Lawrence Berkeley National Laboratory
LBL Publications

Title
The Search for Neutrinoless Double Beta Decay with 130Te with CUORE-0:

Permalink
https://escholarship.org/uc/item/47k613b2

Author
Ouellet, Jonathan, L

Publication Date
2017-12-05

The Search for $0\nu\beta\beta$ Decay in ^{130}Te with CUORE-0

by

Jonathan Loren Ouellet
Ph.D. Thesis

Department of Physics
University of California, Berkeley
Berkeley, Ca 94720

and

Nuclear Science Division
Ernest Orlando Lawrence National Laboratory
Berkeley, Ca 94720

Spring 2015

DISCLAIMER

The Search for $0\nu\beta\beta$ Decay in ^{130}Te with CUORE-0

by

Jonathan Loren Ouellet

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 Kolomensky, Chair
Professor Gabriel Orebi Gann
Professor Eric B. Norman

Spring 2015

The Search for $0\nu\beta\beta$ Decay in ^{130}Te with CUORE-0

Abstract

The Search for $0\nu\beta\beta$ Decay in ^{130}Te with CUORE-0

by

Jonathan Loren Ouellet

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Yury Kolomensky, Chair

This thesis describes the design, operation and results of an experimental search for neutrinoless double beta decay ($0\nu\beta\beta$) of ^{130}Te using the CUORE-0 detector.

The discovery of $0\nu\beta\beta$ would have profound implications for particle physics and our understanding of the Universe. Its discovery would demonstrate the violation of lepton number and imply that neutrinos are Majorana fermions and therefore their own anti-particles. Combined with other experimental results, the discovery of $0\nu\beta\beta$ could also have implications for understanding the absolute neutrino mass scale as well as the presently unknown neutrino mass hierarchy.

The CUORE experiment is a ton-scale search for $0\nu\beta\beta$ in ^{130}Te expected to begin operation in late 2015. The first stage of this experiment is a smaller 39-kg active-mass detector called CUORE-0. This detector contains 11 kg of ^{130}Te and operates in the Laboratori Nazionali del Gran Sasso lab in Italy from 2013 – 2015.

The results presented here are based on a $^{\mathrm{nat}}$TeO$_2$ exposure of 35.2 kg·yr, or 9.8 kg·yr exposure of ^{130}Te collected between 2013 – 2015. We see no evidence of $0\nu\beta\beta$ and place an upper limit on the $0\nu\beta\beta$ decay rate of $\Gamma_{0\nu\beta\beta} < 0.25 \times 10^{-24}$ yr^{-1} (90% C.L.), corresponding to a lower limit on the half-life of $T^{0\nu}_{1/2} > 2.8 \times 10^{24}$ yr (90% C.L.).

We combine the present result with the results of previous searches in ^{130}Te. Combining it with the 1.2 kg·yr ^{130}Te exposure from the Three Towers Test run we place a half-life limit of $T^{0\nu}_{1/2} > 3.3 \times 10^{24}$ yr (90% C.L.). And combining these results with the 19.75 kg·yr ^{130}Te exposure from Cuoricino, we place the strongest limit on the $0\nu\beta\beta$ half-life of ^{130}Te to date, at $T^{0\nu}_{1/2} > 4.5 \times 10^{24}$ yr (90% C.L.). Using the present nuclear matrix element calculations for ^{130}Te, this result corresponds to a 90% upper limit range on the effective Majorana mass of $m_{\beta\beta} < 250 - 710$ meV.

To my family

Contents

6 CUORE-0 Analysis and $0\nu\beta\beta$ Fit   107

7 CUORE and Beyond   140

Bibliography   145

A CUORE-0 Dataset Data   155

B Generalized Amplitude Evaluation   165

C Bayesian Nuisance Parameters   176

D The Coldest Cubic Meter in the Universe   179

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor, Yury Kolomensky, whose guidance and suggestions have been extremely valuable. I am especially grateful for the long and insightful debates (occasionally over glasses of wine) and for support to follow my own pursuits and the encouragement to see them through.

I would also like to acknowledge Stuart Freedman, whose sudden passing was a tragic loss to our group and to the scientific community. I only had a brief opportunity to work with Stuart, but his lessons left a lasting impression. His many scientific accomplishments leave us the highest standard to which we must aspire.

I would like to thank the other members of my committee, Gabriel Orebi Gann and Eric Norman, who provided very useful guidance and feedback in the writing of this thesis.

The experiment upon which this dissertation is based is the result of years of work by the members of the CUORE collaboration and builds upon decades of experience. I have been very fortunate to have been able to learn from this wealth of expertise and to be a part of such an ambitious enterprise.

I would like to acknowledge Marisa Pedretti and Lucia Canonica for their tireless efforts and leadership during the construction and operation of the CUORE-0 detector. I would like to acknowledge all of the on-site shifters who keep the CUORE-0 detector running.

I would like to acknowledge the members of the CUORE-0 analysis group that I have had the opportunity to work closely with over the past years and whose work contributed to this thesis: Ke Han, Tom Banks, Tommy O'Donnell, Matteo Biassoni, Jeremy Cushman, Luca Gironi, Kyungeun Lim, Maria Martinez, Reina Maruyama, Gabriele Piperno, Stefano Pozzi, Silvia Capelli, Claudia Tomei, Marco Vignati and Brian Zhu.

I would like to thank the current and former Berkeley and LBL postdocs and staff, Raul Hennings-Yeoman, Yuan Mei, Vivek Singh, Jake Feintzeig, Richard Kadel and Brian Fujikawa for many useful conversations, collaborations, and support. Particularly, I would like to thank Tom Banks and Tommy O'Donnell for their invaluable feedback during the writing of this document. I would like to acknowledge my fellow graduate students at Berkeley, Alexey Drobizhev and Sachi Wagaarachchi with whom I have worked very closely while setting up and operating the Berkeley dilution refrigerator.

I would like to thank Paolo Gorla and Carlo Bucci, from whom I have learned much about the dark art of dilution refrigerators as well as Tom Wise with whom I worked closely in Italy.

I would like to acknowledge Don Orlando and the 111 lab. All told, I spent 3.5 years in the 111 lab and never stopped learning new things. I would like to acknowledge the multitude of people and resources in the Berkeley Physics Department and LBL, including the Weak Interactions group, Anne Takazawa and Donna Sakima.

Finally, I would like to thank my family and friends for their love and support over the years.

Chapter 1

Introduction

Contrary to popular belief, the moon is *not* made out of cheese. It is, however, made out of matter. This is a non-trivial statement because it could have just as easily been made out of anti-matter. So why is matter preferred?

Since its formulation in the late 1960s, the Standard Model (SM) of particle physics has been extremely successful in explaining nearly all of the phenomena that we observe on a microscopic scale. Its five decades of success was capped with the 2012 discovery of the Higgs Boson at the Large Hadron Collider (LHC) [1, 2]. The Standard Model of Cosmology (referred to as ΛCDM) is to the Universe on large scales what the SM is to the Universe on small scales. Combined together, the SM and ΛCDM have been able to explain much of the Universe on scales both small and large, with each one relying on the other. Even from an experimental perspective, measurements on the largest scales have placed limits on processes at the smallest scales, and vice versa. But for all of these successes, there are still many unanswered questions that neither theory has yet been able to adequately address. And though there are many unanswered questions, we will focus on one:

Why is there something instead of nothing?

As far as we can tell, everything that we see in the Universe is composed primarily of matter — all of the galaxies, the stars, the Earth, and even the Moon. But the SM tells us that it equally could have been composed of anti-matter and the Universe would look identical. So it is natural to ask, why matter? Why not anti-matter? But the problem is even worse than this. Our current picture of the Universe (SM+ΛCDM) tells us that all the matter and energy we observe in the Universe were created fractions of a second after the Big Bang. However, our present understanding dictates that matter and anti-matter should have been created in equal abundances. As the Universe cooled, eventually the anti-matter and matter would have paired off and annihilated, leaving nothing to create the galaxies and stars. Clearly, this is not what happened. Somehow, matter was preferred over anti-matter. After all the anti-matter paired off and annihilated, there was a remainder of matter; and this remainder produced all of the structure that we see in the Universe. This is known as the Baryon Asymmetry Problem.

   This thesis does not take on this question in whole. Rather, we focus on a small part of one possible way to answer this question. In this chapter, we make this question more concrete, and describe the components that any solution to this question will require. We then describe the role neutrinos could play in resolving the Baryon Asymmetry Problem, and the evolution of our understanding of these ghostly particles over the past 50 years.

1.1   Baryon Asymmetry and the Sakharov Conditions

The Baryon Asymmetry Problem is characterized by the baryon asymmetry parameter

$$\eta_B \equiv \frac{n_B - n_{\bar{B}}}{n_\gamma} \tag{1.1}$$

where n_B is the number density of baryons (e.g. number of protons per cubic meter), $n_{\bar{B}}$ is the number density of anti-baryons (e.g. number of anti-protons per cubic meter) and n_γ is the number density of photons. As far as we can tell, anti-matter in the Universe is extremely sparse; it only exists in significant quantities as the by product of other high-energy processes between matter. The baryon asymmetry has been measured through cosmological observations and Big Bang Nucleosynthesis modeling to be $\eta_B \approx n_B/n_\gamma = (6.047 \pm 0.074) \times 10^{-10}$ [3].

   Naïvely, we can interpret this number in the following way: in the early universe, for every 1,000,000,000 anti-protons, there were roughly 1,000,000,001 protons. Eventually, all the anti-protons paired off with the protons and annihilated, leaving a remainder of 1 proton per 2 billion photons. So the question is not, "Why is there so much more matter than anti-matter?", the question is actually "Why does the Universe have this tiny preference for matter?"

   We do not actually expect that *all* matter and anti-matter would have paired off and annihilated. The Universe was expanding rapidly, and the distance between neighboring particles was growing rapidly; so we expect that some relic amount of matter and anti-matter would have frozen out and survived the annihilation phase. But this only accounts for a baryon density of $n_B/n_\gamma \sim 10^{-20}$ [4], or roughly enough for one galaxy in the observable Universe — not the 10 billion that we see.

   One possibility is that it is only our local Universe that is an upward fluctuation in matter density. Perhaps some other portion of the Universe is dominated by anti-matter in such a way that the average baryon asymmetry is much smaller than it is in our local neighborhood. While we cannot fully rule this out, we have reasonably solid evidence that this is not the case. By looking for the annihilation that would inevitably occur at the boundaries between matter and anti-matter dominated regions we can place bounds on this possibility within our *observable* Universe [5, 6].

