different detectors and different datasets. These detector response functions are how we characterize our detectors' finite energy resolutions: a physical energy deposit of magnitude E in a detector results in a random response sampled from the function $f(E)$. For an ideal detector this is just a Gaussian of some width centered on E, but for CUORE we model the detector response as a sum of 3 Gaussians [67, 68]. This consists of one central Gaussian centered near E and two smaller subpeak Gaussians below and above E, which was empirically determined to yield the best fit to data:

$$f(E) \propto \mathrm{Gaus}(E+\delta,\sigma) + A_1\mathrm{Gaus}(\rho_1(E+\delta),\sigma) + A_2\mathrm{Gaus}(\rho_2(E+\delta),\sigma)$$

Here, $\mathrm{Gaus}(\mu,\sigma)$ indicates a Gaussian with mean μ and standard deviation σ, and the proportionaliy of $f(E)$ indicates a normalization factor so that it acts as a probability distribution function. δ is the energy bias and σ the resolution of the response. The coefficients A_1, A_2 are both restricted to be less than 1, and $\rho_1 < 1, \rho_2 > 1$ place the subpeaks below and above

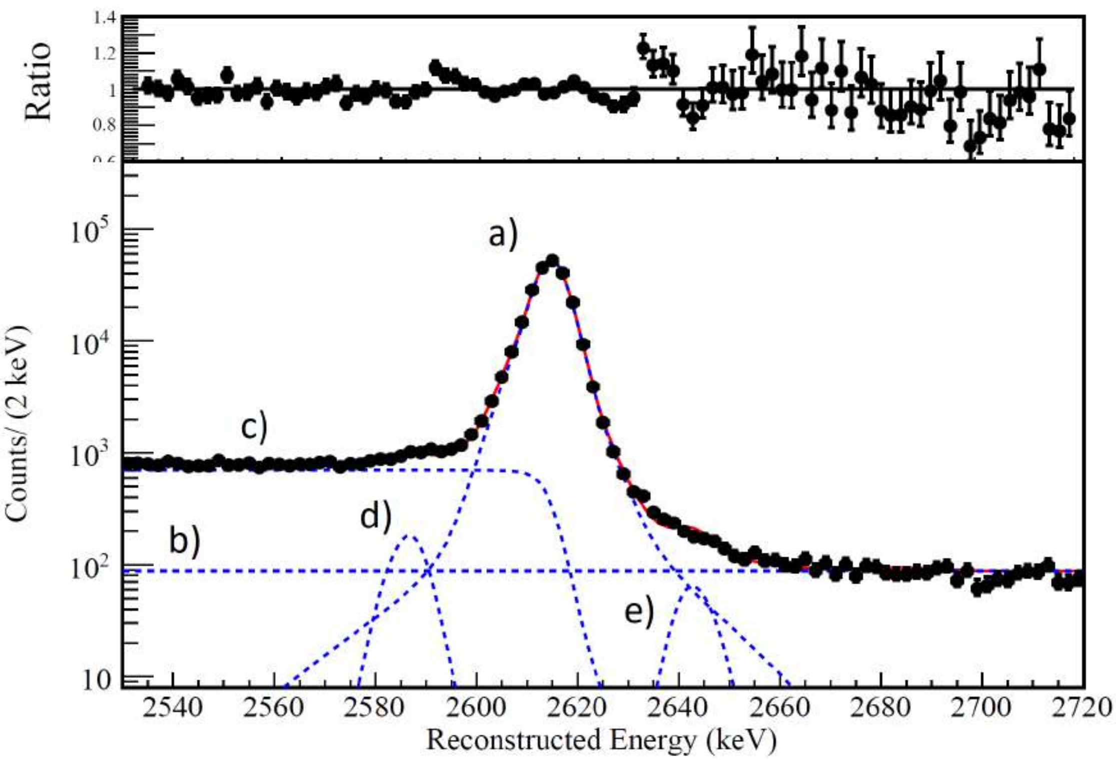


Figure 6.2: Fit to the calibration peak at 2615 keV for a single dataset. This plot shows the sum of the fits and data of all 19 towers, but the fits are performed for each tower separately. The 3-Gaussian detector response function at the center labeled a) is what we extract from this fit, and the other components only exist to model the rest of this spectrum so that we can get an accurate estimate of the detector response function. The components are described in detail in the text. The detector response function for each detector is independent, but the magnitudes of the other components are determined on a tower-by-tower level.

the main peak. These subpeaks account for uncertainties in the stabilization and calibration procedures which cause deviation from the ideal Gaussian behavior. The detector response function is extracted through a fit to the 2615 keV γ peak in calibration data and then has its energy-dependence obtained through scaling to the physics data.

The fit to the calibration peak is a multi-component fit shown in Fig. 6.2. This is performed as a simultaneous fit to all detectors of an entire tower, which tend to have similar calibration statistics due to their shared position. The fit components are:

a) 3-Gaussian detector response function $f(E)$, specified by the 6 parameters described above (the amplitude of the central peak is just a normalization, and is not part of the detector response function).

b) Flat background. A linear background was used in some of CUORE's earlier results, but we replaced it with the flat background after finding it was sufficient, with no significant evidence that a linear background was necessary for the fit.

c) Multi-Compton shelf, which is flat beneath the peak and rolls off as a complementary error function with δ, σ determined by the detector response function

d) X-ray escape peak, corresponding to a 2615-keV γ deposit followed by the escape of one of the Te x-rays near 30 keV. The shape of this peak is a copy of the 3-Gaussian detector response function using the same δ, σ, but with energy shifted down according to the energy of the x-ray escape and with a floating amplitude.

e) X-ray coincidence peak, corresponding to a 2615-keV γ ray fully absorbed at the same time as a Te x-ray emitted from a different nearby crystal. This is a copy of the 3-Gaussian detector response function just like with the x-ray escape peak, but with the position shifted above the 2615 keV peak instead of below.

The sizes of the various components besides the detector response function are fit with shared rates for the whole tower, since individual detectors generally do not have the statistics to estimate the magnitudes of these extra components. The first datasets that used the internal calibration systems also featured an additional Gaussian near 2687 keV in their lineshape fit, corresponding to a 2615-keV γ pair producing and having one of the resulting 511-keV γ rays escape, while also having the 583-keV γ from the ^{208}Tl decay deposit its energy in the same crystal at the same time. With the newer datasets using the external calibration systems, the probability of this occurring became small enough that this peak no longer appears in the lineshape.

Once we have the detector response functions, we can calculate the full-width half max (FWHM) resolution of each detector in each dataset. This distribution is shown in Fig. 6.3, where we can see the bulk of the channels have FWHMs under 10 keV but also see that there is a trail of poor-performing detectors, likely due to variations in the NTD thermistors, detector assembly specifics, and sensitivity to external factors like vibrational noise. We eliminate the worst-performing detectors in each dataset with FWHM > 19 keV from the $0\nu\beta\beta$ analysis to minimize the chance of overlap between natural radioactive peaks and a $0\nu\beta\beta$ peak. This cut value was chosen to correspond to a nominal 0.5% loss in sensitivity, which is less than the precision with which we can establish a final limit. We additionally eliminate a further $\sim$0.15% of detector-datasets that have $\delta > |1\sigma|$, which indicates calibration difficulties causing the response to be significantly offset from the actual energy of an event. After these cuts, we repeat the whole fit procedure with the remaining detector-datasets and find a harmonic mean FWHM of 7.78 keV at the calibration peak. Since ΔE shows up in the denominator of the $0\nu\beta\beta$ sensitivity expression, the harmonic mean serves as a characterization of the overall performance of the CUORE detectors.

With the detector response functions determined on calibration data, we must next account for possible differences in detector performance between calibration data and physics

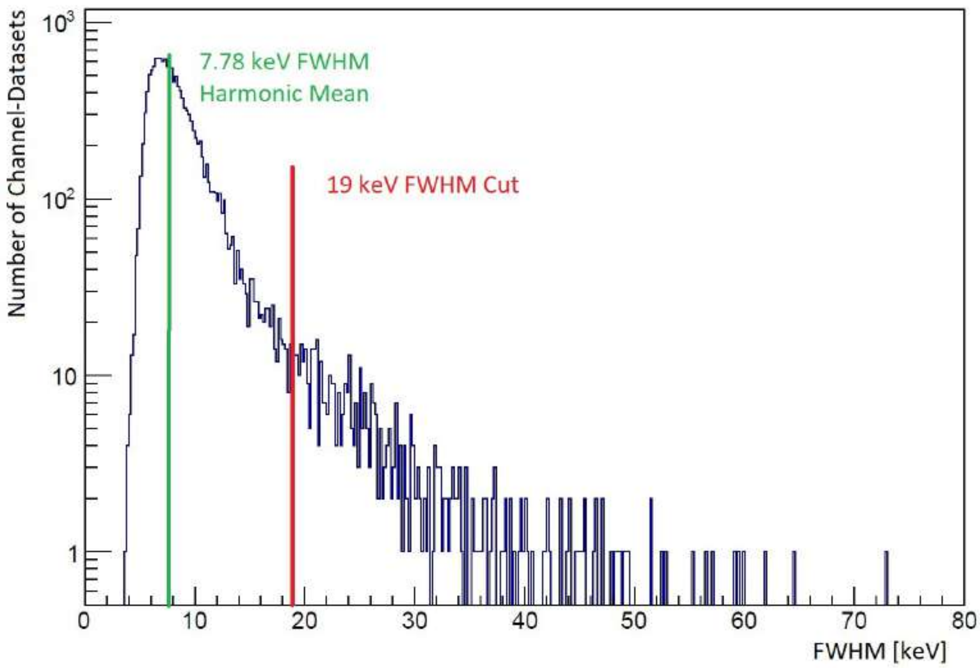


Figure 6.3: Distribution of the FWHM at the 2615 keV calibration peak for each channel-dataset pair, since each detector can have different performance between different datasets. The 19-keV FWHM cut is marked, showing it mostly cuts a trailing tail of poor performing detectors. The marked harmonic mean resolution is calculated after these poor-performing channels are eliminated from the analysis.

data. The natural background peaks in physics data usually do not have sufficient statistics to be fit for a single detector at a time, so we instead introduce a scaling function for each dataset. The response in physics data $f_{Physics}(E)$ is also normalized as a probability distribution function and is given by:

$$\begin{aligned}
f_{Physics}(E) \propto\ & \text{Gaus}(E + \delta\frac{E}{Q_{Tl208}} + Bias(E), Scaling(E) * \sigma) \\
& + A_1\text{Gaus}(\rho_1(E + \delta\frac{E}{Q_{Tl208}} + Bias(E)), Scaling(E) * \sigma) \\
& + A_2\text{Gaus}(\rho_2(E + \delta\frac{E}{Q_{Tl208}} + Bias(E)), Scaling(E) * \sigma) \\
Bias(E) =\ & a_0 + a_1E + a_2E^2 \\
Scaling(E) =\ & b_0 + b_1E
\end{aligned}$$

The function $Bias(E)$ is a measure of how the calibration function may incorrectly reconstruct certain energies in physics data, on top of any corrections we already accounted for in calibration data with the δ parameter. This is parametrized as a quadratic function with free parameters a_0, a_1, a_2. The function $Scaling(E)$ is a measure of the energy-dependence of the detectors' energy resolution, implemented as a scaling factor on the nominal resolution σ obtained from the calibration lineshape fit. This is parameterized as a linear function of energy with free parameters b_0, b_1. Previous results from CUORE used a quadratic function for this as well, but further studies this time have found that the quadratic fits for $Scaling(E)$ do not yield meaningfully better reduced χ^2 values than a linear fit. The forms of both $Bias(E)$ and $Scaling(E)$ are empirically determined, in the absence of any strong physical reason to believe they should scale with energy in a specific way. The form of $f_{Physics}(E)$ notably assumes that the detector response function becomes closer to an ideal Gaussian as the energy E decreases, with the positioning of the 2 subpeaks determined by ratios ρ_1, ρ_2 that multiply E instead of by constant offsets from E. This is supported in our data, with lower energy peaks tending to be more Gaussian than the peak at 2615 keV, but it is not obvious why this is the case. If the subpeaks in the detector response function are caused by imperfections in stabilization and calibration, then this would indicate that there is an energy-dependent uncertainty associated with them; alternatively, this could indicate that the non-Gaussian response of our detectors is due to some actual physical effect that scales with energy. For now, the root cause of the 3-Gaussian shape of the detector response function remains under investigation, and we continue to model it empirically.

The energy bias and scaling functions are obtained by finding the best-fit bias and scaling values for each of several natural γ peaks that show up in the physics data, assuming that the values are the same for all detectors and fitting them all simultaneously (but using the individual detector response functions for each detector obtained from the calibration fits). An example of the results for one dataset is shown in Fig. 6.4. Interpolating to the $0\nu\beta\beta$ energy region, we find that the bias is in general small, at < 0.7 keV at $Q_{\beta\beta}$. The resolution scaling to $Q_{\beta\beta}$ in physics data is also not significantly different from the calibration resolution, with an exposure-weighted harmonic mean FWHM of (7.8 ± 0.5) keV at $Q_{\beta\beta}$. The uncertainties on $Bias(E)$ and $Scaling(E)$ are correlated and are treated together in the final $0\nu\beta\beta$ fit. The overall breakdown of the detector response function results in the calibration and physics data for each dataset in this analysis is shown in Table 6.1.

Efficiencies

We apply a number of cuts on the way from transforming the raw CUORE data into the energy spectrum we finally analyze, and each of these cuts has an associated sub-100% efficiency for good events. The effects of these cuts on the physics spectrum are shown in Fig. 6.5. These efficiencies have to be accounted for in the $0\nu\beta\beta$ analysis since they constitute the probability that we have eliminated an actual $0\nu\beta\beta$ event from our data. The first efficiency is the reconstruction efficiency, which is the probability that a good event has been triggered, had its energy reconstructed properly, and passed the basic data-quality cuts. The first of these basic cuts is a pileup cut requiring that there is only one "peak" in

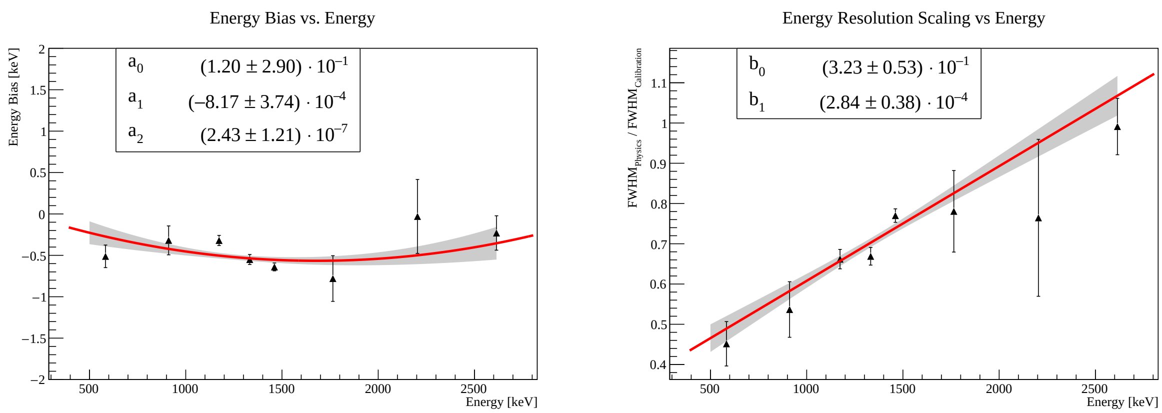


Figure 6.4: The functions for $Bias(E)$ (**left**) and $Scaling(E)$ (**right**), which characterize how the calibration results scale onto physics data. The values of the fit parameters for these functions, described previously in the text, are shown as well. The points come from the natural γ peaks in the physics data, and the second-order polynomial fit for the bias and the linear fit for the resolution scaling are obtained by fitting to these points. The shaded region indicates the 1σ uncertainty on the bias and scaling functions.

the event window[1], and the second is a requirement that the event window does not overlap with a response from the detector's heater being activated. These efficiencies are evaluated using the heater events periodically injected at the same energy throughout a dataset. Since these are externally flagged, we know the total number of heater events, allowing us to calculate the efficiencies of each of these steps. The component efficiencies for each detector are calculated as:

- Trigger efficiency: number of triggered heater events / number of injected heater events
- Energy reconstruction efficiency: number of triggered heater events within 3σ of their mean energy / number of triggered heater events within 10σ of their mean energy, where σ is determined by a fitting a Gaussian to the energies of all triggered heater events
- Pileup cut efficiency: number of triggered heater events passing the pileup cut / number of triggered heater events
- No-heater efficiency: this one obviously cannot be estimated with the heater events themselves, but we instead just calculate it as $(1 - \frac{\text{Event window width}}{\text{Time between heater events}})$. This is valid

[1] As mentioned before, this cut does not actually reject all cases of pileup, since it is not easy to algorithmically identify how many true "peaks" there are in an event window. This cut thus errs on the side of being too loose, and we let the PSD handle the cases that slip through.

Table 6.1: List of the detector performance results by dataset. The number of active channels is the number that pass all analysis cuts, including the cuts on FWHM and bias mentioned in the text, and are used in the $0\nu\beta\beta$ analysis. The FWHM values for both calibration and physics data are the exposure-weighted harmonic means for the corresponding dataset. Uncertainties for the calibration data results come from the detector response function fit uncertainties, and the uncertainties for the physics data FWHM and energy bias come from the uncertainties on $Scaling(E)$ and $Bias(E)$. The final row summarizes the exposure-weighted harmonic mean FWHMs for all datasets.

Dataset	Number of Active Channels	Calibration FWHM at 2615 keV (keV)	Physics FWHM at $Q_{\beta\beta}$ (keV)	Physics Energy Bias at $Q_{\beta\beta}$ (keV)
1	876	8.70 ± 0.03	6.2 ± 2.1	-0.34 ± 0.42
2	936	7.14 ± 0.02	7.6 ± 2.2	-0.03 ± 0.38
3	943	7.95 ± 0.01	8.4 ± 2.0	-0.33 ± 0.40
4	940	7.68 ± 0.11	9.2 ± 1.8	-0.19 ± 0.30
5	937	7.39 ± 0.02	7.7 ± 3.5	-0.31 ± 0.70
6	939	7.93 ± 0.16	7.9 ± 2.0	-0.42 ± 0.36
7	936	7.99 ± 0.03	7.2 ± 1.8	-0.43 ± 0.34
8	957	7.48 ± 0.18	8.0 ± 1.7	-0.40 ± 0.31
9	941	7.56 ± 0.07	7.3 ± 1.8	-0.11 ± 0.36
10	929	8.06 ± 0.10	7.4 ± 1.7	-0.13 ± 0.35
11	924	8.63 ± 0.21	8.8 ± 1.7	$+0.25 \pm 0.31$
12	924	7.51 ± 0.11	7.8 ± 1.7	-0.39 ± 0.29
13	935	7.60 ± 0.04	8.1 ± 1.6	-0.46 ± 0.30
14	950	7.70 ± 0.18	7.1 ± 1.7	-0.73 ± 0.31
15	942	7.62 ± 0.09	7.5 ± 1.6	-0.33 ± 0.28
Combined		7.78 ± 0.03	7.8 ± 0.5	

since the width of our event windows (10 seconds) is much smaller than the frequency with which we inject heater events (about once every 10 minutes).

The results for each detector are then averaged to obtain an overall reconstruction efficiency for each dataset. The only component of this efficiency that could in principle be systematically different among certain detectors is the pileup cut efficiency, since crystals with higher internal radioactive contaminations and crystals located closer to the outside of the cryostat can have higher event rates. However, even in these crystals, the event rate is still overall very low and the pileup cut efficiency is still very high, so the variation in the reconstruction efficiency tends to be small. We can thus reasonably expect the reconstruction efficiency to be similar among detectors, so this averaging allows us to calculate this efficiency for the detectors that do not have heaters. This also makes the resulting efficiency numbers more tractable for the $0\nu\beta\beta$ fit procedure, only having to account for one efficiency and its

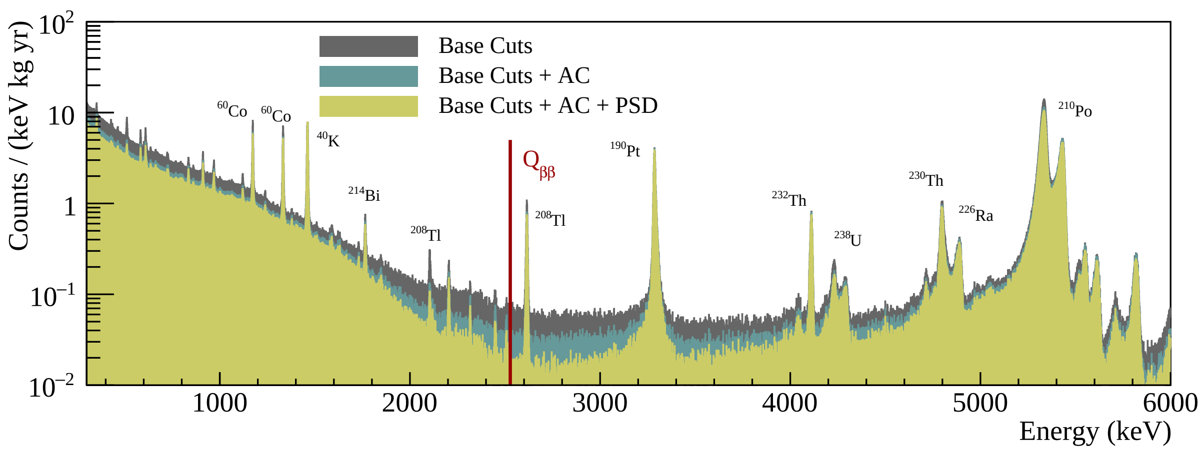


Figure 6.5: The physics spectrum for 1038.4 kg·yr of TeO_2 exposure after the base cuts, the anticoincidence (AC) cut, and the PSD cut. Prominent natural radioactive peaks are labeled, as is the ^{130}Te $Q_{\beta\beta}$. The double-peak structures in the higher energy α peaks are the result of surface α contaminations, with the lower peak corresponding to the nuclear recoil being deposited in a different detector and the higher peak corresponding to the nuclear recoil and α daughter being in the same detector. One can see that the anticoincidence and PSD cuts mostly preserve events in the peaks (other than a few peaks known to be correlated with other deposits, such as the ^{60}Co peaks or the ^{208}Tl single-escape peak at 2104 keV, for which the AC cut has a noticeable effect), but both cuts substantially reduce the number of events in the continuum. In the case of the anticoincidence cut, this is because the continuum events are more likely to be multi-scattering events, and in the case of the PSD cut this is because the noise-like events tend to show up evenly across the spectrum instead of being clustered in the peaks.

uncertainty for each dataset instead of one for each of the 900+ detectors.

The next efficiency is the anticoincidence efficiency, also calculated on a dataset level, which is the probability that a good multiplicity-1 event is not eliminated by the anticoincidence cut due to an accidental coincidence with an unrelated event. The anticoincidence cut takes advantage of CUORE's source=detector setup by considering only the multiplicity-1 events as defined in Chapter 4, since higher multiplicity events are more likely to be backgrounds. An accidental coincidence is a pair of events that register as being in coincidence in our analysis due to their proximity in time but which were actually due to two unrelated physical sources. An accidental coincidence thus has the ability to eliminate a fully contained $0\nu\beta\beta$ event from our analysis. This efficiency is evaluated using events from the 1460-keV peak, which mostly correspond to fully absorbed γ particles from the decay of ^{40}K and thus serve as one of the few "pure" samples of multiplicity-1 events available to us. These events

Table 6.2: Summary of exposure-weighted average efficiencies for the data used in this $0\nu\beta\beta$ analysis. The total analysis efficiency is the product of the reconstruction, anticoincidence, and PSD efficiencies. The containment efficiency is evaluated from MC simulations.

Total analysis efficiency	92.4(2)%
Reconstruction efficiency	96.418(2)%
Anticoincidence efficiency	99.3(1)%
PSD efficiency	96.4(2)%
Containment efficiency	88.35(9)% [63]

are uncorrelated with any other physical events, so any of these events rejected by the anticoincidence cut will be due to accidental coincidences (this can be seen in Fig. 4.6, where there is a notable absence of a vertical or horizontal band at 1460 keV). The efficiency is calculated by fitting the peak with a Gaussian + flat background both before and after the anticoincidence cut. In both cases, the number of signal events is determined by counting the number of events within $4.5\sigma_{precut}$ of the mean of the Gaussian and subtracting off the number contributed by the flat background in that region, with the final efficiency determined by the ratio of signal events before and after the cut. The uncertainty is determined by generating toy Monte Carlo samples while varying the number of signal and background events according to Poisson statistics and checking the median variations in efficiency.

The third efficiency is the PSD efficiency, whose basics were described already in Sec. 5.3. We make the modification that the PSD efficiency is calculated using an exposure-weighted sum of the detector data instead of the direct sum. For each peak used in the PSD efficiency calculation, the events contributed by each detector are normalized by its exposure / the total number of events in that peak's energy region in that particular detector. This avoids disproportionate contributions to the efficiency calculation from detectors that see naturally higher rates of events in one of those γ peaks as a result of their positioning. Since we are using this efficiency for the $0\nu\beta\beta$ fit in the end, and since the $0\nu\beta\beta$ rate for each detector is proportional to its exposure, we use this exposure-weighted sum for the PSD efficiency calculation. This is also evaluated on a per-dataset level.

Lastly, we have the containment efficiency, which is the probability that a $0\nu\beta\beta$ event from one of our crystals is fully contained within that crystal. This accounts for the $0\nu\beta\beta$ events that deposit energy in multiple crystals that we are purposefully excluding from our analysis for the sake of simplicity. In principle, a more sophisticated analysis could account for the possibility of higher multiplicity $0\nu\beta\beta$ events and slightly improve our sensitivity, but the containment efficiency is high enough that we do not have a significant loss in sensitivity by considering only the multiplicity-1 events. This efficiency is obtained through Monte Carlo simulations. A summary of all relevant efficiencies is shown in Table 6.2.

Systematic Uncertainties

The first systematic uncertainties we must account for are from our analysis procedure. The statistical uncertainty on the overall analysis efficiency is treated as a systematic, as are the correlated uncertainties in the energy bias and resolution scaling functions. We additionally impose a systematic uncertainty on the PSD efficiency to account for how the efficiencies may vary among detectors, since the PSD procedure is performed at a detector level but we evaluate its efficiency at a dataset level. For each dataset, we calculate the difference in the overall PSD efficiency estimate if we use the exposure-weighted sum method or the direct-sum method for determining the efficiency on each γ peak. We use the RMS of this difference to put an overall systematic uncertainty of $\pm 0.3\%$ on the PSD efficiency for all datasets.

The other uncertainties come from components of the fit drawn from external measurements. There is the containment efficiency of $(88.35 \pm 0.09)\%$, whose precision is limited by the Monte Carlo simulation used to calculate it [63]. Besides that, we also account for the uncertainties in the values of $Q_{\beta\beta} = (2527.515 \pm 0.013)$ keV (error-weighted average of the available measurements [45, 69, 70]) and its natural isotopic abundance of $(34.167 \pm 0.002)\%$ [71]. All systematics are included as nuisance parameters in the Bayesian fit, taking priors according to their uncertainties. This ends up being a multivariate prior for the detector response function scaling, where the energy bias and resolution scalings are correlated, and Gaussian priors for everything else.

It is worth noting that in our previous result, the PSD efficiency systematic uncertainty took a flat prior instead of a Gaussian prior [72]. This was because in that analysis we obtained different PSD efficiency numbers depending on whether we evaluated it using multiplicity-1 or multiplicity-2 events, according to the methods described in Chapter 5. Since we did not understand why the results were different, we used a flat prior to express our ignorance about what the correct value should be and instead say that we just know it should be somewhere in the range between the two results. We more recently determined that our multiplicity-2 events sometimes have high contamination rates from false coincidences, caused by either correlated noise or processing problems. The false coincidences cause bad events to enter the signal sample we would use to calculate PSD efficiency, artificially lowering the resulting number when the PSD eliminates these events as it is supposed to. The solution would require a background subtraction procedure for the multiplicity-2 events or further developments in the multiplicity analysis. Although these should not be complicated to do, we have not yet developed the appropriate machinery for them, so in this analysis we simply chose to use only the PSD efficiency numbers evaluated on the multiplicity-1 events, whose validity we have high confidence in. The proper prior for the PSD efficiency systematic is thus now a Gaussian, since we believe we know what the nominal correct value should be.

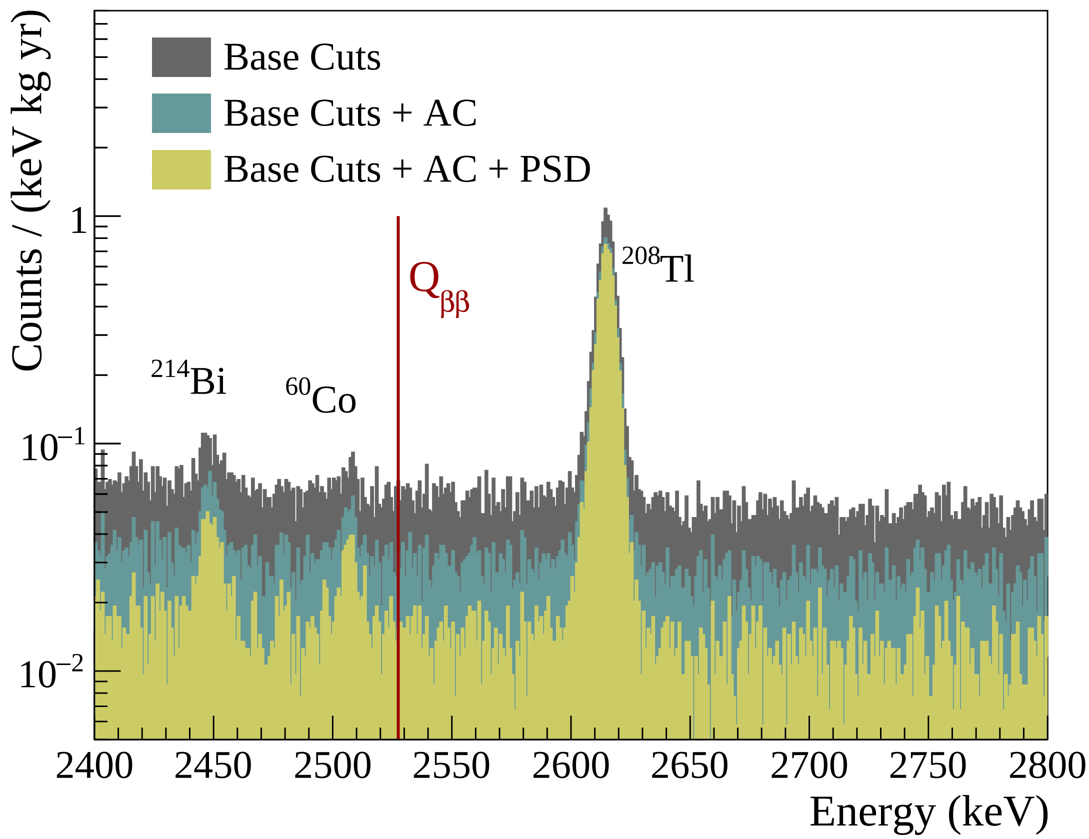


Figure 6.6: Energy spectrum near the ROI after each of the analysis cuts. The ROI itself is taken from 2490 to 2575 keV, which excludes contributions from the ^{214}Bi and ^{208}Tl peaks while being wide enough to estimate the background levels. The ^{60}Co peak is too close to $Q_{\beta\beta}$ to reasonably exclude, so it is included as part of the fit to the ROI. We can see the anticoincidence cut has a notable effect on the ^{208}Tl peak, which is sometimes in coincidence with the 583 keV γ from the same decay, but does not disproportionately cut events from the other peaks. The PSD cuts an approximately flat distribution of events across this energy range, corresponding to noise-like events that show up across the whole spectrum.

6.2 $0\nu\beta\beta$ Fit Procedure

We take a Bayesian approach to the analysis, performing an unbinned fit over all datasets looking at the 2490-2575 keV region. This energy region, which we refer to as the region of interest (ROI), captures any potential $0\nu\beta\beta$ signal while also allowing us to estimate the surrounding background. The estimate of the surrounding background is necessary so that we can determine whether the number of events near $Q_{\beta\beta}$ is in excess of the expected number due to other backgrounds, which would indicate a possible $0\nu\beta\beta$ signal. The energy spectrum near the ROI after each of the analysis cuts is shown in Fig. 6.6. The fit is performed using the Bayesian analysis toolkit (BAT) software package [73], which uses a Markov Chain Monte Carlo (MCMC) using the Metropolis-Hastings algorithm to sample the permitted parameter space [74]. The fit parameters are:

- $0\nu\beta\beta$ **rate** $\Gamma_{0\nu}$: the $0\nu\beta\beta$ decay rate of ^{130}Te in [yr^{-1}], common to all datasets.
- **Background index (BI)**: the average background rate in [counts / (keV·kg·yr)], independent for each dataset. This accounts for differences in noise levels, analysis efficiencies, and detector selection that can cause slight differences between datasets, even though the radioactive isotopes contributing to the background remain mostly constant on the timescale of CUORE's lifetime.
- **Background slope**: a linear slope to the background rate in [counts / (keV·kg·yr) / keV] over the range of the ROI, common to all datasets. The source of this slope is expected to be due to some behavior of natural radioactive backgrounds, which would be the same between datasets.
- ^{60}**Co Rate** Γ_{Co60}: the rate of ^{60}Co events, implemented as only one free parameter but scaled between the datasets according to the time that has elapsed between them by taking into account the ^{60}Co half-life of 5.3 years, which is not negligible on the timescale of years for which CUORE has taken data. The 1173.2-keV and 1332.5-keV γ rays from ^{60}Co decay add up to a peak at 2505.7 keV on the rare instances where they are both fully absorbed in the same crystal, which lies in our ROI[2].

These parameters are all assigned uninformative uniform priors, with cutoff values wide enough not to exclude any possible values given the data. The background slope is permitted to be negative or positive, but the $0\nu\beta\beta$ rate, ^{60}Co rate, and background rates are all constrained to physical nonnegative values. The other inputs to the fit are the various efficiencies and the detector response functions $f_{Physics}$ for each channel-dataset. The systematic uncertainties described in the prior section are straightforward to account for in this Bayesian framework - all of them are simply assigned Gaussian priors, with the exception of

[2] One may wonder why these events are not eliminated by our pileup or PSD cuts, since these seem like pileup events by definition. These 2 γ rays occur within less than 1 ps of each other as part of a single ^{60}Co decay, so our detectors do not have the timing resolution to separate them.

the energy bias and scaling functions, which are treated together with multivariate correlated priors. These additional components are treated as nuisance parameters and marginalized over to extract the final results for $\Gamma_{0\nu}$.

For any individual channel in a single dataset with a total of n events in the ROI with energies $\{E_i\}$, the likelihood function that we must optimize is:

$$\mathcal{L}_{CH,DS}(\{E_i\}) = \frac{e^{-\lambda}\lambda^n}{n!}\prod_i \left\{ \frac{s}{\lambda} f_{Physics}(Q_{\beta\beta})|_{E_i} + \frac{c}{\lambda} f_{Physics}(2505.7)|_{E_i} \right.$$
$$\left. + \frac{b}{\lambda} * \frac{1}{\Delta E}\left[1 + \mathrm{BI}_{\mathrm{Slope}} * (E - E_{mid})\right] \right\}$$

where s is the number of expected $0\nu\beta\beta$ signal events, c is the number of expected ^{60}Co events, b is the number of expected background events, and $\lambda = s + c + b$ is the total number of expected events. The prefactor on the likelihood is simply the Poisson probability of obtaining n events out of an expectation of λ. $E_{mid} = 2532.5$ keV is the middle of the ROI, included so that the BI is the average background level for the whole ROI, and $\Delta E = 85$ keV is the width of our ROI. The detector response function $f_{Physics}$ for the channel-dataset is evaluated at E_i to determine the probability that an event with energy E_i could have been the detector's response to a $0\nu\beta\beta$ or ^{60}Co event. The values of s, c, b are determined by the detector's exposure and the fit parameters mentioned above:

$$s = \epsilon_{analysis} \cdot \epsilon_{containment} \cdot \Gamma_{0\nu} \cdot N_A \cdot a \cdot (\text{Detector exposure})$$
$$c = \epsilon_{analysis} \cdot \Gamma_{Co60} \cdot e^{-t_{DS}/\tau} \cdot (\text{Detector exposure})$$
$$b = BI \cdot \Delta E \cdot (\text{Detector exposure})$$

Here, $\epsilon_{analysis}$ is the overall efficiency from the various analysis cuts, and $\epsilon_{containment}$ is the Monte Carlo derived efficiency for full containment of a $0\nu\beta\beta$ event in its detector. $\epsilon_{analysis}$ also shows up in the expression for c since these events are expected to be signal-like, but this does not affect the fit since it only scales the Γ_{Co60} value. N_A is Avogadro's number and a is the fraction of the detector mass that is ^{130}Te mass, together giving the conversion from detector exposure (counting the whole mass of the ^{nat}TeO$_2$ crystals) to ^{130}Te exposure. In the expression for the number of cobalt events c, τ is the decay time for ^{60}Co and t_{DS} is the time of the dataset relative to the beginning of all CUORE data. The overall likelihood for all data is given by a product of likelihoods over all channels in all datasets $\prod_{Datasets}\prod_{Channels}\mathcal{L}_{Ch,Ds}$. Our task is thus to maximize this likelihood by tuning the fit parameters for the $0\nu\beta\beta$, ^{60}Co, and background rates.

6.3 $0\nu\beta\beta$ Analysis Results

Bayesian Result

The final physics spectrum in the ROI and the corresponding fit result are both shown in Fig. 6.7. We measure a best-fit rate of:

$$\Gamma_{0\nu} = (0.9 \pm 1.4) \cdot 10^{-26} \text{ yr}^{-1}$$

which is compatible with 0. From the marginalized posterior probability distribution function (PDF) for $\Gamma_{0\nu}$, we can integrate to find that 90% of the PDF falls beneath $3.2 \cdot 10^{-26}$ yr^{-1}. This corresponds to a half-life limit at a 90% credibility interval (C.I.) of:

$$T_{1/2}^{0\nu} > 2.2 \cdot 10^{25} \text{ yrs}$$

To study the background, we repeat the fit under the background-only hypothesis, fixing $\Gamma_{0\nu}$ to 0. From this we can extract the background indices for each dataset, which are shown in Fig. 6.8. We measure an overall exposure-weighted background index of $(1.49 \pm 0.04) \cdot 10^{-2}$ counts / (keV·kg·yr) at $Q_{\beta\beta}$, which is close to but slightly above the original CUORE goal of $1 \cdot 10^{-2}$ counts / (keV·kg·yr). The best-fit slope to the BI is $(-1.8 \pm 1.9) \cdot 10^{-5}$ counts / (keV·kg·yr) / keV, resulting in the noticeable but small slope in the fit result, but the compatibility with 0 indicates the weakness of the effect. We can also see that the background does not vary significantly between datasets, as expected[3]. By looking at the 2700-3100 keV region, we obtain an exposure-weighted estimate of the degraded α background rate of $(1.40 \pm 0.02) \cdot 10^{-2}$ ckky. This suggests that $> 90\%$ of the background in the ROI is due to degraded α events, further motivating the upgrade to CUPID that will be discussed in the second half of this dissertation.

We study the effects of our systematic uncertainties by first performing the fit without any of them included and then adding them in one at a time. Overall, we find that the inclusion of the systematics shifts the value of the global mode of the posterior PDF of $\Gamma_{0\nu}$ by 0.8%, which also results in a 0.8% effect on the limit on $T_{1/2}^{0\nu}$. A summary of the effects is shown in Table 6.3. The uncertainties on the various efficiencies do not affect the quality of the fit, since they simply scale the value of $\Gamma_{0\nu}$ up or down. Their effects on $\Gamma_{0\nu}$ are thus reported as percentage effects, and we find that their posterior PDFs are almost the same as their prior PDFs. The uncertainties on $Q_{\beta\beta}$ and the detector response function scaling affect the fit itself, and so we report their additive effects on $\Gamma_{0\nu}$ instead. The uncertainty on $Q_{\beta\beta}$ is small, but the uncertainties on the detector response function are fairly significant and are responsible for the majority of the impact on $\Gamma_{0\nu}$. In total, our limit on $T_{1/2}^{0\nu}$ is unchanged by the inclusion or exclusion of the systematic uncertainties, up to the two significant figures with which we report the results.

[3]Since the relevant background sources are expected to be about the same over all datasets, a dataset with a significantly different background would indicate anomalies in its analysis. Possible contributors would include stabilization, calibration, and PSD issues.

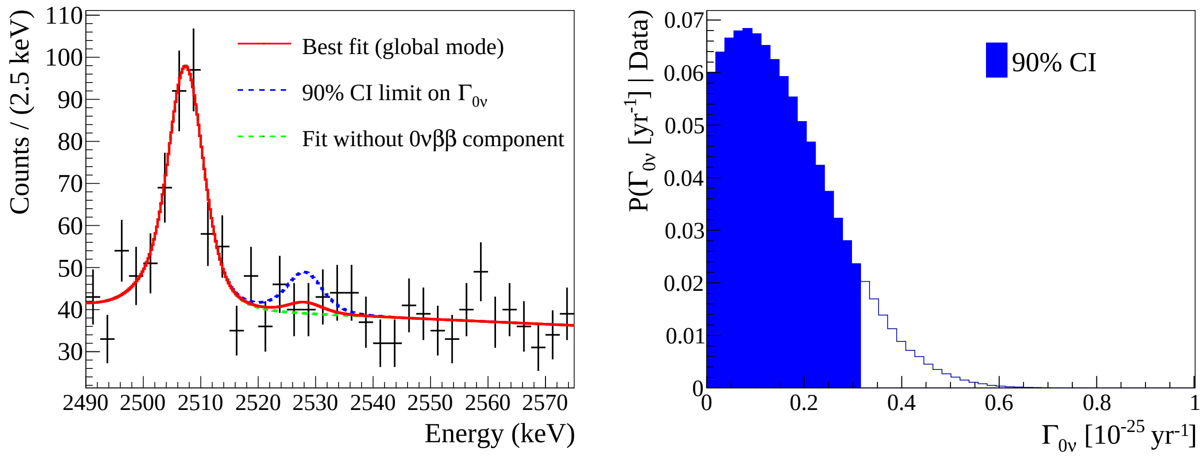


Figure 6.7: **Left**: results of the $0\nu\beta\beta$ fit, drawn with and without the $0\nu\beta\beta$ signal component, as well as with the $0\nu\beta\beta$ rate fixed to its 90% credibility upper limit. **Right**: corresponding marginalized posterior probability distribution function for $\Gamma_{0\nu}$, with the 90% credibility interval highlighted.

Table 6.3: Effects of the systematic uncertainties affecting the $0\nu\beta\beta$ decay analysis. The total analysis efficiency is the product of all the efficiencies listed in Table 6.2 except containment. The PSD efficiency is the additional systematic uncertainty from possible variations between detectors described in the text. The first 4 systematics are multiplicative effects and their impacts are presented as percentages. The last 2 are additive systematics, for which we cite their absolute effect on the signal $\Gamma_{0\nu}$. The third column indicates the variation induced on the marginalized 90% C.I. limit, while the last column indicates the effect they have on the posterior global mode $\hat{\Gamma}_{0\nu}$.

Fit parameter systematics			
Systematic	Prior	Effect on the Marginalized $\Gamma_{0\nu}$ Limit	Effect on $\hat{\Gamma}_{0\nu}$
Total analysis efficiency	Gaussian	0.2%	$< 0.1\%$
PSD efficiency	Gaussian	0.3%	$< 0.1\%$
Containment efficiency	Gaussian	0.2%	$< 0.1\%$
Isotopic abundance	Gaussian	0.2%	$< 0.1\%$
$Q_{\beta\beta}$	Gaussian	$< 0.1 \cdot 10^{-27}$ yr^{-1}	$< 0.1 \cdot 10^{-27}$ yr^{-1}
Energy bias and Resolution scaling	Multivariate	$0.2 \cdot 10^{-27}$ yr^{-1}	$0.1 \cdot 10^{-27}$ yr^{-1}

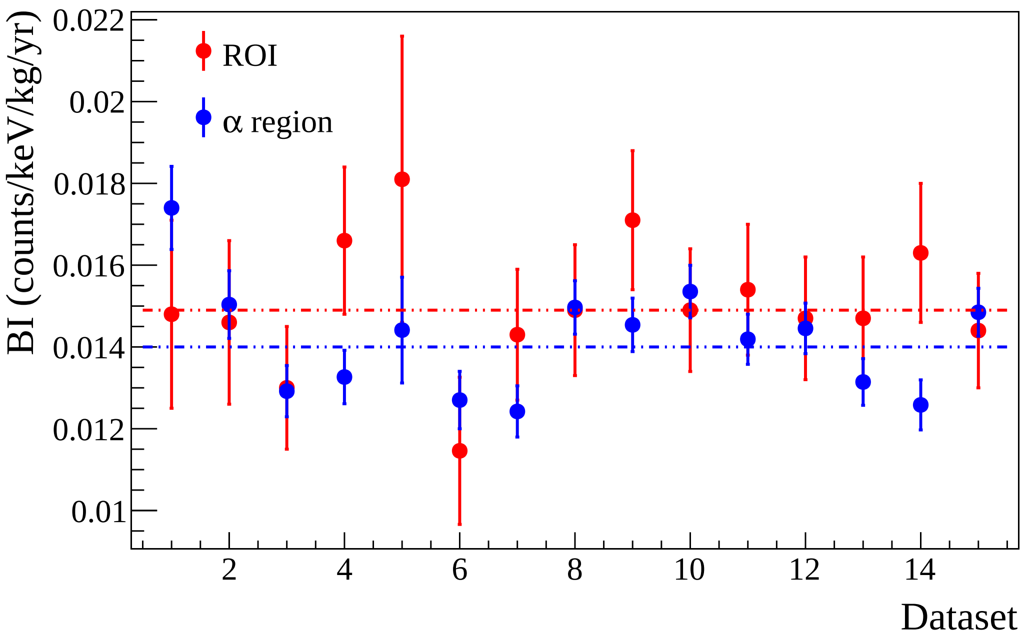


Figure 6.8: Background indices in the ROI and in the α region for each dataset. The ROI BIs are obtained through the $0\nu\beta\beta$ fit, and the BIs in the α region are obtained by fitting a constant background to the 2700-3100 keV energy region, which is populated almost entirely by degraded α events. The exposure-weighted average over all datasets is also drawn for each one as a horizontal line. The similarity between the two provides evidence that the ROI backgrounds are mostly comprised of degraded α events as well, and we see that the backgrounds do not vary significantly between datasets.

Given our data and the results of the fit, we can also calculate our expected exclusion sensitivity to $0\nu\beta\beta$ decay, accounting for normal statistical fluctuations. This is done by generating 10^4 toy experiments, split into 15 datasets with the same exposures as our actual data. For each of these datasets, we randomly populate the ROI with background and ^{60}Co events, with the background event energies sampled from the linear background we obtained from the background-only fit and with the ^{60}Co energies sampled from the detector response function of the appropriate detector. The number of background and ^{60}Co events for each toy experiment is determined by first considering the best-fit rates and their errors from the $0\nu\beta\beta$ fit, and then additionally imposing Poisson statistic fluctuations on them. After the toys are generated, we fit each of the results using the same procedure that we used for the actual data and extract a limit on $T_{1/2}^{0\nu}$ from each toy. Since we did not generate any $0\nu\beta\beta$ events in the toy experiments, this gives us a distribution of expected exclusion sensitivities under the assumption that $0\nu\beta\beta$ decay does not occur. This distribution is shown in Fig.

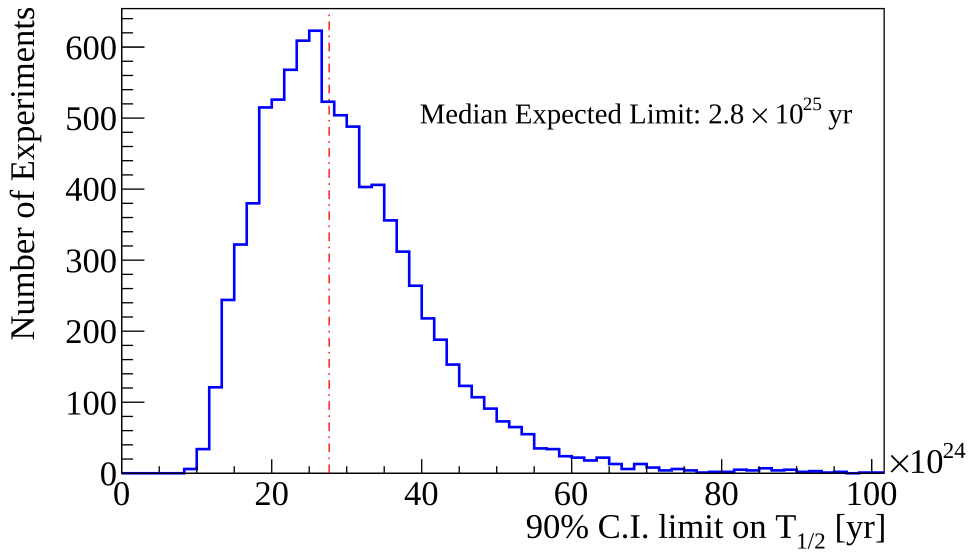


Figure 6.9: Distribution of 90% C.I. exclusion limits on $T^{0\nu}_{1/2}$, generated with 10^4 toy experiments that were all fit using the same fitting procedure we used for the actual result. The median exclusion limit of $2.8 \cdot 10^{25}$ yrs is indicated, and we can see that our actual limit of $T^{0\nu}_{1/2} > 2.2 \cdot 10^{25}$ yrs lies within the bulk of this distribution, telling us that our results were not unusual.

6.9, with a median expected 90% C.I. exclusion limit of $T^{0\nu}_{1/2} > 2.8 \cdot 10^{25}$ yrs. We see that our actual limit of $T^{0\nu}_{1/2} > 2.2 \cdot 10^{25}$ yrs is lower than 72% of the expected possible results; this tells us that statistical fluctuations have caused our limit to be slightly weaker than the expected result, but that our result is well within the bulk of the expected possible outcomes under the null hypothesis.

Frequentist Limit

In addition to the Bayesian result, we also calculate a limit in the Frequentist style for comparison against other experiments that only use Frequentist methods, which are still considered "classical" in much of the field. We do this using the Rolke method [75], wherein we consider the profile likelihood function of the $0\nu\beta\beta$ decay rate, $\lambda(\Gamma_{0\nu})$. The expression $-2\log\lambda$ is then approximated by a χ^2 distribution with 1 degree of freedom, and we can extract a 90% confidence limit by simply considering the values within $\Delta\chi^2 < 2.71$. This method and approximation are valid when we are considering a small signal on top of a background with some associated uncertainty and trying to set a limit. The profile likelihood function is extracted with the Markov Chain Monte Carlo already used for the Bayesian fit.

We divide the space of sampled $\Gamma_{0\nu}$ values into slices of finite width and within each slice we profile over all the other parameters to find the configuration that yields the highest likelihood for $\Gamma_{0\nu}$, which gives us the value of $\lambda(\Gamma_{0\nu})$ in that slice.

We choose slices of width $0.05 \cdot 10^{-26}$ yr^{-1} to evaluate the profile likelihood. This is small enough to match the precision with which we cite our results but wide enough to smooth out effects from the statistics of the MCMC, which cannot perfectly sample the available parameter space. The profile negative log likelihood function is shown in Fig. 6.10, along wih a quadratic fit to smooth out its behavior. This gives a Frequentist best-fit rate for $\Gamma_{0\nu}$ of $(1.1 \pm 0.9) \cdot 10^{-26}$ yr^{-1}, and a corresponding 90% confidence level (C.L.) half-life lower limit of $T^{0\nu}_{1/2} > 2.6 \cdot 10^{25}$ yrs. The Frequentist result is more sensitive than the Bayesian analysis to the effects of our systematic uncertainties; when the systematics are not accounted for, the Frequentist 90% C.L. half-life limit is $T^{0\nu}_{1/2} > 2.1 \cdot 10^{25}$ yrs. We can also compute an expected sensitivity by using the same toy experiments described above and analyzing them with the Frequentist method, yielding a median exclusion limit of $T^{0\nu}_{1/2} > 2.9 \cdot 10^{25}$ yrs.

Conclusions

A summary of the parameters of interest we obtain out of this analysis is shown in Table 6.4. Taking our Bayesian result as the "official" one, if we assume that $0\nu\beta\beta$ decay would be mediated by the exchange of a light Majorana neutrino, we can convert the limit on the $0\nu\beta\beta$ half-life of ^{130}Te into a limit on the neutrino's effective Majorana mass $m_{\beta\beta}$. Using the current range of nuclear matrix element calculations [29], this gives us a 90% C.I. limit of[4]:

$$\mathrm{m}_{\beta\beta} < 90 - 305 \text{ meV}$$

The comparison of this result against the limits from other experiments using different isotopes can be seen in Fig. 2.7. It is worth noting that this analysis actually presents a weaker limit than CUORE's previous result using 372.5 kg·yr of TeO_2 exposure (corresponding to 103.6 kg·yr of ^{130}Te exposure) [72], which had set a 90% C.I. lower limit on the ^{130}Te $0\nu\beta\beta$ decay half-life of $T^{0\nu}_{1/2} > 3.2 \cdot 10^{25}$ yrs. This was due to a strong statistical underfluctuation near $Q_{\beta\beta}$ in the previous analysis that led to a much stronger limit than expected. With the addition of more data this time and a re-analysis of the old data, this fluctuation went away and we ended up with a mild positive statistical fluctuation near $Q_{\beta\beta}$ instead, resulting in our weaker limit. A more detailed discussion of this can be found in Appendix A, but this was not an unexpected possibility, and we consider this more recent result with more data to be the one with higher confidence. While it is not the strongest limit on the $0\nu\beta\beta$ decay rate of ^{130}Te ever presented, we can say that with its large exposure of 1038.4 kg·yr of TeO_2 (288.8 kg·yr of ^{130}Te exposure), this result is the most sensitive measurement of the ^{130}Te $0\nu\beta\beta$ decay rate to date.

[4]Out of the NMEs described in the cited review, we exclude the result from the Chapel Hill QRPA calculation, which is an outlier and has not been included in other experiments' $m_{\beta\beta}$ conversions.

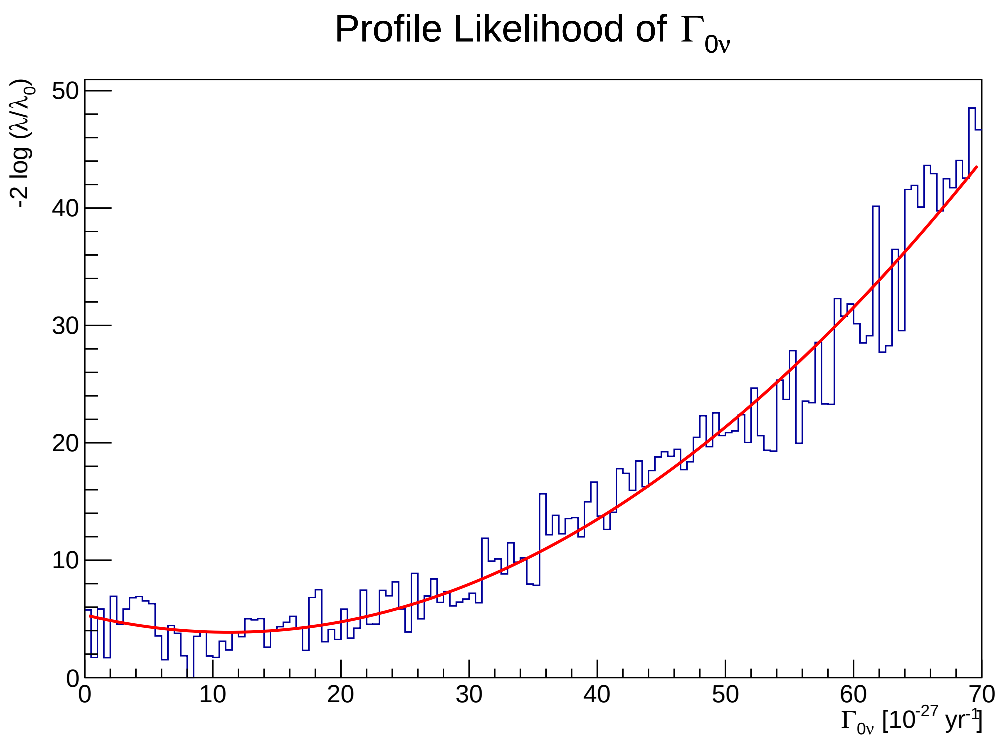


Figure 6.10: Profile negative log likelihood function of $\Gamma_{0\nu}$, obtained by sampling the Markov Chain Monte Carlo used to calculate the Bayesian result. The plotted value is $-2\log(\lambda/\lambda_0)$, where the factor of two is so it follows a χ^2 distribution and where it is normalized by the minimum negative log likelihood λ_0. The function is binned in slices of width $0.05 \cdot 10^{-26}$ yr to smooth out fluctuations due to the imperfect statistics of the MCMC sampling. The profile likelihood is further smoothed out with a quadratic fit drawn in red, which is used to extract the Frequentist best-fit rate and corresponding limit.

CUORE continues to take data now, with all of the data not analyzed here still blinded to a $0\nu\beta\beta$ analysis. While we wait for more exposure before we unblind more data to update our $0\nu\beta\beta$ results, there are many other analyses that can be conducted on the already unblinded data. These include higher multiplicity analyses, which could help us recover the remaining 12% of $0\nu\beta\beta$ events that are not fully contained in one crystal, as well as background studies and low-energy studies that could be used to conduct other exotic physics searches. In the meantime, we are also already working towards CUPID, the proposed next-generation upgrade to CUORE, which shall be the topic of the remainder of this dissertation.

Table 6.4: Summary of the results of interest from the $0\nu\beta\beta$ analysis. The important parameters for any $0\nu\beta\beta$ experiment, as described back in Sec. 2.2, are shown in the top half: the isotopic exposure, analysis efficiencies, energy resolution, and background index. The best-fit rates, 90% C.I. half-life limits, and median exclusion sensitivities for both the Bayesian and Frequentist analyses are shown in the bottom half.

$0\nu\beta\beta$ Analysis Summary	
^{130}Te Exposure	288.8 kg·yr
Analysis Efficiency	92.4(2)%
Containment efficiency	88.35(9)%
FWHM at 2615 keV in calibration data	7.78(3) keV
FWHM at $Q_{\beta\beta}$ in physics data	7.8(5) keV
Background Index at $Q_{\beta\beta}$	$1.49(4) \cdot 10^{-2}$ ckky
Bayesian Analysis	
Best-fit $\Gamma_{0\nu}$	$(0.9 \pm 1.4) \cdot 10^{-26}$ yr^{-1}
Marginalized 90% C.I. limit	$T^{0\nu}_{1/2} > 2.2 \cdot 10^{25}$ yr
Median 90% exclusion sensitivity	$T^{0\nu}_{1/2} > 2.8 \cdot 10^{25}$ yr
Frequentist Analysis	
Profiled best-fit $\Gamma_{0\nu}$	$(1.1 \pm 0.9) \cdot 10^{-26}$ yr^{-1}
Profiled 90% C.L. limit	$T^{0\nu}_{1/2} > 2.6 \cdot 10^{25}$ yr
Median 90% exclusion sensitivity	$T^{0\nu}_{1/2} > 2.9 \cdot 10^{25}$ yr

Part II

CUPID

Chapter 7

The CUPID Experiment

CUORE ultimately aims to reach a total lifetime of ∼5 years, reaching an exclusion sensitivity of $T^{0\nu}_{1/2} > 9.0 \times 10^{25}$ yrs for ^{130}Te, assuming a full-width half max (FWHM) resolution of 5 keV and background index of 0.01 counts / (keV·kg·yr) at $Q_{\beta\beta}$. Looking at these experimental parameters of exposure, energy resolution, and backgrounds, one can consider how CUORE can still be improved. The exposure is limited by the fact that CUORE uses unenriched tellurium, so one "easy" way to improve sensitivity is to just spend the money to use more enriched material, but even at 100% enrichment we would only get a factor of 3 improvement. CUORE has not yet reached its energy resolution goals, generally still operating at around 7-8 keV FWHM, so this is a possible area of improvement, but we do not expect to get much better than 5 keV FWHM with the cryogenic calorimetric method. It is then clear that the backgrounds are the area with the largest potential for improvement. Background models suggest that over 90% of the background near the ^{130}Te Q-value is from α events [58], and this is evident from CUORE's most recent $0\nu\beta\beta$ analysis as well. If these α events can be identified and rejected, we would be able to reduce our backgrounds by more than an order of magnitude.

This is the idea behind CUPID (CUORE Upgrade with Particle ID)[1], a proposed next-generation experiment building upon CUORE's experience. As CUORE approaches its intended lifetime of 5 years, it will see increasingly diminishing returns from continuing to acquire data due to the $\sqrt{Mt}$ scaling caused by its nonnegligible backgrounds. However, the competitiveness of CUORE's already-attained $0\nu\beta\beta$ results demonstrate the viability of its approach and show that the CUORE cryostat is meeting its required technical specifications. CUPID will thus aim to reuse this cryostat, but with new detectors and a new payload to drastically enhance its sensitivity to $0\nu\beta\beta$ decay, mostly through the use of particle ID to substantially reduce background levels. As soon as CUORE finishes taking data and the cryostat is warmed up and reopened, installation of the CUPID detectors can begin. The current timeline aims to have CUPID begin taking data by 2029.

[1]Given "cuore" means "heart" in Italian, one can only imagine how clever the person who came up with the acronym CUPID must have felt.

With the commissioning of the CUPID experiment still a few years away at minimum, there are a number of design decisions that have not been finalized yet, pending results from small-scale studies being performed by the many collaborating groups. This chapter will discuss the basic principles behind CUPID, some of the design decisions that have been made and still have to be made, and how these play into its sensitivity goals.

7.1 The Particle ID Technique of CUPID

As mentioned before, α events are mostly indistinguishable from β/γ events when looking at only the heat channel, where there is little sensitivity to anything other than the total number of phonons produced. However, they become distinguishable if we add a means of detecting light as well. There are two types of light emission that are relevant here: Cherenkov radiation and scintillation light. Cherenkov radiation is emitted by any charged particle traveling with speed greater than c/n in a material with index of refraction n. For a β particle with a kinetic energy of 1 MeV, this threshold condition is satisfied for $n > 1.12$. The Cherenkov radiation is very weak if the threshold is only barely met, but all crystals that have been used as cryogenic calorimeters comfortably clear this threshold. Li_2MoO_4 on the low end has $n = 1.44$ [76], while something like TeO_2 has $n > 2.5$ [77]. Although they are not themselves charged, γ particles will similarly generate Cherenkov radiation either by electron-positron pair creation or by scattering off an electron in the material. By contrast, α particles are not relativistic at the MeV energy scale we are looking at and will never produce Cherenkov light in our applications. Scintillation light works by a very different mechanism; for inorganic scintillators, scintillation light is the result of electrons de-exciting in the material after being initially excited by some ionizing radiation, generally resulting in light emissions in or near the visible wavelengths[2]. Energy deposits from α particles do cause scintillation emissions, but their light tends to be "quenched" relative to the light yields from β/γ particles. This means that the scintillation light yield from an α deposit is not the same as the light yield from a β/γ deposit of the same energy. The quenching factor is a phenomenologically determined number to describe this difference, which we cannot yet predict from theory. We thus see that in the case of either Cherenkov or scintillation light, the light yields will differ for α particles and β/γ particles.

Building on this idea, Fig. 7.1 is a schematic for a basic detector arrangement that could allow us to distinguish between α heat deposits and β/γ heat deposits. By instrumenting some sort of light detector near each crystal, one can look for the coincidence of a light signal with the heat signal in the crystal. Studying the amount of light detected alongside a heat signal would then allow for active discrimination between α and β/γ events. The role of the light detector is generally served by a silicon or germanium wafer, which is then equipped

[2] Organic scintillators work by yet another mechanism, wherein an entire molecule is excited instead of just the electrons. They are frequently used in many other experiments, but all materials we would use for cryogenic calorimeters are inorganic and would follow the theory of inorganic scintillators if they scintillate at all.

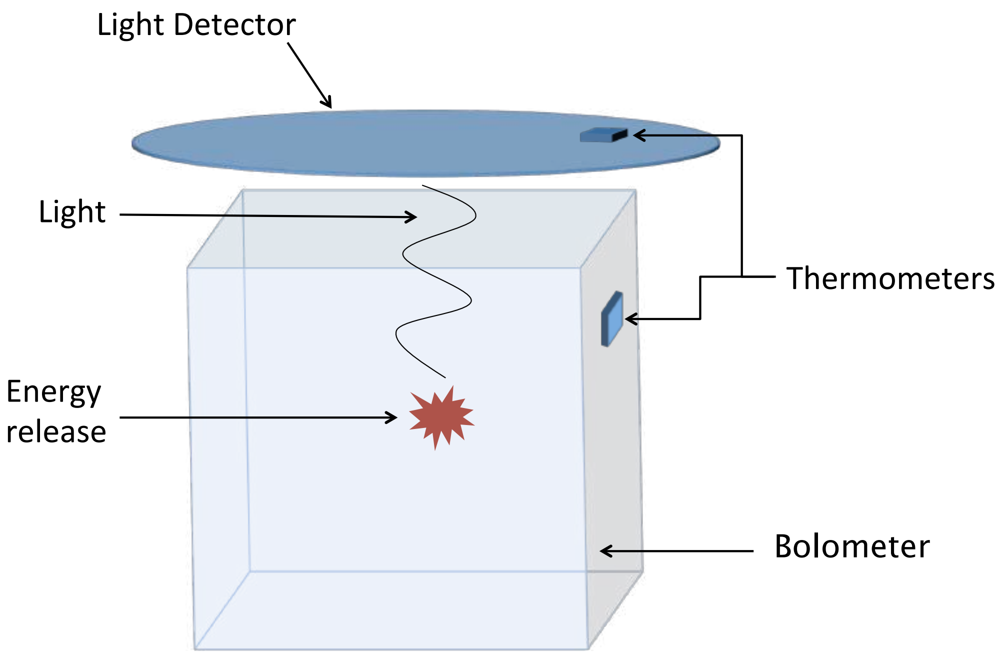


Figure 7.1: Schematic of what a detector unit for CUPID would look like. The crystals containing the $\beta\beta$ isotope would continue to act as calorimeters, but each one would also have a light detector nearby to be able to detect light signals in coincidence with the crystal's heat signal.

with some kind of sensor to measure the total light deposit on it. A number of smaller experiments have already successfully demonstrated this principle. For TeO_2, the light yield is expected to be dominated by a low quantity of Cherenkov radiation, but it has been shown that using Neganov-Luke amplification on a Ge wafer can increase the size of the signal enough to distinguish between α and β/γ events [78]. The CUPID-0 collaboration attempted a separate demonstration of particle discrimination with light detectors using scintillating ZnSe crystals, which exhibit the unusual property that their α scintillation quenching factor is greater than 1 [79]. A simple analysis of the light yields is thus insufficient for particle discrimination in ZnSe, but CUPID-0 was still able to achieve strong $\alpha/\beta(\gamma)$ discrimination by analyzing the shapes of the light pulses, allowing them to set the current leading limit on $0\nu\beta\beta$ decay of ^{82}Se [39]. The CUPID-Mo collaboration tried using scintillating Li_2MoO_4 crystals instead, surrounding them with reflecting foils to optimize the light yield seen by light detectors above and below the crystals. Using just the light yield differences we were also able to show almost complete $\alpha/\beta(\gamma)$ discrimination, allowing us to set the leading $0\nu\beta\beta$ limit for ^{100}Mo [40]. Further details of CUPID-Mo will be discussed in chapter 9, but its success was a major factor in determining the baseline material choice for CUPID.

7.2 CUPID Experimental Design

While CUPID will be reusing the infrastructure provided by the CUORE cryostat, it will be replacing the entire cryogenic payload used for the actual $0\nu\beta\beta$ search. This naturally provides an opportunity to change the isotope used. The advantages of ^{130}Te mentioned in Chapter 3 still hold true; namely, it has a very high natural isotopic abundance and it has a fairly high Q-value. However, being contained in the form of TeO_2, it has a natural disadvantage that TeO_2 has extremely weak light yields. More details on the light yields of TeO_2 are presented in Chapter 8, but this means that a TeO_2-based CUPID would need very high light collection efficiency and possibly a method of light amplification. This makes it quite tempting to consider other isotopes that can be incorporated into scintillating crystals, which would have much higher light yields and probably be easier to perform particle discrimination in. The two primary alternative options would be ^{82}Se and ^{100}Mo, which as mentioned have been tested in the CUPID-0 and CUPID-Mo demonstrators[3]. The enrichment costs of these two isotopes and the particle discrimination abilities shown with ZnSe and Li_2MoO_4 are comparable, and their Q-values of 2998 keV for ^{82}Se [80] and 3034 keV for ^{100}Mo [81] are both above the 2615 keV γ line from ^{208}Tl backgrounds. However, the energy resolution obtained with ZnSe was significantly worse with a FWHM of 20 keV at $Q_{\beta\beta}$, while Li_2MoO_4 saw a FWHM of 7.6 keV at $Q_{\beta\beta}$. This difference, thought to be caused by crystal imperfections associated with the difficulties of growing large ZnSe crystals, tips the scale in favor of ^{100}Mo.

The baseline isotope choice for CUPID is now ^{100}Mo, enriched to $> 95\%$ abundance and contained in the form of Li_2MoO_4 (LMO). Relative to the option of using enriched ^{130}Te in TeO_2, this has the advantage of having more proven $\alpha/\beta(\gamma)$ discrimination capabilities, given what CUPID-Mo has already been able to accomplish with 20 LMO crystals. In addition, the higher Q-value of ^{100}Mo will eliminate the concern of multi-Compton scattered ^{208}Tl γ rays, which would otherwise still contribute a background that cannot be rejected with CUPID's particle ID abilities. The surprising drawback of using ^{100}Mo is its unusually "fast" $2\nu\beta\beta$ decay rate, with a half-life of 7.1×10^{18} years [82]. While the energy resolution of CUPID is expected to be good enough that the $2\nu\beta\beta$ spectrum will not spill into the $0\nu\beta\beta$ search region, the exposure of CUPID will be large enough that there is a risk of two $2\nu\beta\beta$ events happening in the same crystal at almost the same time. When this happens, the pileup of the two events of energy E_1 and E_2 can be registered as a single event of energy close to $E_1 + E_2$ if the detector has insufficient timing resolution to discern that there were actually two energy deposits. Pileup of $2\nu\beta\beta$ events can thus emulate a $0\nu\beta\beta$ signal, becoming an important new background for CUPID once environmental backgrounds have been mostly eliminated through a combination of light-based particle ID and ^{100}Mo's higher Q-value.

[3] Another interesting option is ^{48}Ca, which with its particularly high Q-value of 4263 keV is above pretty much every natural β/γ background. Combined with α particle ID using a scintillating crystal like CaF_2, this would make a $0\nu\beta\beta$ search with ^{48}Ca effectively background free. However, there are significant technical difficulties associated with enriching ^{48}Ca, which has a particularly low natural isotopic abundance of 0.187%, and so it is currently not suitable for a tonne-scale experiment like CUPID.

There is active work to develop analysis methods to better reject this pileup background [65], but there are soft limitations associated with the speed of our detectors and sensors.

This raises the question of sensor and detector choice. While the heat absorbers will obviously be the LMO crystals themselves, there are options for the light detector material and the sensors used in both the crystals and the light detectors. The current baseline choice for light detectors is thin germanium wafers, detecting the total amount of absorbed light by the temperature rise that they induce in the wafers. This approach for cryogenic light detection has been successfully used in experiments like CUPID-0 and CUPID-Mo already, and it is in general a well-understood technique [83]. The baseline choice for the thermistors that actually measure the total deposits in both the crystals and the light detectors is NTDs, which have historically been extensively used for cryogenic calorimetric experiments. However, there are active efforts to develop transition edge sensors (TES) as a possible alternative to NTDs for the light detectors. TES have smaller thermal footprints and benefit from negative electrothermal feedback, meaning that the heating provided by their electrical bias naturally decreases when their temperature rises, helping facilitate the return to the baseline state. Their nature as low-impedance sensors also results in much smaller RC time constants compared to NTDs when taking into account parasitic cable capacitances. These characteristics give TES-equipped detectors faster response times than NTD-equipped detectors, which would improve timing resolution and potentially help with $2\nu\beta\beta$ pileup rejection.

Additionally, while the cold volume provided by the CUORE cryostat cannot be enlarged, there are many options for how to pack the detectors in. The 19-tower CUORE arrangement leaves a fair amount of empty space, so CUPID can aim to fit in even more $\beta\beta$ isotope by picking a denser packing. Other nontrivial questions include whether to use cylindrical or cubic crystals, how and where to deploy the light detectors in the detector arrangements, and whether to include reflecting foils to increase the light collection efficiency. There is also a question of whether to include cryogenic frontend electronics instead of using entirely room temperature electronics like was done in CUORE, which will be discussed in detail in Chapter 10. These sorts of design questions are all being investigated with smaller scale experiments before the final decisions for CUPID will be made.