   More commonly, we believe that the Universe began in a state with no baryon asymmetry — anti-matter and matter on equal footing — and evolved a baryon asymmetry through a
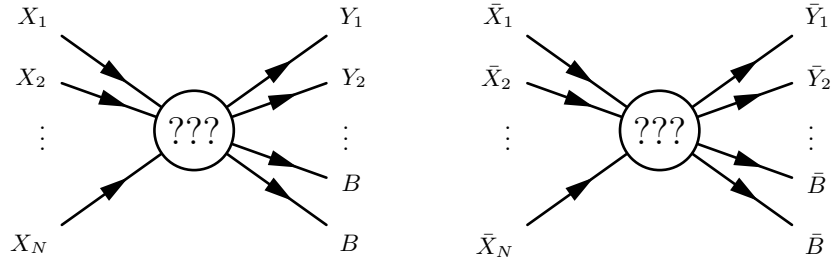
Figure 1.1: A schematic of a generic process which produces 2 baryons, BB, (*left*) and the CP conjugate process that produces 2 anti-baryons, $\bar{B}\bar{B}$, (*right*). The Sakharov conditions say that in order to produce an excess of baryons, such processes must exist and the process on the left must outpace the process on the right.

process known as Baryogenesis. There are many models for how Baryogenesis could occur, but any model must satisfy three conditions known as the Sakharov conditions [7]:

1. Any model must have a mechanism that violates baryon number symmetry.

2. That mechanism must also violate C-symmetry and CP-symmetry.

3. The mechanism must be active while the Universe was not in thermal equilibrium.

The first of these three conditions is easily understood: if we want to produce an excess of baryons over anti-baryons, we must have a way of creating baryons without also creating anti-baryons. But the presence of such a mechanism would also imply that the analogous process for creating anti-baryons without baryons is also possible (assuming CPT is conserved). The second condition (that C and CP-symmetry be violated) just implies that the process must happen faster for baryons than anti-baryons, otherwise the two processes will cancel each other out. The third condition is slightly less obvious, but boils down to the fact that in thermal equilibrium, the process for creating baryons comes into equilibrium with the process destroying them, so that the net ratio of baryons to anti-baryons cannot change; we thus need to break thermal equilibrium in order to slow the rate of destroying baryons compared to the rate of creating them. Together, these three conditions specify a minimum set of requirements in order to produce an excess of baryons over anti-baryons in an evolving Universe.

Within the SM all three Sakharov conditions are actually met, and so a baryon asymmetry could have been generated in the early Universe *without* the need for any new physics beyond the SM physics. However, it turns out not to be enough. Specifically, the SM only has enough CP-violation to account for a baryon asymmetry of $|\eta_B| \lesssim 10^{-26}$ [8]. This is ~ 6 orders of magnitude smaller than the relic baryon density we expect to freeze out, $n_B/n_\gamma \sim 10^{-20}$. To

solve the Baryon Asymmetry Problem, we need physics beyond the SM.

Leptogenesis

There are many proposals which extend the SM to introduce a mechanism for Baryogenesis [9–13] (just to cite a few). A problem is that many of them introduce new physics at very high energies — so high as to be beyond the realm of experimental reach. So while these models of Baryogenesis can nicely explain the baryon asymmetry problem, testing their veracity is difficult.

One theory that has gained popularity recently is Baryogenesis via Leptogenesis [13]. In this approach, the baryon asymmetry is not generated directly by producing baryons (e.g. protons) over anti-baryons (e.g. anti-protons), but rather by producing an asymmetry in the leptons (e.g. electrons and neutrinos) and converting that into an asymmetry in the baryons.

While so far we have only discussed the matter asymmetry of the Universe in terms of the baryon asymmetry (e.g. the preference of protons over anti-protons), we also expect that there is an analogous lepton asymmetry (e.g. a preference for electrons over positrons). We typically focus on the baryons because we have cosmological constraints that allow us to say something about their asymmetry. The lepton asymmetry, on the other hand, is dominated by the Cosmic Neutrino Background (CνB) which has never been directly detected and could carry a very wide range of lepton asymmetry values[1]. Given that we see it in the baryons, it is natural to expect that the leptons carry a non-zero asymmetry as well, and thus it is tempting to link their origins to a common process in the early Universe.

In order for Leptogenesis to occur, the analogous Sakharov conditions must be satisfied. We need a process which violates lepton number symmetry. We need this process to violate C-symmetry and CP-symmetry. And we need a departure from thermal equilibrium. Neutrinos have long been thought to play a role in models of Leptogenesis, but the relatively recent discovery of neutrino mass has brought many of these questions within the reach of experiment. We shift our discussion to neutrinos and the role they could play in answering these questions.

1.2 Neutrinos and Neutrino Mass

When the SM was originally formulated, neutrinos were thought to be massless. They came in three flavors, ν_e, ν_μ, and ν_τ, with their associated anti-particles, $\bar{\nu}_e$, $\bar{\nu}_\mu$, and $\bar{\nu}_\tau$, and they were always associated with their corresponding charged leptons, e, μ, or τ.

This picture began to break down in the late 1960s with the Homestake Experiment run by Raymond Davis Jr. The experiment was located in the Homestake Mine in South Dakota and was a large tank of perchloroethylene (dry-cleaning fluid). The experiment detected

---

[1]Averaged over the whole Universe, we expect $\sim 1/4$ protons per cubic meter; for neutrinos we expect something more like $\sim 3 \times 10^8$ per cubic meter.
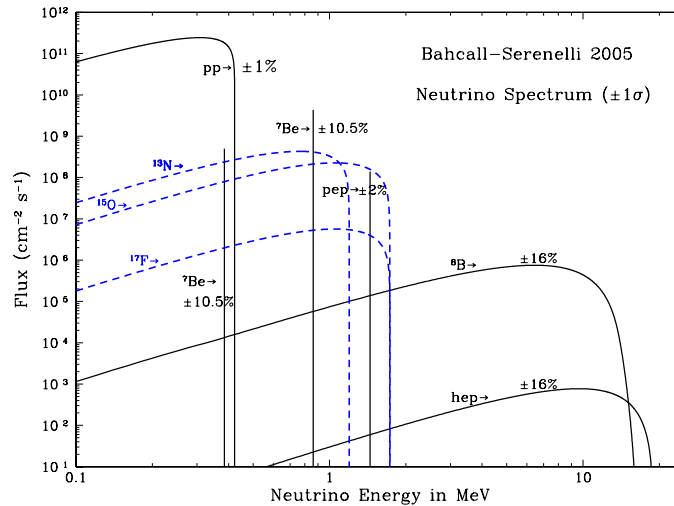
Figure 1.2: Predicted spectrum of solar neutrinos and their sources. (Figure from John Bahcall's website, `http://www.sns.ias.edu/~jnb/`.)

electron neutrinos, ν_e, — primarily from the sun — via the reaction

$$^{37}\text{Cl} + \nu_e \rightarrow \,^{37}\text{Ar} + e^-$$
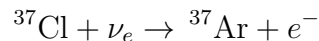
By collecting the buildup of ^{37}Ar in the tank at the end of every month, they could measure the solar electron neutrino flux and compare this number to the flux expected from solar modeling calculations being done by John Bahcall. Their measured flux was consistently low by a factor of ~ 3 [14]. This became known as the "Solar Neutrino Problem".

At first, the discrepancy was blamed on experimental systematics or miscalculations in the solar modeling. However, other solar measurements based on solar acoustics supported the solar models, and when other neutrino experiments began to see similar discrepancies in the flux of solar ν_e [15–21], it became increasingly clear that the problem lay with our understanding of neutrinos. By 2002, the Sudbury Neutrino Observatory (SNO) had been able to measure the total neutrino flux (not just the electron neutrino flux). They showed that while the electron neutrino flux was low by a factor of ~ 3, the flux of ν_μ and ν_τ made up for the deficit [22, 23]. This solved the Solar Neutrino Problem by showing that the missing ν_e were "oscillating" into ν_μ and ν_τ between the sun and the detector! By the end of 2002, the KamLAND experiment published its results on nuclear reactor $\bar{\nu}_e$ disappearance [24, 25], which, when combined with the solar neutrino results, indicated a unique solution to the Solar Neutrino Problem: neutrinos have mass. Since then, many experiments have confirmed and refined this understanding and built a very robust picture of massive neutrinos (see [26] for a review).
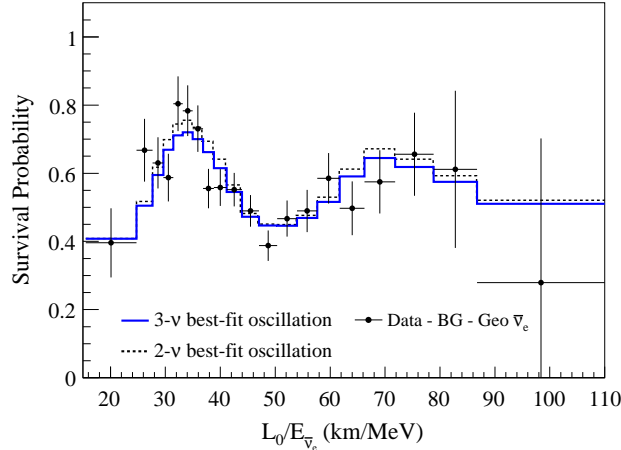
Figure 1.3: The $\bar{\nu}_e$ survival probability measured in the KamLAND detector. (Figure from [25].)

Neutrino Mass and Leptogenesis

The discovery of neutrino oscillations opens up two very fundamental questions about neutrinos. The first is in regards to their mass scale and its origin. The second question regards the extent of their oscillations and ties neutrinos back into the discussion about the Baryon Asymmetry Problem.

It is natural to ask why the neutrinos were massless in the SM to begin with. There are several answers to this question, but the simplest is that neutrinos are a factor of about a million to a trillion times lighter than the other particles in the SM. Thus in almost every experiment they appear massless[2]. Even today we know that the neutrinos have mass, but we have not yet been able to accurately measure that mass.

This massive discrepancy between the charged fermions and the neutrinos in the SM is made clear in Fig. 1.4 and its cause is still an open question in particle physics. There are many theoretical models which seek to explain the mass scale of the neutrinos via a variety of different mechanisms. Indeed, the small mass of the neutrinos may very well be a window into some new physics.