Projected Sensitivity

CUPID will aim to be able to probe the entire inverted-hierarchy region with its $0\nu\beta\beta$ search, meaning that if neutrinos turn out to exist in the inverted mass hierarchy and if neutrinoless double beta decay mediated by a light Majorana neutrino is a process that exists in nature, then CUPID should be able to see it with 10 years of data. With the deployment of 472 kg of LMO containing 253 kg of ^{100}Mo, and assuming a FWHM resolution of 5 keV at $Q_{\beta\beta}$ and a background index of 10^{-4} counts / (keV·kg·yr) around $Q_{\beta\beta}$, CUPID's stated goal is a $0\nu\beta\beta$ half-life 90% exclusion sensitivity of 1.5×10^{27} years and a 3σ discovery sensitivity of 1.1×10^{27} years [84]. These experimental parameters are all picked to be easily achievable given what has already been accomplished in both CUORE and the smaller-scale demonstrators for CUPID technology. The background index assumes effectively perfect α rejection and modest coverage by a muon veto system, with remaining backgrounds coming from $2\nu\beta\beta$ pileup and conservative estimates of β/γ contaminants from the experience of CUPID-Mo for the LMO crystals and from the background model of CUORE for the cryostat and its shields. A more ambitious background goal could be to reach a background index of 2×10^{-5} counts / (keV·kg·yr), which would render CUPID effectively background-free. This would require additional innovations in material and crystal radiopurity, reduction in surface backgrounds, and improvements in timing resolution to reject a higher percentage of $2\nu\beta\beta$ pileup.

Fig. 7.2 shows the $0\nu\beta\beta$ discovery sensitivity of CUPID compared against other next-generation $0\nu\beta\beta$ experiments that have been proposed on comparable timescales to CUPID. The column for CUPID-reach indicates the possible sensitivity of CUPID if the aforementioned innovations in background reduction are successfully achieved, allowing it to operate free of background near $Q_{\beta\beta}$. CUPID-1T is not an experiment actually being proposed right now, but is instead a thought experiment considering the ultimate capabilities of the cryogenic calorimetric method, operating an even larger cryostat to contain 1000 kg of ^{100}Mo while remaining background free. One can see that all next-generation experiments are aiming to have full sensitivity to the inverted mass hierarchy, using a variety of approaches and a variety of isotopes, though all of these experiments are still using projected performance numbers. Having a diverse field of experiments like this will serve as way for them to validate each other's results in the event that one actually detects a hint of $0\nu\beta\beta$ decay.

In the meantime, there is still work to do developing the techniques that will be used in these next-generation experiments. The subject of the remainder of this dissertation will be the work I have done contributing towards the realization of CUPID, both through results that have influenced CUPID's design decisions and through work that will improve CUPID's performance.

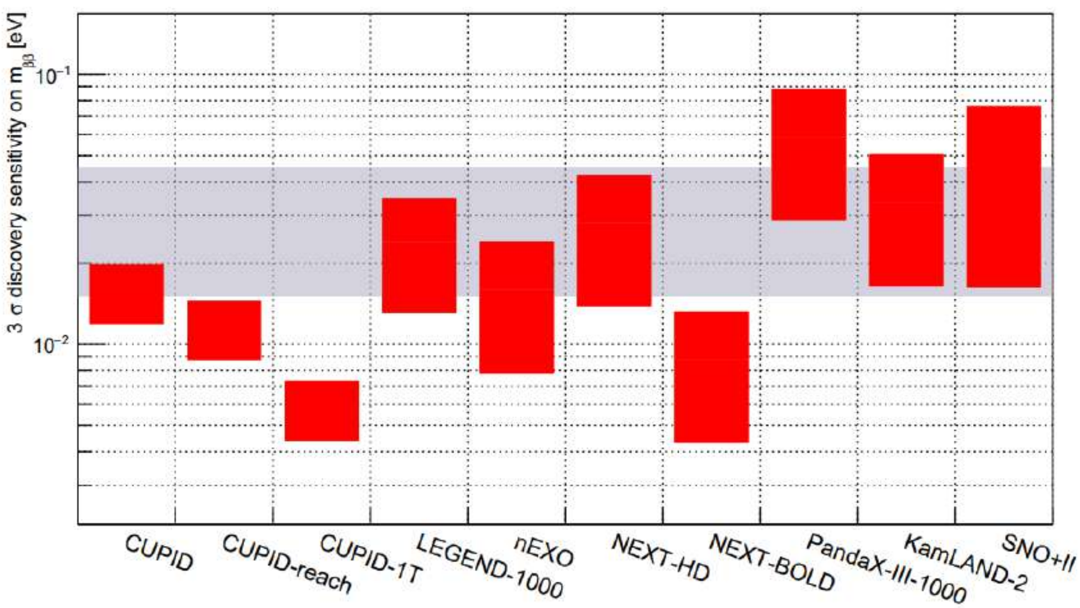


Figure 7.2: $m_{\beta\beta}$ values that could be reached with 3σ $0\nu\beta\beta$ discovery sensitivity for CUPID and an assortment of other next-generation $0\nu\beta\beta$ experiments, assuming that $0\nu\beta\beta$ decay is mediated by the exchange of a light Majorana neutrino. The red bars correspond to current uncertainties on the nuclear matrix elements for each experiment's $\beta\beta$ isotope and indicate the $m_{\beta\beta}$ values that the experiments could "discover", and the gray shaded band corresponds to the $m_{\beta\beta}$ values permitted under the inverted mass hierarchy with 3 light neutrinos. The column labeled CUPID assumes the conservative parameters discussed in this section (253 kg of ^{100}Mo, a 5 keV FWHM resolution, a background index of 10^{-4} ckky, and a livetime of 10 years). CUPID-reach supposes additional innovations to reach the background-free regime, and CUPID-1T imagines the reach of a far-future cryostat that could hold 1.8 tonnes of LMO and still operate as a background-free experiment. Reprinted from [84].

Chapter 8

TeO_2 Light Yield Characterization

In order for the principle of using light signals to perform particle identification in CUPID to be viable, the light yield of β/γ particles depositing energy in the crystals containing the $0\nu\beta\beta$ isotopes must be large enough to be easily detectable. Crystals that have relatively smaller light yields can still be used with improved light detectors with lower thresholds or by arranging the geometry of the detector configuration to optimize light collection, but this requires accurate modeling of the quantity and nature of the light emissions.

Out of the candidate materials originally considered for CUPID, TeO_2 has the lowest light yield. It exhibits little to no scintillation light, and so a TeO_2-based CUPID would have to rely on detection of the much weaker Cherenkov light instead. Although light yields in TeO_2 were generally understood to be dominated by Cherenkov radiation [85], some other experiments reported seeing minor luminescence at cryogenic temperatures [78, 86]. This chapter describes a set of measurements focused on determining how much scintillation-like light, if any, is present in TeO_2, as well as how much of an impact the imperfect optical surfaces of a TeO_2 crystal could have on total expected light yields. This was done by using a sophisticated light detector setup called CHESS and matching the observed results to detailed Monte Carlo simulations. This chapter will describe the nature of the light we're trying to measure, the CHESS experimental setup, the calibration and analysis procedure employed, and the final results setting tight constraints on the scintillation yields of TeO_2 and demonstrating the importance of crystal surface effects on the expected light yield.

8.1 Cherenkov Radiation

We begin with a brief theoretical overview of the properties of Cherenkov radiation. Consider a charged particle traveling with a velocity of $v = \beta c$ through a medium with a dielectric constant $\epsilon(\omega)$, which in general depends on the angular frequency ω of the electromagnetic wave. There is normally some incoherent energy loss as a result of the charged particle's electromagnetic interactions with the surrounding medium, with the energy lost to each area of the medium decaying exponentially with the distance from the particle's actual path. In particular, the energy deposit at a perpendicular distance a away from the particle's path contains a term of the form $e^{-(\lambda+\lambda^*)a}$, where $\lambda^2 \equiv \frac{\omega^2}{v^2}[1-\beta^2\epsilon(\omega)]$. However, consider then the scenario in which $\epsilon(\omega)$ is real and:

$$\beta^2\epsilon(\omega) > 1$$

In such a case, λ will be purely imaginary and the exponential suppression with distance a vanishes. This implies the loss of energy to an infinite distance away, corresponding to a radiative loss. This is what is known as Cherenkov radiation, occurring when the charged particle has a velocity greater than the phase velocity of light in the medium that it is passing through. This can be thought of as analagous to the sonic boom that occurs when objects travel faster than the speed of sound, with electromagnetic waves instead of sound waves. The fact that the perturbations to the local electromagnetic field induced by the particle travel slower than the particle itself results in the formation of a coherent wavefront that is the Cherenkov radiation. The Frank-Tamm formula gives the total energy radiated in this form by the uniform motion of a particle with charge ze:

$$\frac{dE}{dx} = \frac{z^2e^2}{c^2}\int_{\epsilon(\omega)>(1/\beta^2)} \omega\left[1-\frac{1}{\beta^2\epsilon(\omega)}\right]d\omega$$

This radiation also has the characteristics of being entirely transversely polarized and being emitted at an angle θ_C relative to the particle's direction of motion, with:

$$\cos\theta_C = \frac{1}{\beta\sqrt{\epsilon(\omega)}}$$

The strong directionality of Cherenkov radiation and the fact that it can occur in any material with index of refraction $n > 1$ make it very useful for particle detectors. The first trait allows the extraction of an incident particle's direction of travel and even its velocity by measuring θ_C. The second provides a mechanism for particle detection in even simple materials like water or ice, which are cheap to use in large quantities. This was famously used in Super-Kamiokande's discovery of atmospheric neutrino oscillation with a giant water tank, where the directional information provided by Cherenkov radiation allowed them to compare neutrino fluxes coming from above and below their detector [11].

From the point of view of studying the light yields from Cherenkov radiation, we're often more interested in the number of photons of each wavelength we can expect instead of the total energy deposit. Even for light detectors that work by measuring total energy deposit instead of counting photons, such as the cryogenic light detectors that will be employed in CUPID, simulations of light propagation and light collection efficiency will rely on tracking the individual photons. The Frank-Tamm formula can be straightforwardly rearranged to determine the number N of emitted photons. With the substitutions $dE = \hbar\omega dN$, $n^2 = \epsilon$, $d\omega = cd\lambda/\lambda^2$, and $\alpha =$ the fine structure constant, we obtain the formula:

$$\frac{d^2N}{dxd\lambda} = \frac{2\pi\alpha z^2}{\lambda^2}\left[1-\frac{1}{\beta^2n^2(\lambda)}\right]$$

This tells us that most Cherenkov photons are emitted at lower wavelengths, explaining the characteristic blue glow typically associated with Cherenkov radiation[1]. In the case of TeO_2, one can expect that a 2.5-MeV electron would emit around 280 Cherenkov photons with a total energy of 780 eV [85], which is significantly smaller than the yields that we get from scintillating materials. It is also notable that electrons at this energy traveling through TeO_2 experience strong scattering and do not follow straight paths, which greatly weakens the directional info that is in principle contained in the Cherenkov light.

8.2 The CHESS Setup

The CHESS (CHErenkov/Scintillation Separation) experiment is an array of photomultiplier tubes (PMTs) arranged to capture information about both the quantity and spatial distribution of light emitted in a target medium [87]. It was originally designed to demonstrate the ability to distinguish between scintillation and Cherenkov light from a water-based liquid scintillator, with potential applications for future large-scale neutrino detectors such as THEIA [88]. For studies with solid cubic targets, we collect data in two distinct configurations that we call the cosmic-muon configuration and radioactive-source configuration, shown in Fig. 8.1. In both cases, the entire setup is enclosed in a dark box that is surrounded by four scintillator panels arranged to provide 4π solid angle coverage, allowing the veto of any events with light caused by stray cosmic muons.

The cosmic-muon configuration features 12 PMTs arranged in a cross beneath the target material being tested, with 3 PMTs in each of the 4 arms. Two cylindrical scintillator tags are placed directly above the target and below the acrylic medium, allowing for the identification of cosmic muons passing straight down through the target, which are then used for analysis. To help optimize the percentage of light from the target that is able to reach the PMTs, the target is placed on top of a large acrylic block that serves as an optical propagation medium. Since the direction of Cherenkov light emitted by a downward-traveling muon is known, the propagation medium's index of refraction of $n \approx 1.5$ helps guide this light towards the PMTs.

In the radioactive-source configuration, either a ^{241}Am or ^{90}Sr source can be deployed on top of the target block. In the case of ^{241}Am, we look for light coming from its 5.64 MeV α decay, and for ^{90}Sr we look for the 2.28 MeV β decay of its short-lived daughter ^{90}Y, which is assumed to be in secular equilibrium. Two PMTs are directly coupled to the target, which we call trigger PMT 1 and trigger PMT 2. We study the charge distribution and trigger

[1] The reason that the glow appears blue instead of the even lower wavelength of violet is because of biology rather than physics, just like the reason we see the sky as blue even though it is a phenomenon stemming from the λ^{-4} dependence of Rayleigh scattering intensity. It is an artifact of how color is just our mental interpretation of the combined signals we get from the 3 types of cones in our eyes, which are stimulated in different intensities by different wavelengths of light. Violet and blue wavelengths both primarily stimulate the blue cone, but the combination ends up being mostly perceived as blue.

rates in these two PMTs instead of the light collected in the PMT ring beneath, due to the significantly lower light yields compared to those expected from cosmic muons.

Monte Carlo Simulation

To accompany the CHESS measurements, we build a GEANT4-based Monte Carlo simulation using the RAT-PAC framework [90]. This includes simulation of the optical properties of each of the elements in the darkbox, as well as a simulated response of each PMT to the impact of a photon, accounting for detection efficiencies and differing gains and noise among the PMTs obtained from calibration data. The index of refraction and optical absorption lengths of TeO_2 were extracted from [85], which also notes that TeO_2 exhibits mild birefringence. Although we did not have the ability to simulate birefringent effects in our framework, we tested for its possible effects by varying the index of refraction of TeO_2 from its ordinary to extraordinary values and found that its effects were negligible in all of our setups. Our optical framework also allows the addition of scintillation yields to any material according to the GLG4Scint model implemented in RAT-PAC, specified by the total photons released per energy deposit, the characteristic time constant of the emissions, and the energy spectrum of the scintillation photons. Since TeO_2 is not known to have any appreciable scintillation yield and since we do not use photon timing information in this analysis, we assign it a scintillation time constant that is arbitrairly short and a scintillation spectrum that is flat in the sensitive region of our PMTs. The total light yield can then be tuned as appropriate to fit the observed results.

Effects from having optical surfaces that are not perfectly smooth are also important in this application. As can be partially seen in Fig. 8.2, CUORE-style TeO_2 crystals have 4 matte faces and 2 glossy faces, and by visual inspection we can see that they merit different optical treatments in simulation. We use the *glisur* model implemented in Geant4 to simulate these imperfect optical surfaces. In this model, every optical surface is treated as being made up of some distribution of microfacets chacterized by a "polish" parameter ranging from 0 to 1. A polish of 1 corresponds to a perfect optical surface with normal Snell's law refraction and reflection, while a polish of 0 is maximally rough. From a quantitative perspective, for every surface interaction, the model generates a random vector on a sphere of radius (1-polish) and adds it to the nominal normal vector to "deflect" it, and the resulting sum vector is used for the refraction/reflection calculations. To avoid overparameterizing our model of the TeO_2 crystal, we assume the 4 matte faces share a polish parameter and that the 2 glossy faces share another polish parameter.

8.3 Calibration and Analysis Technique

As with any experimental setup, there are a number of components that have to be calibrated. The responses of the PMTs are the obvious ones, but we are also sensitive to the specific geometry of the interior of the container of our ^{90}Sr source, which can attenuate the energy

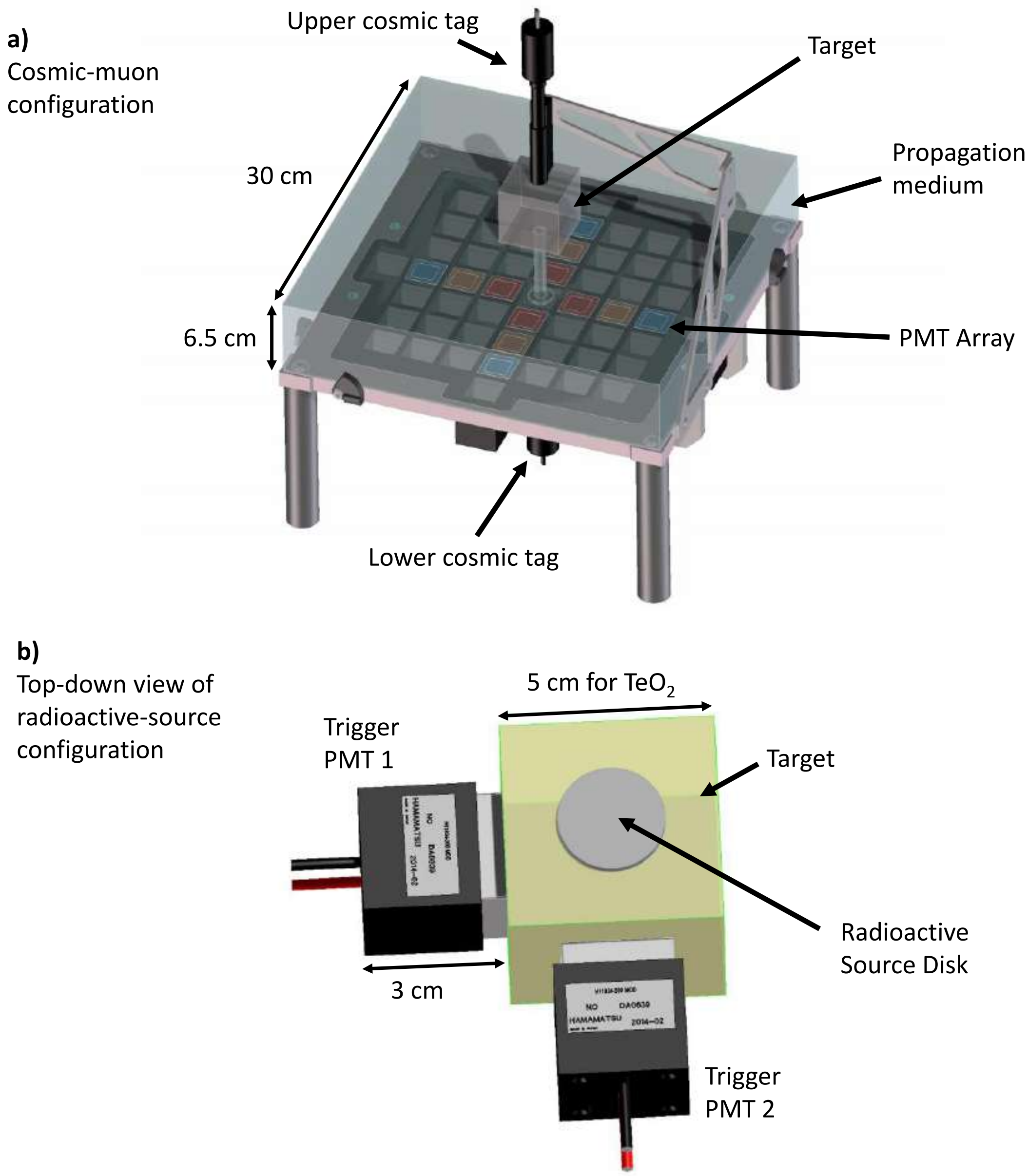


Figure 8.1: Schematics of the two CHESS detector configurations used here. The cosmic-muon configuration in a) uses muons passing straight down through the target, and the radioactive-source configuration in b) is triggered on events coming from α or β decays coming from a source deployed on top of the target. Reprinted from [89].

of the outcoming electrons. Since we are using the trigger rates as part of our analysis, we also have to account for the 29-ms deadtime after each triggered event built into the DAQ system, which will have varying effects on the resulting overall trigger efficiency. The details of these calibrations are described in this section.

The general goal of our analysis using the CHESS setup is to disentangle the effects of crystal surface roughness effects from a scintillation-like light yield that adds to the expected Cherenkov light of TeO_2, which are generally correlated if we look only at the total light yield from the crystal. Cosmic muons are high enough energy that they are not strongly scattered from passing through one of our targets, so using the cosmic-muon configuration of CHESS and looking at the spatial distribution of light in the rings of PMTs beneath the target, we can take advantage of the directional Cherenkov light from the downward traveling muons to obtain information about the crystal's surface roughness. This is then combined with data taken with a ^{90}Sr button source in the radioactive-source configuration to obtain a best fit result for both the total scintillation yield and the surface polish parameters. The procedure is performed on a cubic ultraviolet-transmitting (UVT) acrylic target first to confirm that it yields reasonable results.

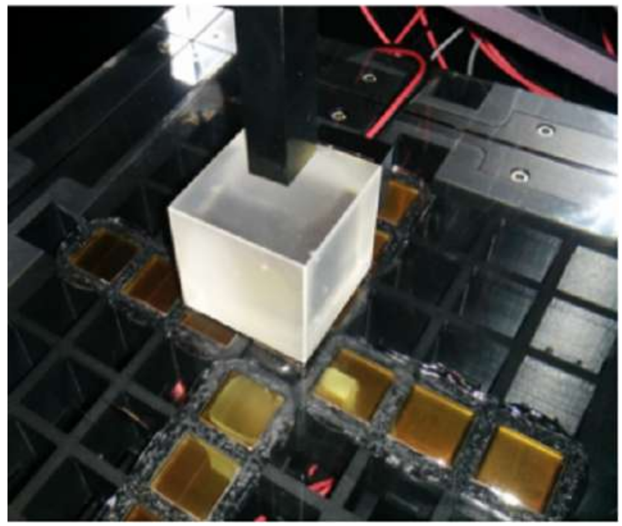


Figure 8.2: A CUORE-style TeO_2 crystal placed in the CHESS setup under the cosmic-muon configuration. The matte surface finishes for some crystal faces are visible.

Calibration

Each of the PMTs is calibrated with the presence of a low-activity β source and a UVT acrylic target, allowing for the single photo-electron peaks in the PMT charge distributions to be visible. The PMT charge distributions are fit with a sum of Gaussians described by:

$$A_e \exp\left[-\frac{(x-\mu_e)^2}{2\sigma_e^2}\right] + \sum_n A_n \exp\left[-\frac{(x-\mu_n)^2}{2\sigma_n^2}\right]$$

Here, $\mu_n = \mu_e + n * \mu_{SPE}$ and $\sigma_n = \sqrt{\sigma_e^2 + n * \sigma_{SPE}^2}$. The floating parameters in the fit are the electronic noise mean μ_e, electronic noise peak width σ_e, single photoelectron mean μ_{SPE}, single photoelectron peak width σ_{SPE}, and the amplitudes A_e, A_n of each Gaussian. This model assumes a constant electronic noise for each PMT and assumes that the width of the photoelectron peaks are dominated by the Poisson statistics of the cascading electrons in the

PMT. Calibration uncertainties for the PMTs in the cosmic-muon configuration are negligible for the analysis, but the uncertainties in the two trigger PMTs used in the radioactive-source configuration are included as systematics.

To accurately simulate the rate and energy spectrum of β decays from our ^{90}Sr source while accounting for the container geometry, we collect data with the source deployed on top of a calibrated cadmium-zinc-telluride (CZT) crystal designed for radiation measurements. We record the event rate and energy spectrum measured by the CZT crystal with the ^{90}Sr source right-side up and upside down and adjust the geometry in our simulation to match the results. The outer dimensions of the container are fixed in the simulation according to the actual physical dimensions and we assume the container is radially symmetric, but the interior is otherwise adjustable. We use the difference between the right-side up and upside down configuration CZT measurements to break any possible degeneracies in the interior geometry of the source. The best-fit results along with the corresponding best-fit rates for the source activity are shown in Fig. 8.3, where we can see that the upside-down configuration of the source results in a lower endpoint of the β energy spectrum due to greater attenuation of the βs before they can escape the container. The uncertainty on our best-fit rate of (3235 ± 25) Bq comes from performing the fit while varying the energy threshold from 0.7 to 1.0 MeV. This number is compatible with the stated activity of the source provided by the manufacturer.

To calculate the trigger efficiency resulting from the DAQ deadtime, we look at the distribution of the time between triggered events in the radioactive-source configuration. We treat the physical events as a Poisson process, so that the time between them in the absence of deadtime follows an exponential distribution. We then assume the DAQ deadtime follows a Gaussian distribution, so that if we define A as a normalization factor, λ as the "true" event rate, B as the variation in the deadtime, $t_{deadtime}$ as the nominal deadtime, and Erf(z) as the normal error function, we can fit the overall distribution of time between events as:

$$Ae^{-\lambda t} * Erf\left(B[t - t_{deadtime}]\right)$$

We perform this fit for every measurement done in the radioactive-source configuration, including background measurements, since the efficiency depends on the "true" event rate. We obtain trigger efficiencies of $(6.4 \pm 0.1)\%$ for UVT acrylic β data, $(10.6 \pm 0.1)\%$ for TeO_2 β data, and $(24.5 \pm 0.1)\%$ for background data. The uncertainties from all these calibrations are summarized in Table 8.1, and are treated as nuisance parameters in the final analysis.

Analysis Method

To measure the amount of scintillation light caused by electron-like energy depositions, we collect data in both the cosmic-muon configuration and the radioactive-source configuration using a ^{90}Sr β source. These two datasets complement each other to allow the disentangling of effects from varying surface polish and surplus scintillation light. We analyze each set of data separately and then combine the results to obtain our final conclusions. Our free

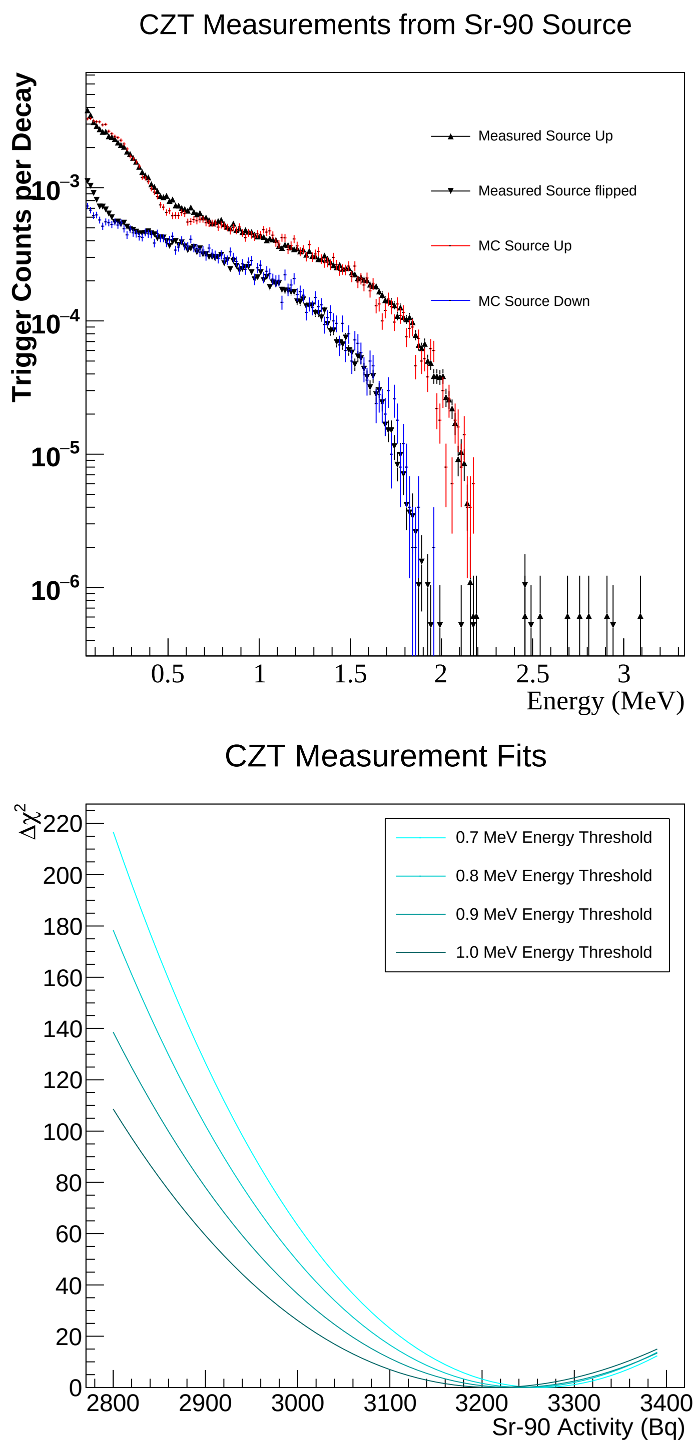


Figure 8.3: **Top**: overlaid simulated and measured energy spectra in a CZT crystal for a ^{90}Sr button source deployed right-side up and upside down, using the best-fit geometry of the source container. **Bottom**: $\Delta\chi^2$ curves for the source activity using different energy thresholds for the fit.

parameters are the Monte Carlo parameters of ℓ scintillation photons per MeV of energy deposit and two polish parameters p_1, p_2 corresponding to the glossy and matte faces of the TeO_2 crystal. We scan over these parameters and simulate the expected results for each parameter configuration, comparing them against the measured data to calculate the likelihood values $\mathcal{L}(\mu \text{ Data}|p_1, p_2, \ell)$ and $\mathcal{L}(\beta \text{ Data}|p_1, p_2, \ell)$. The negative log-likelihoods of these two sets of data are then summed to find the best-fit polish and scintillation values.

We calculate $\mathcal{L}(\mu \text{ Data}|p_1, p_2, \ell)$ by considering the ratios between the number of photoelectrons observed in each of the three radial groupings of PMTs in the array (shown n Fig. 8.1a) on an event-by-event basis. This gives two independent ratios, for which we perform a Kolmogorov-Smirnov (KS) test between the Monte Carlo predicted distributions and the actual measured distributions. We use toy Monte Carlo to generate the expected distribution of KS test statistic results if the measured data actually were sampled from the same distribution as the Monte Carlo prediction, given the number of cosmic muon events we actually measured. From this distribution of KS test results, we can convert the actual KS test statistic for each set of parameters p_1, p_2, ℓ into a likelihood value.

Since the cosmic-muon analysis only uses the ratios between the rings of PMTs, uncertainties in calibration largely cancel out, and the primary systematic uncertainty is just whether there exists some discrepancy in light collection efficiency between the rings. If such a discrepancy exists, it would be a result of a geometrical effect that we did not properly account for. To test for this possibility, we re-run the Monte Carlo simulation for a UVT acrylic target while varying the light collection efficiency for each ring of PMTs. We reject any values that worsen the best fit by more than 1.35 log-likelihood units (the 90% confidence level) or exclude the expected true value of 0 scintillation in acrylic at more than the 90% level. Said in plain language, this means that we reject any light collection efficiency values that significantly worsen our overall fit quality or that would result in a conclusion that is known to be nonphysical. After this procedure, we find that there's plausibly a 15% uncertainty in light collection efficiency. We account for this in the TeO_2 analysis by repeating the analysis many times while randomly sampling the light collection efficiency from within this range and averaging the resulting likelihoods for each p_1, p_2, ℓ configuration.

By contrast, we calculate $\mathcal{L}(\beta \text{ Data}|p_1, p_2, \ell)$ by looking at the total number of events that pass our trigger cuts. Our hardware trigger threshold in the DAQ system requires 2 photoelectrons in trigger PMT 1 in the radioactive-source configuration. In offline analysis, we impose an additional cut of requiring at least 3 PEs in trigger PMT 1 so that we don't have to worry about detector effects near the trigger threshold, and we additionally require in coincidence at least 1 PE in trigger PMT 2 to minimize any contributions from dark noise in trigger PMT 1. The likelihood is then calculated by just comparing the number of actual measured counts against the number of counts predicted by Monte Carlo, following Poisson statistics. This comparison is sensitive to a number of systematic uncertainties coming from the various calibrations mentioned in the previous section, summarized in Table 8.1.

Combining these datasets gives us a likelihood $\mathcal{L}(\mu, \beta \text{ Data}|p_1, p_2, \ell)$, from which we can extract the best fit polish parameters and scintillation yield. To obtain a limit on the amount of electron scintillation present in the target material, we profile this likelihood over the polish

Table 8.1: Systematic uncertainties that affect the β and α analyses. Constraints on the ^{241}Am source activity come from the manufacturer's specifications, while the rest come from the calibration procedures described in this section.

Source of Systematic	Fractional Uncertainty
SPE calibration (trigger PMT 1)	1.7%
SPE calibration (trigger PMT 2)	0.6%
^{90}Sr source activity	0.8%
^{241}Am source activity	3.1%
	1.6% for UVT acrylic
Trigger efficiency	0.9% for TeO_2
	0.4% for Background

parameters to obtain the value $\mathcal{L}(\mu, \beta \text{ Data}|\ell)$, which is a function only of the scintillation light yield ℓ.

In addition to these measurements of electron scintillation, we use data collected in the radioactive-source configuration with a ^{241}Am α source to set a limit on α scintillation, which is in general not the same as the β scintillation. Since the α particles don't have enough energy to even emit Cherenkov light, we expect the total light yield to be very close to 0. We fix the polish parameters p_1, p_2 in Monte Carlo to the best-fit values obtained from the μ and β analyses and look at the rate of events passing the same cuts on the two trigger PMTs as in the β data, comparing the background-subtracted actual rate against the predicted rates for each scintillation yield ℓ to set a limit on the α scintillation that can be present.

Validation with Acrylic Target

To validate this analysis approach, we first perform the electron scintillation analysis involving cosmic muon and ^{90}Sr data with a $6 \times 6 \times 6$ cm^3 cubic acrylic target, which closely emulates the shape of a $5 \times 5 \times 5$ cm^3 CUORE TeO_2 crystal. We follow the procedure the same way as with the TeO_2 target, with the only difference being the use of a single polish parameter p for all 6 faces of the acrylic cube instead of using TeO_2's two polish parameters p_1, p_2. A validation procedure like this lets us confirm that we obtain reasonable results using a more well-understood material, in which we know what "reasonable" should mean. In the case of the acrylic cube, we expect that we should obtain a fairly high polish value for the surfaces and a scintillation yield that is compatible with 0.

Fig. 8.4 shows the comparisons of Monte Carlo predictions against actual observations for the acrylic target, using the best-fit polish and scintillation light yield values coming from the negative log-likelihood curves shown in Fig. 8.5. We obtain a best-fit scintillation yield that is consistent with the expected value of 0, and the best-fit polish value of 0.86 is also reasonably high, matching what we expect from visual inspection of the acrylic surfaces. From this we can extract a 90% confidence upper limit of 5.7 scintillation photons / MeV

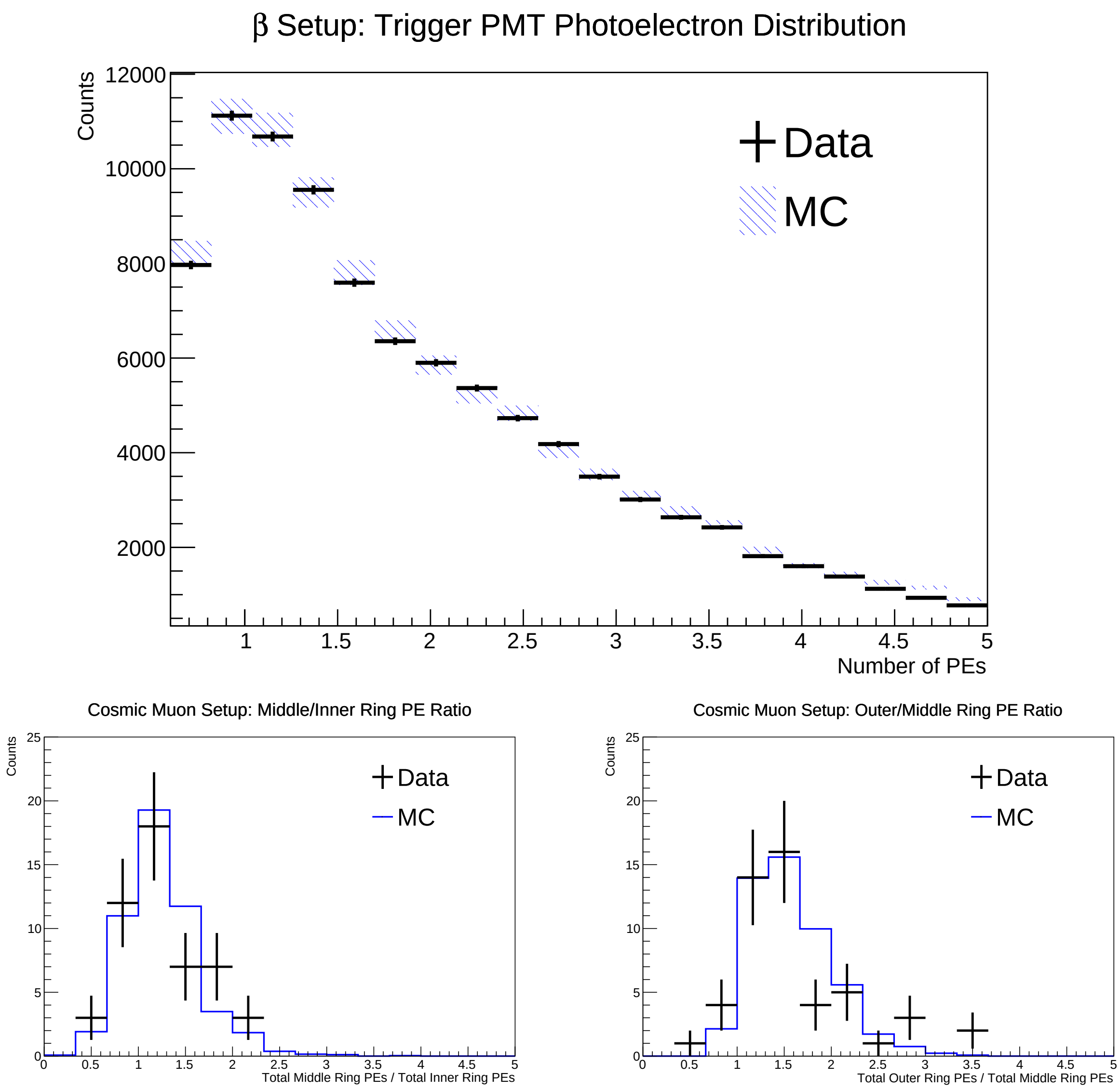


Figure 8.4: Comparison of Monte Carlo predictions to observed results for the ^{90}Sr data and cosmic muon data with an acrylic target, using the best-fit polish and scintillation values. **Top**: expected and observed photoelectron distributions in the trigger PMT for the ^{90}Sr data. Shaded regions correspond to 1σ systematic uncertainties incorporated into the Monte Carlo simulation. **Bottom**: expected and observed ratios of light between the PMT rings for the cosmic muon data. Reprinted from [89].

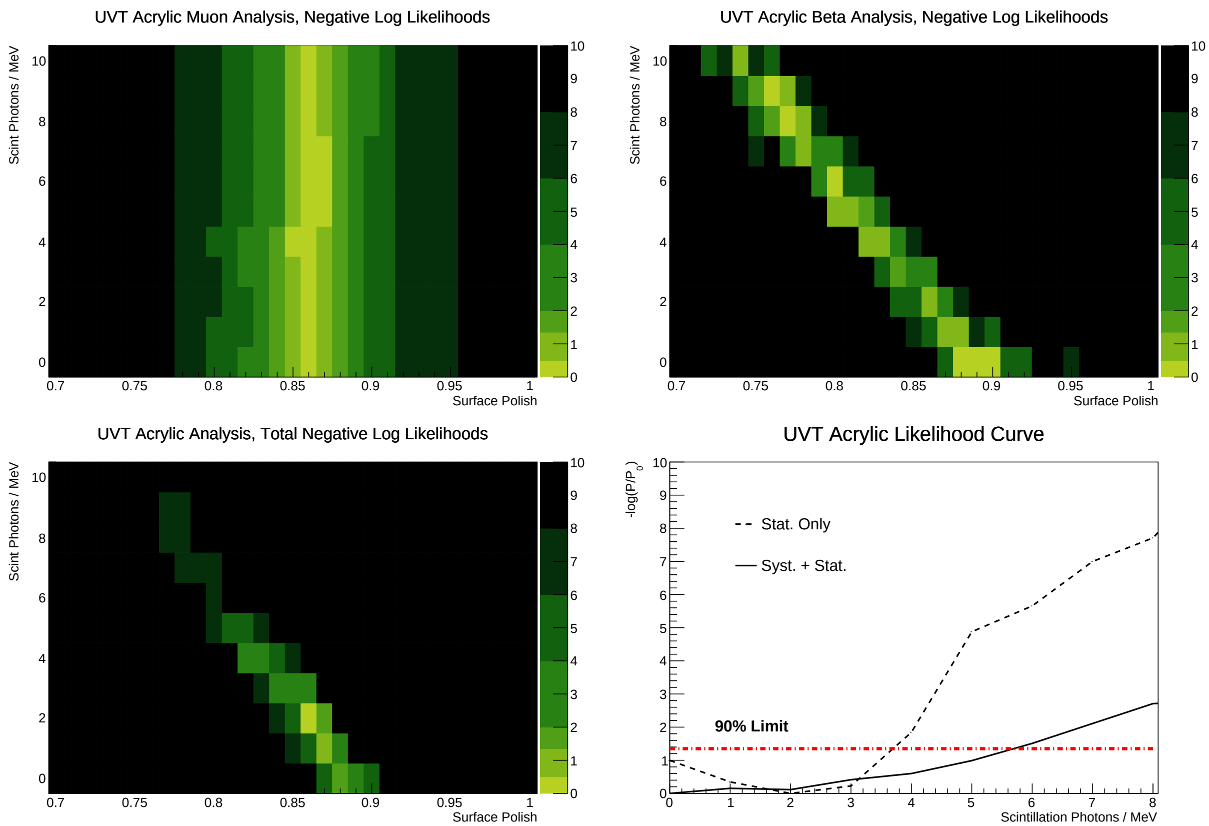


Figure 8.5: **Top:** negative log-likelihood curves for the cosmic muon (left) and ^{90}Sr β (right) analyses with an acrylic target. **Bottom:** results of the acrylic analysis obtained by combining the μ and β information, showing both the negative log-likelihood in the 2-dimensional space of polish and scintillation yield (left), as well as the negative log-likelihood curve of only scintillation yields marginalized over the polish parameter (right). Plots are all normalized to a minimum negative log-likelihood of 0. Reprinted from [89].

in acrylic, giving us an idea of the capability of this setup and procedure to constrain scintillation yields of materials that we expect to be nonscintillating.

The shape of the negative log-likelihood curves in Fig. 8.5 demonstrate the power of combining cosmic muon and β source measurements in the CHESS setup. Cosmic muons are high enough energy that they emit significant Cherenkov light passing through a 5-6 cm cubic target like what we use, so the cosmic-muon setup is not very sensitive to small additional amounts of scintillation light. However, since we select for muons traveling almost straight downwards through our target, the resulting Cherenkov light is highly directional in a way that we can predict. This makes the cosmic-muon configuration very sensitive to surface roughness effects, which change the expected distribution of light in the rings of PMTs beneath the target. On the other hand, β radiation from the ^{90}Sr source is mostly in the 1-2 MeV energy region, so any associated Cherenkov radiation is expected to be quite weak and small amounts of additional scintillation light are quite noticeable. In the radioactive-source configuration of the CHESS setup, the effect of surface roughness is to change the rate at which light inside the target is either internally reflected or allowed to escape, which changes the percentage of the light that is captured by the trigger PMTs directly coupled to the target. This results in a degeneracy between the effects of surface roughness and scintillation light yield in the radioactive-source configuration, but this degeneracy is broken by incorporating the cosmic muon data, giving the clear result in the bottom figures of Fig. 8.5.

8.4 TeO_2 Results

We turn now to the result of interest with this measurement, using a CUORE-style TeO_2 crystal as the target. As already described, we apply the same analysis procedure that was validated on the acrylic target, with the notable difference being that the TeO_2 crystal has two polish parameters p_1, p_2 for the glossy and matte faces respectively. The crystal is deployed with the 2 glossy faces on the top and bottom and the 4 matte faces on the sides, for the sake of symmetry in the CHESS configuration. We obtain best-fit parameters of a polish of 0.85 for the glossy faces, a polish of 0.55 for the matte faces, and a scintillation yield of 0, with the corresponding prediction from Monte Carlo plotted with the observed results in Fig. 8.6. The corresponding negative log-likelihood distributions are shown in Fig. 8.7, where for readability the results are all marginalized over the glossy face polish already. We see again that the cosmic-muon configuration provides information about the favored surface roughness parameters, which combined with the β data lets us narrow the permitted scintillation yields. After marginalizing over all polish parameters, we obtain a 90% confidence upper limit on β scintillation in TeO_2 of 5.3 scintillation photons / MeV. This confirms that scintillation-like light is negligible for TeO_2 compared to the 105 to 108 photons / MeV expected from Cherenkov radiation [85].

Building on the results of the μ/β analysis for TeO_2, we can also analyze data collected with a ^{241}Am α source in the radioactive-source configuration. We use Monte Carlo simula-

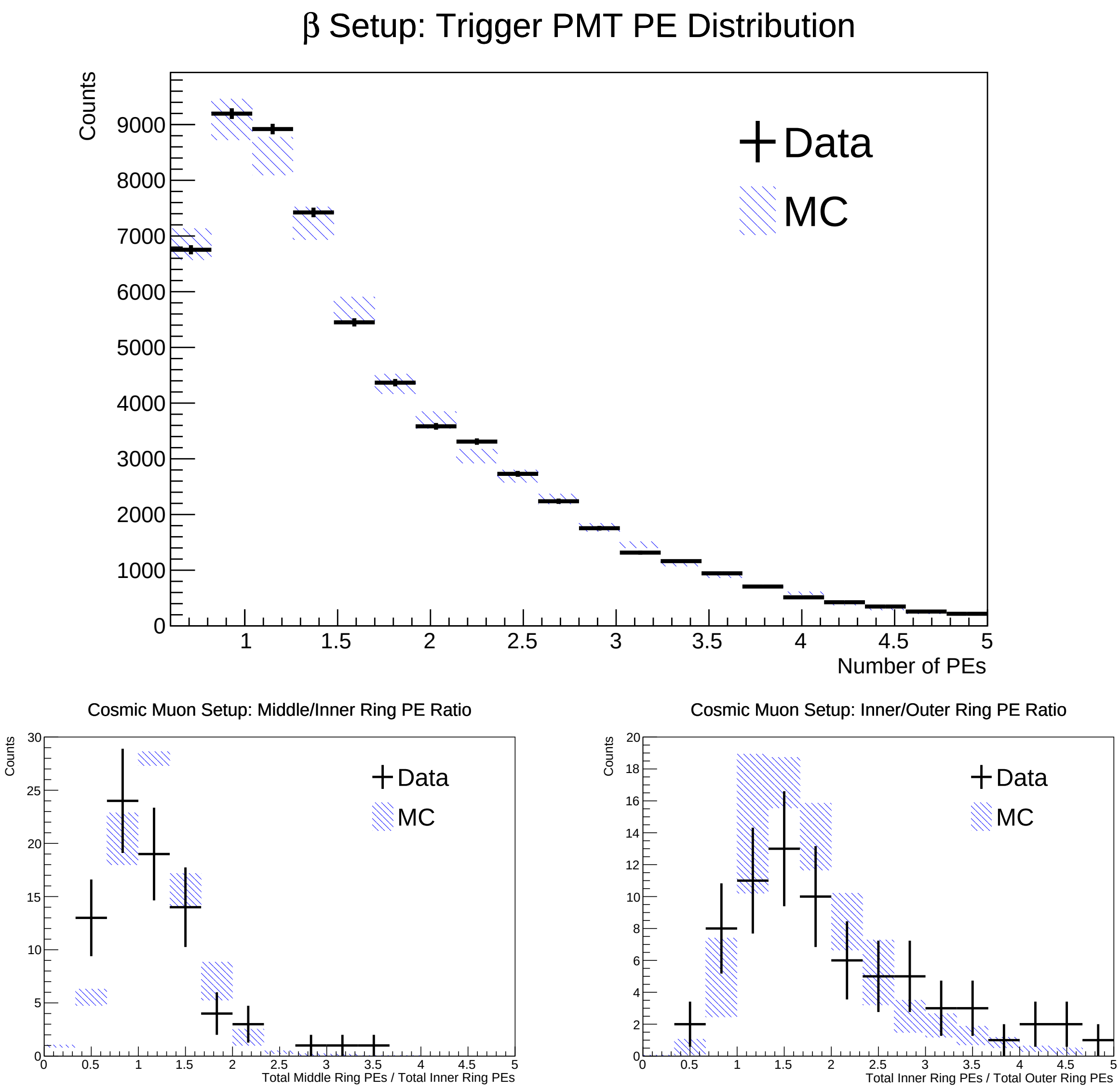


Figure 8.6: Comparison of Monte Carlo predictions to observed results for the ^{90}Sr and cosmic muon data using a TeO_2 target, where the Monte Carlo simulations are done with the best-fit polish and scintillation values obtained from the analysis. **Top:** the photoelectron distributions in the trigger PMT for the ^{90}Sr data that pass all the cuts described in Sec. 8.3. **Bottom:** ratios of the total light in each ring of PMTs for the cosmic-muon configuration. Reprinted from [89].

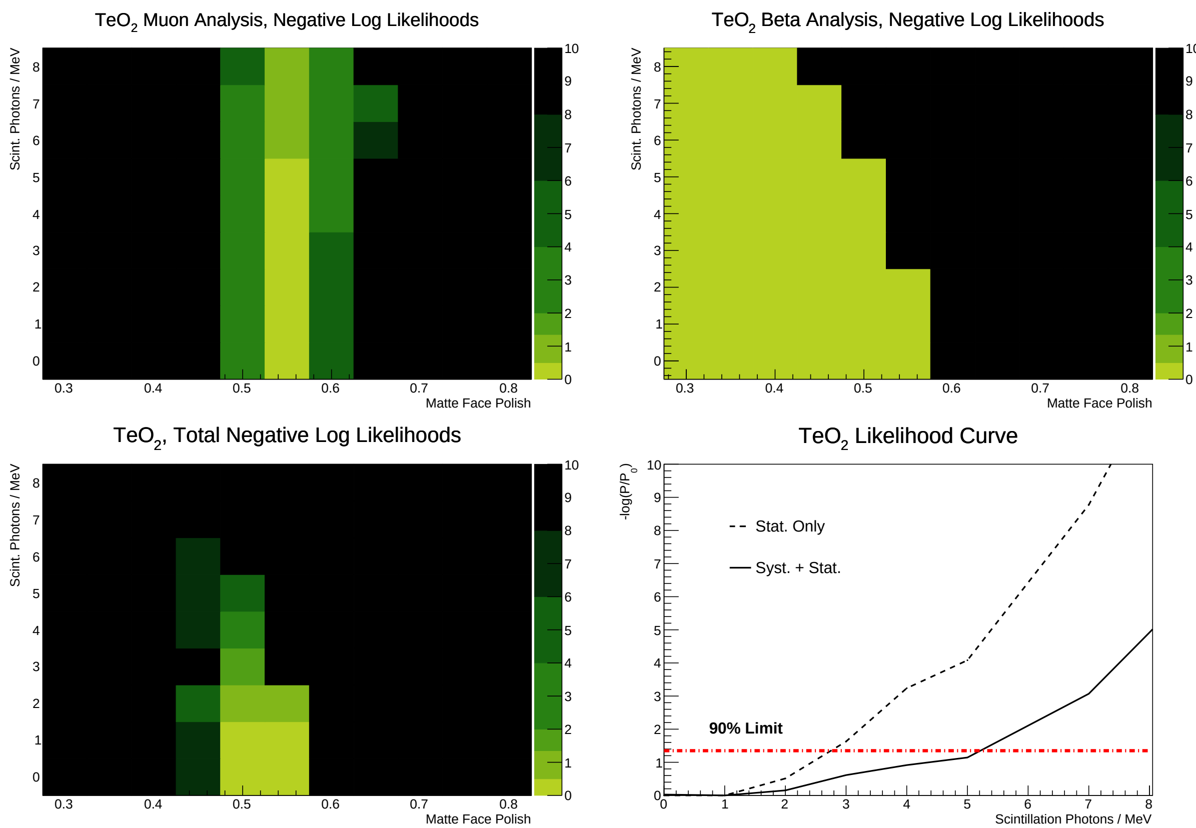


Figure 8.7: **Top:** negative log-likelihood of scintillation and matte face polish parameters obtained from the cosmic muon and ^{90}Sr data with a TeO_2 target, marginalized over the glossy face polish parameter for readability. We can see the cosmic muon data narrows the permitted polish parameters, and the ^{90}Sr data provides a band of permitted polish and scintillation parameters. **Bottom:** negative log-likelihood of the scintillation and matte face polish parameters obtained by combining the μ and β data (left), along with the total negative log-likelihood curve for TeO_2 β scintillation marginalized over both polish parameters (right). Reprinted from [89].

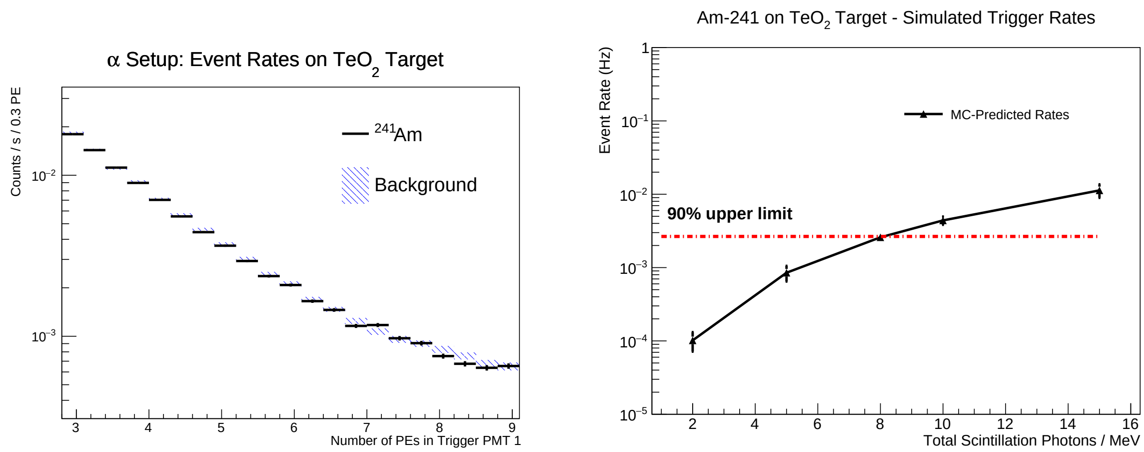


Figure 8.8: **Left:** event rates in the trigger PMT for the radioactive-source configuration, both with a ^{241}Am α source and without any source deployed. It can be seen that they are statistically indistinguishable. **Right:** event rates for possible α scintillation yields predicted by Monte Carlo, as well as 90% upper limit on the observed background-subtracted event rate, corresponding to a limit of < 8 scintillation photons / MeV. Reprinted from [89].

tion to predict the event rate expected for possible α scintillation yields in the radioactive-source configuration, fixing the polish parameters to those obtained from the μ/β best fit. We also assume the α particles from our source are unattenuated by the geometry of the source itself, which will result in a conservative limit. The observed rates are shown in Fig. 8.8, where we see that the presence of an α source has no discernible impact on the event rate compared to when no source is deployed at all. We account for the possibility of hardware-caused fluctuations in the event rate over time by splitting the datasets into n time chunks, including a final systematic uncertainty of $1/\sqrt{n}$ times the standard deviation of the rate over these time chunks. This gives us an event rate of 0.1610 ± 0.0004 (stat.) $\pm\, 0.0006$ (syst.) Hz for the α data and an event rate of 0.1609 ± 0.0005 (stat.) $\pm\, 0.0014$ (syst.) Hz for the background data after applying all of the cuts described in Sec. 8.3. Our final background-subtracted rate is then:

$$[0.1 \pm 0.6 \text{ (stat.)} \pm 1.5 \text{ (syst.)}] \times 10^{-3} \text{ Hz}$$

This corresponds to a 90% confidence upper limit of 2.8×10^{-3} Hz on the background-subtracted rate, which gives a limit on TeO_2 α scintillation of 8 photons / MeV. This confirms that α scintillation in TeO_2 is negligible compared to even the small amounts of Cherenkov radiation expected from β/γ particles.

A summary of the analysis results using these CHESS measurements is shown in Table 8.2. These results demonstrate how we can use the CHESS setup's ability to capture information about the spatial distribution of light to build more sophisticated optical models,

Table 8.2: Best-fit polish values and corresponding scintillation limits for UVT acrylic and TeO_2, obtained using the analysis procedure described in this section combining the cosmic muon measurements and radioactive source measurements in the CHESS setup. We reasonably find that the glossy TeO_2 faces have a similar best-fit polish value to the acrylic block, while the matte faces have a notably lower polish value. We also see that the β/γ scintillation limits are similar for both materials, which are both expected not to scintillate.

Material	Polish Value	β/γ Scintillation Limit (90% CL)	α Scintillation Limit
UVT Acrylic	0.86	$<$ 5.7 photons / MeV	-
TeO_2	0.85 (glossy face) 0.55 (matte face)	$<$ 5.3 photons / MeV	$<$ 8 photons / MeV

incorporating factors such as the varying surface roughnesses of crystals. In particular, the validated model we have obtained from these measurements tells us not only that both β/γ and α scintillation in TeO_2 are negligible, but that surface roughness is in general an important factor for light collection efficiency. Due to confirmations that TeO_2 light yields are very low, CUPID now plans to use scintillating Li_2MoO_4 crystals, which have much higher light yields and was extensively tested in the CUPID-Mo demonstrator discussed in the next chapter. These CHESS results indicate that crystal surface roughness will still be an important consideration when it comes to the question of optimizing light collection efficiency.

We do have to note that the polish parameter used in the Monte Carlo simulations here do not directly correspond to some real-life measure of the roughness of a surface, and we can only empirically match the polish parameter to observation as we have done in this analysis. But this still allows us to use Monte Carlo to make qualitative determinations such as whether we should prefer a smoother or rougher surface for each face of the crystal. More significant limitations are the fact that the CHESS setup is at room temperature and uses PMTs as the light detectors. The optical emission and transmission properties of Li_2MoO_4 are known to change with temperature [91, 92], and so a room temperature setup may not be sufficient to model the material's cryogenic optical behavior. Cryogenic light detectors also generally use a calorimetric detection method, with a Si or Ge wafer as the absorber, so they detect the total energy of all deposited photons instead of a count of the number of photons in some sensitive wavelength range like a PMT measures. However, a benchtop setup like CHESS could still be used in combination with cryogenic test setups to build a validated optical model that can be used to inform the geometrical design of the CUPID detector array, allowing us to explore different possibilities more efficiently.

Chapter 9

The CUPID-Mo Demonstrator

While the idea of using light detectors for particle discrimination in CUPID is straightforward, we of course want to test the design in a smaller-scale CUPID-like setup before spending tens of millions of dollars to commission the full experiment. This allows us to measure details such as the quality and radiopurity achievable with the crystals, the actual light yields we see, the strength of the particle discrimination we can achieve, and the energy resolution that can be obtained. A demonstrator should also use multiple towers containing multiple crystals, simulating the eventual CUPID setup, so that challenges that arise from operating multiple detectors simultaneously can be made apparent. The clearest way to test all of these aspects is to just do an actual $0\nu\beta\beta$ search, which requires the experiment to be conducted in a proper underground lab for shielding from cosmic muons.

CUPID-Mo was one such demonstrator, which aimed to show that ^{100}Mo contained in the form of Li_2MoO_4 was a good option for the final design of CUPID. It was located at the Laboratoire Souterrain de Modane (LSM) underneath the Alps in France, providing a rock overburden equivalent to about 4800 meters of water. This rock shielding reduces the muon flux in the underground experimental halls at LSM to about 3.8×10^{-7} counts / cm^2 / minute, over 6 orders of magnitude lower than at sea level [93]. CUPID-Mo operated 20 LMO crystals containing highly enriched Mo, for a total of 4.158 kg of LMO with 2.264 kg of ^{100}Mo [94]. They were placed into spare space in the EDELWEISS-III cryostat, a custom-made reversed wet dilution fridge [95]. The reversed geometry places the colder stages on top of the warmer ones, with the helium dewar below the cryostat. As it is a wet fridge, the liquid helium bath must be periodically replenished, but losses are minimized with the use of 3 cryocoolers that reliquefy the cold vapor. The CUPID-Mo detectors were installed in early 2018, took data until a cryogenic maintenance in fall of 2018, and then continued to stably take data from early 2019.

CUPID-Mo ultimately was a success, demonstrating consistently high performance in its LMO crystals from both the energy resolution and radiopurity perspectives, as well as effectively complete $\alpha/\beta(\gamma)$ separation with its light detectors. This helped set the decision for CUPID's baseline design to use scintillating LMO crystals. CUPID-Mo completed its data collection in summer of 2020, with a final exposure of 2.71 kg·yr of ^{100}Mo. This chapter presents its data collection and analysis procedure, as well as the $0\nu\beta\beta$ analysis results coming from an analysis of its first 1.17 kg·yr of ^{100}Mo exposure.

9.1 Data Collection and Analysis

Detector Setup and Analysis

CUPID-Mo uses cylindrical LMO crystals, measuring 44 mm in diameter and 45 mm in height. Each crystal is contained within a copper structure and surrounded with reflecting foil on the lateral sides, with a germanium wafer placed below to serve as the light detector, as shown in Fig. 9.1. The light detectors are coated in SiO to reduce their reflectivity and increase their light collection efficiency. Both the LMO crystals and Ge light detectors are equipped with NTD thermistors to measure the energy deposits in them. The detector modules are arranged into 5 towers for installation in the cryostat, with each tower containing 4 crystals stacked vertically. Each of the lower crystals is also able to see the light detector from the detector module directly above it, with the uppermost crystal in each tower being capped by reflective copper on top. 15 of the LMO crystals thus effectively have two light detectors each, while the 5 topmost crystals only have one each.

This analysis uses data collected at operating temperatures of 20.7 and 22 mK. Periods with excessive noise or otherwise poor conditions are manually rejected, and one of the LMO detectors is rejected for exceptionally poor performance[1]. The data streams for both the LMO detectors and the light detectors are continuously saved with a sampling rate of 500 Hz, and events are analyzed as 3-second time windows with 1 second before the trigger and 2 seconds after. The data are organized into datasets containing at least one calibration period each, with the physics data in each dataset being collected under generally similar operating conditions. Calibrations are performed with mixed U/Th sources deployed outside of the cryostat, with the most prominent and highest-energy calibration line coming from 2615 keV γ rays from ^{208}Tl decays.

The general structure of the CUPID-Mo analysis chain is similar to that of CUORE described in Chapter 4, using a software-based derivative trigger for live data monitoring and an optimum trigger for later full analysis, with the events then going through optimum filter-based amplitude evaluation, stabilization, calibration, and finally multiplicity and pulse shape cuts before yielding an energy spectrum that can be used for a $0\nu\beta\beta$ analysis. CUPID-Mo obviously has the additional step of calculating the light yields observed in coincidence with any particular heat event in a LMO crystal and using these values to perform α rejection; this step is done at the amplitude evaluation stage, so that all following steps can take

[1]The single poor-performing LMO crystal was an old crystal reused from the LUMINEU experiment, where it exhibited a good energy resolution similar to the other LMO crystals [96]. Its poor performance in CUPID-Mo is thus not considered a flaw of the LMO crystal-growing procedure in general, since all other LMO crystals performed well. Possible explanations for this one detector's performance include a problem with the specific sensors or electronics used for that crystal, or a result of an experimental water-etching procedure used to treat that specific crystal after the LUMINEU tests.

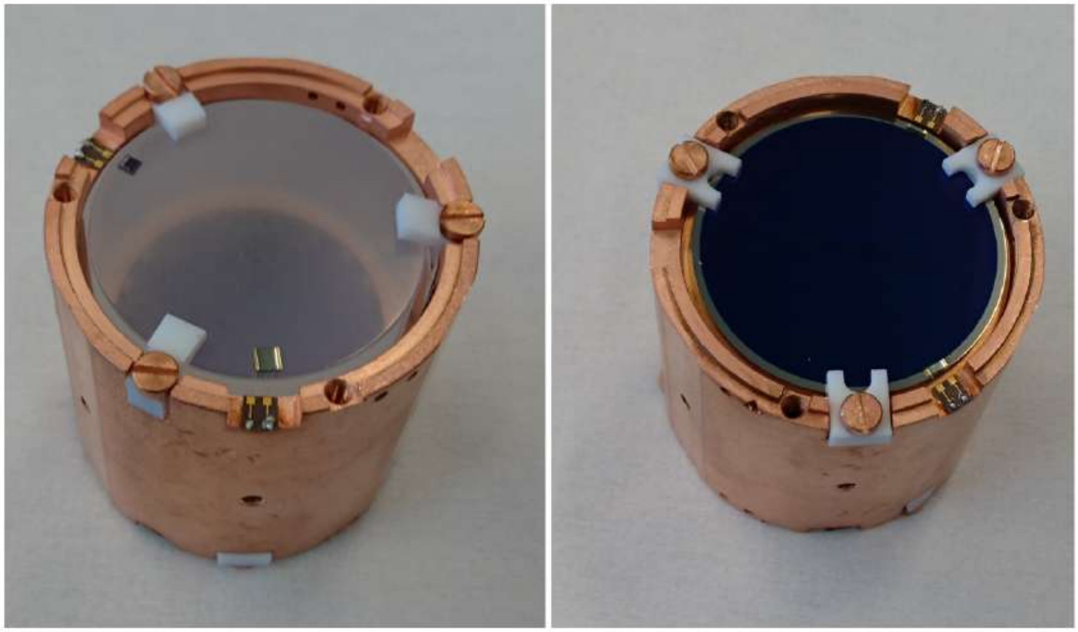


Figure 9.1: **Left**: one of the CUPID-Mo detector modules as seen from the top, with the cylindrical LMO crystal contained in its copper structure. The crystal's transparency to light can be seen here. **Right**: the same detector module from the bottom, where the Ge-based light detector is visible. The light detector's placement allows it to capture most of the light escaping out of the bottom face of the crystal. These cylindrical modules are stacked on top of each other to form the CUPID-Mo towers. Photos reprinted from [94].

advantage of the light yield cut to "clean" the data where needed. I will detail here the primary differences in the CUPID-Mo procedures compared to the CUORE ones.

Light Yield Analysis

The most significant difference between the CUPID-Mo analysis and the CUORE analysis is the existence of the light detectors. Each time a LMO detector is triggered, we analyze a 3-second window from the corresponding light detector(s) at the same time. The time windows from the light detectors are also passed through optimum filters and have their optimum-filtered amplitudes extracted. These light amplitudes can then be plotted against the heat energies registered in the corresponding LMO detector to see the band of expected light yields from β/γ events. For the LMO detectors that face two light detectors, we use a resolution-weighted average of the signals in the two detectors as the measure of the light yield for an event. The resulting plot of the observed light yield against the total energy of the heat deposit in the LMO crystal is shown in Fig. 9.2 for a typical detector. From here

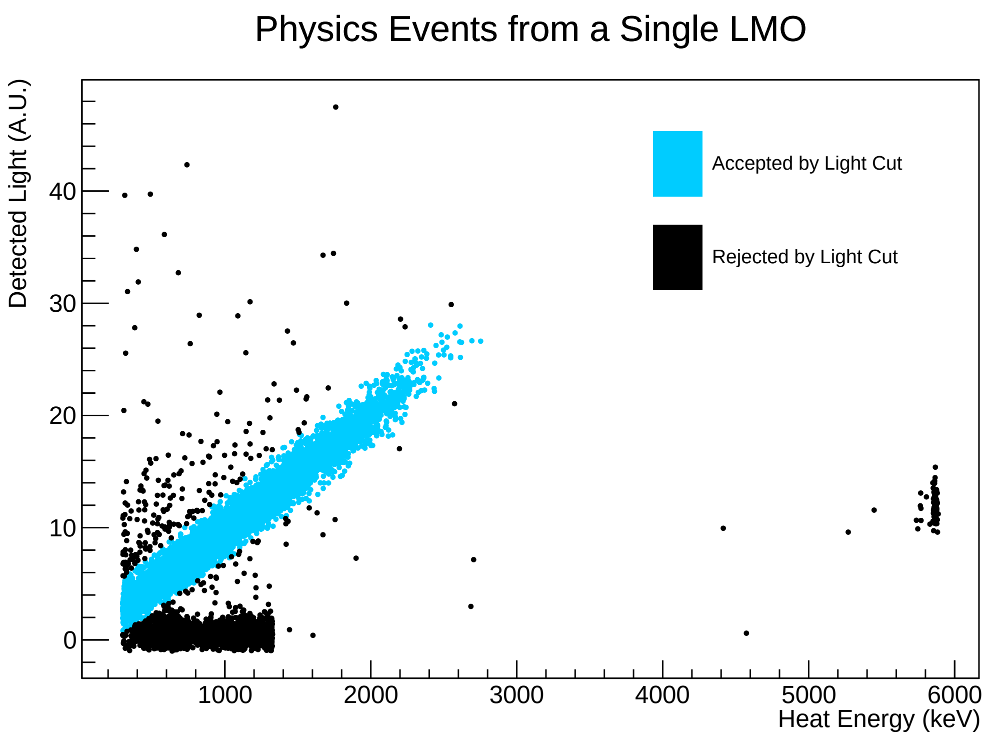


Figure 9.2: The detected light versus detected heat for one of the LMO crystals from one of the datasets in CUPID-Mo. The light detectors are not calibrated here, so the amount of detected light is in arbitrary units. The events that pass the light cut are highlighted in blue, with all events falling outside this band being rejected as non-β/γ events. Some clusters of rejected events at low energies are from noise or heater events, while the clusters at higher energies can be identified as natural α backgrounds.

we perform a robust fitting procedure to eliminate the outlier populations, such as those with near 0 light yield. This allows us to determine the average light yield of a "good" event as a function of its energy, as well as the typical variation in light yield for an event of any particular energy. The light yield cut is then applied by eliminating events that deviate by more than 3σ from the expected light yield of an event at that energy, giving the resulting band of accepted β/γ events. This procedure and the light cut thresholds are determined separately for each LMO detector.

Looking at events that are eliminated by the light yield cut in Fig. 9.2, we see a few classes of rejects. There are the events that lie along a band near 0 light but have heat energies in the β/γ region, which are mostly either heater events[2] or isolated noise. Next,

[2]Unlike CUORE, CUPID-Mo did not have the ability to externally flag its heater events, so they had to be removed from the physics spectrum during processing instead.

there are the scattered events that lie above the band of expected light yields, which can be caused by correlated noise across detectors or occasionally by scattered γ particles that deposit large amounts of energy in the light detector. Lastly, there are the events which have some light yield but are still below the band of expected β/γ yields. These are the α events, which exhibit a quenched light yield of about 20% the amount seen with β/γ events. A cluster of these events can be seen above 5 MeV, coming from ^{210}Po contaminations[3].

During the initial processing we also saw events whose heat energies fell within the range expected from high-energy natural α peaks, but which nonetheless passed the light yield cuts and seemed to be good multiplicity-1 events. However, closer inspection showed that when these types of events occurred in crystals with two light detectors, one of the light detectors registered a much larger energy deposit than the other. This suggests that these events are caused by α decays where the nuclear recoil or a very degraded α particle reaches the light detector instead of being fully absorbed in the LMO crystal, resulting in an energy deposit in the light detector that emulates a light signal. These events are eliminated by an additional cut on crystals with two light detectors that requires the light yields seen in each light detector to be not too different. This is determined with the use of another robust fitting procedure to find the typical difference in light yields between the two light detectors for events already accepted by the previous light yield cut, from which we then eliminate events where the difference in light yields is more than 3σ away from the typical value.

Stabilization

The LMO crystals in CUPID-Mo are equipped with heaters similar to those of CUORE for thermal gain stabilization, but the CUPID-Mo heater events could not be externally flagged and were often unstable. As a result, the heaterTGS procedure described in Chapter 4 was not used, and stabilization in CUPID-Mo was done entirely with the calibrationTGS procedure. Furthermore, some of the datasets in CUPID-Mo only had one calibration period, instead of the typical structure of one calibration at the beginning and one calibration at the end that CUORE follows. When there are two calibration periods, the calibrationTGS procedure generally has at least two distinct populations of events at different baselines to perform the fitting procedure, which usually results in a well-constrained linear fit of the unstabilized amplitude against baseline. However, when there is only one calibration period, the calibrationTGS procedure is much less well-constrained, leaving a possibility of systematic offsets in the stabilized amplitudes of events in the physics data, corresponding to an additional systematic uncertainty on our energy calibration. In extreme cases, a low statistics calibration could even result in obtaining a stabilization function that has the opposite slope of the "true" stabilization function.

To quantify this effect, we generate toy Monte Carlo samples to test for how off our stabilization procedure could be for these datasets with only one calibration period. For

[3]The heat energies of this cluster of events are not at the expected values for ^{210}Po decays because this plot uses our β/γ calibration. α events in LMO have quenched yields and need a separate calibration to have proper energy estimates.