So far, we have discovered neutrino oscillations in the context of conversion between flavors in neutrinos and anti-neutrinos separately

$$\nu_\alpha \leftrightarrow \nu_\beta \qquad\qquad \bar{\nu}_\alpha \leftrightarrow \bar{\nu}_\beta$$

[2]For reference, measuring the mass of a neutrino, $m_\nu \lesssim 0.1$ eV, relative to the mass of the next lightest particle, the electron, $m_e = 511$ keV, is analogous to measuring the mass of person by measuring the change in weight of the Empire State Building as they walk in the door.
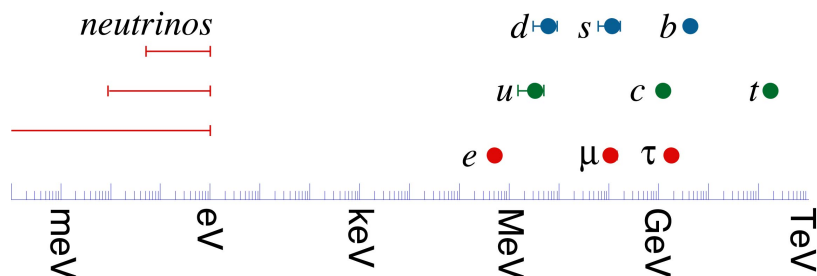
Figure 1.4: The masses of the fermions in the SM. While the charged fermions have masses on the order of MeV to 100 GeV, the neutrinos are some 6 to 12 orders of magnitude lighter. Why this is the case is one of the open questions of particle physics. (Image from H. Murayama, `http://hitoshi.berkeley.edu/`.)

where α, β represent the various neutrino flavors, (e, μ, τ). However, this discovery opens up the prospect for oscillations of the form

$$\nu_\alpha \leftrightarrow \bar{\nu}_\beta$$

This type of oscillation is not allowed for any other particle in the SM (referring to the fermions), and this can be easily understood since every other particle in the SM carries an electric charge. An oscillation of the form $e^- \to e^+$ would violate charge conservation. However, neutrinos carry no electric charge, and thus a conversion from a neutrino to an anti-neutrino (or vice versa) is not a priori forbidden. This would make neutrinos unique among the SM particles as they could convert into their own anti-particles.

This prospect is of interest to the discussion of Leptogenesis because this type of oscillation manifestly violates lepton number — i.e. it carries a lepton into an anti-lepton or vice versa. If this type of oscillation is possible — and discovered — it would prove that lepton number is not a conserved quantity in nature. This would provide a possible mechanism for Leptogenesis in the early Universe and could help explain the baryon asymmetry.

Neutrinos can even provide the CP-violation necessary for Leptogenesis. If the oscillation probability for the (lepton number-conserving) process $\nu_\alpha \to \nu_\beta$ differs from its CP-conjugate process $\nu_\alpha \to \nu_\beta$, this would violate CP-symmetry. Further, if lepton number-violating oscillations are also possible, there could be *additional* sources of CP-violation in the oscillations between neutrinos and anti-neutrinos.

What makes these questions so presently interesting — and relevant to the discussion of the baryon asymmetry — is that the discovery of neutrino mass has shown us that many of them may be within experimental reach. This gives us a path to probe questions about physics beyond the SM and about the processes that occurred in the very early Universe by carefully studying the properties of the weakest and lightest known particles.

1.3   Neutrinoless Double Beta Decay

In this introduction we have briefly introduced a major open question in particle physics today: the origin of the Baryon Asymmetry of the Universe. We then outlined some of the questions that can be experimentally probed that might help address this issue: lepton number conservation, absolute neutrino mass scale, and CP-violation. This thesis focuses on an experiment that addresses the question of lepton number violation — specifically whether or not the conversion between $\nu \leftrightarrow \bar{\nu}$ is possible.

The CUORE experiment searches for a process called Neutrinoless Double Beta Decay. This is a hypothesized nuclear decay mode that violates lepton number and would indicate that the conversion from $\nu \leftrightarrow \bar{\nu}$ is indeed possible. Whereas oscillation experiments look for neutrinos of one flavor produced in one location to oscillate to another flavor when they reach a detector in a different location, in neutrinoless double beta decay searches we can think of the oscillation $\nu \leftrightarrow \bar{\nu}$ as occurring inside a single decaying nucleus. So rather than producing and detecting neutrinos, neutrinoless double beta decay experiments simply observe a collection of decaying atoms and search for the signature of the lepton number-violating decay.

This decay, if possible, is one of the slowest known processes. Given the present experimental limits, in order to see a single atom undergo this decay one would expect to wait longer than 10 trillion trillion years — or a million billion times the age of the Universe. Thus all experiments that search for it share several common themes. They are large, in order to simultaneously observe as many atoms as possible. They are clean, in order to get rid of any other sources of radioactive decay. And they are underground, in order to shield from the cosmic rays that could swamp the detector.

Chapter 2

Double Beta Decay

In the previous chapter we outlined some of the successes and shortcomings of the SM and motivated why we are interested in studying neutrino physics. In this chapter, we focus on the process of neutrinoless double beta ($0\nu\beta\beta$) Decay. We begin by grounding some of the neutrino physics from the previous chapter in a mathematical framework. We lay out an extension to the SM that predicts $0\nu\beta\beta$ decay, and connect this model to the experimentally measurable decay rate. This model will serve as the baseline model for the whole of this thesis. But to emphasize that the baseline model is just one of many possible, we briefly outline several alternative mechanisms by which $0\nu\beta\beta$ decay could occur. Finally, we briefly outline the present experimental status of the search for $0\nu\beta\beta$ decay.

2.1   Neutrinos in the Standard Model

The SM Lagrangian contains three massless left-handed neutrinos, ν_e, ν_μ, ν_τ and three associated charged particles e, μ, τ. We define neutrinos such that ν_e are produced with e, ν_μ with μ, and ν_τ with τ. We expect that we should never see a ν_μ produced with a τ. This is made concrete by saying that each neutrino, ν_α, along with its charged partner, ℓ_α, obey a U(1) symmetry which leaves $\mathcal{L}_{\text{SM}}$ unchanged. For a set of arbitrary angles, θ_α, the transformation

$$\begin{pmatrix} \nu_\alpha \\ \ell_{L\alpha} \end{pmatrix} \rightarrow e^{i\theta_\alpha} \begin{pmatrix} \nu_\alpha \\ \ell_{L\alpha} \end{pmatrix} \qquad\qquad \ell_{R\alpha} \rightarrow e^{i\theta_\alpha} \ell_{R\alpha}$$

has no effect on the theory. Here, the Greek indices run over e, μ, τ and indicate that each flavor can transform independently. The conserved charge associated with this symmetry is called lepton flavor. In other words, in every interaction, the SM predicts that the number of each flavor lepton is separately conserved, N_e, N_μ and N_τ. So, if an electron is produced in a reaction, then a $\bar{\nu}_e$ is produced to zero the resulting change in lepton flavor number.

This picture was shattered by the discovery of neutrino oscillations, which definitively demonstrated the conversion of neutrinos among the flavor states $\nu_\alpha \leftrightarrow \nu_\beta$ and that lepton

flavor is not, in fact, conserved. The picture that emerged was that of three neutrino flavor eigenstates that are a superposition of at least three neutrino mass eigenstates.

Neutrino Mixing

In a simplified two flavor mixing model we can write the neutrino flavor eigenstates, ν_α and ν_β, as a linear combination of neutrino mass eigenstates, ν_1 and ν_2, with masses m_1 and m_2:

$$\begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \tag{2.1}$$

This means that neutrinos, which are produced in the flavor eigenstates that we defined, actually propagate in a different set of states. This means that the linear combination of mass states $\cos(\theta)\nu_1 + \sin(\theta)\nu_2$ can couple to the charged lepton ℓ_α but *not* to ℓ_β.

Because the neutrinos are not produced in an exact energy eigenstate, their wavefunction evolves, and the relative phases between the mass states change with time. In the neutrino rest frame, the time evolution of the state is given by

$$|\psi(t)\rangle = \cos(\theta)e^{im_1 t}|\nu_1\rangle + \sin(\theta)e^{im_2 t}|\nu_2\rangle$$

We can define the "survival probability" as the probability that a neutrino produced in one location in a flavor eigenstate will interact in another location in the same flavor eigenstate. For an α neutrino of energy E, having traveled a distance L, the survival probability is given by

$$P_{\alpha\alpha} = 1 - \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2 L}{4E}\right) \tag{2.2}$$

where $\Delta m^2 \equiv m_2^2 - m_1^2$. By measuring this survival probability as a function of L/E, experiments are able to determine the mass squared splittings Δm^2 and mixing angles.

Expanding to a generic mixing model with N mass eigenstates, we write the relation between the three flavor eigenstates, ν_α, and N mass eigenstates, ν_i, as

$$|\nu_\alpha\rangle = \sum_{i=1}^{N} U_{\alpha i}|\nu_i\rangle \tag{2.3}$$

where U is a 3×N unitary matrix. Presently, most data are consistent with $N = 3$, however, there are some tentative indications for $N > 3$ [19, 27–33].

In an $N = 3$ model the mixing matrix U, called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, can be parameterized in the standard way:

$$\begin{aligned} U = & \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{\text{CP}}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{\text{CP}}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{\text{CP}}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{\text{CP}}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{\text{CP}}} & c_{23}c_{13} \end{pmatrix} \\ & \times \begin{pmatrix} e^{i\alpha_1/2} & 0 & 0 \\ 0 & e^{i\alpha_2/2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned} \tag{2.4}$$

Here we use the shorthand s_{ij} and c_{ij} for $\sin(\theta_{ij})$ and $\cos(\theta_{ij})$ respectively. This model is parameterized by three angles θ_{12}, θ_{23} and θ_{13}, three CP-violating phases δ_{CP}, α_1 and α_2 and three neutrino masses, m_1, m_2 and m_3. All of these parameters are, in principle, measurable quantities[1]. The best-fit parameters from a global fit to all currently running oscillation experiments are listed in Table 2.1. Since the survival probabilities are only sensitive to mass-squared differences, not the masses themselves, the neutrino mass scale is currently unknown. We know the sign of Δm_{12}^2 from solar measurements, but the sign of Δm_{13}^2 is presently unknown. There are two possibilities for the spectrum of neutrino masses: the normal hierarchy, $m_1 < m_2 < m_3$, and the inverted hierarchy, $m_3 < m_1 < m_2$, as shown in Fig. 2.1.

The remaining parameters: the absolute neutrino mass scale, the neutrino mass hierarchy, and the CP violating phases are all currently the focus of intense interest. However, there is another another question that is opened by the discovery of neutrino mixing.

From the point of view of fundamental symmetries, the discovery of neutrino oscillations reduced the $U(1)_e \times U(1)_\mu \times U(1)_\tau$ flavor symmetries of the SM to a single overall $U(1)$ symmetry that we call Lepton number, L:

$$
\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} \to e^{i\theta} \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} \qquad\qquad e_R \to e^{i\theta} e_R
$$
$$
\begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix} \to e^{i\theta} \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix} \qquad\qquad \mu_R \to e^{i\theta} \mu_R \qquad\qquad (2.5)
$$
$$
\begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \to e^{i\theta} \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \qquad\qquad \tau_R \to e^{i\theta} \tau_R
$$

Notice that in Eqn. 2.1, we have multiple transformation angles θ_α, whereas here all particles transform together. This symmetry implies that even with neutrino oscillations, the *total* number of leptons is conserved. In other words, N_e, N_μ and N_τ may change individually, but $N_e + N_\mu + N_\tau$ should remain unchanged. This lepton number symmetry is similar to the baryon number symmetry in the quark sector. Establishing if these are indeed fundamental symmetries of nature is one key to understanding the origin of the Baryon Asymmetry of the Universe.

Fermion Masses in the Standard Model

Knowing that neutrinos are not massless, we must find a way to accommodate this in the SM. We begin with the standard Higgs mechanism. All massive particles in the SM acquire their mass through an interaction with the Higgs field during Electroweak Symmetry Breaking

---

[1]The CP-violating Majorana phases are, at least in theory, measurable quantities. Oscillation experiments that search for $\nu_\alpha \leftrightarrow \bar{\nu}_\beta$ would be sensitive to α_1 and α_2 through the oscillation asymmetry $(P(\nu_\alpha \to \bar{\nu}_\beta) - P(\bar{\nu}_\alpha \to \nu_\beta))/(P(\nu_\alpha \to \bar{\nu}_\beta) + P(\bar{\nu}_\alpha \to \nu_\beta))$ [34]. But these signals are typically suppressed by several orders of magnitude relative to typical backgrounds. More realistic approaches are sensitive to $\alpha_1 - \alpha_2$ for a narrow range of parameter spaces [35].
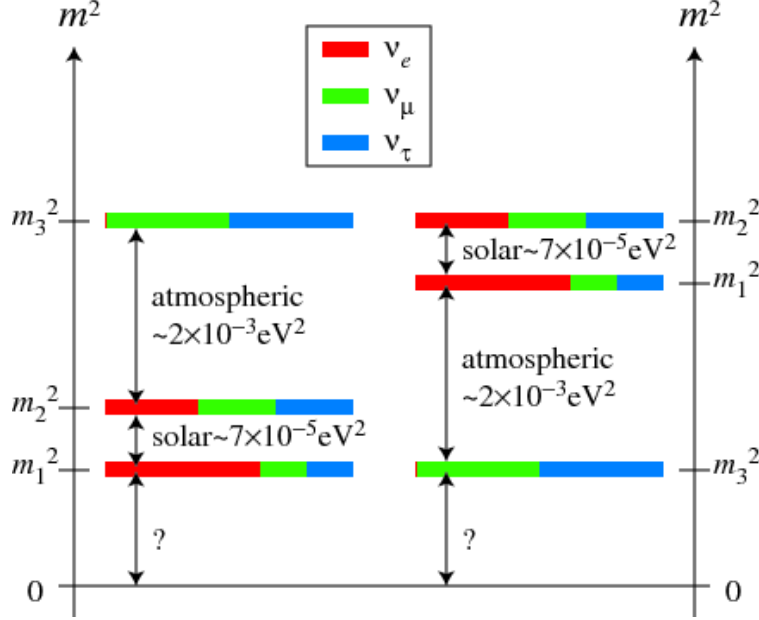
Figure 2.1: Two possible mass orderings for the neutrino masses. The "Normal Hierarchy" on the left or the "Inverted Hierarchy" on the right. The unknown size of the splitting between the lightest mass state and zero represents the unknown absolute mass scale of the neutrino. (Figure from [36].)

Table 2.1: Best-fit and 3σ intervals for neutrino mixing parameters from a global fit to oscillation data [37]. Table taken from [26]. Values are evaluated for normal hierarchy, while numbers in parentheses are evaluated for inverted hierarchy. The value Δm^2 is the splitting between m_3 and the average of m_1 and m_2 and is given by $\Delta m^2 \equiv m_3^2 - (m_2^2 + m_1^2)/2$.

Parameter	Best Fit ($\pm 1\sigma$)	3σ Interval
Δm_{12}^2 [10^{-5} eV^2]	$7.54^{+0.26}_{-0.22}$	$(6.99 - 8.18)$
$\|\Delta m^2\|$ [10^{-3} eV^2]	2.43 ± 0.06 (2.38 ± 0.06)	$2.23 - 2.61(2.19 - 2.56)$
$\sin^2\theta_{12}$	0.32 ± 0.017	$0.259 - 0.359$
$\sin^2\theta_{23}$	$0.437^{+0.033}_{-0.023}(0.455^{+0.039}_{-0.031})$	$0.374 - 0.628(0.380 - 0.641)$
$\sin^2\theta_{13}$	$0.0234^{+0.0020}_{-0.0019}(0.0240^{+0.0019}_{-0.0022})$	$0.0176 - 0.0295(0.0178 - 0.0298)$
δ/π (2σ range)	$1.39^{+0.38}_{-0.27}(1.31^{0.29}_{-0.33})$	$(0.00 - 0.16) \oplus (0.86 - 2.00)$
		$((0.00 - 0.02) \oplus (0.70 - 2.00))$

(EWSB). For example, the charged leptons acquire their mass via a Yukawa term in the Lagrangian:

$$\mathcal{L}_Y = \lambda_\alpha \Phi^T \bar{L}_\alpha \ell_{\alpha R} + \text{h.c.} \tag{2.6}$$

where $\Phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ is the complex scalar Higgs doublet. $L_\alpha \equiv \begin{pmatrix} \nu_{\alpha L} \\ \ell_{\alpha L} \end{pmatrix}$ is the left-handed doublet of the charged lepton, ℓ, and its corresponding neutrino, $\nu_{\alpha L}$. When the Higgs field gains a non-zero expectation value, $\langle \Phi \rangle = \begin{pmatrix} 0 \\ v \end{pmatrix}$, this term becomes

$$\mathcal{L}_{\text{mass},\ell} = m_\ell \bar{\ell}_L \ell_R + \text{h.c.} \tag{2.7}$$

where $m_\ell \equiv \lambda_\ell v$. This type of mass term is referred to as a Dirac-type mass, because it couples a left-handed particle field, ℓ_L, to a right-handed field, ℓ_R. A term of this type exists for all the charged fermions.

Within the SM, the neutrinos are massless because they do not have a right-handed component field, ν_R. They are all, by definition, left-handed fields and thus cannot have a Dirac-type mass. Thus, the neutrino oscillation experiments present us with a conundrum, "How do we extend the Standard Model to accommodate neutrino mass?"

The problem is more complicated than just introducing neutrino mass into the SM. As we saw before, not only do neutrinos have mass, but they are significantly lighter than all the other fermions in the SM. So besides just giving the neutrinos mass we also seek to explain why they are so light. One possibility is that they just *are*. However, this is not a very satisfactory answer, and so we search for another.

Type I Seesaw Mechanism

The most straightforward approach is to add right-handed components for the neutrinos, N_{Ri}, and add new terms to the SM Lagrangian. These right-handed neutrino fields are not already present in the SM because they have zero charge under all of the SM interactions, thus cannot be directly seen in any possible experiment. These neutrinos are said to be "sterile". Using these new right-handed neutrino fields, N_{Ri}, we can now write two new mass terms into the Lagrangian. For three generations of left-handed neutrinos and N generations of right-handed neutrinos, we write

$$\mathcal{L}_{\text{mass},\nu} = -\lambda^D_{\alpha i} \bar{L}^T_\alpha \tilde{\Phi} N_{Ri} - \frac{1}{2} M^R_{ij} \bar{N}_{Ri} N^c_{Rj} + \text{h.c.} \tag{2.8}$$

This first term is analogous to the other fermion mass terms in Eqn. 2.6; where $\lambda^D_{i\alpha}$ are Yukawa coupling constants and $\alpha = e, \mu, \tau$, and i, j run over the N hypothetical right-handed neutrino states. After EWSB, the first term leads to a standard Dirac type mass term as in Eqn. 2.7. The second term in the above equation is a Majorana type mass term. Rather than coupling a left-handed field to a right-handed field, a Majorana type mass term couples a right-handed field, N_R, to its left-handed charge conjugate field, $N^c_R \equiv C\bar{N}^T_R$ — connecting a particle to its own anti-particle. This is only allowed for N_R because these particle fields are completely uncharged in the SM, and so the Majorana mass term carries no SM charges.

After EWSB, the Lagrangian becomes

$$
\begin{aligned}
\mathcal{L}_{\text{mass},\nu} &= -\bar{\nu}_{\alpha L} M^D_{\alpha i} N_{Ri} - \tfrac{1}{2} \bar{N}_{Ri} M^R_{ij} N^c_{Rj} + \text{h.c.} \\
&= -\tfrac{1}{2} \begin{pmatrix} \bar{\boldsymbol{\nu}}^{\boldsymbol{c}}_{\boldsymbol{L}} & \bar{\boldsymbol{N}}_{\boldsymbol{R}} \end{pmatrix} \begin{pmatrix} 0 & \boldsymbol{M_D} \\ \boldsymbol{M_D}^T & \boldsymbol{M_R} \end{pmatrix} \begin{pmatrix} \boldsymbol{\nu_L} \\ \boldsymbol{N^c_R} \end{pmatrix} + \text{h.c.}
\end{aligned}
\tag{2.9}
$$

Here we have defined $(M_D)_{\alpha i} \equiv \lambda_{\alpha i} v$ as a $3 \times N$ matrix of Dirac masses and rewritten the equation in matrix form. The zero in this mass matrix may seem strange, but is a direct result of the fact that the left-handed neutrino carries a weak charge whereas N_R is uncharged. There are various extensions to the SM and higher dimension operators that will generate non-zero values but, for simplicity, we ignore them here.

We now make a very particular assumption: we assume that $M_D \ll M_R$. This choice is somewhat ad hoc, but we justify this by citing the origin of these mass terms. The Dirac mass term, M_D, was generated during EWSB by the usual Higgs mechanism, so it's reasonable to assume that this term is typical of the Higgs mass scale. But the M_R terms came *before* EWSB — they were already present as mass terms in Eqn. 2.8. We therefore assume that this term is generated independently of the Higgs mechanism, and so we might think that it is representative of a mass scale significantly larger than the EWSB scale.