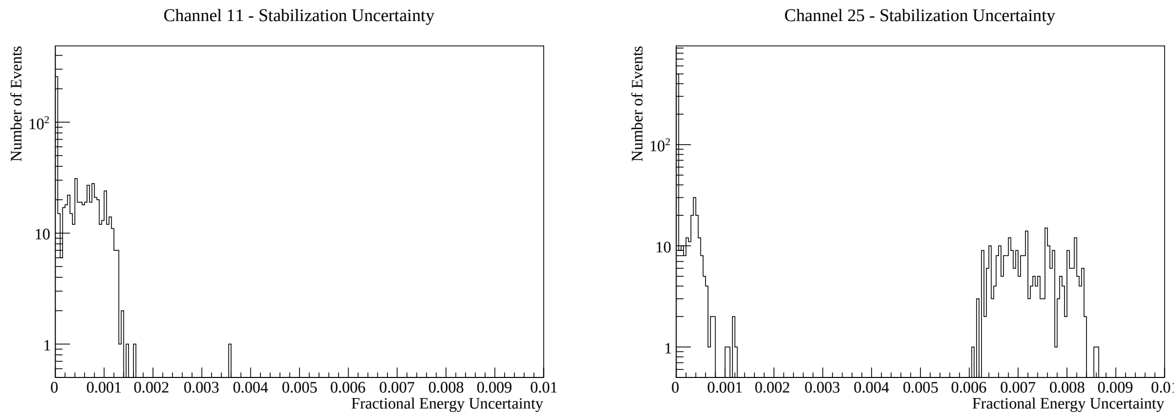


Figure 9.3: Uncertainties on the energies of physics events from two selected detectors from a dataset that had only one calibration period. Only events with energy > 500 keV and passing the light cut are considered, to obtain an estimate of how much the stabilization uncertainty affects signal-like events. **Left**: an example of a channel that does not suffer heavily from stabilization uncertainties, either because the stabilization had high statistics or because the physics runs did not deviate much from the operating conditions of the calibration runs. **Right**: a channel that has a population of events with significant energy uncertainties due to stabilization, caused by physics data collected at baselines far from the operating conditions of the calibration.

each detector, the calibrationTGS procedure gives a stabilization function slope $a_{observed}$ obtained from a fit over n data points that were identified as 2615 keV γ events, and which occurred over some range of baseline values. We then assume a FWHM energy resolution of 7 keV at 2615 keV and assume the true stabilization function is linear. We scan a range of possible true stabilization function slopes, and for each test value a_{test} of the slope, we randomly generate n points with (Baseline, Amplitude) values. The baselines are sampled from a uniform distribution over the actual range of baselines seen in the calibrationTGS procedure. The amplitudes are sampled from a Gaussian distribution with mean μ given by the expected amplitude at that baseline given a_{test}, and with $\sigma = 0.001\mu$ given by the assumed energy resolution. We then fit the resulting collection of n points and extract the corresponding stabilization function slope. We repeat this procedure 1000 times for each a_{test} and obtain a distribution of possible observed values of the slope a given a true value a_{test}, forming a Gaussian distribution centered on a_{test}. If the stabilization function slope $a_{observed}$ falls within 3σ of a_{test} in this distribution, we say that a true slope of a_{test} is compatible with our observed value of $a_{observed}$. We then take the range of compatible a_{test} values as the uncertainty on our stabilization slope $a_{observed}$.

We can then convert this uncertainty on the stabilization function into uncertainties on the energies of events from physics data, depending on the baselines of each event. Events from physics runs that were collected at baselines far from the baselines of the calibration runs will tend to have higher energy uncertainty, and detectors with low calibration statistics will tend to have higher uncertainties on the stabilization as well. Examples of the resulting energy uncertainties for physics events in two detectors are shown in Fig. 9.3. We see in all cases that the stabilization uncertainty leads to energy uncertainties much less than 0.1% in the vast majority of events, which is insignificant compared to our normal energy resolution. There is some small portion of events with higher uncertainties, coming from physics runs collected at significantly different baselines compared to the calibration runs, but the magnitude of their uncertainties varies between detectors.

We cut events with unacceptably high stabilization uncertainties by tuning a figure of merit $\epsilon/\sqrt{B}$. ϵ is the efficiency of the cut and calculated with the number of events we're cutting compared to the total number of physics events in the detector, making the assumption that event rates are approximately constant across all physics runs. The background is determined as the number of background events that could spill into the region around $Q_{\beta\beta}$ as a result of this energy misestimation, allowing them to be confused for $0\nu\beta\beta$ events. The resulting cuts follow what would be naively expected, eliminating the clusters of events with high uncertainties. The resulting loss in efficiency is negligible and has no effect on our reported exposure.

Pulse Shape Discrimination

CUPID-Mo uses a PCA-based procedure for PSD similar to that employed by CUORE. Like in CUORE, the goal of the PSD procedure is to eliminate pileup and noise-like events that could pollute the $0\nu\beta\beta$ search region. Unlike in CUORE, there are light detectors that already serve to purify our selected events, eliminating both α events and many noise events. We train the PCA components for each detector in CUPID-Mo using the multiplicity-1 events from the 1000-2000 keV energy range in the physics data that pass the light yield cut. These events should be mostly $2\nu\beta\beta$ events and γ backgrounds, which suffice as a mostly pure sample of "good" events. In total, this training sample selection procedure gives us over a thousand events per detector per dataset, which provides enough statistics to obtain a proper result. The principal components we obtain from this training for one detector are shown in Fig. 9.4. Unlike in CUORE, we can see that the signal-like features are not contained entirely in the leading principal component. While noise-like features become more and more prominent as we consider the subleading components, these components cannot be completely ignored if we want to fully describe the signal-like behavior of events. This can be quantified by the explained variance per component, shown in Fig. 9.5, where we see that the leading component captures only 82% of the variance, compared to the $> 90\%$ seen in Fig. 5.2.

We compute the reconstruction error with a similar formula to that used in CUORE as described in Chapter 5, but using these trained PCA components instead of the average pulse. For CUPID-Mo with its events defined by 3-second time windows sampled at 500 Hz, the waveform data for an event can be written as a vector $\boldsymbol{x}$ with dimension $D = 1500$. If

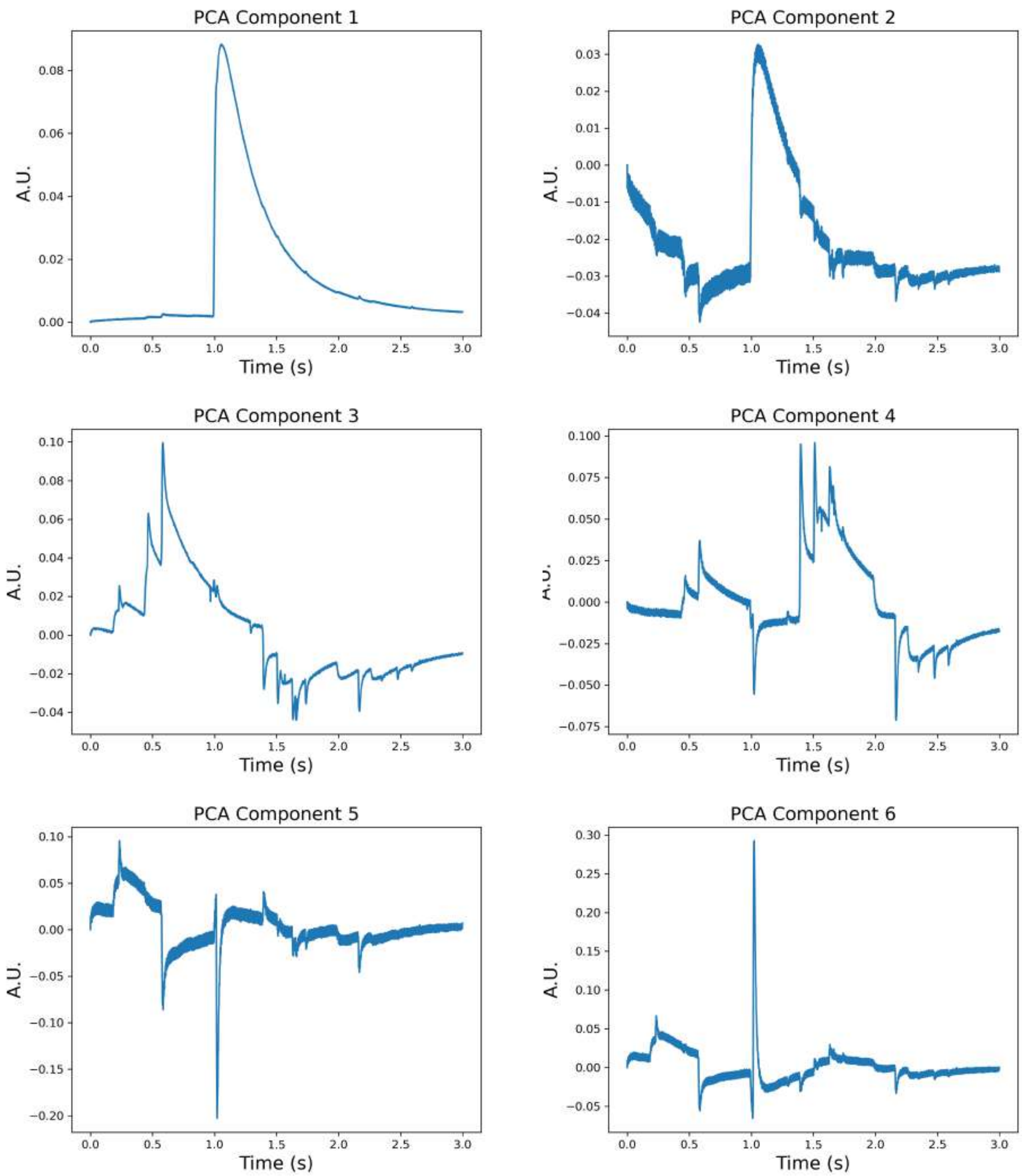


Figure 9.4: Leading principal components obtained by the PCA procedure trained on events from physics data for one detector in one dataset in CUPID-Mo. We see that the leading component most closely emulates the ideal response of a detector, while the subleading components still contain some signal-like behavior but contain increasing amounts of noise-like behavior.

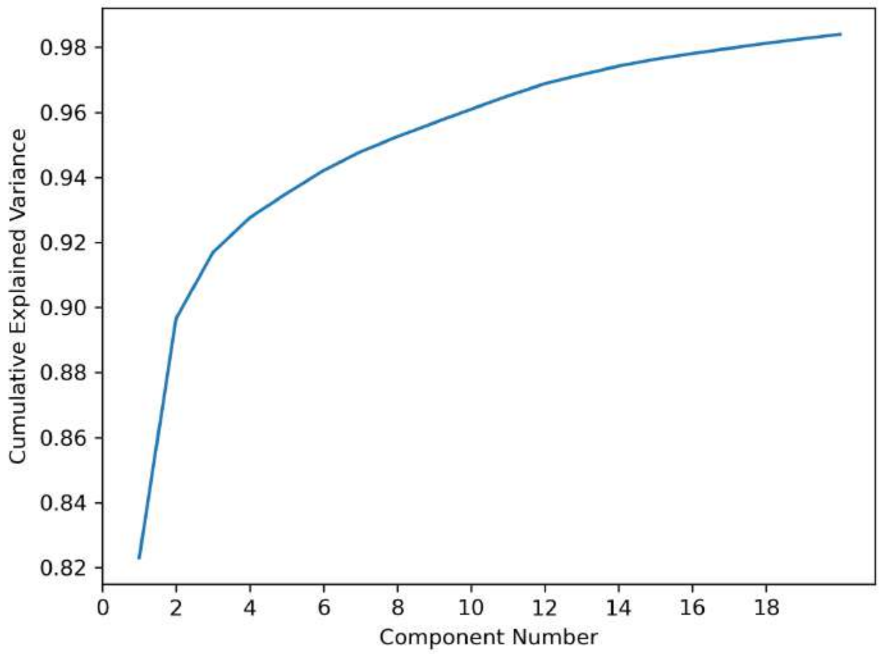


Figure 9.5: Cumulative fractional explained variance as a function of the number of PCA components used, for the same detector and dataset shown in Fig. 9.4. Using more than one component allows us to explain a non-negligibly greater fraction of the variance.

we use the N principal components $\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_N$ from the PCA training procedure, then the reconstruction error is defined as:

$$RE = \sqrt{\sum_{d=1}^{D} \left[x_d - \sum_{i=1}^{N} (\boldsymbol{x} \cdot \boldsymbol{u}_i) u_{i,d} \right]^2}$$

Here, x_d indicates the value at index d of $\boldsymbol{x}$, and $u_{i,d}$ indicates the value at index d of $\boldsymbol{u}_i$. After the reconstruction errors are calculated, they are normalized in the same fashion as in CUORE, and we can choose a cut based on the normalized reconstruction errors for each dataset. We compute these normalized reconstruction errors for a range of possible values of N, the number of principal components we use in the reconstruction error calculation. To compare their performances, we can look at the standard receiver operating characteristic (ROC) curves, plotting the signal efficiency against the background rejection capability.

While the signal efficiency can be estimated in CUPID-Mo in the same way that is done in CUORE, looking at the 2615-keV γ peak, there is a question of how we can estimate the background rejection efficiency. In CUORE, the degraded α region in 2700-3100 keV can be used as an effective proxy for the ROI background. In CUPID-Mo, with its particle discrimination capabilities using its light detectors, there are no background events expected in the ROI. We instead postulate that our PSD procedure for CUPID-Mo should only be removing pileup and noise-induced events, which are the remaining uncontrolled backgrounds that could still pollute the $0\nu\beta\beta$ ROI. To estimate this background, we consider the number of events in the energy region from 2750 to 2980 keV in calibration data. Since the highest energy peak in our calibration sources is the 2615 keV γ peak from ^{208}Tl, events in this region are mostly caused by pileup or other kinds of noise.

The resulting ROC curves for different numbers of principal components used in the reconstruction error calculation are shown in Fig. 9.6. In these plots, reaching further into the top left indicates better performance, as this corresponds to high signal acceptance with high background rejection. There is noticeable improvement when we use more than one principal component, but the performance levels off after we get to 4 or so. The superiority of the 4-component cut over the 1-component cut can be seen in the calibration peak shown in Fig. 9.7, where the 4-component cut preserves similar numbers of events within the peak while cutting more from the tail. The overall efficiency of the 4-component PSD cut is shown in Fig. 9.8, evaluated on a number of γ peaks from the physics data.

The performance of the PCA method for PSD in CUPID-Mo is markedly different from its performance in CUORE in how it prefers the use of multiple principal components for the definition of the reconstruction error. This makes the average pulse based approach of CUORE less useful for CUPID-Mo as well. This difference is likely caused by the presence of baseline and high-frequency noise in the CUPID-Mo waveforms, which inflate the reconstruction error of events even though we would not want to qualify those types of events as "bad" enough to be rejected by PSD. Using too few principal components in the calculation thus results in the procedure having difficulty distinguishing between small pileup features and constant amounts of high frequency noise, causing the corresponding ROC curve to be pushed downwards and to the right. One way to address this issue could be to apply the PCA procedure and calculate reconstruction errors on the optimum-filtered waveforms instead of the unfiltered waveforms. Another approach could be to consider windowed reconstruction errors, evaluated only over some subset of the full D dimensions of a data vector $\boldsymbol{x}$. For instance, if we only look at the waveform at the trigger time $\pm$100 ms, we would have extra sensitivity to pileup features in this time frame that would affect our energy estimates, without having to deal with constant noise across the whole time window of the event. Neither of these techniques were used in CUPID-Mo's results, but could be useful for improving the PSD in CUPID, where pileup will be one of the dominant backgrounds.

9.2 $0\nu\beta\beta$ Search Results

The physics spectrum from an exposure of 2.16 kg·yr of LMO, corresponding to 1.17 kg·yr of ^{100}Mo, is shown in Fig. 9.9. The base and anticoincidence cuts eliminate most noise-induced events and cosmic muon events. The light yield cut then eliminates almost all

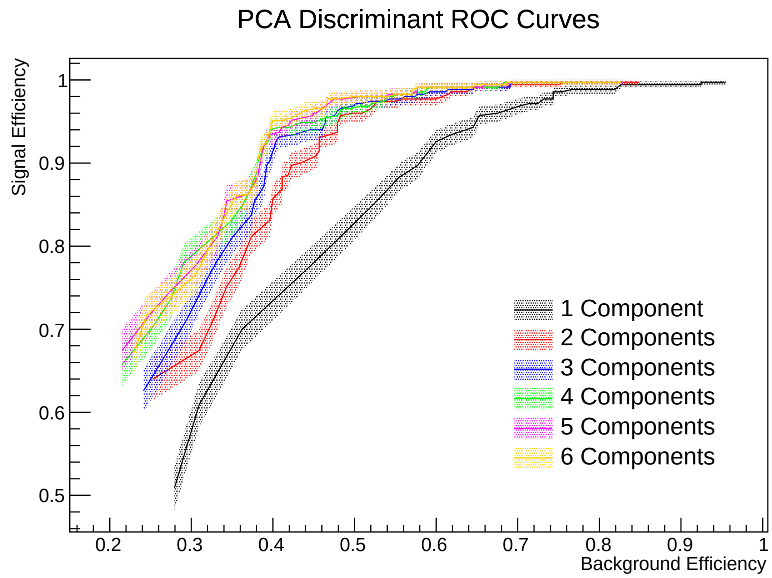


Figure 9.6: ROC curves showing the signal efficiency evaluated on the ^{208}Tl 2615 keV peak in the physics data against the background efficiency evaluated on the 2750-2980 keV events in calibration. The curves shown are for the reconstruction error calculated using from 1 to 6 leading principal components from the PCA training procedure, showing that we see improvement using more than one component. Shaded bands correspond to the 1σ statistical uncertainties on the efficiencies. Reprinted from [97].

events above 3 MeV, which mostly consist of α decays, though there do exist β decays from primordial backgrounds that can reach this energy region, such as the decays of ^{208}Tl (Q-value of 5001 keV) and ^{210}Tl (Q-value of 5489 keV). Finally, the PSD cut eliminates trace amounts of pileup or other noise that escaped the previous cuts. In comparison to the CUORE spectrum shown in Fig. 6.5, the PSD cut in CUPID-Mo has a much less noticeable effect on the backgrounds relevant to a $0\nu\beta\beta$ search. This is because most of these events are already eliminated through the use of the light detectors in CUPID-Mo, but we can see that the PSD does eliminate a few events that would otherwise populate the high-energy region.

We specify the region of interest for the $0\nu\beta\beta$ search to be 2984 to 3084 keV, taking 50 keV on either side of the Q-value of 3034 keV for ^{100}Mo. Since we originally expected to have close to zero events in the ROI, we blind the data by simply hiding all events with energies in the ROI. After the analysis procedure is finalized, we unblind by revealing any events that pass the cuts that we decided upon before unblinding. We found no events close enough to $Q_{\beta\beta}$ to be $0\nu\beta\beta$ candidate events, but there was one event near the upper edge of the ROI. Further inspection showed that this event could be cut with a delayed coincidence analysis, specifically looking at the ^{212}Bi $\rightarrow$ ^{208}Tl $\rightarrow$ ^{208}Pb decay chain. The ^{212}Bi decay includes an

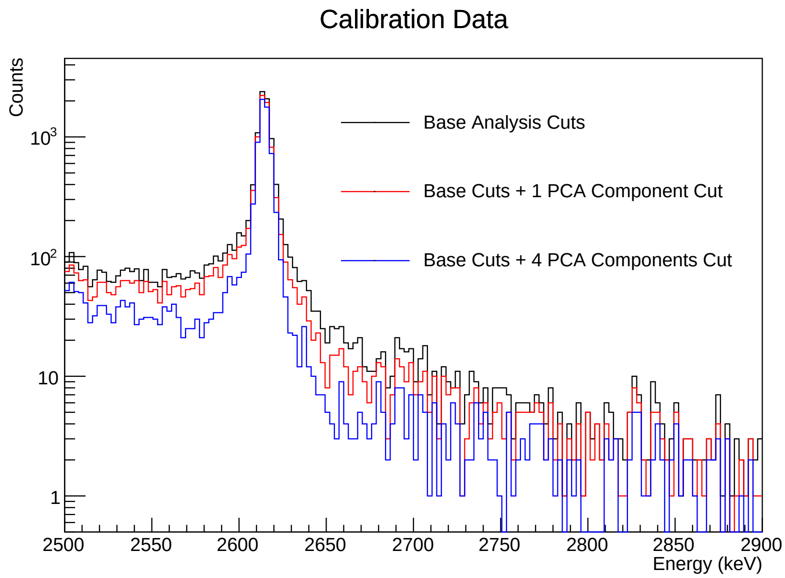


Figure 9.7: 2615-keV calibration peak for all detectors in one dataset. The base analysis cuts include basic data quality cuts (eliminating noisy periods and applying a simple pileup cut like the one described in Chapter 6), the anti-coincidence cut, and the light yield cut. We see that the PCA-based PSD disproportionately cut events from the sidebands of the peak, where the pileup or noisy events that slip through the base analysis cuts constitute a higher proportion of the total number of events in the energy region. Reprinted from [97].

α daughter with 6.2 MeV of energy, while the ^{208}Tl decay in the second step includes a β daughter with energy up to 5 MeV, having the potential to pass our light yield cut and fall within the ROI. However, the second decay only has a half-life of 183.2 seconds. We can thus almost entirely eliminate this background if we veto events that occur soon after a detected ^{212}Bi decay. This is done by eliminating all events that occur within 1832 seconds after an α event in the 6 to 6.3 MeV energy range. Since ^{212}Bi events are quite rare in CUPID-Mo's detector environment anyway, this results in a negligibly small 0.02% loss in exposure, but it eliminates the one event near the upper end of our ROI.

With no events in the ROI, we can extract a naive approximate limit on the $0\nu\beta\beta$ decay rate of ^{100}Mo with simple counting statistics by considering this to be a background-free experiment. The total number of expected $0\nu\beta\beta$ events follows a Poisson distribution with $\lambda = N_A Mt\Gamma_{0\nu}/A$, with N_A being Avogadro's number, $A = 100$ being the atomic mass of ^{100}Mo, and Mt being the ^{100}Mo exposure. With an overall analysis efficiency $\epsilon = 90.6\%$, the probability of observing 0 events follows from the Poisson probability distribution:

$$\sum_{k=0} (1-\epsilon)^k \frac{\lambda^k e^{-\lambda}}{k!}$$

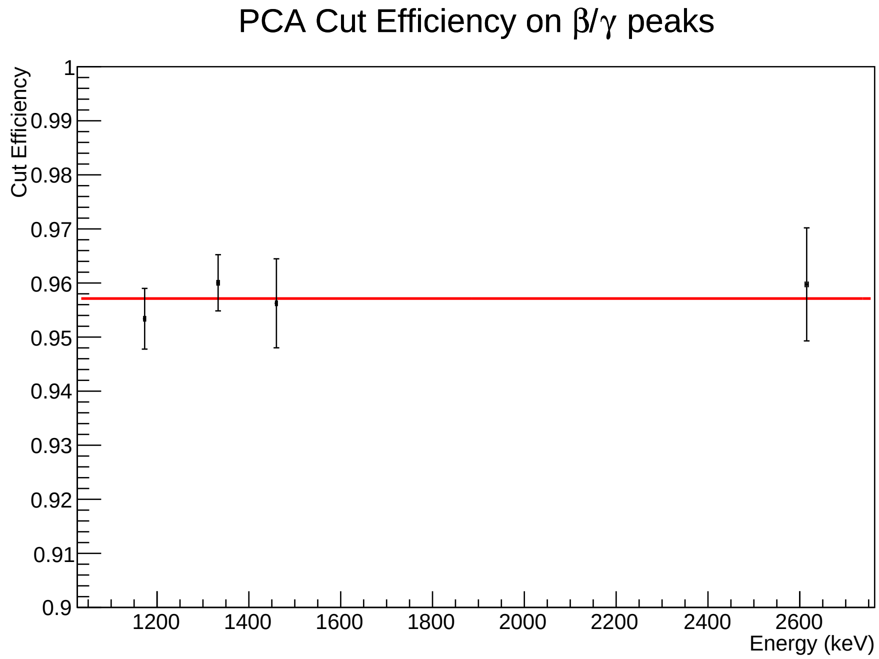


Figure 9.8: Efficiency of the PCA-based PSD cut on the multiplicity-1 events from the natural ^{60}Co, ^{40}K, and ^{208}Tl γ peaks for a collection of datasets, fitting each peak with a Gaussian signal plus flat background in the same style as CUORE (see Chapter 5). A constant fit is drawn in red to estimate the overall efficiency, since no significant energy-dependence is observed.

The 90% C.L. limit on $\Gamma_{0\nu}$ is then determined by the value that gives a 10% chance of observing 0 events from this expression. This comes out to a 90% C.L. limit of $\Gamma_{0\nu} < 3.6\cdot 10^{-25}$ yr^{-1}, which can be considered a very simplified Frequentist limit. This corresponds to a half-life limit of:

$$T_{1/2}^{0\nu} > 1.9 \cdot 10^{24} \text{ yrs}$$

Of course, this does not account for any systematic uncertainties on the isotopic abundance of ^{100}Mo in our enriched crystals and the efficiencies involved in our analysis. More significantly, it does not account for the fact that we do expect the possibility of backgrounds in our ROI, though the expected number of background events is low enough that our observation of 0 is an unsurprising result. A full Bayesian analysis accounting for these factors is done in [40], yielding a 90% CI limit of $T_{1/2}^{0\nu} > 1.5 \cdot 10^{24}$ yrs. This corresponds to a limit on the effective Majorana mass of $m_{\beta\beta} < 310 - 540$ meV, using the current range of available nuclear matrix elements [29].

CUPID-Mo has now completed its data-taking and has been decommissioned so that its detectors can be used for studies elsewhere. Its full exposure will be used for another

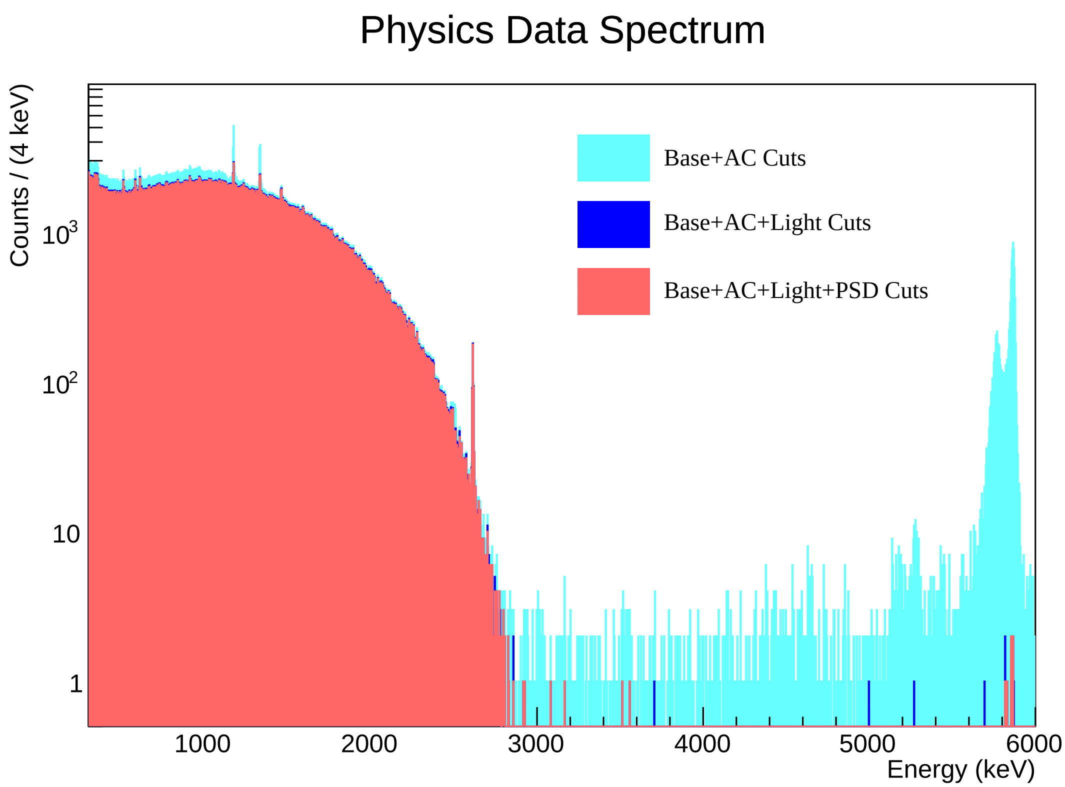


Figure 9.9: Energy spectrum for physics data from 2.16 kg·yr of LMO exposure in CUPID-Mo. The base and anticoincidence (AC) cuts are our basic data-quality cuts, and the light yield cuts eliminate almost all events above the β/γ region. The PSD cut eliminates some trace amounts of remaining events above the β/γ region, which were the result of pileup or other noise.

$0\nu\beta\beta$ analysis, a higher precision measurement of the $2\nu\beta\beta$ half-life of ^{100}Mo, and other excited-states and exotic-physics analyses. This limit on the $0\nu\beta\beta$ decay half-life of ^{100}Mo using a subset of CUPID-Mo's full dataset has already surpassed the previous limit set by NEMO-3 while using a significantly smaller exposure [98], demonstrating the power of both the high efficiency of the cryogenic calorimetric method and the background rejection capabilities of using the dual readout of heat and light signals. It has also served as a demonstration of the types of backgrounds CUPID may still face and how we can address them. For instance, certain β backgrounds can be rejected with delayed coincidence analyses tagging the decays to their parent nuclei, and pileup backgrounds can be rejected with more sophisticated PSD techniques. The all-around excellent performance of the LMO crystals and light detectors in CUPID-Mo allow us to say that using the same technologies for CUPID is a "conservative" baseline design choice, providing confidence that even in the absence of significant new developments we can reach the basic requirements of CUPID using this existing tested technology.

Chapter 10

Cryogenic Electronics for CUPID

While the physical technique and conditions used for a particle detector are obviously important, the electronics system used for readout is of utmost importance as well. Electronic noise can easily spoil the fidelity of a physical signal if not properly handled. The CUORE electronics for biasing the NTD sensors and performing signal amplification and filtering are tightly controlled for stability and are tuned to be able to optimally capture the range of a CUORE pulse in the energies of interest [52]. As mentioned before, these front-end electronics are all located at room temperature outside of the cryostat. From a reliability standpoint this provides a number of advantages. For one, we don't have to worry about effects that appear when electronics are operated at cryogenic temperatures, which are in general not fully understood. Electronics also naturally dissipate some amount of power when in use, which contributes an additional heat load to the cryostat if they're at the cold stages. Lastly, any electronics located at the cryogenic stages of the experiment cannot be physically replaced without warming up the entire cryostat and opening it. Given the time it takes to warm and cool the CUORE cryostat in its entirety, this is a time-prohibitive measure that is avoided unless absolutely necessary, so any cryogenic electronics that encountered problems or turned out to be suboptimal would just have to be left there, whereas room temperature electronics can be replaced, fixed, or adjusted in short order.

There do exist a number of benefits to using cryogenic electronics as well. At a basic level they are advantageous because in a dilution fridge, colder means closer to the sensors that are placed at the coldest stage, and the closer our front-end electronics are to the sensors we're trying to read out the better. These benefits are not just limited to reading sensors on cryogenic calorimeters either; the field of quantum computing also has significant interest in developing classical cryogenic electronics at 4 K or lower temperatures to interface with qubits. The planned upgrade to CUPID provides us with an opportunity to upgrade the electronics of CUORE, including the option to incorporate cold electronics instead of continuing to use an entirely room-temperature setup.

This chapter will present the work we have done towards developing ASICs (Application-Specific Integrated Circuits) using 180-nm CMOS (Complementary Metal-Oxide Semiconductor) technology that could operate in the 0.01 K to 4 K regime in CUPID. I will discuss the idea behind using cryogenic CMOS and what the general advantages could be, as well as how this would fit into the demands of CUPID. I will also discuss characterization and modeling efforts we have done for the behavior of these devices all the way down to sub-100 mK temperatures, for which there previously existed little to no data. Lastly, I will discuss progress towards being able to operate a CMOS-based amplifier at 100 mK, which will serve as our first step towards demonstrating feasibility for CUPID.

10.1 Advantages of Cold Electronics

The possibility of classical CMOS-based electronics operated at cryogenic temperatures down to 4 K has recently acquired significant interest from the quantum computing community [99, 100], which similarly deals with the problem of interfacing with objects placed at the coldest stage of dilution fridges. In their case, qubits are operated at these cold temperatures in order to prevent thermal fluctuations from breaking their quantum coherence, and the question is how to both control and read out these qubits in a fashion satisfying the requirements of an actual quantum algorithm. Dilution fridges are naturally designed in a way that puts the coldest stage deepest in the cryostat, with any connections into it getting thermalized by first passing through the warmer stages in succession. Any wires that reach from outside the cryostat into the coldest stage thus have a fairly long minimum length in order to be able to traverse this distance. The resulting increased wire capacitance can introduce latency undesirable for advanced operations [101] and even pick up vibrational noise that can cause quantum decoherence [102]. As quantum computers scale up the number of qubits they operate, the wiring required to access each of them from room temperature also becomes prohibitive from the perspective of both the associated thermal load on the fridge and the physical complexity.

Conventional CMOS-based electronics operated at the cold stages of the fridge offer a solution to all of these problems. By being operated at cold temperatures, they can be placed closer to the qubits and thus offer more precise control and higher fidelity readouts. This also allows the possibility of multiplexing the signals at a stage close to the sensors, with one CMOS-based ASIC reading and controlling multiple qubits at the same time, so that the amount of wiring that must travel down to the cold stages from outside is reduced. The use of conventional CMOS processes allows for easy scaling by taking advantage of the maturity of the semiconductor industry, as long as cryogenic operation requires no special modifications to the devices. The limiting factors are properly modeling cryogenic CMOS behavior so that the cryogenic circuits behave as desired, as well as limiting the heat dissipation imposed by the cold electronics so that they do not exceed the cryostat's cooling power. These factors have resulted in cryogenic CMOS-based ASICs typically being placed at the 4 K stage of dilution fridges, which serves as a healthy compromise. At this stage, they're close enough to the qubits in the coldest stage of the cryostat to obtain the desired benefits, but also at a warm enough stage that the cryostat has plenty of cooling power to sustain the continuous operation of these electronics.

Application to CUPID

The advantages of cryogenic front-end electronics for quantum computing turn out to all apply to the CUORE/CUPID setup as well. We don't have to worry about quantum decoherence, but sending an NTD signal all the way from 10 mK to room temperature before preamplification does allow for additional noise pickup. It is possible that performing preamplification closer to the actual detectors could help CUPID reach its energy resolution goal of a 5 keV FWHM, which CUORE has not yet achieved. The lower cable capacitance from having the NTD closer to the preamplifier would also improve our pulse shape fidelity, which could in turn improve our pileup rejection capabilities. In addition, CUORE currently has to send individual wires from room temperature down to the mixing chamber stage for each of the heaters and NTD thermistors on each of the 988 TeO_2 crystals in order to perform biasing and readout. The thermal load and complexity of this wiring scheme is acceptable under CUORE's present operating conditions, but could become problematic in CUPID, where the total number of crystals will likely increase and where the addition of light detectors would further double the number of wires. Since calorimetric signals are slow, they don't need to be digitized at a rate much higher than a few kHz. It is thus quite achievable to time-domain multiplex the detector signals with cryogenic electronics; even a modest multiplexing factor of 4 or 8 would reduce the required number of wires to a very manageable number.

A schematic of how cryogenic CMOS-based ASICs could be incorporated into CUPID is shown in Fig. 10.1. The front-end preamplifiers should ideally be as close to the sensors as possible, so we would aim to place those somewhere beneath the still; for instance, if not in the mixing chamber itself, then they could still be at 50 mK on the heat exchanger plate in the CUORE cryostat (see for reference Fig. 3.4). This would largely be determined by the power budget required by the amplifiers, as even with the $> 3\ \mu$W of cooling power that the CUORE cryostat provides at its coldest stage [55], this would restrict the power consumption of each amplifier to O(1 nW) if we were to include O(1000) of them. Additional electronics for multiplexing, control, and digitization, which do not need to be as close to the detectors to serve their purposes, can be placed slightly higher at the still or 4 K stage. At 1 K there are many mW of cooling power available, and at 4 K it goes up to several watts of cooling power, so these electronics would have substantially more freedom in their design with the less restrictive power constraints. From there, they provide the ability to drive the signal up the remaining one to two meters of cable to reach the room temperature electronics that will actually store the data.