We now rewrite this equation in terms of a diagonalized mass matrix:

$$
\begin{aligned}
\mathcal{L}_{\text{mass},\nu} &= -\tfrac{1}{2} \begin{pmatrix} \bar{\boldsymbol{\nu}}^{\boldsymbol{c}}_{\boldsymbol{L}} & \bar{\boldsymbol{N}}_{\boldsymbol{R}} \end{pmatrix} U \left[\begin{pmatrix} \boldsymbol{M_D M_R}^{-1} \boldsymbol{M_D}^T & 0 \\ 0 & \boldsymbol{M_R} \end{pmatrix} + O\left(M_D^4/M_R^3\right) \right] U^\dagger \begin{pmatrix} \boldsymbol{\nu_L} \\ \boldsymbol{N^c_R} \end{pmatrix} \\
&\quad + \text{h.c.} \\
&\approx -\tfrac{1}{2} \begin{pmatrix} \bar{\boldsymbol{\nu}}^{\boldsymbol{c}\prime}_{\boldsymbol{L}} & \bar{\boldsymbol{N}}'_{\boldsymbol{R}} \end{pmatrix} \begin{pmatrix} \boldsymbol{m_\nu} & 0 \\ 0 & \boldsymbol{M_N} \end{pmatrix} \begin{pmatrix} \boldsymbol{\nu'_L} \\ \boldsymbol{N^{c\prime}_R} \end{pmatrix} + \text{h.c.}
\end{aligned}
\tag{2.10}
$$

The mass matrix at the center of this equation describes a set of 3 light Majorana neutrino mass eigenstates, $\boldsymbol{\nu'_L}$ with masses m_ν and N heavy Majorana neutrino mass eigenstates, $\boldsymbol{N'_R}$, with masses M_N, where

$$
\boldsymbol{m_\nu} \equiv \boldsymbol{M_D M_R}^{-1} \boldsymbol{M_D}^T \qquad\qquad \boldsymbol{M_N} \equiv \boldsymbol{M_R}
\tag{2.11}
$$

This introduces a set of light neutrinos whose masses are suppressed by $M_D/M_R \ll 1$ and a set of heavy neutrinos with mass scale M_R. This is an extremely powerful observation. The two terms that we introduced to the SM Lagrangian not only generate a mass term for the neutrinos, but also explain why the neutrino masses are so small! The Higgs mass components *are* naturally the same scale as the other fermions but the neutrino masses are *suppressed* by some larger energy scale theory. This is referred to as a Type I seesaw mechanism.

Looking at the right half of equation 2.10, we can identify the mass eigenstates:

$$
\begin{pmatrix} \boldsymbol{\nu'_L} \\ \boldsymbol{N^{c\prime}_R} \end{pmatrix} = U^\dagger \begin{pmatrix} \boldsymbol{\nu_L} \\ \boldsymbol{N^c_R} \end{pmatrix}
\tag{2.12}
$$

Here, U is the full $(3 + N) \times (3 + N)$ unitary matrix that rotates between the flavor basis and the mass basis. The portion of U that rotates between $\boldsymbol{\nu_L}$ and $\boldsymbol{\nu'_L}$ is exactly the 3×3 matrix that we previously called the PMNS matrix.

An important aspect of this theory is that the Lagrangian that we introduced in Eqn. 2.8 does not conserve any U(1) symmetry. Specifically, the lepton number symmetry that we wrote in Eqn. 2.5 is violated if $M_R \neq 0$. Conversely, if we let $M_R = 0$ then the symmetry is restored. We could have introduced only the Dirac mass terms, which are sufficient to explain neutrino oscillations alone. But once we do this, we have no physical reason to exclude an M_R term. In this sense, the theory which generates Majorana masses for the neutrinos and violates lepton number is considered more "natural" than manually setting M_R to zero to conserve lepton number.

An alternate way to look at this is that the N_R neutrinos are almost completely decoupled from the SM, and we can generate a Majorana mass M_R term for them at a very high energy scale. If indeed $M_R = 0$ then that means that it is either tuned to be exactly zero, or that lepton number symmetry is also a symmetry of that higher-energy theory. But we have reason to suspect that lepton number is not a fundamental symmetry of nature — the Baryon Asymmetry — and thus it seems unnatural for us to insist that $M_R = 0$.

The model described here is one of the simplest of *many* methods by which we can introduce Majorana neutrinos. However, most models follow a similar recipe and many of the conclusions are general. Many models can both introduce Majorana mass and explain the smallness of the neutrino masses. A Majorana neutrino mass, by any mechanism, violates lepton number; in a very basic sense it allows the conversion of a neutrino into an anti-neutrino, which manifestly violates lepton number by two units. In the next section we introduce a specific mechanism by which we can probe a Majorana neutrino mass and we connect it back to a measurement of the Majorana neutrino mass term, m_ν.

2.2 $2\nu\beta\beta$ and $0\nu\beta\beta$

Double beta decay is a second-order weak decay where a nucleus, (Z, A), undergoes two beta decays to its isobar $(Z+2, A)$ in a single step, emitting two electrons in the process. In theory, this process can happen via two different channels: the SM-allowed lepton number conserving two neutrino double beta decay channel, or the lepton-number violating neutrinoless double beta decay channel.

The two-neutrino double beta decay ($2\nu\beta\beta$) channel conserves lepton number by also producing two electron anti-neutrinos in the decay

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

This process was first proposed in 1935 by Maria Goeppert-Mayer [38]. Within the SM all nuclei can undergo $2\nu\beta\beta$ so long as the decay is kinematically allowed. However, because it is a second-order weak process it is usually heavily suppressed relative to the first-order single beta decay to $(Z + 1, A)$ which significantly outpaces any $2\nu\beta\beta$ signal. Because of this, the

process has only been measured in a handful of nuclei. Typically, these nuclei are even-even nuclei where the odd-odd $(Z + 1, A)$ nucleus has a larger mass and the single beta decay along the isobar is thus kinematically forbidden. An example for the $A = 130$ isobar is shown in Fig. 2.2. In this isobar, the nucleus ^{130}Te is unstable and eventually decays to ^{130}Xe, but cannot decay through the heavier nucleus, ^{130}I, and so must decay directly via a $\beta\beta$ decay. The nuclei ^{48}Ca and ^{96}Zr present an alternative single beta decay suppression mechanism. In these isotopes, the single beta decay is kinematically allowed, but is a sixth-order forbidden decay due to a large change in angular momentum.

Because of the heavy suppression, the half-lives for these decays are in the range 10^{19}–10^{24} yr and are actually the slowest processes ever measured — indeed, the $2\nu\beta\beta$ half-lives of ^{128}Te and ^{136}Xe are the longest half-lives ever measured[2]. The measured half-lives of various isotopes are listed in Table 2.2.

Table 2.2: Measured $2\nu\beta\beta$ half-lives for various isotopes. The ^{130}Te $2\nu\beta\beta$ half-life number comes from a 2011 paper, but is expected to be updated shortly after this thesis.

Isotope	$T_{1/2}^{2\nu\beta\beta}$ (yr)	Method	Citation
^{48}Ca	$4.4^{+0.6}_{-0.5} \times 10^{19}$	Direct	[40]
^{76}Ge	$(1.5 \pm 0.1) \times 10^{21}$	Direct	[40]
^{82}Se	$(9.2 \pm 0.7) \times 10^{19}$	Direct	[40]
^{96}Zr	$(2.3 \pm 0.2) \times 10^{19}$	Direct	[40]
^{100}Mo	$(7.1 \pm 0.4) \times 10^{18}$	Direct	[40]
^{116}Cd	$(2.8 \pm 0.2) \times 10^{19}$	Direct	[40]
^{128}Te	$(1.9 \pm 0.4) \times 10^{24}$	Geochem	[40]
^{130}Te	$6.8^{+1.2}_{-1.1} \times 10^{20}$	Direct	[41]
^{136}Xe	$(2.165 \pm 0.016 \pm 0.059) \times 10^{21}$	Direct	[42]
^{150}Nd	$(8.2 \pm 0.9) \times 10^{18}$	Direct	[40]
^{238}U	$(2.0 \pm 0.6) \times 10^{21}$	Radiochem	[40]

Since the neutrinos rarely interact within the detector, their energy is practically never detected. So from an experimental standpoint, the measurable quantity in a double beta decay is typically the total kinetic energy of the electrons produced in the decay. This produces a broad energy spectrum that extends from 0 keV (when the neutrinos carry away all the energy) up to the endpoint energy (when the neutrinos are produced practically at rest). The spectrum of total electron kinetic energies, E_{Tot}, can be analytically derived using the Rosen-Primakoff approximation [43]:

$$\frac{dN}{dE_{\mathrm{Tot}}} \propto E_{\mathrm{Tot}}(Q - E_{\mathrm{Tot}})^5 (1 + 2E_{\mathrm{Tot}} + \frac{4E_{\mathrm{Tot}}^2}{3} + \frac{E_{\mathrm{Tot}}^3}{3} + \frac{E_{\mathrm{Tot}}^4}{30}) \tag{2.13}$$

where E_{Tot} is in units of MeV. This spectrum is shown in Fig. 2.3.