We have opted to use 180-nm CMOS technology for the development of cryogenic electronics for CUPID. This is a much cheaper option compared to newer processes, which will be of benefit when we eventually have to scale to thousands of devices to fully instrument CUPID's thousands of sensors. However, while there has been work characterizing CMOS devices down to 4 Kelvin temperatures for quantum computing applications, very little work going down to the sub-Kelvin level exists. The challenge for CUPID is thus to understand the behavior of CMOS devices at the sub-100 mK temperatures we may be operating them at, and then to design circuits which will satisfy the power budget of CUPID while also being

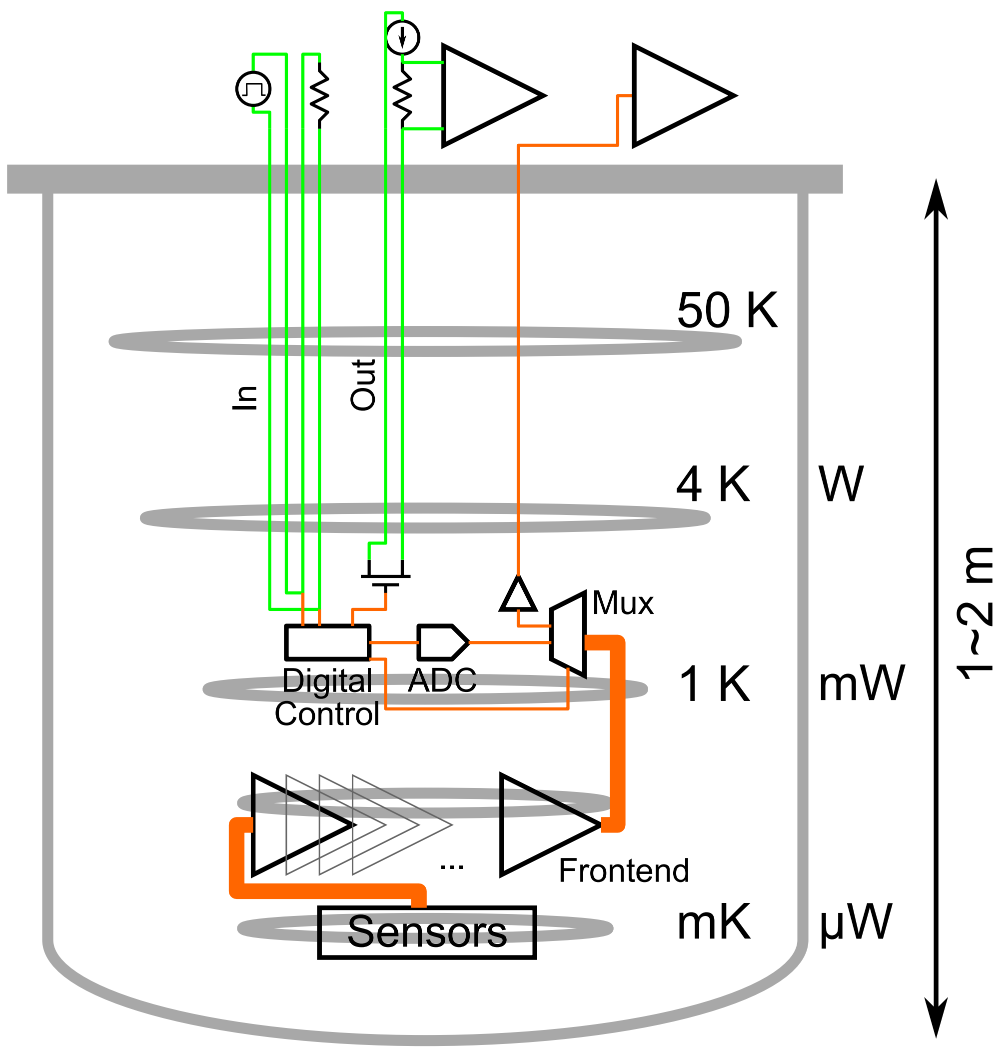


Figure 10.1: Schematic of how cryogenic electronics could be integrated into CUPID. Front-end preamplifiers can be located at one of the innermost stages below the still, in order to minimize their distance to the sensors. Additional supporting electronics can be placed slightly higher at the 1 K stage, where more cooling power is available, to send the signal the rest of the way up to room temperature.

able to deliver the signals that we want.

10.2 180-nm CMOS Characterization

As a brief introduction to the basic physics of a MOSFET (Metal Oxide Semiconductor Field Effect Transistor), the general structure of a n-channel MOSFET (NMOS) is shown in Fig. 10.2. The source and drain terminals are at regions that are locally highly n-doped, which sit in a p-doped substrate called the body that is often tied to the same voltage as the source terminal. Under the normal "off" conditions, even when a positive voltage V_{ds} is applied from the source to drain, (almost) no current is able to flow because of the depletion zones formed between the p-substrate and the n-doped source/drain regions. The gate is kept insulated from the rest of the device by means of a metal oxide layer, but when a voltage is applied to the gate it still has an effect on the body of the transistor through the resulting electric field. When the voltage V_{gs} between the source/body and the gate becomes high enough and exceeds what is called the threshold voltage V_{th}, an inversion layer forms beneath the metal oxide layer, pushing away p-type carriers that were already there in the substrate and pulling in n-type carriers from the source and drain. This inversion layer, populated by n-type carriers like the source and drain terminals, acts as the channel for conducting current between them. In the so-called linear regime where $V_{gs} > V_{th}$ and $V_{ds} < V_{gs} - V_{th}$, the source-drain current I_{ds} of the MOSFET exhibits an ohmic response to V_{ds} just like a normal resistor.

Once the drain voltage V_{ds} reaches the saturation voltage $V_{sat} \approx V_g - V_{th}$, the channel region becomes pinched off, as the gate no longer has sufficient potential to pull n-type carriers away from the drain terminal to fully form the channel[1]. Once V_{ds} is increased beyond this saturation voltage, the device enters the saturation regime, where the drain current I_{ds} stops increasing and stays at the value it attained at $V_{ds} = V_{sat}$. The pinch-off region continues to increase in size as V_{ds} increases further, but the current flowing through the inversion layer is still able to pass through the pinchoff region and reach the drain. This gives rise to the characteristic shape of a MOSFET's $I_{ds} - V_{ds}$ curves. A p-channel MOSFET (PMOS) operates on basically the same principles, but the body is n-doped and the source/drain terminals are p-doped instead. NMOS and PMOS manufactured using the same process can have different response characteristics depending on the width and length of the conduction channel that would connect the source and drain when the device is turned on, depending on the doping levels in the body of the device, and depending on the thickness of the metal oxide layer beneath the gate.

As the temperature at which we try to operate NMOS and PMOS goes down, we can expect important characteristics such as the threshold voltages, the saturation currents, and the transconductances $g_m = (\partial I_{ds}/\partial V_{gs})|_{V_{ds}}$ to all change in meaningful fashions. These effects must all be accounted for if we wish to use these devices in integrated circuits at the

[1]The saturation voltage tends not to be exactly $V_g - V_{th}$, and models account for this with tunable prefactors that come from factors like body effects.

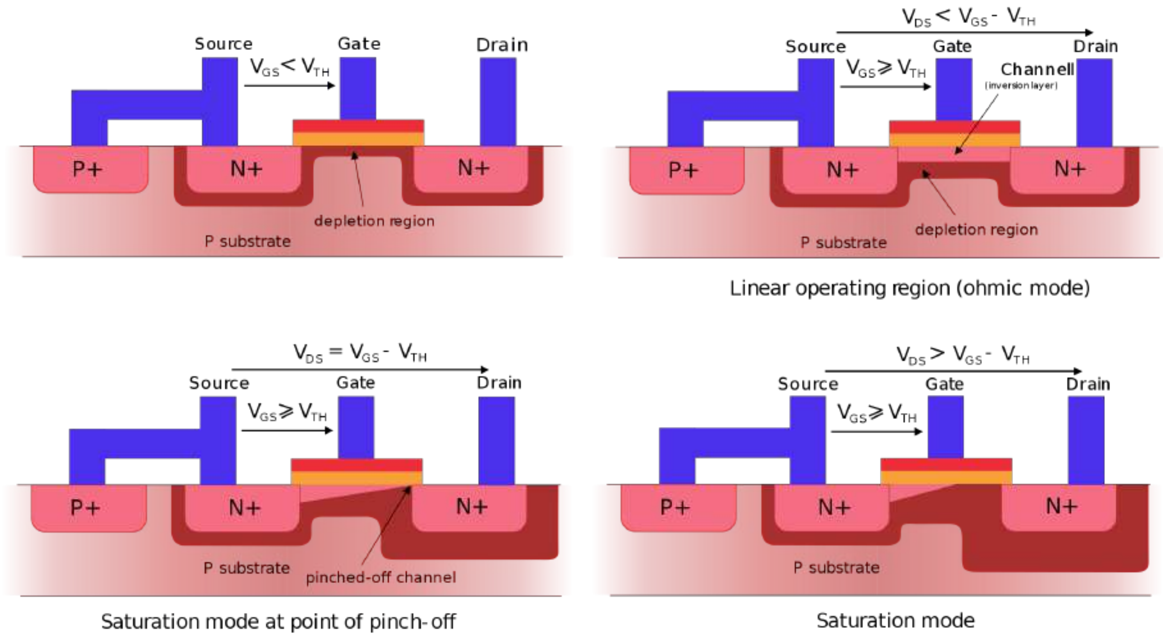


Figure 10.2: Basic structure of a n-channel MOSFET in enhancement mode, which is the typical operation mode for most applications. In its off mode (**top left**), a depletion region between the p-substrate and the n-doped source/drain regions prevents the flow of current. When a sufficiently large voltage is applied beteween the gate and the body, an inversion layer forms that allows the conduction of current between the source and drain (**top right**). Once the source-drain voltage becomes too large, the channel starts becoming pinched off and the source-drain current saturates (**bottom**). Image from [103].

cryogenic stages of a dilution fridge, ranging in temperature from 10 mK to 4 K. Widely used models such as BSIM3 use large numbers of parameters to predict the behavior of both NMOS and PMOS with different channel sizes and under different bias conditions [104]. These models have some ability to account for temperature effects, but they are generally only intended for use in the 200 to 400 K range [105], which is far above the temperature regime we are considering. In particular, the silicon substrate undergoes carrier freeze-out once the device drops below about 40 K, at which point there are no thermal carriers and any current flow relies on the carriers freed by the electric field effect. There has recently been more work developing ways to extend these models to liquid nitrogen (77 K) and even liquid helium (4 K) temperatures [106], relying on systematic approaches to tune existing model parameters to match observed behavior, but these still require a large set of experimental data to calibrate the model. We thus begin by characterizing a wide array of NMOS and PMOS at the temperatures at which we hope to operate them.

We performed tests with chips featuring standard NMOS and PMOS manufactured using the TSMC 180-nm process. These include regular-V_{th}, medium-V_{th}, and native-V_{th} devices, as well as arrays with thin and thick oxide layers. The channels have widths and lengths ranging from 0.18 μm to 10 μm. We wirebond subsets of the available devices on the chip at a time and sweep a range of V_{ds} and V_{gs} values to characterize their IV responses. We conduct tests with the chip at room temperature ($\sim$300 K), mounted on a jig that is dipped into a liquid nitrogen bath (77 K), and at the mixing chamber stage of a dilution fridge (down to 10 mK). This allows us to track the temperature dependence of the device characteristics at some key temperatures. The mixing chamber can be operated stably at a range of temperatures below 1 K (where the ^{3}He in the mixture condenses), and it can also be held around 15 K (at the limits of the fridge's precooling system) for long enough to conduct measurements. However, any other temperatures are generally only attained in passing during a cooldown or warmup, and so we cannot easily scan a full range of temperatures. While it would be nice to build a full model of the temperature dependence of MOSFET behavior, this is ultimately unnecessary for our applications, where the circuits will be operated at designated temperatures rather than over ranges of temperatures.

An example of the IV characterization results for W/L = 0.5μm/0.5μm regular-V_{th} thin-oxide NMOS and PMOS is shown in Fig. 10.3, with measurements done at 300 K, 77 K, and 0.1 K. The first and most important qualitative observation is that the devices are all fully functional even at the coldest temperatures that we can reach with a dilution fridge. After that, we can identify some key differences between the cold measurements and the room temperature measurements. Looking at the $I_{ds} - V_{gs}$ curves, we can see that the current response I_{ds} "turns on" at higher gate voltages V_{gs}, corresponding to a higher threshold voltage at low temperatures. This is an expected result of the increase in Fermi potential as temperature decreases, causing a need for stronger electric fields in order to form the inversion layer. More sophisticated models for the temperature-dependence of V_{th} must account for other effects as well, and there are observations that V_{th} saturates below some temperature threshold in NMOS while it continues increasing in PMOS, likely as a result of stronger freeze-out effects in hole carriers [107, 108]. Our observations are consistent with this, showing that the increase in V_{th} at cold temperatures is larger for PMOS than it is for the same sized NMOS.

We also see that while V_{th} is higher at cold temperatures, once the MOSFET is turned on the current response I_{ds} tends to be higher. This is another expected effect – lower temperatures increase carrier mobility and suppress lattice vibrations, decreasing the resistance in the channel as long as a conduction path actually exists. Relatedly, we see an increase in the transconductance g_M at lower temperatures too, clearly visible in the slopes of the $I_{ds} - V_{gs}$ curves. Using a very simple square-law model for MOSFET response, we could expect the drain current to have the behavior:

$$I_{ds} = 2k[(V_{gs} - V_{th})V_{ds} - 0.5V_{ds}^2] \text{ (linear region)}$$

$$I_{ds} = k(V_{gs} - V_{th})^2 \text{ (saturation region)}$$

Here, $k \propto \frac{W}{L}T^{-3/2}$ is a prefactor that accounts for the channel width W and length L, as well as the temperature-dependence of the carrier mobility effects. This would suggest

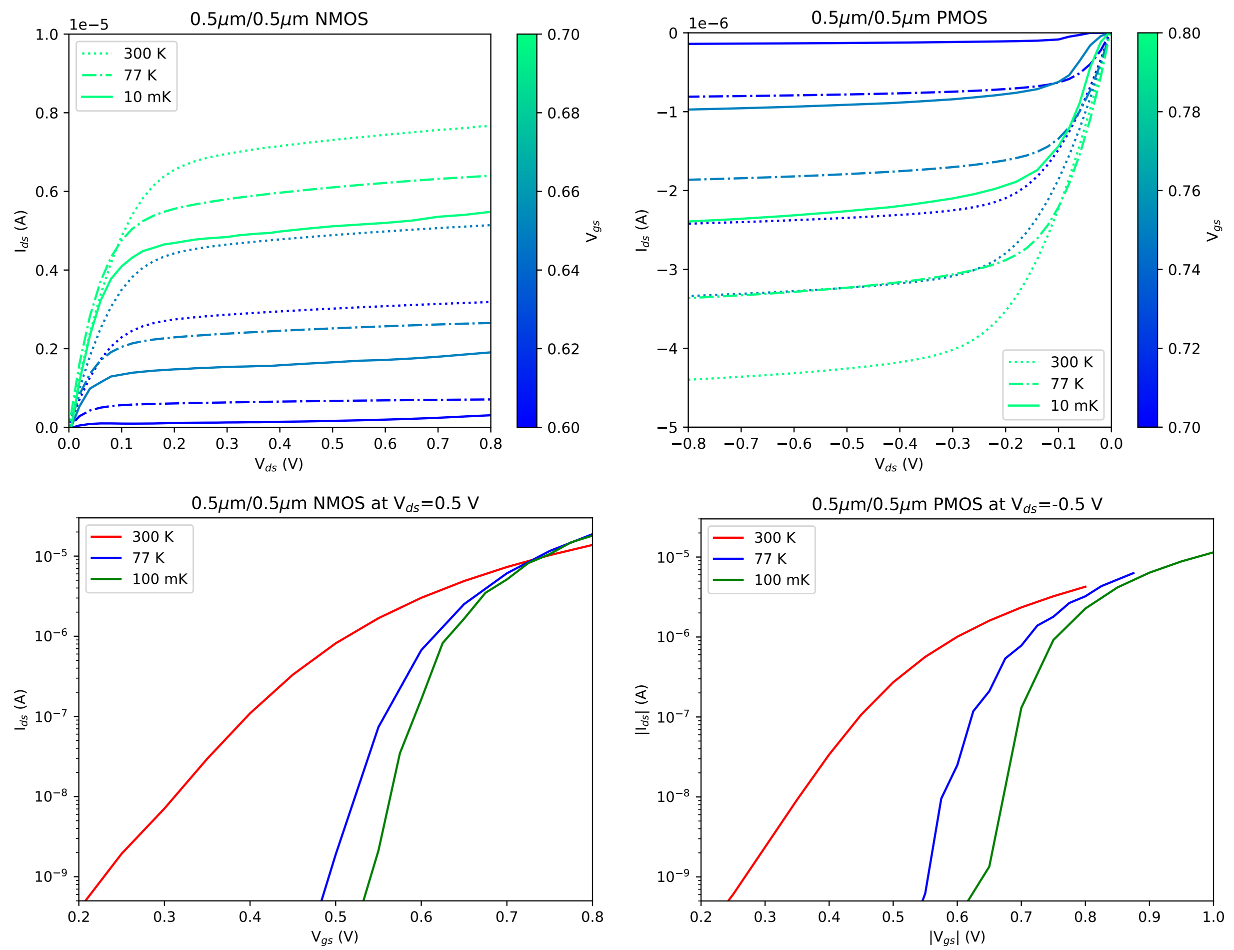


Figure 10.3: **Top**: I_{ds} vs V_{ds} curves for single 0.5μm/0.5μm regular-V_{th} NMOS and PMOS for a range of V_{gs} values, taken at temperatures of 300 K, 77 K, and 0.1 K. The change in response with temperature is clear, but it is also clear that it is not a simple extrapolation to go from 300 to 77 to 0.1 K. **Bottom**: I_{ds} vs V_{gs} curves for the same devices, collected in the saturation regime with $V_{ds} = 0.5$ V. The increase in threshold voltage and transconductance at lower temperatures is clearly visible here.

that g_M is also proportional to $T^{-3/2}$, but we see that this obviously does not hold for low enough temperatures, as a result of more complicated factors coming into play. This is shown for selected devices in Fig. 10.4, where we collected data at a number of temperatures during the warmup of our dilution fridge. Nonetheless, the increased transconductance at low temperatures is promising for the performance of cryogenic amplifiers, where g_M is important to their gain.

It is also interesting to note the behavior of native MOSFETs at cryogenic temperatures.

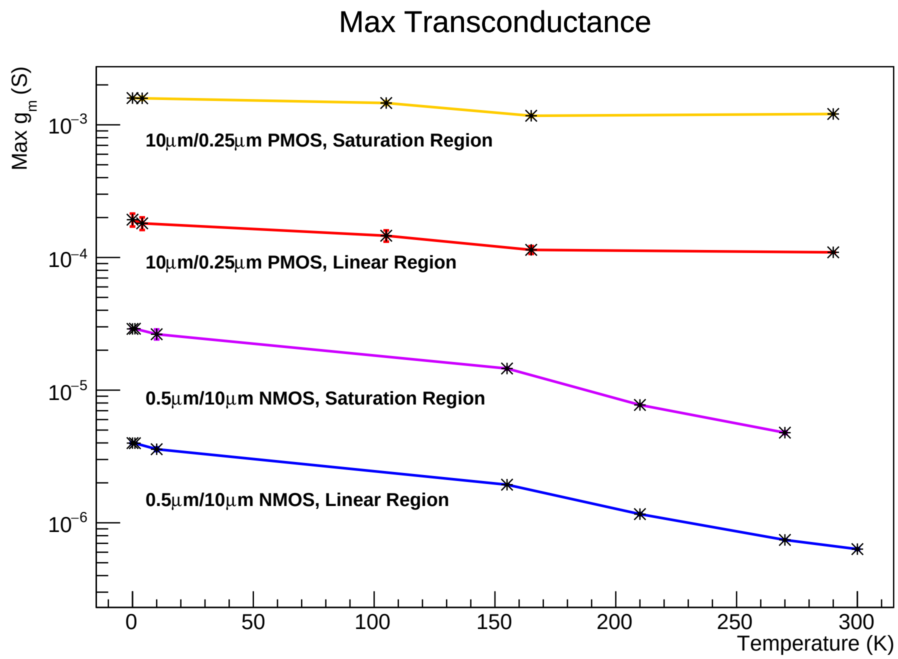


Figure 10.4: Maximum transconductance g_M of selected NMOS and PMOS at different temperatures, in both the linear operating regime ($V_{ds} = 0.04$ V) and saturation regime ($V_{ds} = 0.5$ V). Max g_M is calculated by looking for the value of V_{gs} that maximizes $\partial I_{ds}/\partial V_{gs}$ at the given V_{ds} value. There is a general trend of increased g_M at lower temperatures, but the exact form of the trend is not obvious. Reprinted from [109].

Native NMOS, so named because they are often built on the "native" lightly p-doped body instead of the much more heavily doped p-wells that normal NMOS are typically built on, have threshold voltages near 0, since much weaker fields are needed to form the inversion layer. This can be useful for the low power circuits needed for cryogenic applications, and can also be useful for circumventing the problems associated with high threshold voltages in cryogenic circuits. Results for one native NMOS are shown in Fig. 10.5, where we can see that at room temperature it is "on" even for $V_{gs} = 0$ V, but at 100 mK it acquires a small positive threshold voltage. Carrier freeze-out causes the need for some nonzero electric field from the gate to pull enough carriers to form the conduction channel.

Hysteresis Effects

At cold temperatures, we observe a notable hysteresis effect in the NMOS and PMOS with the largest channel widths and lengths. This is shown in Fig. 10.6, where the $I_{ds} - V_{ds}$ curve is not monotonic when we scan from small V_{ds} to large V_{ds}, but it takes the expected shape when we scan from large V_{ds} to small V_{ds}. This effect starts to appear below about 40 K, but

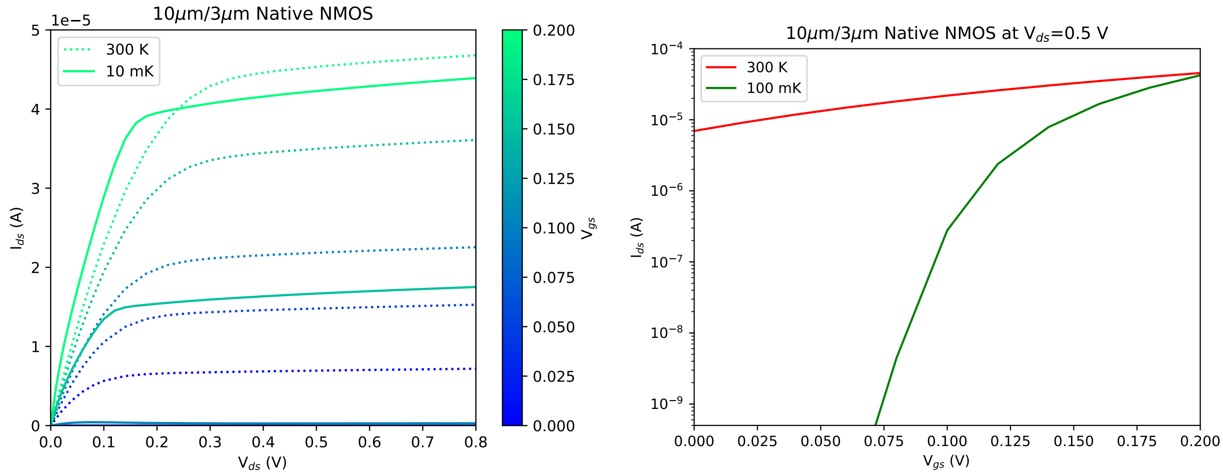


Figure 10.5: **Left**: $I_{ds} - V_{ds}$ curves for a W/L=10μm/3μm native-V_{th} NMOS, at both 300 K and 100 mK. The separate curves correspond to a scan on V_{gs} from 0 to 0.2 V in 0.05 V intervals. **Right**: $I_{ds} - V_{gs}$ curves for the same device in the saturation regime, with $V_{ds} = 0.5$ V. We can see here that at 300 K the device has a threshold voltage < 0, but it picks up a positive threshold voltage at 100 mK.

the specific point varies by device. The MOSFETs with smaller channel sizes don't exhibit this effect at all, and it is generally stronger in NMOS than in PMOS. The effect is also time-dependent, with the kink in the IV curve becoming less wide if the scan is conducted very slowly.

This same effect has in fact been seen previously at liquid helium temperatures, where it was explained by the time constants associated with depletion layer formation in the body, driven by avalanche-generated substrate currents [110]. In this model, we note that we cannot ignore the time it takes for the MOSFET to reach its equilibrium state when a gate voltage V_{gs} is applied. With most charge carriers frozen out at the cryogenic temperatures under consideration, the electric field imposed by the gate voltage will gradually rearrange the carrier distribution in the substrate. This manifests as a substrate current that flows until the equilibrium state is reached, with the appropriately sized inversion and depletion layers in the body. The time it takes for this process to occur is calculated to be proportional to $\exp(-V_1/(V_{ds} - V_{sat}))$ for some constant V_1. This is a result of impact ionization near the pinchoff region of the transistor. For $V_{ds} < V_{sat}$, there is no pinchoff yet, and the drain current is mostly confined to the inversion layer with minimal ability to impact the substrate current. Once V_{ds} becomes large enough, impact ionization in the pinchoff region becomes possible, freeing electron-hole pairs in the substrate. In the case of a NMOS, the newly freed electrons continue on into the drain, and the newly freed holes are pushed by the gate voltage down into the substrate, aided by the increased conductivity of the low-temperature

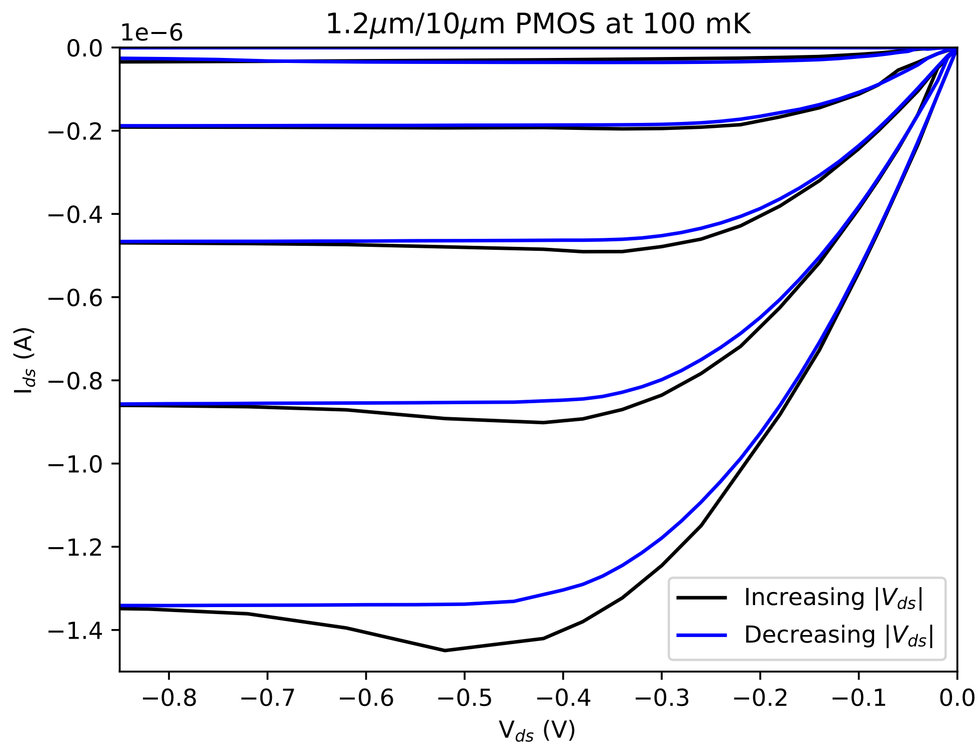


Figure 10.6: $I_{ds} - V_{ds}$ curves for a W/L=1.2μm/10μm PMOS, with the different curves corresponding to a scan on V_{gs} from 1.0 V to 1.6 V in 0.1 V steps. When we scan from small V_{ds} to high V_{ds}, a distinctive kink appears in the curves that is not seen in normal MOSFET operation. When scanning from high V_{ds} to low V_{ds}, this kink is not present, and the IV curves take on their expected shape.

material. This results in the forced formation of a new depletion region, with the drain voltage assisting by ionizing charge carriers and allowing a more rapid change of state.

Looking back at Fig. 10.6 again, this model now explains our observations. When we scan from low V_{ds} to high V_{ds}, the substrate current is low at first and the rearrangement of charge carriers in the body is slow. Once V_{ds} becomes high enough, impact ionization suddenly frees many carriers from the pinchoff region, resulting in an avalanche substrate current. This causes a surge in the drain current as electrons/holes from the substrate are swept in, while the holes/electrons are pushed deeper into the substrate, depending on whether it is a NMOS/PMOS. This process ends once the new depletion region has been formed and there are no more available carriers, at which point the drain current settles back to its saturation value, yielding the kink shape in the IV curves. On the other hand, when we scan from high V_{ds} to low V_{ds}, we start in the regime where the depletion region forms quickly. As a result, the IV curve smoothly decreases as we decrease V_{ds}. The depletion region

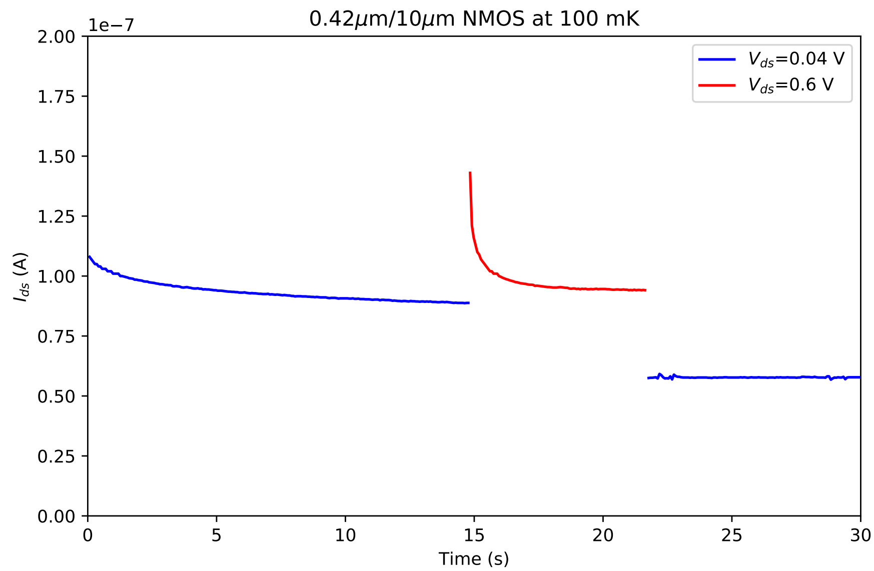


Figure 10.7: I_{ds} response of a W/L=0.42μm/10μm NMOS at 100 mK operated with V_{gs} = 0.7 V, with the gate voltage turned on from 0 at $t = 0$ seconds. The color of the response indicates whether V_{ds} was set to 0.04 V in the linear regime or at 0.6 V in the saturation regime. We see that there are two distinct conduction levels for the $V_{ds} = 0.04$ V, depending on whether the device has experienced a high V_{ds} yet. Once the NMOS enters the "low" conduction regime, it remains there until V_{gs} is toggled off and on.

formation becomes slow once we reach low V_{ds} values, but this does not have a noticeable effect on the shape on the IV curve, with no kind of surge in current like what we see when scanning in the opposite direction. This also explains why we only notice the effect in devices with large widths and lengths, as large channels offer a larger volume of the substrate that can be ionized by the drain voltage, leading to a larger surge in the drain current.

A separate memory effect can be seen in certain devices below about 4 K where there are transient surges in current when we adjust V_{ds}, and where the channel conductance for low V_{ds} appears to depend on whether the device has previously been operated with $V_{ds} > V_{sat}$, shown in Fig. 10.7. When the gate voltage is first turned on from 0 to a value $V_{gs} > V_{th}$, the drain current response to a small V_{ds} is unusually high. Increasing V_{ds} to a large value causes a short surge in I_{ds} before the device settles back to a stable value. After that, setting V_{ds} to a small value again will show a smaller current response than before. This state persists until V_{gs} is toggled off and on, at which point the I_{ds} response to a small V_{ds} will once again be high. This effect has been seen in both NMOS and PMOS.

This can be summarized by classifying the device as being in a "low" or "high" conduction state for small V_{ds} biases, corresponding to the smaller and larger I_{ds} responses. When V_{gs}

is first turned on, the device starts in the high-conduction state. Once V_{ds} is increased to a large value $> V_{sat}$ for the first time, the device switches to the low-conduction state and remains there until V_{gs} is turned off again. The transient spikes in I_{ds} are reminiscent of the other hysteresis effect described earlier, but it is not obvious how this specific phenomenon could be explained by the same model, and this effect occurs in a much wider range of devices compared to the previous effect. A similar kind of two-state behavior has previously been seen in 180-nm CMOS at 3 K [111], where it was explained as the result of impact ionization causing a voltage buildup in the body, which effectively modifies the threshold voltage. This seems consistent with our observations, where our "high" conduction state corresponds to a lower effective V_{th}, and the larger V_{ds} discharges the body to cause the return to the "normal" threshold voltage and "low" conduction regime. Investigation of this effect is still ongoing to see if our observations are fully caused by the same underlying mechanisms.

Circuit designs generally do not expect these sort of transient effects in the MOSFETs that they use, although it is not obvious how exactly these effects would impact the resulting circuit behavior. The first described effect occurs in a limited set of devices, but the second effect is seen in a wider range of NMOS and PMOS. Further testing will be required to see how these hysteresis effects affect the behavior of cryogenic CMOS-based circuits and how we can avoid, manage, or take advantage of them.

Simulation

We can attempt some simple modifications of typical out-of-the-box models in order to simulate some of the basics of cryogenic CMOS behavior, setting aside the transient and history-dependent behaviors for now. Using an array of different sized NMOS as an example, we confirm that the BSIM3 model accurately replicates their IV responses at room temperature and properly accounts for the dependence on the channel width and length. We then try tuning the model to replicate the behavior of NMOS at 100 mK. The BSIM3 model has quite a lot of parameters that allow it to describe subtleties in NMOS and PMOS behavior [104], but we begin by just tuning a few important ones.

To account for the change in threshold voltage, we tune V_{TH0} and K_1. V_{TH0} is the threshold voltage when the body and source terminals are tied together for devices with large channel lengths, which we can tune as a basic measure of how the threshold voltage is different at cryogenic temperatures. K_1 is the first-order body effect coefficient, which accounts for how substrate currents in the body modify the threshold effects. We include this in our parameter tuning since we know that changes in substrate currents at < 4 K are definitely not negligible, as evidenced by the observed hysteresis effects. To account for the change in material conductivity at cryogenic temperatures, we tune U_A and U_B, which are respectively the first and second order mobility degradation coefficients that modify the carrier mobility values. In general terms, V_{TH0} lets us modify the threshold voltage, K_1 lets us modify the shape of the transition from the linear regime to saturation regime, and U_A, U_B let us modify the magnitude of the current responses.

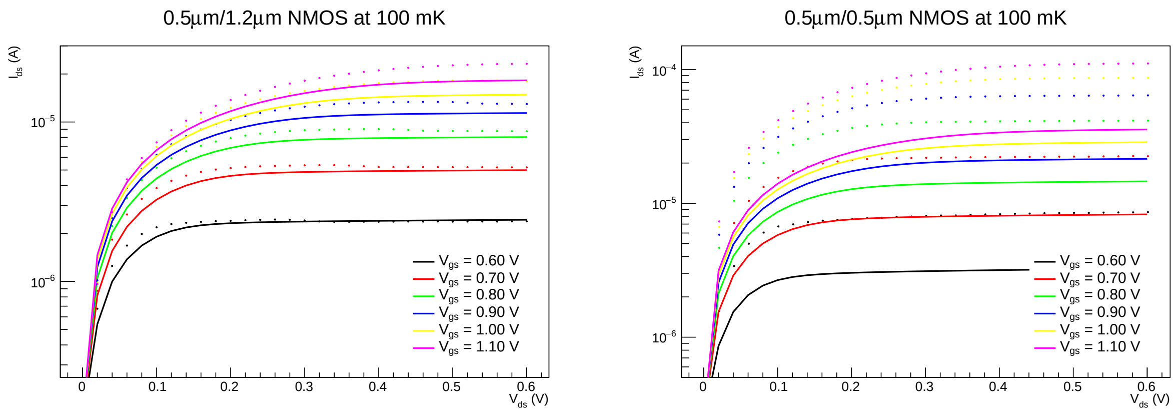


Figure 10.8: **Left**: a W/L=0.5μm/1.2μm NMOS to which we tune a BSIM3 model using the procedure described in the text. The measured responses at 100 mK are plotted as points, and the solid curves are the output of the tuned BSIM3 model. Although the agreement is not perfect, we see that we are able to model the overall trends, and we are within a factor of 2 of the actual values. **Right**: a NMOS of the same type, but with a different channel size of W/L=0.5μm/0.5μm. The measured responses at 100 mK are again plotted as points, and the solid curves are the predictions of the BSIM3 model using the parameters obtained from tuning to the device on the left. We see that the model does a fairly poor job of extrapolating to other device sizes after being tuned on one size. Reprinted from [109].

It should be noted that we do not expect to fully replicate cryogenic CMOS behavior by only tuning these 4 parameters, but this offers a starting point. An example of the results that we can achieve is shown in Fig. 10.8, where we tune the parameters using measurement data from a W/L=0.5μm/1.2μm NMOS at 100 mK. We then apply the obtained parameters to simulate the behavior of a W/L=0.5μm/0.5μm NMOS with the same doping levels and compare against the measurements from that device. We find that this simple tuning procedure allows us to replicate the behavior of a single device fairly well, but it does not at all properly predict how the response should change for other channel sizes. A full model would have to tune other parameters that govern channel size effects in the BSIM3 model as well.

In principle, we should be able to build a full model by taking into account our full array of NMOS and PMOS measurements at 100 mK with many different channel widths and lengths and tuning a wider set of parameters, and we do intend to do this in future work. However, this will be done with the aid of more sophisticated software, as opposed to the simple fitting procedures done here that can handle only a small number of parameters. For now, being able to replicate the behavior of just one device in simulation is sufficient to build more complicated circuits, as long as we only use NMOS and PMOS for which we have actually conducted IV characterization measurements under cryogenic conditions.

10.3 Cryogenic CMOS-based Circuits

After characterization measurements of single NMOS and PMOS, the next step is to test iterations of circuits that use the tested devices to perform the functions that we want for our cryogenic electronics. This includes switches that we will use for time-domain multiplexing, line drivers that we will use to drive the cold signals out of the fridge, and preamplifiers that will be placed next to NTDs or TESs to amplify their signals. We start with characterization of single circuits at a time, using manually injected signals to characterize their behavior. After that, we will plan to put them together to read out actual thermal signals from an NTD at the cold stage of a dilution fridge, and eventually we will plan to demonstrate multiplexed readout of multiple NTDs at once. For now, we are on the first step and are iterating on circuit designs based on the observed behavior. This section presents some of these preliminary results.

Our amplifiers and line drivers need to be able to support the bandwidth we need for signals from our cryogenic sensors. For NTDs, the signals are quite slow, so we actually only need a bandwidth of up to 10 kHz or so, while TES signals are faster and could take advantage of higher bandwidths. The frequency response of one of our line drivers is shown in the Bode plot in Fig. 10.9, where we can see it can sustain a signal up to over 1 MHz, with comparable performance at 77 K and at room temperature. We also observe a power consumption of O(100 μW) at temperatures < 1 K, which is far too high for the mixing chamber but is perfectly manageable at the still stage or higher in the dilution fridge, where we intend to place it.

An example of one of our NTD amplifiers and its response at 100 mK is shown in Fig. 10.10, using a manually injected input voltage for characterization instead of an actual NTD signal. The amplifier is biased so that it sits in the middle of the sharp transition, and simulation of its expected cryogenic behavior shows that a very small change in voltage coming from the NTD can be amplified by a factor of up to 200. At 100 mK, our tests of the amplifier response show that a gain of that magnitude seems possible with a power consumption of O(100 nW), which is manageable at the mixing chamber stage of the fridge. However, the response under cryogenic conditions unexpectedly loses its sharpness and tapers off about halfway between the maximum and minimum V_{out} values, which limits its dynamic range. This effect does not exist in the room temperature behavior, so it is some sort of unexplained cryogenic effect. Even more concerningly, it exhibits the unusual feature that the response at 100 mK is different when we scan V_{in} going up versus going down.

Looking at the amplifier's circuit diagram, this behavior seems to be the result of the second described hysteresis effect from the previous section. With our amplifier design, scanning V_{in} up and down is equivalent to approaching V_{ds} from high or low values for a specific PMOS in the circuit, given the constant bias current. This corresponds to operating that PMOS in either the low-conduction or high-conduction regime, depending on the direction from which we are approaching V_{in}, which causes the transition point of the amplifier response to differ. This hysteresis in the amplifier is not necessarily prohibitive to its use in our cryogenic system; since thermal signals from NTDs are always in the same direction,

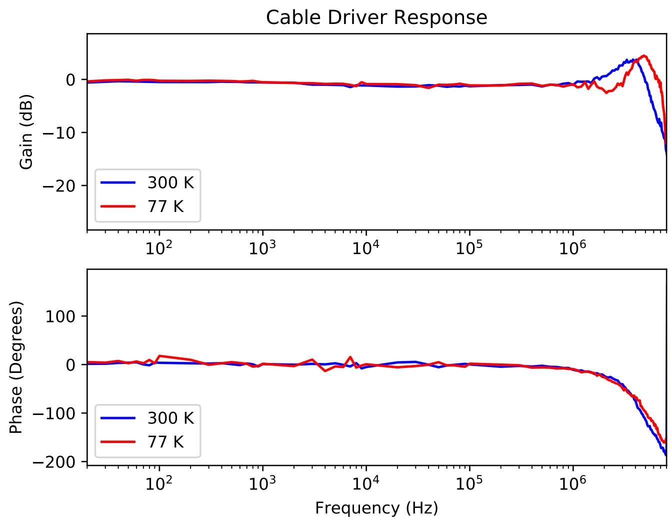


Figure 10.9: Bode plot for a line driver using Miller compensation, operated at room temperature and at 77 K. We see that it is able to drive signals up to over 1 MHz at the expected unity gain.

we can bias it so that it always operates in the same regime. Furthermore, since we have an understanding of its source, the effect could possibly be eliminated with modifications to the amplifier design. The effect could even be useful if exploited somehow, but regardless of which approach we take, it will have to be properly accounted for.

Further tests with these cryogenic CMOS-based circuits are still underway, but the results so far seem promising, with the desired gains, bandwidths, and power consumption limits all looking achievable, although we are contending with hysteresis effects in cryogenic conditions. With CUPID still several years away, 180-nm CMOS remain a viable option for cryogenic front-end electronics in its final design, and our developments here hold promise for a wide range of other applications that would benefit from cold front-end electronics.

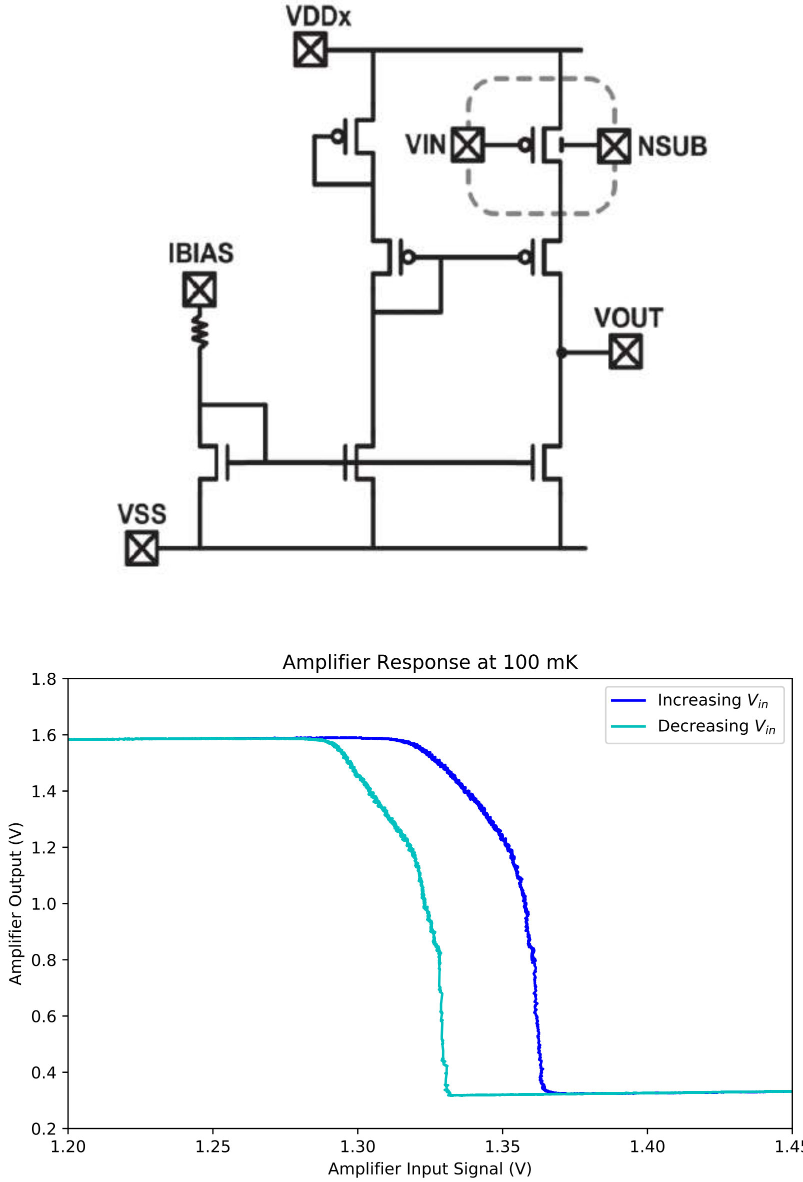


Figure 10.10: **Top**: circuit diagram of a cryogenic amplifier designed for NTD signals, operated with a 25 nA bias current and with V_{dd} set to 1.8 V above V_{ss}, which is the maximum operating voltage for these MOSFETs. **Bottom**: output voltage response for this amplifier operated at 100 mK. We slowly sweep the input voltage and estimate an achievable gain of > 100, close to the expected gain of 200 obtained from simulation. However, we see that the response is different depending on whether we scan V_{in} going up or going down. We also see that the response loses its sharpness about halfway from the minimum V_{out} to the maximum V_{out}, which does not occur in room-temperature behavior.

Chapter 11

Conclusions

Since the neutrino was first postulated almost a century ago to explain an anomaly in observations of the humble beta decay, we have learned much about its properties and come to understand its significance to a wide range of phenomena, but there is still much we do not know about it. The search for neutrinoless double beta decay is one of the centerpieces of modern efforts to look for new physics beyond the Standard Model, providing one of the most promising avenues to investigate the fundamental nature of the neutrino. The scope of $0\nu\beta\beta$ experiments has reached the point where fairly large collaborations are necessary for each experiment, and this trend will only continue as we move to the next generation of tonne-scale experiments. This dissertation has presented my personal contributions to one set of experiments that are part of this global effort: analysis developments for CUORE that culminated in its most recent limit on the $0\nu\beta\beta$ decay of ^{130}Te, and analysis, simulation, and hardware developments for a variety of projects that have been building towards the eventual CUPID experiment.

At this point in time, CUORE has demonstrated mastery of its most significant technical challenge, the stable operation of the largest dilution fridge in the world over the course of years. While its analysis procedure is now fairly mature too, there are still additional improvements that can be made. More sophisticated coincidence analyses will allow us to recover additional $0\nu\beta\beta$ sensitivity from the 12% efficiency that we are losing by only considering single-crystal events. Marginal levels of further background reduction will also be possible with delayed coincidence analyses like the simple one employed in CUPID-Mo. Most significantly, CUORE's energy resolution of 7-8 keV FWHM at $Q_{\beta\beta}$ is still above its original design goal of a 5-keV FWHM, which had been successfully attained in CUORE-0 [59]. We still hope to improve the energy resolution through a combination of tuning the operating temperature and NTD bias conditions, tuning the external electronics, and using analysis techniques to decorrelate detector signals from mechanical noise in the cryostat. CUORE is also developing analysis mechanisms to unlock its sensitivity to other exotic physics interactions, which generally involve the study of either low-energy or high-multiplicity events. All of these developments will not only be useful for CUORE, but will be informative for CUPID too.

While CUORE continues to collect data, CUPID is now in its design stages, with isotope enrichment and crystal growth efforts anticipated to begin soon so that the CUPID payload will be ready once CUORE reaches the end of its lifetime. Using a new isotope, new crystals, new detectors, and new electronics, CUPID seeks to improve upon CUORE in every way: more $\beta\beta$ isotope, lower backgrounds, and better energy resolution. One of CUPID's most significant advantages will be its ability to fully reutilize the tried and tested CUORE cryostat, and although there are new challenges associated with the number of detectors and drastic levels of background reduction that CUPID is aiming for, this means that CUPID is realizable even using only the technologies that we have already tested in previous experiments and demonstrators. CUPID is projected to be competitive with other proposed next-generation $0\nu\beta\beta$ experiments, with the ability to probe the entire inverted mass hierarchy region [84]. As the prospect of a $0\nu\beta\beta$ discovery is imaginable with this ensemble of next-generation experiments [112], the isotope flexibility of the cryogenic calorimetric approach used by CUPID could become useful as well. If a discovery were to be claimed, theory immediately provides us with corresponding estimates of the true value of $m_{\beta\beta}$ by assuming that the process was mediated by a light Majorana neutrino in our current 3 light neutrino paradigm. This would yield a range of predictions of the $0\nu\beta\beta$ decay rates for other isotopes, which we would want to check as confirmation. Ge and Xe experiments are not so flexible with their isotope choice, but the CUORE/CUPID cryostat could be outfitted with crystals containing other $0\nu\beta\beta$ isotopes to perform this cross-validation. This capability could become invaluable in the event that we observe inconsistencies between results from the isotopes that we will already be studying.

The developments for CUORE and CUPID described in this dissertation are not limited to these specific experiments. The operation of large-scale cryogenic facilities will be useful for dark matter and other rare event searches, gravitational wave detectors, and quantum computing setups [113], and the deep cryogenic front-end electronics we are developing for CUPID could be similarly useful in these other applications. Analysis techniques for dimensionality reduction and feature extraction are also crucial to many modern experiments, which continue to collect more and more detailed data and larger and larger datasets.

To end with some perspective, even as I note the accomplishments of CUORE and the promising future of CUPID, it is of course true that other $0\nu\beta\beta$ experiments have been comparably successful and have comparably promising future upgrades as well. And even with the diversity of methods that different $0\nu\beta\beta$ experiments are employing, all of them together are still only studying one specific property of the neutrino, while a wide assortment of other experiments are studying the many other unanswered questions about neutrinos. The collective effort to better understand the neutrino is also part of an even greater effort to push our understanding of fundamental physics, involving precision experiments, high-energy colliders, astrophysical particle detectors, and cosmological studies, and this too is only one subset of all the research that we classify as "physics". My work I have discussed in this dissertation is only a piece^n of humanity's pursuit of a better understanding of the natural world, but I eagerly anticipate seeing what comes next.

Bibliography

[1] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[2] P. von Ballmoos. “Antimatter in the Universe: constraints from gamma-ray astronomy”. In: *Hyperfine Interact.* 228 (Feb. 2014), pp. 91–100. DOI: 10.1007/s10751-014-1024-9.

[3] A. D. Sakharov. “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Soviet Physics Uspekhi* 34.5 (May 1991), pp. 392–393. DOI: 10.1070/pu1991v034n05abeh002497.

[4] R. Lehnert. “CPT Symmetry and Its Violation”. In: *Symmetry* 8.11 (2016). ISSN: 2073-8994. DOI: 10.3390/sym8110114.

[5] W. Buchmüller, R. D. Peccei, and T. Yanagida. “Leptogenesis as the Origin of Matter”. In: *Annu. Rev. Nucl. Part. Sci.* 55.1 (2005), pp. 311–355. DOI: 10.1146/annurev.nucl.55.090704.151558.

[6] V. A. Rubakov and M. E. Shaposhnikov. “Electroweak baryon number non-conservation in the early Universe and in high-energy collisions”. In: *Physics-Uspekhi* 39.5 (May 1996), pp. 461–502. DOI: 10.1070/pu1996v039n05abeh000145.

[7] C. L. Cowan et al. “Detection of the Free Neutrino: a Confirmation”. In: *Science* 124.3212 (1956), pp. 103–104. ISSN: 0036-8075. DOI: 10.1126/science.124.3212.103.

[8] L. W. Alvarez. “A Proposed Experimental Test of the Neutrino Theory”. Tech. rep. UCRL-328. Lawrence Berkeley National Lab, Apr. 1949. DOI: 10.2172/929771.

[9] R. Davis, D. S. Harmer, and K. C. Hoffman. “Search for Neutrinos from the Sun”. In: *Phys. Rev. Lett.* 20 (21 May 1968), pp. 1205–1209. DOI: 10.1103/PhysRevLett.20.1205.

[10] Y. Fukuda et al. “Measurements of the Solar Neutrino Flux from Super-Kamiokande’s First 300 Days”. In: *Phys. Rev. Lett.* 81 (6 Aug. 1998), pp. 1158–1162. DOI: 10.1103/PhysRevLett.81.1158.

[11] Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[12] B. Aharmim et al. "Combined analysis of all three phases of solar neutrino data from the Sudbury Neutrino Observatory". In: *Phys. Rev. C* 88 (2 Aug. 2013), p. 025501. DOI: `10.1103/PhysRevC.88.025501`.

[13] K. Eguchi et al. "First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance". In: *Phys. Rev. Lett.* 90 (2 Jan. 2003), p. 021802. DOI: `10.1103/PhysRevLett.90.021802`.

[14] S. Fukasawa, M. Ghosh, and O. Yasuda. "Complementarity between Hyperkamiokande and DUNE in determining neutrino oscillation parameters". In: *Nucl. Phys. B* 918 (2017), pp. 337–357. ISSN: 0550-3213. DOI: `10.1016/j.nuclphysb.2017.02.008`.

[15] M. G. Aartsen et al. "Combined sensitivity to the neutrino mass ordering with JUNO, the IceCube Upgrade, and PINGU". In: *Phys. Rev. D* 101 (3 Feb. 2020), p. 032006. DOI: `10.1103/PhysRevD.101.032006`.

[16] X. Qian and P. Vogel. "Neutrino mass hierarchy". In: *Prog. Part. Nucl. Phys.* 83 (2015), pp. 1–30. ISSN: 0146-6410. DOI: `10.1016/j.ppnp.2015.05.002`.

[17] The ALEPH Collaboration et al. "Precision electroweak measurements on the Z resonance". In: *Phys. Rep.* 427.5 (2006), pp. 257–454. ISSN: 0370-1573. DOI: `10.1016/j.physrep.2005.12.006`.

[18] S. Böser et al. "Status of light sterile neutrino searches". In: *Prog. Part. Nucl. Phys.* 111 (2020), p. 103736. ISSN: 0146-6410. DOI: `10.1016/j.ppnp.2019.103736`.

[19] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: `10.1016/j.physletb.2012.08.020`.

[20] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett. B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: `10.1016/j.physletb.2012.08.021`.

[21] H. Murayama. "Origin of Neutrino mass". In: *Physics World* (May 2002), pp. 35–39. URL: `http://hitoshi.berkeley.edu/neutrino/PhysicsWorld.pdf`.

[22] Y. Cai et al. "Lepton Number Violation: Seesaw Models and Their Collider Tests". In: *Front. Phys.* 6 (2018), p. 40. ISSN: 2296-424X. DOI: `10.3389/fphy.2018.00040`.

[23] V. Brdar et al. "Type I seesaw mechanism as the common origin of neutrino mass, baryon asymmetry, and the electroweak scale". In: *Phys. Rev. D* 100 (7 Oct. 2019), p. 075029. DOI: `10.1103/PhysRevD.100.075029`.

[24] O. G. Miranda and J. W. F. Valle. "Neutrino oscillations and the seesaw origin of neutrino mass". In: *Nucl. Phys. B* 908 (2016). Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015, pp. 436–455. ISSN: 0550-3213. DOI: `10.1016/j.nuclphysb.2016.03.027`.

[25] M. Goeppert-Mayer. "Double Beta-Disintegration". In: *Phys. Rev.* 48 (6 Sept. 1935), pp. 512–516. DOI: `10.1103/PhysRev.48.512`.

[26] S. R. Elliott, A. A. Hahn, and M. K. Moe. "Direct evidence for two-neutrino double-beta decay in ^{82}Se". In: *Phys. Rev. Lett.* 59 (18 Nov. 1987), pp. 2020–2023. DOI: 10.1103/PhysRevLett.59.2020.

[27] S. Y. F. Chu, L. P. Ekstrom, and R. B. Firestone. "The Lund/LBNL Nuclear Data Search, A=130 Isobar Diagram". http://nucleardata.nuclear.lu.se/toi/sumframe.htm. Accessed: 2021. 1999.

[28] W. H. Furry. "On Transition Probabilities in Double Beta-Disintegration". In: *Phys. Rev.* 56 (12 Dec. 1939), pp. 1184–1193. DOI: 10.1103/PhysRev.56.1184.

[29] J. Engel and J. Menéndez. "Status and future of nuclear matrix elements for neutrinoless double-beta decay: a review". In: *Reports on Progress in Physics* 80.4 (Mar. 2017), p. 046301. DOI: 10.1088/1361-6633/aa5bc5.

[30] J. Kotila and F. Iachello. "Phase-space factors for double-β decay". In: *Phys. Rev. C* 85 (3 Mar. 2012), p. 034316. DOI: 10.1103/PhysRevC.85.034316.

[31] J. Schechter and J. W. F. Valle. "Neutrinoless double-β decay in SU(2)$\times$U(1) theories". In: *Phys. Rev. D* 25 (11 June 1982), pp. 2951–2954. DOI: 10.1103/PhysRevD.25.2951.

[32] M. Duerr, M. Lindner, and A. Merle. "On the quantitative impact of the Schechter-Valle theorem". In: *J. High Energ. Phys.* 91 (2011). DOI: 10.1007/JHEP06(2011)091.

[33] S. R. Elliott and P. Vogel. "Double Beta Decay". In: *Ann. Rev. Nucl. Part. Sci.* 52.1 (2002), pp. 115–151. DOI: 10.1146/annurev.nucl.52.050102.090641.

[34] A. Barabash et al. "SuperNEMO double beta decay experiment". In: *J. Phys. Conf. Ser.* 375.4 (July 2012), p. 042012. DOI: 10.1088/1742-6596/375/1/042012.

[35] H. V. Klapdor-Kleingrothaus et al. "Evidence for Neutrinoless Double Beta Decay". In: *Mod. Phys. Lett. A* 16.37 (2001), pp. 2409–2420. DOI: 10.1142/S0217732301005825.

[36] I. Esteban et al. "The fate of hints: updated global analysis of three-flavor neutrino oscillations". In: *J. High Energ. Phys.* 178 (2020). DOI: 10.1007/JHEP09(2020)178.

[37] M. Agostini et al. "Final Results of GERDA on the Search for Neutrinoless Double-β Decay". In: *Phys. Rev. Lett.* 125 (25 Dec. 2020), p. 252502. DOI: 10.1103/PhysRevLett.125.252502.

[38] A. Gando et al. "Search for Majorana Neutrinos Near the Inverted Mass Hierarchy Region with KamLAND-Zen". In: *Phys. Rev. Lett.* 117 (8 Aug. 2016), p. 082503. DOI: 10.1103/PhysRevLett.117.082503.

[39] O. Azzolini et al. "Final Result of CUPID-0 Phase-I in the Search for the ^{82}Se Neutrinoless Double-β Decay". In: *Phys. Rev. Lett.* 123 (3 July 2019), p. 032501. DOI: 10.1103/PhysRevLett.123.032501.

[40] E. Armengaud et al. "New Limit for Neutrinoless Double-Beta Decay of ^{100}Mo from the CUPID-Mo Experiment". In: *Phys. Rev. Lett.* 126 (18 May 2021), p. 181802. DOI: 10.1103/PhysRevLett.126.181802.

[41] M. Aker et al. "First direct neutrino-mass measurement with sub-eV sensitivity". 2021. arXiv: 2105.08533 [physics.hep-ex].

[42] Planck Collaboration et al. "Planck 2018 results - VI. Cosmological parameters". In: *A&A* 641 (2020), A6. DOI: 10.1051/0004-6361/201833910.

[43] M. J. Dolinski, A. W. P. Poon, and W. Rodejohann. "Neutrinoless Double-Beta Decay: Status and Prospects". In: *Ann. Rev. Nucl. Part. Sci.* 69.1 (2019), pp. 219–251. DOI: 10.1146/annurev-nucl-101918-023407.

[44] G. Bellini et al. "Cosmic-muon flux and annual modulation in Borexino at 3800 m water-equivalent depth". In: *J. Cosmol. Astropart. Phys.* 2012.05 (May 2012), pp. 015–015. DOI: 10.1088/1475-7516/2012/05/015.

[45] S. Rahaman et al. "Double-beta decay Q values of ^{116}Cd and ^{130}Te". In: *Phys. Lett. B* 703.4 (2011), pp. 412–416. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2011.07.078.

[46] C. Arnaboldi et al. "Production of high purity TeO_2 single crystals for the study of neutrinoless double beta decay". In: *J. Cryst. Growth* 312.20 (2010), pp. 2999–3008. ISSN: 0022-0248. DOI: 10.1016/j.jcrysgro.2010.06.034.

[47] V. Alenkov et al. "First results from the AMoRE-Pilot neutrinoless double beta decay experiment". In: *Eur. Phys. J. C* 79.791 (2019). DOI: 10.1140/epjc/s10052-019-7279-1.

[48] T. Iida et al. "The CANDLES experiment for the study of Ca-48 double beta decay". In: *Nucl. Part. Phys. Proc.* 273-275 (2016). 37th International Conference on High Energy Physics (ICHEP), pp. 2633–2635. ISSN: 2405-6014. DOI: 10.1016/j.nuclphysbps.2015.10.013.

[49] M. Barucci et al. "Measurement of Low Temperature Specific Heat of Crystalline TeO_2 for the Optimization of Bolometric Detectors". In: *J. Low Temp. Phys.* 123 (2001), pp. 303–314. DOI: 10.1023/A:1017555615150.

[50] E. Andreotti et al. "Production, characterization, and selection of the heating elements for the response stabilization of the CUORE bolometers". In: *Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip.* 664.1 (2012), pp. 161–170. ISSN: 0168-9002. DOI: 10.1016/j.nima.2011.10.065.

[51] E. E. Haller et al. "NTD Germanium: A Novel Material for Low Temperature Bolometers". In: *Neutron Transmutation Doping of Semiconductor Materials*. Ed. by Robert D. Larrabee. Boston, MA: Springer US, 1984, pp. 21–36. ISBN: 978-1-4613-2695-3. DOI: 10.1007/978-1-4613-2695-3_2.

[52] C. Arnaboldi et al. "A front-end electronic system for large arrays of bolometers". In: *J. Instrum.* 13.02 (Feb. 2018), P02026. DOI: 10.1088/1748-0221/13/02/p02026.

[53] I. C. Bandac et al. "The $0\nu2\beta$-decay CROSS experiment: preliminary results and prospects". In: *J. High Energ. Phys.* 18 (2020). DOI: 10.1007/JHEP01(2020)018.

[54] F. Pobell. "Matter and Methods at Low Temperatures". Springer-Verlag Berlin Heidelberg, 1996. DOI: 10.1007/978-3-662-03225-1.

[55] C. Alduino et al. "The CUORE cryostat: An infrastructure for rare event searches at millikelvin temperatures". In: *Cryogenics* 102 (2019), pp. 9–21. ISSN: 0011-2275. DOI: 10.1016/j.cryogenics.2019.06.011.

[56] A. D'Addabbo et al. "An active noise cancellation technique for the CUORE Pulse Tube cryocoolers". In: *Cryogenics* 93 (2018), pp. 56–65. ISSN: 0011-2275. DOI: 10.1016/j.cryogenics.2018.05.001.

[57] L. Pattavina et al. "Radiopurity of an archaeological Roman lead cryogenic detector". In: *Eur. Phys. J. A* 55.127 (2019). DOI: 10.1140/epja/i2019-12809-0.

[58] C. Alduino et al. "The projected background for the CUORE experiment". In: *Eur. Phys. J. C* 77.543 (2017). DOI: 10.1140/epjc/s10052-017-5080-6.

[59] K. Alfonso et al. "Search for Neutrinoless Double-Beta Decay of ^{130}Te with CUORE-0". In: *Phys. Rev. Lett.* 115 (10 Sept. 2015), p. 102502. DOI: 10.1103/PhysRevLett.115.102502.

[60] J. S. Cushman et al. "The detector calibration system for the CUORE cryogenic bolometer array". In: *Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip.* 844 (2017), pp. 32–44. ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.11.020.

[61] S. Di Domizio et al. "A data acquisition and control system for large mass bolometer arrays". In: 13.12 (Dec. 2018), P12003. DOI: 10.1088/1748-0221/13/12/p12003.

[62] S. Di Domizio, F. Orio, and M. Vignati. "Lowering the energy threshold of large-mass bolometric detectors". In: *J. Instrum.* 6.02 (Feb. 2011), P02007. DOI: 10.1088/1748-0221/6/02/p02007.

[63] C. Alduino et al. "Analysis techniques for the evaluation of the neutrinoless double-β decay lifetime in ^{130}Te with the CUORE-0 detector". In: *Phys. Rev. C* 93 (4 Apr. 2016), p. 045503. DOI: 10.1103/PhysRevC.93.045503.

[64] I. T. Jolliffe and J. Cadima. "Principal component analysis: a review and recent developments". In: *Phil. Trans. R. Soc. A* 374.20150202 (Apr. 2016). DOI: 10.1098/rsta.2015.0202.

[65] A. Armatol et al. "Novel technique for the study of pileup events in cryogenic bolometers". In: *Phys. Rev. C* 104 (1 July 2021), p. 015501. DOI: 10.1103/PhysRevC.104.015501.

[66] T. Bouwmans and E. H. Zahzah. "Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance". In: *Comput. Vis. Image Underst.* 122 (2014), pp. 22–34. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2013.11.009.

[67] A. Drobizhev. "Searching for the $0\nu\beta\beta$ decay of ^{130}Te with the ton-scale CUORE bolometer array". PhD thesis. University of California, Berkeley, 2018.

[68] L. Marini. "The CUORE experiment: from the commissioning to the first $0\nu\beta\beta$ limit". PhD thesis. Università degli Studi di Genova, 2018.

[69] M. Redshaw et al. "Masses of ^{130}Te and ^{130}Xe and Double-β-Decay Q Value of ^{130}Te". In: *Phys. Rev. Lett.* 102 (21 May 2009), p. 212502. DOI: 10.1103/PhysRevLett.102.212502.

[70] N. D. Scielzo et al. "Double-β-decay Q values of ^{130}Te, ^{128}Te, and ^{120}Te". In: *Phys. Rev. C* 80 (2 Aug. 2009), p. 025501. DOI: 10.1103/PhysRevC.80.025501.

[71] M. A. Fehr, M. Rehkamper, and A. N. Halliday. "Application of MC-ICPMS to the precise determination of tellurium isotope compositions in chondrites, iron meteorites and sulfides". In: *Int. J. Mass Spectrometry* 232 (2004), pp. 83–94. DOI: 10.1016/j.ijms.2003.11.006.

[72] D. Q. Adams et al. "Improved Limit on Neutrinoless Double-Beta Decay in ^{130}Te with CUORE". In: *Phys. Rev. Lett.* 124 (12 Mar. 2020), p. 122501. DOI: 10.1103/PhysRevLett.124.122501.

[73] A. Caldwell, D. Kollár, and K. Kröninger. "BAT – The Bayesian analysis toolkit". In: *Computer Physics Communications* 180.11 (2009), pp. 2197–2209. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2009.06.026.

[74] C. P. Robert. "The Metropolis-Hastings Algorithm". 2016. arXiv: 1504.01896 [stat.CO].

[75] W. A. Rolke and A. M. López. "Confidence intervals and upper bounds for small signals in the presence of background noise". In: *Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip.* 458.3 (2001), pp. 745–758. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(00)00935-9.

[76] N. I. Tkhashokov, R. S. Mirzoev, and S. B. Zhilova. "Li_2Mo_4-$(NH_4)_2MoO_4$-H_2O system at 25°C". In: *Russian Journal of Inorganic Chemistry* 54 (2009), pp. 1655–1661.

[77] N. Uchida. "Optical Properties of Single-Crystal Paratellurite (TeO_2)". In: *Phys. Rev. B* 4 (10 Nov. 1971), pp. 3736–3745. DOI: 10.1103/PhysRevB.4.3736.

[78] L. Bergé et al. "Complete event-by-event $\alpha/\gamma(\beta)$ separation in a full-size TeO_2 CUORE bolometer by Neganov-Luke-magnified light detection". In: *Phys. Rev. C* 97 (3 Mar. 2018), p. 032501. DOI: 10.1103/PhysRevC.97.032501.

[79] C. Arnaboldi et al. "Characterization of ZnSe scintillating bolometers for Double Beta Decay". In: *Astropart. Phys.* 34.6 (2011), pp. 344–353. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2010.09.004.

[80] D. L. Lincoln et al. "First Direct Double-β Decay Q-Value Measurement of ^{82}Se in Support of Understanding the Nature of the Neutrino". In: *Phys. Rev. Lett.* 110 (1 Jan. 2013), p. 012501. DOI: 10.1103/PhysRevLett.110.012501.

[81] S. Rahaman et al. "Q values of the ^{76}Ge and ^{100}Mo double-beta decays". In: *Phys. Lett. B* 662.2 (2008), pp. 111–116. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2008.02.047.

[82] E. Armengaud et al. “Precise measurement of $2\nu\beta\beta$ decay of ^{100}Mo with the CUPID-Mo detection technology”. In: *Eur. Phys. J. C* 80.674 (2020). DOI: 10.1140/epjc/s10052-020-8203-4.

[83] J. W. Beeman et al. “Characterization of bolometric light detectors for rare event searches”. In: *J. Instrum.* 8.07 (July 2013), P07021. DOI: 10.1088/1748-0221/8/07/p07021.

[84] The CUPID Interest Group. “CUPID pre-CDR”. 2019. arXiv: 1907.09376 [physics.ins-det].

[85] N. Casali. “Model for the Cherenkov light emission of TeO2 cryogenic calorimeters”. In: *Astropart. Phys.* 91 (2017), pp. 44–50. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2017.03.004.

[86] N. J. Coron et al. “Highly sensitive large-area bolometers for scintillation studies below 100 mK”. In: *Opt. Eng.* 43.7 (2004), pp. 1568–1576. DOI: 10.1117/1.1758730.

[87] J. Caravaca et al. “Experiment to demonstrate separation of Cherenkov and scintillation signals”. In: *Phys. Rev. C* 95 (5 May 2017), p. 055801. DOI: 10.1103/PhysRevC.95.055801.

[88] M. Askins et al. “THEIA: an advanced optical neutrino detector”. In: *Eur. Phys. J. C* 80 (416 May 2020). DOI: 10.1140/epjc/s10052-020-7977-8.

[89] R. G. Huang et al. “Characterization of light production and transport in tellurium dioxide crystals”. In: *J. Instrum.* 14.10 (Oct. 2019), P10032. DOI: 10.1088/1748-0221/14/10/p10032.

[90] “RAT User’s Guide”. https://rat.readthedocs.io/. Accessed: 2021.

[91] X. Chen et al. “Luminescence properties of large-size Li_2MoO_4 single crystal grown by Czochralski method”. In: *J. Cryst. Growth* 558 (2021), p. 126022. ISSN: 0022-0248. DOI: 10.1016/j.jcrysgro.2020.126022.

[92] D. A. Spassky et al. “Low temperature luminescence and charge carrier trapping in a cryogenic scintillator Li_2MoO_4”. In: *J. Lumin.* 166 (2015), pp. 195–202. ISSN: 0022-2313. DOI: 10.1016/j.jlumin.2015.05.042.

[93] B. Schmidt et al. “Muon-induced background in the EDELWEISS dark matter search”. In: *Astropart. Phys.* 44 (2013), pp. 28–39. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2013.01.014.

[94] E. Armengaud et al. “The CUPID-Mo experiment for neutrinoless double-beta decay: performance and prospects”. In: *Eur. Phys. J. C* 80.44 (2020). DOI: 10.1140/epjc/s10052-019-7578-6.

[95] E. Armengaud et al. “Performance of the EDELWEISS-III experiment for direct dark matter searches”. In: *J. Instrum.* 12.08 (Aug. 2017), P08010. DOI: 10.1088/1748-0221/12/08/p08010.

[96] E. Armengaud et al. “Development of ^{100}Mo-containing scintillating bolometers for a high-sensitivity neutrinoless double-beta decay search”. In: *Eur. Phys. J. C* 77.785 (2017). DOI: 10.1140/epjc/s10052-017-5343-2.

[97] R. Huang et al. “Pulse shape discrimination in CUPID-Mo using principal component analysis”. In: *J. Instrum.* 16.03 (Mar. 2021), P03032. DOI: 10.1088/1748-0221/16/03/p03032.

[98] R. Arnold et al. “Results of the search for neutrinoless double-β decay in ^{100}Mo with the NEMO-3 experiment”. In: *Phys. Rev. D* 92 (7 Oct. 2015), p. 072011. DOI: 10.1103/PhysRevD.92.072011.

[99] B. Patra et al. “Cryo-CMOS Circuits and Systems for Quantum Computing Applications”. In: *IEEE Journal of Solid-State Circuits* 53.1 (2018), pp. 309–321. DOI: 10.1109/JSSC.2017.2737549.

[100] S. J. Pauka et al. “A cryogenic CMOS chip for generating control signals for multiple qubits”. In: *Nat. Electron.* 4 (2021), pp. 64–70. DOI: 10.1038/s41928-020-00528-y.

[101] N. T. Bronn et al. “Fast, high-fidelity readout of multiple qubits”. In: *J. Phys. Conf. Ser.* 834 (May 2017), p. 012003. DOI: 10.1088/1742-6596/834/1/012003.

[102] R. Kalra et al. “Vibration-induced electrical noise in a cryogen-free dilution refrigerator: Characterization, mitigation, and impact on qubit coherence”. In: *Rev. Sci. Instrum.* 87.073905 (2016). DOI: 10.1063/1.4959153.

[103] O. Deleage. “Modes of operation of an N-type MOSFET”. https://commons.wikimedia.org/wiki/File:MOSFET_functioning.svg. Accessed: 2021. 2008.

[104] W. Liu et al. “BSIM 3v3.2 MOSFET Model Users’ Manual”. Tech. rep. UCB/ERL M98/51. EECS Department, University of California, Berkeley, Aug. 1998. URL: http://www2.eecs.berkeley.edu/Pubs/TechRpts/1998/3486.html.

[105] A. Kabaoğlu and M. Berke Yelten. “A cryogenic modeling methodology of MOSFET I-V characteristics in BSIM3”. In: *2017 14th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*. 2017, pp. 1–4. DOI: 10.1109/SMACD.2017.7981578.