---

[2]The present limits on proton decay imply that should protons decay, they would have a significantly longer half-life. But, at present, only limits exist.
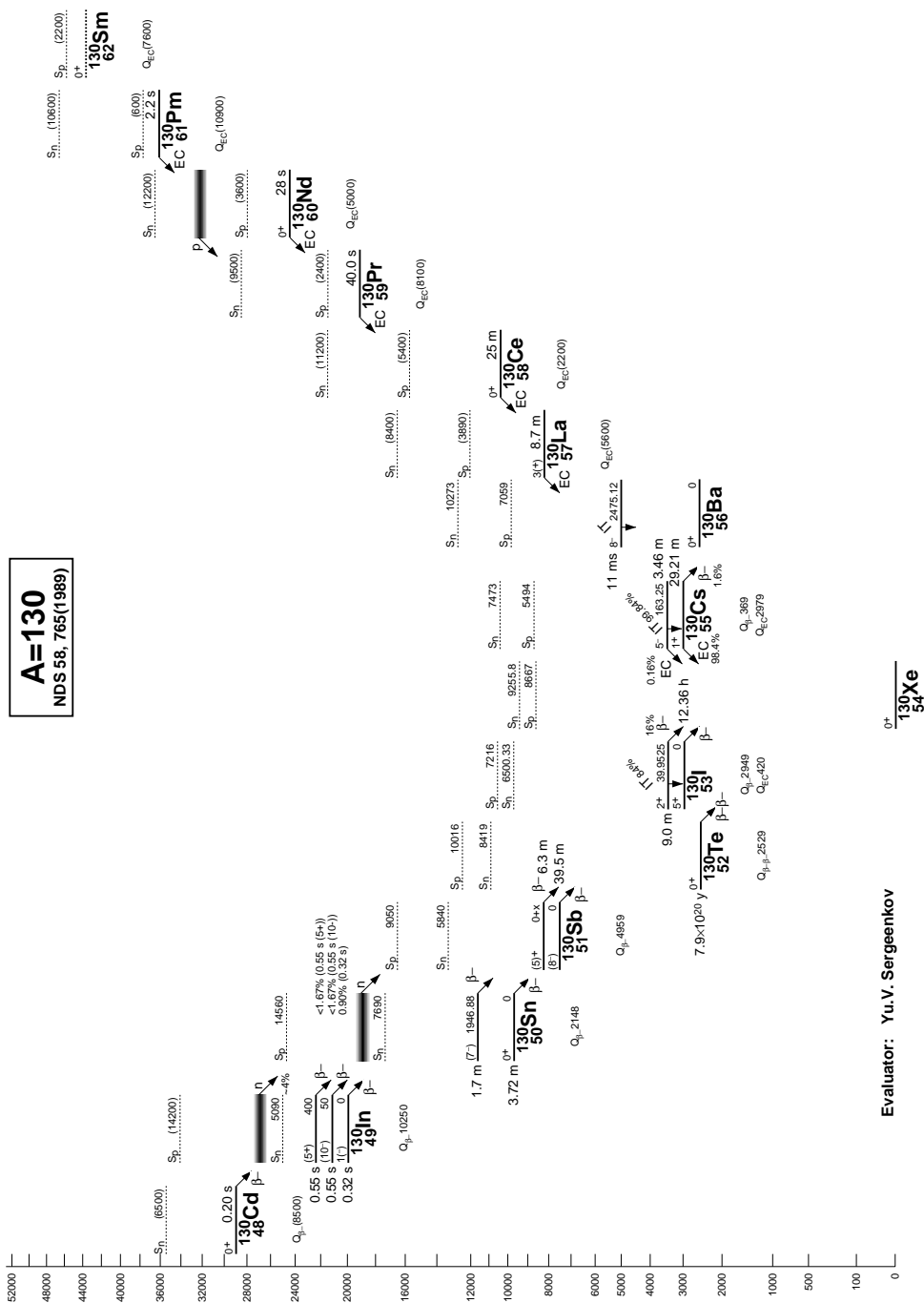
Figure 2.2: The $A = 130$ isobar. The $0\nu\beta\beta$ candidate isotope is ^{130}Te, which is kinematically forbidden from decaying to ^{130}I, but will slowly decay to ^{130}Xe. The nucleus ^{130}Ba could similarly undergo a $\beta^+\beta^+$ decay. Such a decay has never been observed. Figure taken from the Table of Radioactive Isotopes [39].
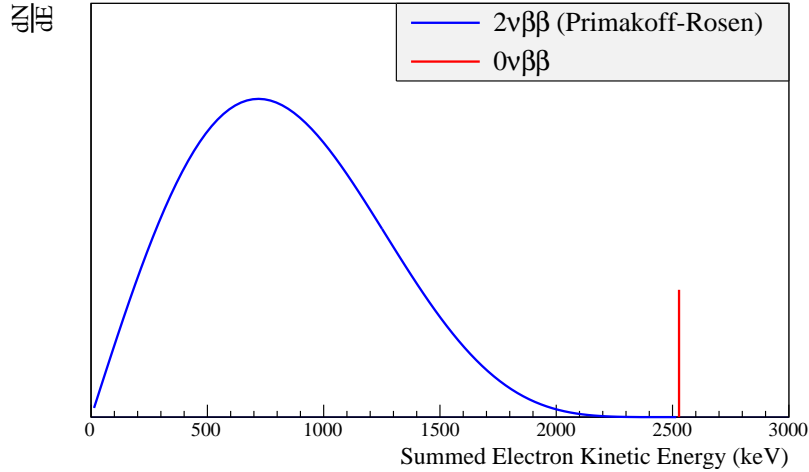
Figure 2.3: Schematic of $2\nu\beta\beta$ and $0\nu\beta\beta$ decay spectra. The height of the $0\nu\beta\beta$ spectrum is exaggerated for clarity. The $2\nu\beta\beta$ spectrum uses the Rosen-Primakoff Approximation.

The $0\nu\beta\beta$ Decay Mode

In 1939, building on the theories proposed by Majorana and Racah [44, 45], Furry pointed out that if indeed neutrinos are Majorana fermions (i.e. $\nu = \bar{\nu}$), then double beta decay could occur without the production of any neutrinos at all [46]. This is the so called neutrinoless double beta decay ($0\nu\beta\beta$) mode:

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

This decay is forbidden in the SM because it violates lepton number by two units. However, if neutrinos are indeed Majorana fermions then lepton number cannot be a symmetry of nature anyway. In fact, the search for this decay channel is presently one of the most sensitive methods to test whether or not lepton number is a conserved quantity of nature, and is also the only practical known method to answer the question of whether or not neutrinos are Majorana fermions. Since $0\nu\beta\beta$ produces no neutrinos to carry away any energy, its experimental signature is a monoenergetic line at the endpoint of the $2\nu\beta\beta$ spectrum.

There are many extensions to the SM which predict $0\nu\beta\beta$ decay via a variety of different mechanisms. In section 2.1, we added two terms to the SM Lagrangian that generated a light Majorana mass. Any mechanism which gives the neutrinos a Majorana mass is itself enough to allow for $0\nu\beta\beta$ decay (we discuss this further in the next section). However, some models predict the possibility of $0\nu\beta\beta$ decay *without* the need to insert a Majorana mass for the neutrinos (e.g. scalar triplet mechanism in Fig. 2.8). This seems to contradict the statement that $0\nu\beta\beta$ decay can only occur if neutrinos are Majorana fermions, however, the statement actually goes the other way: the occurrence of $0\nu\beta\beta$ *implies* that neutrinos are Majorana fermions.
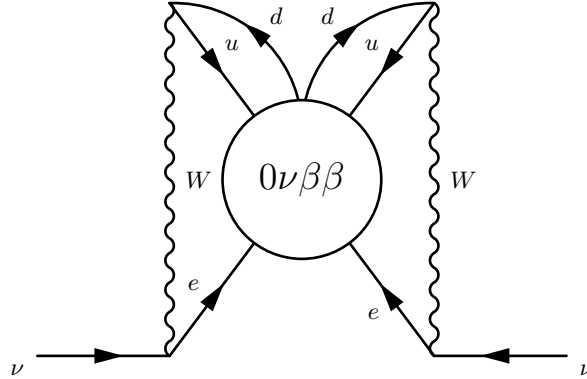
Figure 2.4: The butterfly diagram for generation of a Majorana neutrino mass via whatever mechanism is responsible for $0\nu\beta\beta$. This implies that the discovery of $0\nu\beta\beta$ implies a non-zero Majorana mass for the neutrino — though, it need not be the dominant mass.

Schechter and Valle pointed out that the discovery of $0\nu\beta\beta$ *by any mechanism* would imply that neutrinos can acquire a non-zero Majorana mass [47]. Their observation was that, even if a model does not explicitly give neutrinos a Majorana mass, one can be *generated* via a diagram like that in Fig. 2.4. This simply rearranges the typical $0\nu\beta\beta$ diagram placing all the unknown particle physics into a black box. What results is a modification to the neutrino propagator which converts a ν to *barν* — which is exactly a Majorana mass term. This should not be misunderstood as saying that the $0\nu\beta\beta$ mechanism is the dominant mechanism by which neutrinos generate mass, but rather that it is a minimal mechanism.

From an experimental standpoint, the quantity that we can actually measure is the $0\nu\beta\beta$ decay rate, $\Gamma_{0\nu\beta\beta}$. We can generalize all of the model-dependent particle physics into a generic parameterization of the decay rate. Following [48] we write this as

$$\Gamma_{0\nu\beta\beta} = \ln(2)G_{0\nu}(Q,Z)|\mathcal{M}_{0\nu}|^2|f_{0\nu}|^2 \tag{2.14}$$

This parameterization roughly splits the decay rate into a kinematics part in $G_{0\nu}(Q,Z)$, a nuclear physics part $|\mathcal{M}_{0\nu}|^2$, and all of the microscopic particle physics is contained in the dimensionless quantity, $|f_{0\nu}|^2$. In reality, the three terms are interconnected but can often be considered separately to a reasonable approximation.

Majorana Neutrino Exchange In Type I Seesaw

The neutrino mass model that we presented in section 2.1 predicts two classes of $0\nu\beta\beta$ decay: one mediated by the light Majorana neutrinos (with mass m_ν), and the other through the
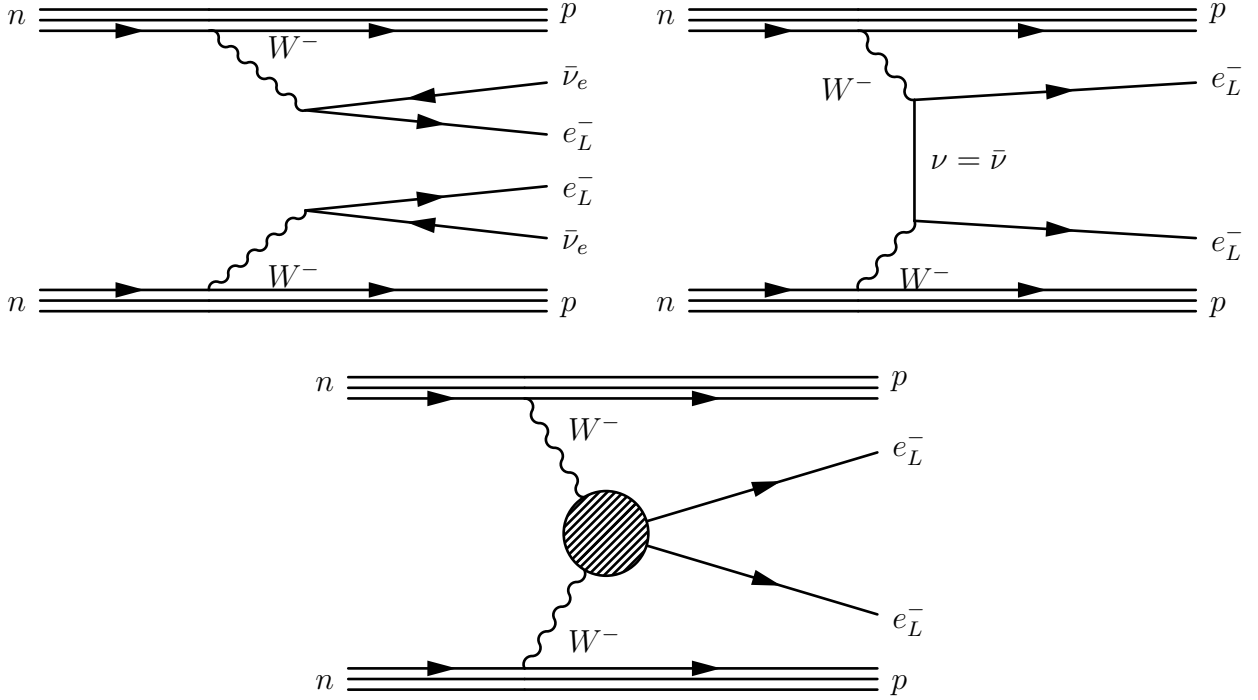
Figure 2.5: *Top Left:* Feynman diagram for $2\nu\beta\beta$. This is a Standard Model second order decay and has been observed in a handful of nuclei. *Top Right:* Diagram for $0\nu\beta\beta$ mediated by the exchange of a light Majorana neutrino ($\nu = \bar{\nu}$). This decay is not allowed in the SM and violates lepton number by $\Delta L = 2$. *Bottom:* Diagram of $0\nu\beta\beta$ with a heavy Majorana neutrino. Unlike the diagram with the light neutrino exchange, this mechanism is a "short range" point like interaction. This decay is also not allows in the SM.