[106] C. Luo et al. “MOSFET characterization and modeling at cryogenic temperatures”. In: *Cryogenics* 98 (2019), pp. 12–17. ISSN: 0011-2275. DOI: 10.1016/j.cryogenics.2018.12.009.

[107] A. Beckers, F. Jazaeri, and C. Enz. “Cryogenic MOSFET Threshold Voltage Model”. In: *ESSDERC 2019 - 49th European Solid-State Device Research Conference (ESSDERC)*. 2019, pp. 94–97. DOI: 10.1109/ESSDERC.2019.8901806.

[108] N. C. Dao et al. “An enhanced MOSFET threshold voltage model for the 6–300K temperature range”. In: *Microelectronics Reliability* 69 (2017), pp. 36–39. ISSN: 0026-2714. DOI: 10.1016/j.microrel.2016.12.007.

[109] R. G. Huang et al. "Cryogenic characterization of 180 nm CMOS technology at 100 mK". In: *J. Instrum.* 15.06 (June 2020), P06026. DOI: 10.1088/1748-0221/15/06/p06026.

[110] B. Dierickx et al. "Model for hysteresis and kink behavior of MOS transistors operating at 4.2 K". In: *IEEE Transactions on Electron Devices* 35.7 (1988), pp. 1120–1125. DOI: 10.1109/16.3372.

[111] A. Zaslavsky et al. "Impact ionization-induced bistability in CMOS transistors at cryogenic temperatures for capacitorless memory applications". In: *Appl. Phys. Lett.* 119.043501 (2021). DOI: 10.1063/5.0060343.

[112] M. Agostini, G. Benato, and J. A. Detwiler. "Discovery probability of next-generation neutrinoless double-β decay experiments". In: *Phys. Rev. D* 96 (5 Sept. 2017), p. 053001. DOI: 10.1103/PhysRevD.96.053001.

[113] D. Q. Adams et al. "CUORE Opens the Door to Tonne-scale Cryogenics Experiments". 2021. arXiv: 2108.07883 [physics.ins-det].

[114] C. Alduino et al. "First Results from CUORE: A Search for Lepton Number Violation via $0\nu\beta\beta$ Decay of ^{130}Te". In: *Phys. Rev. Lett.* 120 (13 Mar. 2018), p. 132501. DOI: 10.1103/PhysRevLett.120.132501.

[115] D. Q. Adams et al. "Measurement of the $2\nu\beta\beta$ Decay Half-Life of ^{130}Te with CUORE". In: *Phys. Rev. Lett.* 126 (17 Apr. 2021), p. 171801. DOI: 10.1103/PhysRevLett.126.171801.

[116] R. D. Cousins. "Why isn't every physicist a Bayesian?" In: *Am. J. Phys.* 63.5 (1995). DOI: 10.1119/1.17901.

[117] D. Q. Adams et al. "High sensitivity neutrinoless double-beta decay search with one tonne-year of CUORE data". 2021. arXiv: 2104.06906 [physics.nucl-ex].

Appendix A

Considerations for the CUORE $0\nu\beta\beta$ Fit

A.1 ROI Fit Range And Components

Excess at 2480 keV

In the previous $0\nu\beta\beta$ result published by CUORE, there were hints of an excess near 2480 keV with a significance of slightly over 2σ [72]. We did not have an explanation for how there could be an excess of that magnitude at that energy, and the significance of the peak was too large to just ignore but too small to be conclusive. We thus decided to exclude it from the ROI entirely. With the addition of more data and a reanalysis of the old data, we can check once more for the significance of this peak. A plot of the rate of events in the 2476-2482 keV energy region over time after all analysis cuts is shown in Fig. A.1, using the 15 datasets unblinded and analyzed in the $0\nu\beta\beta$ analysis described in Chapter 6. We can see that the first 7 datasets corresponding to the data analyzed in [72] do still have an excess of events around 2480 keV in this reanalysis, but the new data sees no excess anymore. One possible explanation is that the original excess was simply a statistical fluctuation, and with the addition of more data we see now that it was nothing of importance.

Another explanation could be that the excess was due to an actual radioactive background, but that its activity has simply decayed away already given the length of time since CUORE first began taking data. In this case, the event rate should follow an exponential decay down to the baseline background rate of the rest of the ROI. Fitting the rates in this fashion gives a half-life of around 180 days, but there are no known possible backgrounds with half-lives on that scale with decay energies in this region, even accounting for possible quenching effects. That suffices as weak evidence against this hypothesis, but we still take the ROI from 2490 to 2575 keV in the $0\nu\beta\beta$ fit so that we don't have to worry about the 2480 keV energy region.

For curiosity, the $0\nu\beta\beta$ fit with the ROI extended down to 2465 keV and including a peak near 2480 keV is shown in Fig. A.2, but this was checked after we had already made

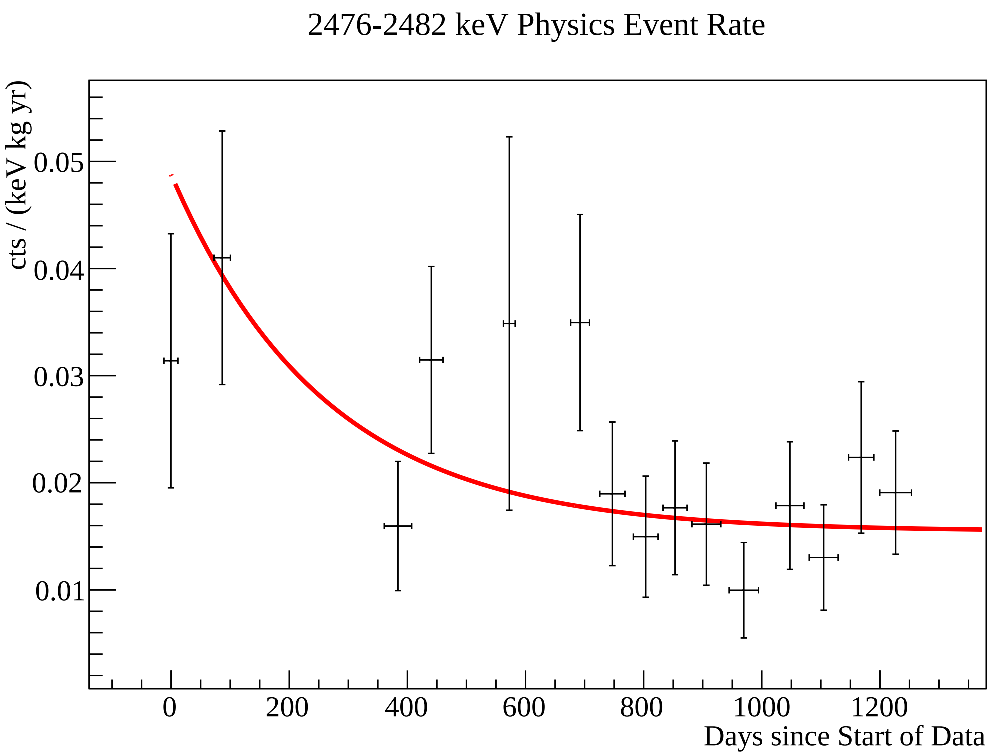
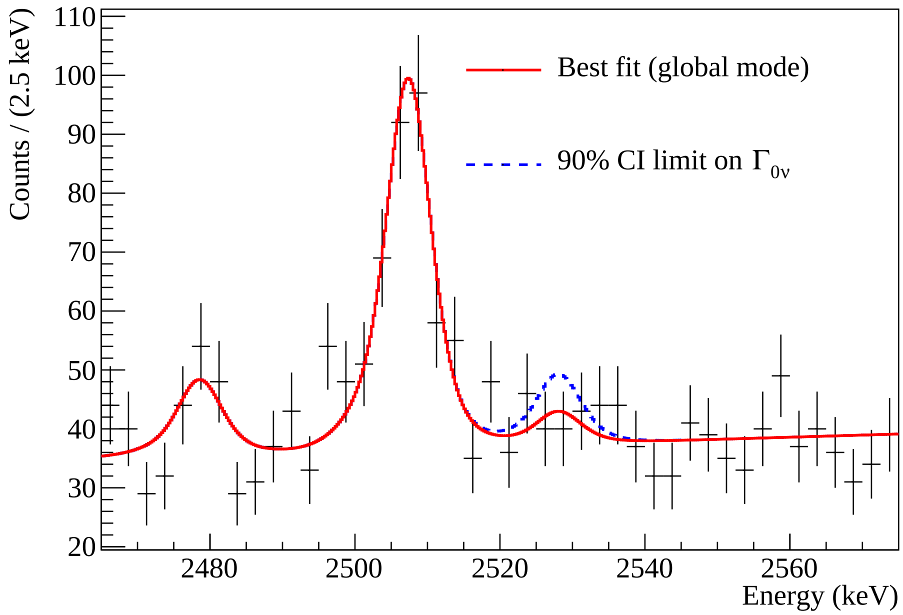


Figure A.1: Event rate in the 2476-2482 keV energy region after all analysis cuts over time, where each point corresponds to one dataset and the uncertainties are just statistical. We can see that in the first 7 datasets, which were analyzed in CUORE's previous result, the rate is higher than usual, but in the most recent datasets the rate is consistent with the expected background rate of 0.0149 ckky obtained from the $0\nu\beta\beta$ fit. The fit drawn in red is a constant plus a decaying exponential, corresponding to a half-life of 180 days, which does not correspond to any known decay at this energy.

the decision looking at blinded data to exclude this region from the fit. The peak near 2480 keV still has a significance of $\sim 2\sigma$ here, but this is lower than it was in the previous result. The resulting limit is slightly weaker, as the background estimate at $Q_{\beta\beta}$ becomes lower and so the statistical excess at $Q_{\beta\beta}$ becomes more prominent.

Background Slope

In both of CUORE's first two $0\nu\beta\beta$ analyses, there were strong statistical underfluctuations near $Q_{\beta\beta}$ leading to stronger than expected results [72, 114]. Although the magnitudes of these fluctuations were not so strong as to be out of the range of reasonable expectations, they were strong enough to make us wonder if they could be caused by misestimations of the



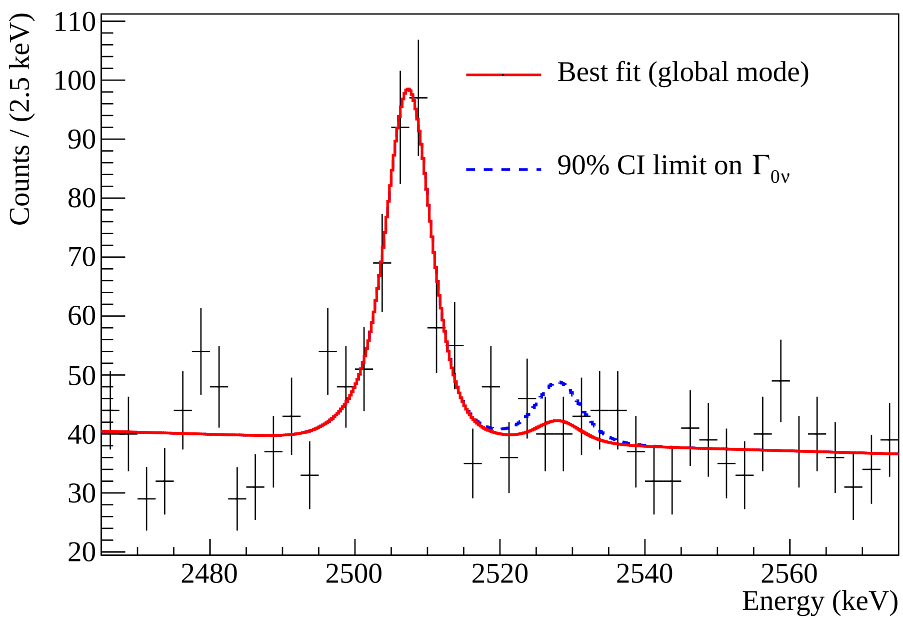


Figure A.2: **Top**: $0\nu\beta\beta$ fit using the CUORE data presented in Chapter 6, but with the ROI extended down to 2465 keV instead of 2490 keV and allowing for a peak near 2480 keV. The rest of the fit components are the same as described in Chapter 6. The fit places the peak at about 2478 keV, with a significance of $\sim 2\sigma$. **Bottom**: the $0\nu\beta\beta$ fit using the ROI extended down to 2465 keV, but without the allowance for a peak near 2480 keV. This results in a slightly higher estimate of the background, resulting in a stronger limit than obtained with the inclusion of this extra peak but still a weaker limit than obtained with the smaller ROI.

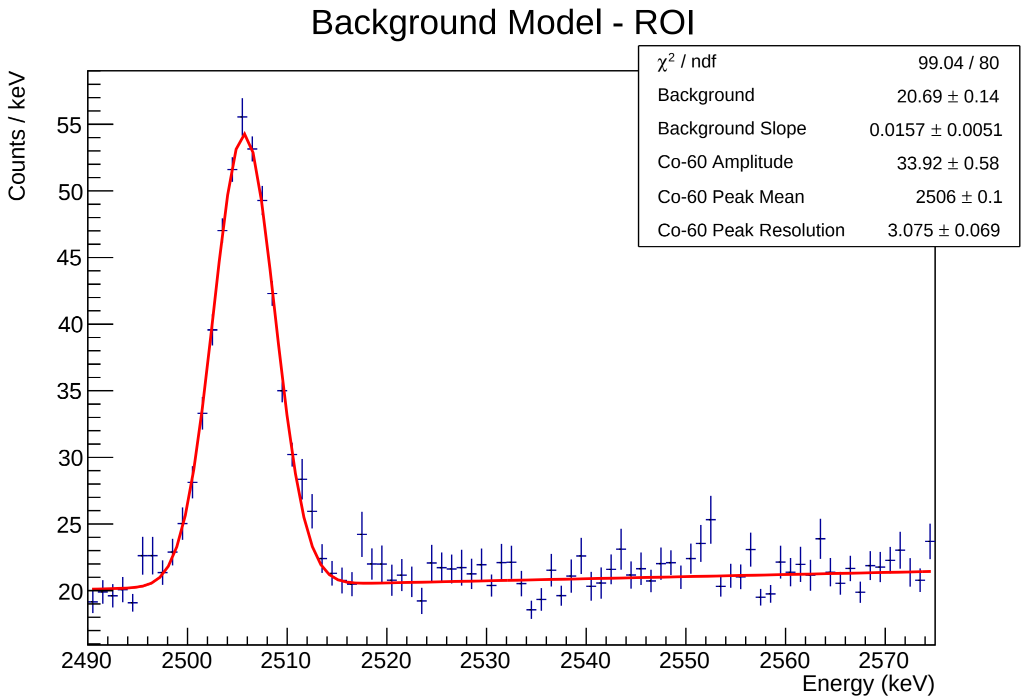


Figure A.3: The ROI generated from the background model assuming no $0\nu\beta\beta$ signal, with errors on each point obtained from the background model Monte Carlo's uncertainty. The fit is a linear background plus a Gaussian centered at the ^{60}Co peak. The fit favors a non-zero background slope with non-negligible significance.

background. Since then, we have developed a more sophisticated background model that was used to obtain a precise measurement of the $2\nu\beta\beta$ half-life of ^{130}Te [115]. Before unblinding, we used the background model to generate an expected background shape in the ROI using the exposure we have in this analysis. We fit it with the ^{60}Co peak plus background, shown in Fig. A.3, and we can see that the fit favors a non-zero slope to the background. This motivated the change to the $0\nu\beta\beta$ fit to allow a linear slope to the background, unlike in CUORE's previous results.

We also conducted a study of whether using a constant fit to the background could result in a bias towards negative values of $\Gamma_{0\nu}$ in our $0\nu\beta\beta$ fit. We generated 10000 toys where the ROI is populated by sampling from the background model prediction, using the associated uncertainty from Monte Carlo for each energy bin. Each toy is fit with a flat background plus ^{60}Co peak plus $0\nu\beta\beta$ component, and we check the resulting distribution of best-fit values for $\Gamma_{0\nu}$. We find no evidence of bias towards either negative or positive values, which provides evidence against the possibility that our previous underfluctuations near $Q_{\beta\beta}$ were the result of improperly fitting the background as a flat component.

A.2 Choice of Bayesian Priors

One of the primary criticisms of Bayesian approaches to data analysis is the dependence on the choice of priors. Our choices of priors for nuisance parameters tend to be quite natural, and in fact one of the advantages of Bayesian analyses is their ability to incorporate systematic uncertainties in a very intuitive manner. Even for nuisance parameters such as the background indices, we tend to have an idea of the order of magnitude we expect to see, and so a uniform prior over a reasonable range suffices as a way to say we have no preference for any particular value. However, when it comes to drawing conclusions about a truly unknown physical parameter of interest, which in our case is the decay rate $\Gamma_{0\nu}$, it becomes much more unclear what the "correct" prior should be.

We want to use an uninformative prior for a $0\nu\beta\beta$ search, meaning that we want our prior to express our (almost) complete ignorance about the true $0\nu\beta\beta$ decay rate. This is usually done with some kind of uniform distribution as the prior, which will nominally not weigh the result towards any particular value. In our analysis, we do this by setting our prior on $\Gamma_{0\nu}$ as a uniform distribution over its possible values. This treats the different possible signal rates as equally likely, and we choose this option because the number of signal events is the actual physical observable in our experiment. However, even if we accept that $\Gamma_{0\nu}$ is the parameter we should be setting a prior on, it is not at all obvious that a uniform prior is the proper uninformative prior. Bayesian statisticians have argued that for a Poisson process where we know nothing about the true rate, the proper uninformative prior should actually be either $P(\Gamma_{0\nu}) \propto 1/\sqrt{\Gamma_{0\nu}}$ or $P(\Gamma_{0\nu}) \propto 1/\Gamma_{0\nu}$, neither of which is the uniform prior that is commonly accepted in particle physics as "uninformative" [116]. The latter option is particularly curious, since it corresponds to a uniform prior on $\ln(\Gamma_{0\nu})$, which is an option that physicists have considered as a way to express our ignorance of the scale of the true value of a physical parameter in nature. This would say that, for instance, a decay rate between 10^{-26} yr^{-1} and 10^{-25} yr^{-1} is *a priori* equally likely as a decay rate between 10^{-25} yr^{-1} and 10^{-24} yr^{-1}. This better reflects how a physicist would probably think about the as-of-yet unknown $0\nu\beta\beta$ decay rates. Phrased casually, this is saying that with our current level of thereotical and experimental results, we are wondering more about what the order of magnitude of $\Gamma_{0\nu}$ is, rather than whether the prefactor is 2 or 3.

The ambiguity in the proper choice of priors becomes even worse when we consider that even if we accept that we should use a uniform prior, there are other reasonable choices for variables to set the prior on rather than $\Gamma_{0\nu}$. An obvious alternative is instead of using the decay rate, we could set our prior on the half-life $T_{1/2}^{0\nu}$. Setting a uniform prior on $T_{1/2}^{0\nu}$ would be the equivalent of setting a uniform prior on $1/\Gamma_{0\nu}$, so this is obviously inconsistent with the option of setting a uniform prior on $\Gamma_{0\nu}$, even though both variables equivalently characterize the Poisson process. Another option is to consider that a more fundamental physical parameter might be $m_{\beta\beta}$, which determines $\Gamma_{0\nu}$ for all $\beta\beta$ isotopes. It is thus reasonable to think of imposing our prior on $m_{\beta\beta}$ instead. This would be the equivalent of imposing a prior on $\sqrt{\Gamma_{0\nu}}$, leading to yet another different result. We can extend this logic even further and say that $m_{\beta\beta}$ isn't fundamental either, since it is determined by the values

Table A.1: Fit results from the $0\nu\beta\beta$ analysis of the CUORE data presented in Chapter 6 for different choices of priors, with the first column showing the value on which we set a uniform prior and the second column showing the equivalent prior PDF on $\Gamma_{0\nu}$. The final column shows the resulting 90% C.I. lower limits on $T_{1/2}^{0\nu}$, obtained by integrating 90% of the marginalized posterior PDFs for $\Gamma_{0\nu}$.

Uniform Prior On	Prior $P(\Gamma_{0\nu})$	90% C.I. $T_{1/2}^{0\nu}$ Limit [yrs]
$\Gamma_{0\nu}$	1	$2.2 \cdot 10^{25}$
$m_{\beta\beta}$	$1/\sqrt{\Gamma_{0\nu}}$	$2.6 \cdot 10^{25}$
$\ln \Gamma_{0\nu}$	$1/\Gamma_{0\nu}$	$3.3 \cdot 10^{25}$
$T_{1/2}^{0\nu}$	$1/\Gamma_{0\nu}^2$	$8.9 \cdot 10^{25}$

of the Majorana phases, δ_{CP} phase, and neutrino masses. Starting with priors on those fundamental values would overly complicate this analysis, but it can be done to postulate the likelihood of possible values of $m_{\beta\beta}$, which would determine discovery probabilities of various $0\nu\beta\beta$ experiments [112].

Given this uncertainty in how to choose a proper prior for a Bayesian analysis with no clear answer, one may wonder why we are using Bayesian techniques at all. Bayesian statistics do offer two notable advantages: a rigorous way to incorporate physical constraints into the priors and an intuitive way to interpret the results. On the first point, we can constrain our priors to forbid negative values of $\Gamma_{0\nu}$, while Frequentist methods tend to use ad hoc methods to avoid unphysical results. On the second point, the proper interpretation of a Bayesian 90% C.I. limit is almost exactly what one would naively expect[1]. On the other hand, the interpretation of a Frequentist limit at 90% confidence is not the way that most people intuitively think about statistics[2]. For these reasons, Bayesian approaches are rising in popularity, but since Frequentist techniques are still considered "classical" in particle physics, they are often treated as a benchmark. The fact that a uniform prior on the physical parameter of interest (in our case, $\Gamma_{0\nu}$) yields results most similar to those obtained by Frequentist methods is thus part of the reason it is the default choice in Bayesian analyses.

To check for the effect that alternative choices of priors would have on our result, we also perform the Bayesian $0\nu\beta\beta$ analysis using uniform priors on $m_{\beta\beta}$, $\ln \Gamma_{0\nu}$, and $T_{1/2}^{0\nu}$. These priors are all undefined at $P(\Gamma_{0\nu} = 0)$, so we must set a lower limit for their probability distribution functions. The choice of lower limit has a strong effect on the resulting limit, as these priors are all weighted towards smaller values of $\Gamma_{0\nu}$. We use 10^{-27} yr^{-1} as the lower limit for these alternative priors on $\Gamma_{0\nu}$, which corresponds to approximately 1 expected signal event with our ^{130}Te exposure.

[1] "Having updated our beliefs in accordance with the experiment's results, we believe there is a 90% probability that the true value falls within this range."

[2] "If an ensemble of similar experiments were repeated many times and constructed their results using this same method, then 90% of the resulting limits would contain the true value of the parameter."

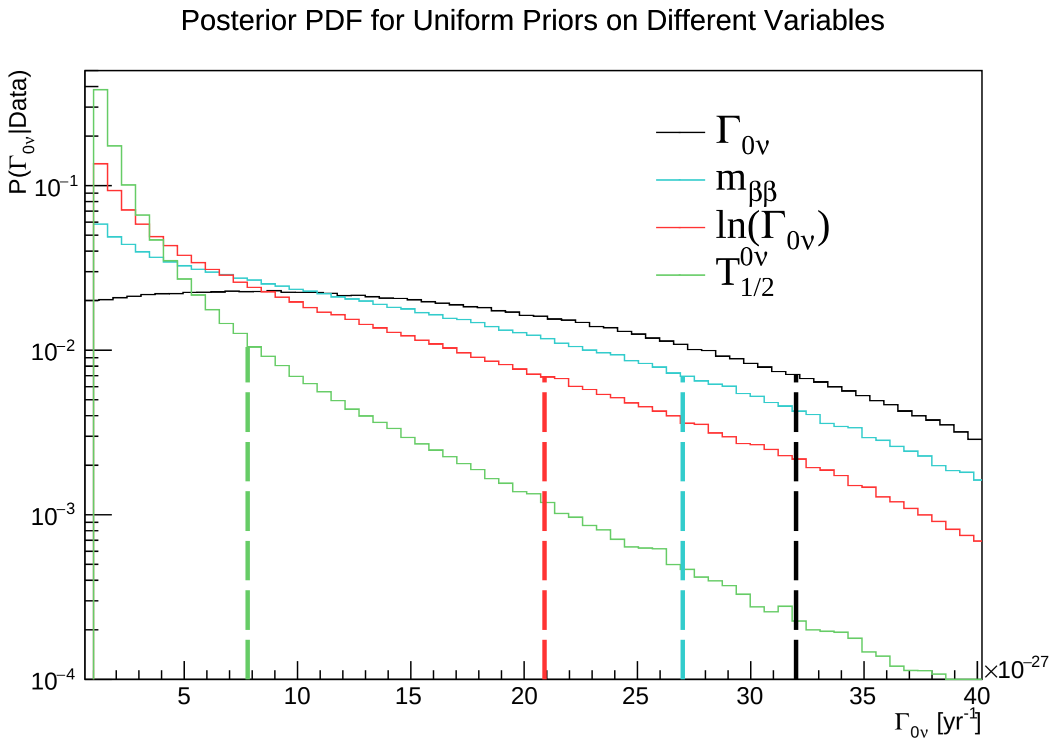


Figure A.4: Posterior PDFs for $\Gamma_{0\nu}$ from analyses of the CUORE data presented in Chapter 6 using different priors for the $0\nu\beta\beta$ rate. The posterior PDFs are labeled by the variable on which a uniform prior was imposed, and the dashed lines indicate the corresponding 90% C.I. upper limits on $\Gamma_{0\nu}$. The best-fit value of $\Gamma_{0\nu}$ is the lower edge of the prior PDF for all cases other than the uniform prior on $\Gamma_{0\nu}$, since our observations are consistent with 0 signal events.

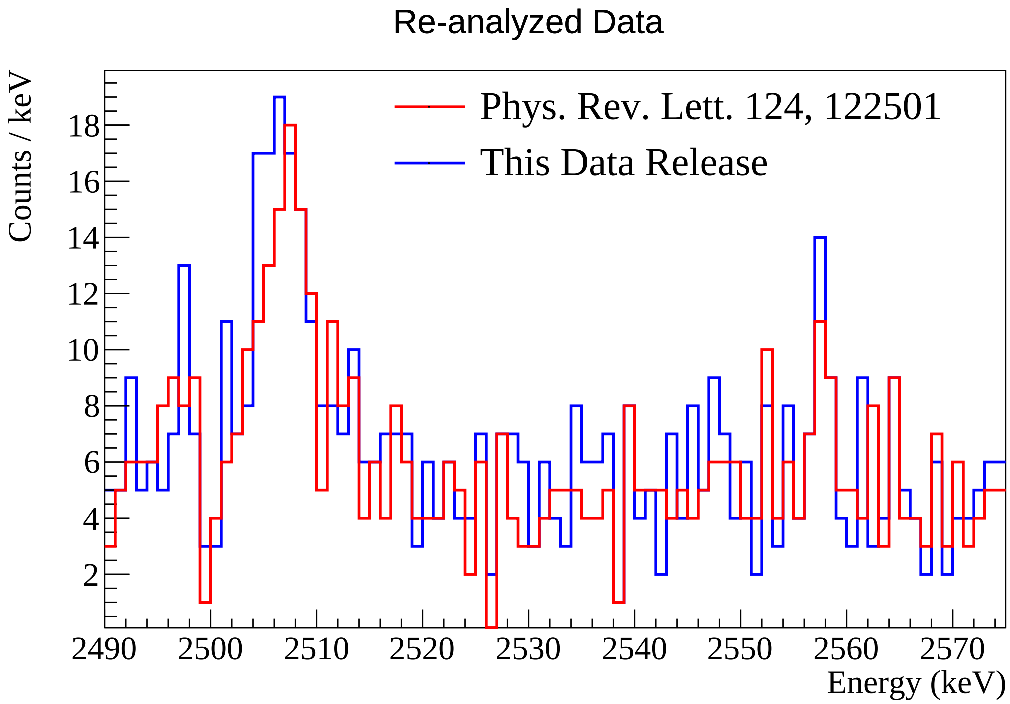
The results from these alternative priors compared to our official result using a uniform prior on $\Gamma_{0\nu}$ are shown in Table A.1, and the comparison of the corresponding marginalized posterior probability distribution functions is shown in Fig. A.4. The limit on the $0\nu\beta\beta$ decay half-life of ^{130}Te improves by just a little in the case of a uniform prior on $m_{\beta\beta}$, but it improves by a factor of over 4 in the case of a uniform prior on $T^{0\nu}_{1/2}$. We can also see that for all of these alternative prior choices, the resulting posterior PDF is peaked at the minimum permitted value of $\Gamma_{0\nu}$. This is the result of having data consistent with 0 signal events and demonstrates the sensitivity of these other priors to the choice of lower limit. Since our data are equally consistent with any value of $\Gamma_{0\nu} < O(10^{-27}\ \mathrm{yr}^{-1})$, the posterior PDF ends up peaked at the minimum value permitted by the prior for these priors that are weighted towards smaller values of $\Gamma_{0\nu}$. This can be considered justification for our choice of a uniform prior when constructing a limit in the presence of no signal, since it is not sensitive to the choice of lower limit in the same way.

A.3 Comparison to Previous Result

The $0\nu\beta\beta$ analysis presented in this dissertation and published in [117] present a weaker limit on the $0\nu\beta\beta$ decay of ^{130}Te than CUORE's previous result in [72]. The new result includes a re-analysis of the data that was used in this previous result, yielding a slightly different set of events in the ROI. The ROI after this re-analysis is shown in Fig. A.5, along with the resulting $0\nu\beta\beta$ half-life limit in comparison with the expected sensitivity and the previous limit. Since we are using the same underlying data, the differences in the ROI are purely the result of differences in the analysis procedure, during which the sub-100% efficiencies cause the loss of some random fraction of good events in the ROI. The re-analysis causes the underfluctuation near $Q_{\beta\beta}$ to significantly weaken. We also see in the re-analysis that the ^{60}Co peak is centered closer to its expected position of 2505.7 keV, while the old result had it slightly misplaced. In addition to giving us greater confidence in the validity of the re-analysis, this also motivated our decision to fix the position of the ^{60}Co peak in the new $0\nu\beta\beta$ fit, instead of floating its position like was done in the previous result.

To check the level of consistency between the previously obtained ROI and the newly obtained ROI from the same data, we generate toy experiments using the ROI spectrum obtained with the new analysis of the old data. Instead of Poisson fluctuating the number of events, we instead take energy bins of 1 keV width and fluctuate the number of events in each bin with a binomial distribution following the difference in analysis efficiencies. Analyzing the resulting toys, we find that the previous limit of $T^{0\nu}_{1/2} > 3.2 \cdot 10^{25}$ yrs was in the top 3% of expected possible results from just varying the events selected by the analysis efficiency. This is interpreted as the likelihood that we could have obtained a limit as strong as the previous one just as a result of the difference in analysis efficiencies between the last result and the new re-analysis, given the result that we see with the re-analysis. Our new limit of $T^{0\nu}_{1/2} > 2.0 \cdot 10^{25}$ yrs using just this old data is no longer as strong of an outlier in the expected possible outcomes.

To show the improvement with the analysis of new CUORE data, a comparison of the expected sensitivities with the previous data release and this data release is shown in Fig. A.6, generated with toys in the fashion described in Chapter 6. Although our actual limit has weakened, the median expected sensitivity improved from $T^{0\nu}_{1/2} > 1.7 \cdot 10^{25}$ yrs to $T^{0\nu}_{1/2} > 2.8 \cdot 10^{25}$ yrs, which is roughly what we expect from a 3-fold increase in exposure. Our actual limit of $T^{0\nu}_{1/2} > 2.2 \cdot 10^{25}$ from this analysis was in the bulk of expected outcomes, being lower than 72% of expected possible results, while the limit of $T^{0\nu}_{1/2} > 3.2 \cdot 10^{25}$ from [72] was in the top 3% of expected outcomes.



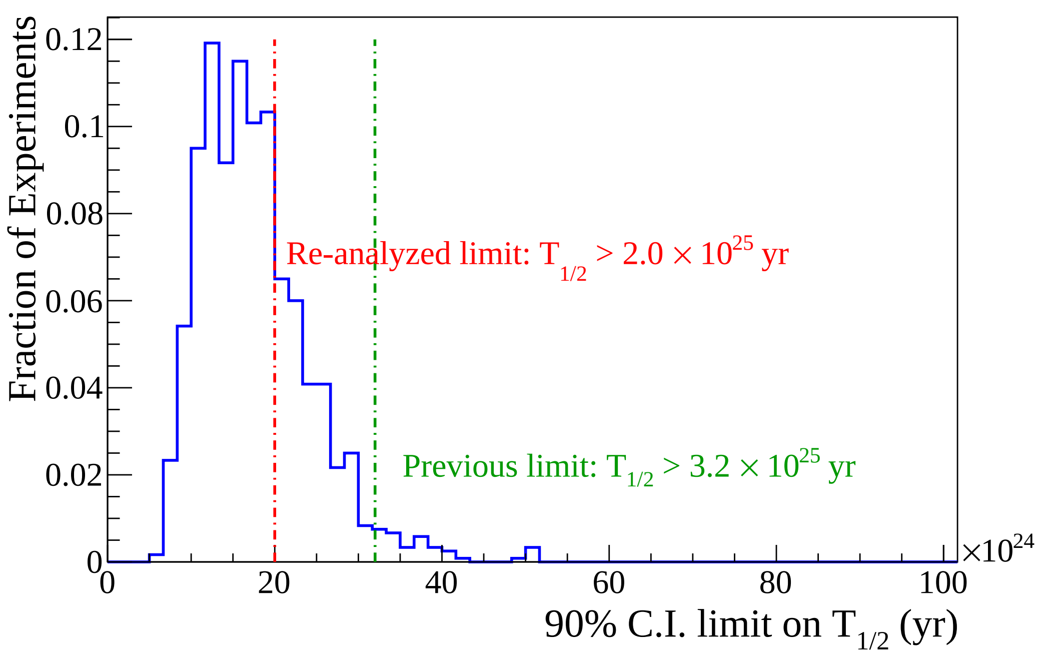


Figure A.5: **Top**: the $0\nu\beta\beta$ ROI that was analyzed in CUORE's previous result plotted with the ROI coming from the new re-analysis of the same data. The differences come from the differences in the analysis procedure and the sub-100% analysis efficiencies. We can see in the re-analysis that the underfluctuation at $Q_{\beta\beta}$ is weaker. **Bottom**: distribution of expected limits on $T_{1/2}^{0\nu}$ using this old data, as determined by generating toy experiments from the ROI spectrum obtained in this new analysis of the old data. The variations between toys are due to the analysis efficiencies instead of Poisson statistics.

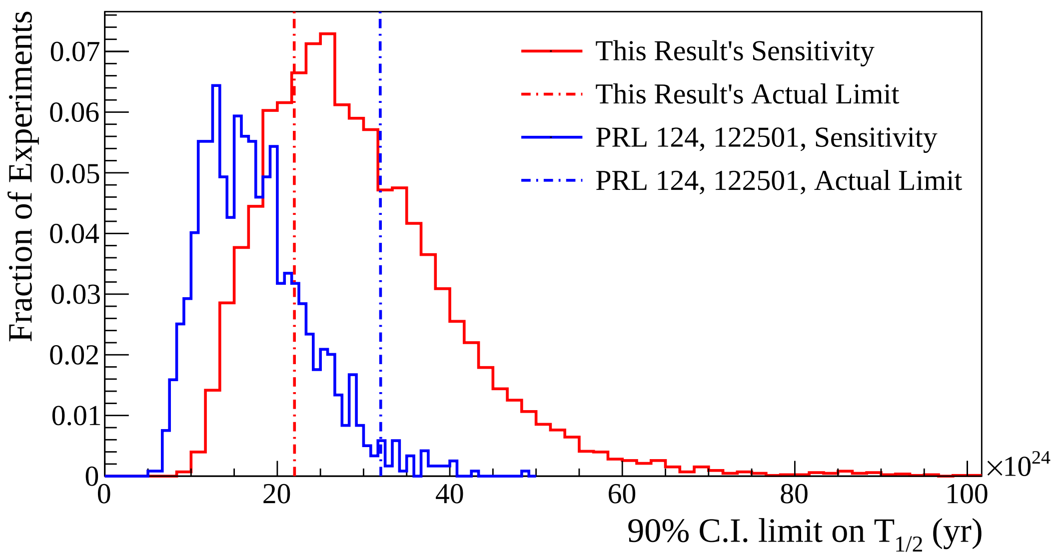


Figure A.6: Distribution of expected 90% C.I. limits on $T^{0\nu}_{1/2}$, as obtained by generating toys while Poisson fluctuating the number of background and ^{60}Co events. The distribution and median expected sensitivity is shown for the set of data corresponding to the previous result in [72], as well as to the full set of data unblinded in the result described in this dissertation. The actual obtained limits in each case are shown as well. We see the new result has an overall improved sensitivity as expected, but it results in a weaker actual limit due to a slight statistical underperformance this time, compared to a strong overperformance in the previous result.