heavy Majorana neutrinos (with mass M_ν). These two channels are shown in the Feynman diagrams in Fig. 2.5. We follow [49] and write the decay rate as

$$\Gamma_{0\nu\beta\beta} = \ln(2)G_{0\nu}(Q,Z)\left|\mathcal{M}_{0\nu}^{\text{Light}}\frac{m_{\beta\beta}}{m_e} + \mathcal{M}_{0\nu}^{\text{Heavy}}\frac{m_p}{M_{\beta\beta}}\right|^2 \tag{2.15}$$

where we define

$$m_{\beta\beta} \equiv \sum_{i\,\in\text{Light}} U_{ei}^2 m_i \qquad \frac{1}{M_{\beta\beta}} \equiv \sum_{j\,\in\text{Heavy}} \frac{U_{ej}^2}{m_j} \tag{2.16}$$

The two sums average separately over the 3 light neutrino mass states and the N heavy neutrino mass eigenstates that we found in Eqn. 2.10. U_{ei} and U_{ej} are the components of the mixing matrix. The two terms correspond to the two extremes that the neutrino masses are either much less than the momentum transfer in the nucleus, $m_{\text{Light}} \ll q \sim 100$ MeV, or

much larger, $m_\text{Heavy} \gg q$. Thus the form of $f_{0\nu}$ stems directly from the two limiting cases of the transition amplitude. This model can be thought of as a β decay at one vertex, releasing an Majorana anti-neutrino, $\bar\nu$, which is the same as a Majorana neutrino (i.e. $\bar\nu = \nu$), which is then absorbed as such at the second vertex.

It's common in the literature to consider only the light mass part of this equation. We follow this approach and take the exchange of light Majorana neutrinos as our baseline model. In this case, the parameter in which we are interested is referred to as the "Effective Majorana Mass", $m_{\beta\beta}$. We can evaluate the sum and write this term as

$$
\begin{aligned}
m_{\beta\beta} &= \left| \sum U_{ei}^2 m_i \right| \\
&= \left| \cos^2\theta_{12}\cos^2\theta_{13}e^{i\alpha_1}m_1 + \sin^2\theta_{12}\cos^2\theta_{13}e^{i\alpha_2}\sqrt{\Delta m_{12}^2 + m_1^2} \right. \\
&\quad \left. + \sin^2\theta_{13}e^{-2i\delta}\sqrt{m_1^2 + \Delta m_{13}^2} \right|
\end{aligned}
\tag{2.17}
$$

Throughout this thesis we will take this as the standard approach and ultimately place a limit on $m_{\beta\beta}$, but we stress that the quantity being measured is only the decay rate, $\Gamma_{0\nu\beta\beta}$; the interpretation is model dependent.

The allowed values of $m_{\beta\beta}$ based on oscillation data are shown in Fig. 2.9, expressed as a function of the lightest neutrino mass: m_1 in the Normal Hierarchy, and m_3 in the Inverted Hierarchy.

2.3   Phase Space and Nuclear Matrix Elements

The kinematic factor in Eqn. 2.15, $G_{0\nu}(Q, Z)$, describes the available phase space for the decay — i.e. the total number of ways the final-state particles can share the kinetic energy of the decay. In most models, this factor is only weakly dependent on the specifics of the decay and can be calculated with reasonably high precision. Generally, this factor scales as the fifth power of the Q-value of the decay[3].

The nuclear physics factors, $|\mathcal{M}_{0\nu}|^2$, are generally more complicated. This is the transition matrix element from the initial nuclear state to the final state. We write this as

$$
\mathcal{M}_{0\nu} \equiv \sum_{n,n'} \left\langle \psi_i \left| \tau_n^\dagger \tau_{n'}^\dagger H_{nn'}(q) \right| \psi_f \right\rangle
\tag{2.18}
$$

where

$$
H_{nn'}(q) \equiv \left[-h^F(q) + h^{GT}(q)\boldsymbol{\sigma}_n \cdot \boldsymbol{\sigma}_{n'} + h^T(q)\left(\boldsymbol{\sigma}_n \cdot \hat{\boldsymbol{q}}\boldsymbol{\sigma}_{n'} \cdot \hat{\boldsymbol{q}} - \boldsymbol{\sigma}_n \cdot \boldsymbol{\sigma}_{n'} \right) \right]
\tag{2.19}
$$

Here, $H_{nn'}(q)$ is the transition operator that converts two neutrons into two protons. The sum over n, n' is over *all* possible nucleon pairs in the nucleus. Different models of $0\nu\beta\beta$ decay

---

[3]This can be calculated easily by dimensional analysis: the decay produces three products, each of which is integrated over all possible momenta, d^3p; however, 4 of these integrals are fixed to conserve momentum and energy, leaving 5 integrals over momentum, each of which has a characteristic scale of Q, yielding Q^5.

predict different forms for $H_{nn'}(q)$, often relying on different simplifying assumptions. For instance, in the case of light Majorana neutrino exchange, the neutrino can carry momentum from one nucleon to another, thus creating a neutrino "potential" that must be included in the evaluation of $H_{nn'}(q)$; while in the case of heavy Majorana neutrino interaction, the interaction is point-like and the momentum transfer is negligible in the calculation and the neutrino potential reduces to a constant.

In this calculation we need to account for the energy levels of the intermediate nucleus. However, we typically use the closure approximation, which replaces our sum over intermediate state energies with a single *effective* energy ($\widetilde{A}$ in Fig. 2.6). This is generally considered a good approximation due to the high momentum transfer in the decay, $q \sim \hbar c/(1 \text{ fm}) \sim 100$ MeV, relative to the excitation energy of the nuclear states $E \sim$ a few MeV.

The matrix element can be thought of as two separate parts: a part that converts *any* two neutrons in the initial state nucleus into *any* two protons in the final state nucleus times an amplitude that *those* two neutrons would decay into *those* two protons in a $0\nu\beta\beta$ decay. We can write this schematically as

$$\mathcal{M}_{0\nu} = \sum_{n,n'} \left\langle \psi_i(n_i, n'_i) \left| \tau_n^\dagger \tau_{n'}^\dagger \right| \psi_f(n_f, n'_f) \right\rangle \times \left\langle n_i n'_i \left| H_{nn'}(q) \right| n_f n'_f \right\rangle \qquad (2.20)$$

Where n_i (n'_i) and n_f (n'_f) are the initial and final spin states of the nucleons in the interaction. The first term can be thought of as the overlap of the final nucleus on the initial nucleus after a specific two neutrons have been converted to two protons. For example, we could think of it as the overlap of the ^{130}Te nuclear wavefunction with the ^{130}Xe wavefunction, with the two specific nucleons undergoing the transition, $(n_i, n'_i) \rightarrow (n_f, n'_f)$. This term is independent of the specifics of the decay model and depends only on the nucleus in question.

Since the nuclei in which we are interested for $0\nu\beta\beta$ are typically large and complicated, the nuclear matrix elements cannot be calculated exactly. While the $H_{nn'}(q)$ terms are model dependent, they are typically easy to calculate; conversely, the amplitude terms, $\langle \psi_i | \tau_n^\dagger \tau_{n'}^\dagger | \psi_f \rangle$, are extremely difficult to calculate and must be approximated. There are a variety of techniques that make varying approximations to this calculation. The most common approaches in the literature are the quasiparticle random phase approximation (QRPA), which sums many possible energy states of the nucleus but only in a limited number of shell configurations, the interacting shell model (ISM), which sums over all possible shell configurations but only over a limited number of state energies, and the interacting boson model (IBM-II) which pairs nucleons in shells and treats them as interacting bosons. Typically, these calculations can differ by factors of a few up to ~ 10.

Often the nuclear matrix element evaluation is broken into three components which are treated separately,

$$\mathcal{M}_{0\nu} = g_A^2 \left[\mathcal{M}_{GT}^{(0\nu)} - \left(\frac{g_V}{g_A} \right)^2 \mathcal{M}_F^{(0\nu)} + \mathcal{M}_T^{(0\nu)} \right] = g_A^2 \mathcal{M}^{(0\nu)} \qquad (2.21)$$
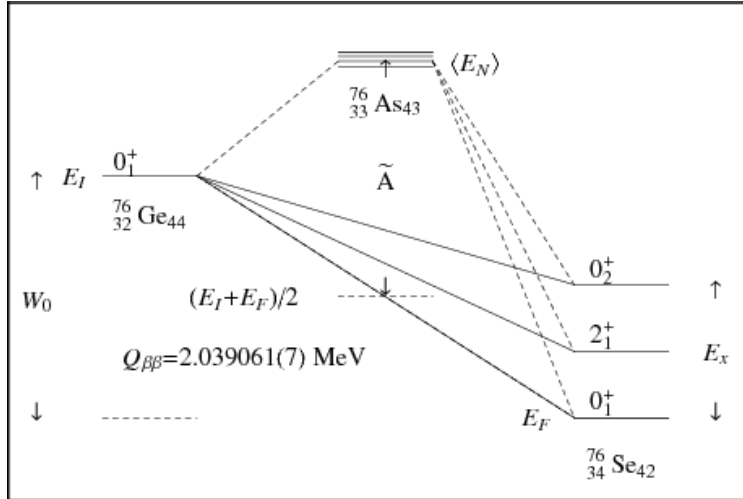
Figure 2.6: A $\beta\beta$ decay can be thought of as a virtual decay to all possible states of the intermediate nucleus followed by a decay to the final state nucleus. These states need to be summed over in calculating the nuclear matrix element. This figure is for decay of ^{76}Ge to ^{76}Se (taken from [48]) and also shows decays to excited final states.

where $\mathcal{M}_{GT}^{(0\nu)}$, $\mathcal{M}_F^{(0\nu)}$, $\mathcal{M}_T^{(0\nu)}$ are the Gamow-Teller transition term, the Fermi transition term and the tensor terms, respectively. For transitions where the initial and final states have $J^P = 0^+$, the Gamow-Teller terms tend to dominate, with the Fermi terms suppressed by a factor of $\gtrsim 2$ and the tensor terms suppressed by a factor of ~ 10.

We have explicitly separated out the axial coupling term, g_A^2, from the rest of the matrix element, $\mathcal{M}^{(0\nu)}$, which has only a weak dependence on g_A. Until recently, it was standard in the literature to use the free neutron axial coupling $g_A = 1.269$ (e.g. [50]). However, [48] recently pointed out that the coupling could be quenched and this would impact the limits placed on $m_{\beta\beta}$.

We often couple the phase space factor and the nuclear matrix element together into a "nuclear figure of merit",

$$F_N^{0\nu} \equiv G_{0\nu}(Q, Z)|\mathcal{M}_{0\nu}|^2 \tag{2.22}$$

When building an experiment to search for $0\nu\beta\beta$, the size of our expected signal will be proportional to a particle physics term, $|f_{0\nu}|^2$, and a term that will depend on our choice of isotope, F_N.

Calculated values for F_N assuming light Majorana neutrino exchange are shown in Fig. 2.7 for a variety of calculation techniques. When choosing a candidate isotope for a $0\nu\beta\beta$ search, we would like to maximize F_N. However, from a practical standpoint, we also need to take into account the isotopic abundance of the candidate. An isotope like ^{48}Ca has a very large Q-value, which makes it a nice candidate, but its tiny isotopic abundance would require a time consuming and expensive enrichment program to produce enough ^{48}Ca nuclei for an experiment. In this thesis, we study ^{130}Te, which has a reasonable F_N, a reasonable Q-
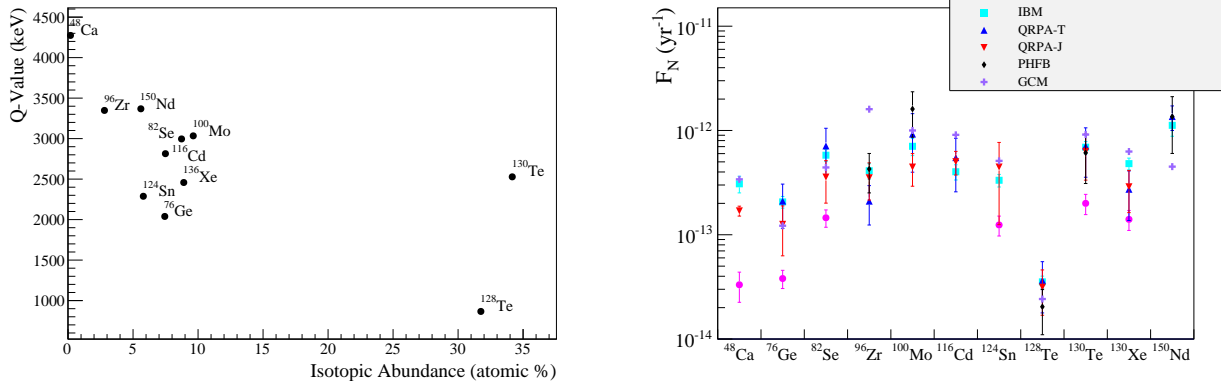
Figure 2.7: *Left:* Decay Q-values vs natural isotopic abundance for various $0\nu\beta\beta$ candidate isotopes. *Right:* Distribution of nuclear figures of merit for various isotopes using a variety of calculations of $|\mathcal{M}_{0\nu}|^2$ assuming the baseline light Majorana neutrino exchange model. The bars represent the minimum and maximum of the range of F_N and the markers represents the central values, they do not represent a "mean" and an "error" in a statistical sense. Phase space factors from [51], nuclear matrix element values from [48, 52–56].

value and the highest isotopic abundance of all of the candidate isotopes. This allows us to build our experiment out of natural Te and avoid the need for a time-consuming enrichment program.

2.4   Alternate $0\nu\beta\beta$ Models

In this section we present several alternate models that lead to $0\nu\beta\beta$ decay. The important thing to note is that there are *many* models which predict $0\nu\beta\beta$ through a variety of mechanisms. The models presented here are each interesting alternatives in their own way.

Type I Inverse Seesaw

The downside to the Type I Seesaw model is that it requires a spectrum of right-handed neutrinos with extremely large masses, M_N. Of course we do not actually know the scale of M_N, but in order to make the light neutrinos light via the seesaw mechanism, we must make the heavy neutrinos heavy enough to be experimentally inaccessible, $\gtrsim 10^{12}$ GeV.

An alternate approach is to introduce $N = 6$ right-handed neutrinos, but split them into two types, ν_R and N_R. Then write the mass matrix of Eqn. 2.9 as

$$\mathcal{L}_{\mathrm{mass},\nu} = -\frac{1}{2} \begin{pmatrix} \bar{\boldsymbol{\nu}}_{\boldsymbol{L}}^{\boldsymbol{c}} & \bar{\boldsymbol{\nu}}_{\boldsymbol{R}} & \bar{\boldsymbol{N}}_{\boldsymbol{R}} \end{pmatrix} \begin{pmatrix} 0 & \boldsymbol{M_1} & 0 \\ (\boldsymbol{M_1})^T & 0 & \boldsymbol{M_2} \\ 0 & (\boldsymbol{M_2})^T & \boldsymbol{\mu} \end{pmatrix} \begin{pmatrix} \boldsymbol{\nu_L} \\ \boldsymbol{\nu_R^c} \\ \boldsymbol{N_R^c} \end{pmatrix} + \mathrm{h.c.} \qquad (2.23)$$

where we define the hierarchy $\mu \ll M_1 \ll M_2$. In this case we are left with a Majorana mass term that looks like

$$\boldsymbol{m_\nu} = \boldsymbol{M_1 M_2}^{-1} \boldsymbol{\mu} (\boldsymbol{M_2}^T)^{-1} \boldsymbol{M_1}^T \tag{2.24}$$

This is called the Type I Inverse Seesaw mechanism. By setting $M_N \sim$TeV, $M_D \sim$GeV and $\mu \sim$keV, we can generate $m_\nu \sim$meV without the need to add particles at experimentally inaccessible energies. Of course, we have added a very particular spectrum of neutrinos for which we have little justification.

Similar to the Type I Seesaw, this model introduces a set of Dirac masses, M_1 and M_2 which by themselves could explain neutrino oscillations; but then adds a parameter μ which violates lepton number and generates Majorana masses. In the Type I Seesaw, we introduced a very large Majorana mass which we justified by saying that it came from a scale much larger than the EWSB scale. Here we introduce a set of Dirac masses which are around the EWSB scale as expected, but then introduced a μ which was assumed to be small. Often we can justify this by generating μ through a spontaneously broken symmetry.

This theory predicts $0\nu\beta\beta$ decay through the same Feynman diagrams as the Type I Seesaw and would leave $m_{\beta\beta}$ unchanged. But since our spectrum of heavy neutrinos is now different, $M_{\beta\beta}$ would be different, as would our calculation of $\mathcal{M}_{0\nu}^{\text{Heavy}}$.

Type II Seesaw

Another possible mechanism which leads to $0\nu\beta\beta$ decay is to introduce a Higgs SU(2) triplet field. Following [58], we write this field as

$$\Delta = \begin{pmatrix} \Delta^-/\sqrt{2} & \Delta^{--} \\ \Delta^0 & \Delta^-/\sqrt{2} \end{pmatrix} \tag{2.25}$$

This field couples directly to the SM fields through the term

$$\mathcal{L}_\Delta = i h_{\alpha\beta} \bar{L}_\alpha^c \tau_2 \Delta L_\beta \tag{2.26}$$

The neutrinos generate a Majorana mass directly by giving Δ^0 a vacuum expectation value (VEV), $\langle \Delta^0 \rangle = v_L$.

$$(\boldsymbol{m_\nu})_{\alpha\beta} = h_{\alpha\beta} v_L \tag{2.27}$$

The other components of this model will cause our newly introduced VEV, v_L, to be suppressed relative to the standard Higgs VEV, $\langle \phi^0 \rangle = v$, by some number of powers of v/m_Δ, which we assume to be small. (The exact number of powers depends on the model.) And thus we generate a small neutrino mass by making v_L/v small. This is called the Type II Seesaw.

With this addition, $0\nu\beta\beta$ decay can now proceed via the diagram in Fig. 2.8, which is characterized by the vertex factors, $\sqrt{2}g^2 v_L$ and h_{ee}. In this limit, we can pull out the particle physics factor of Eqn. 2.14:

$$f_{0\nu}^{II} \propto \frac{h_{ee} v_L m_p}{m_\Delta^2} = \frac{(m_\nu)_{ee} m_p}{m_\Delta^2} \tag{2.28}$$
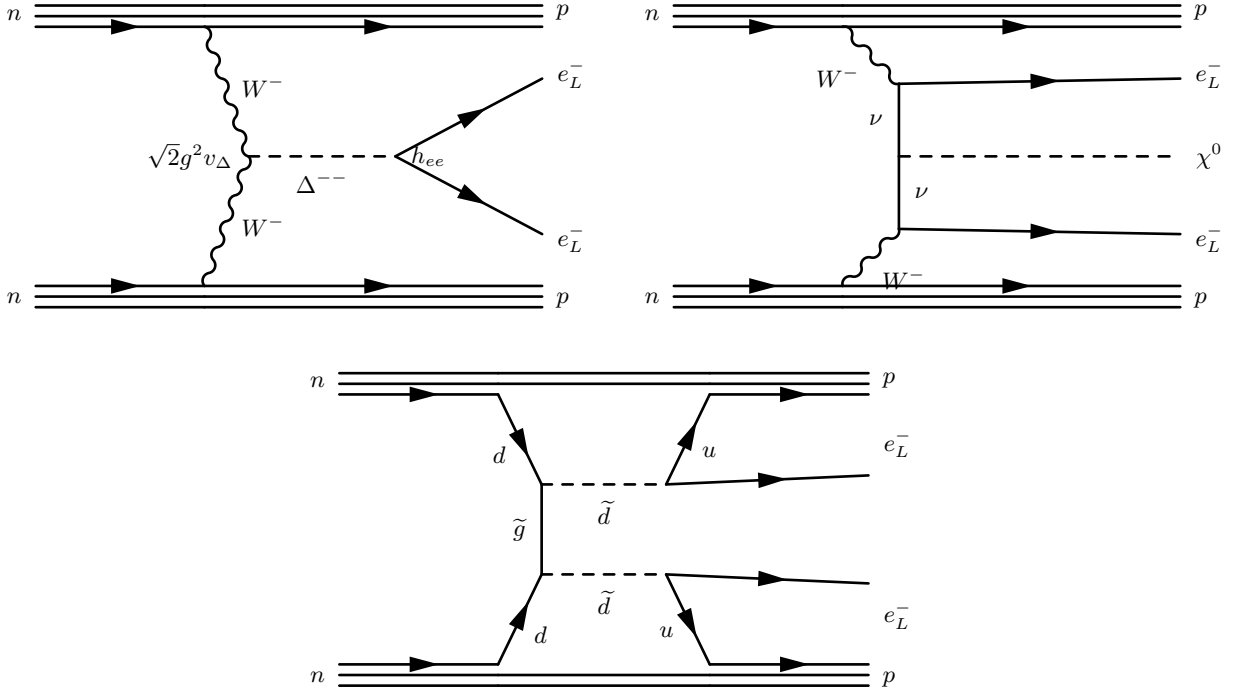
Figure 2.8: Non-standard decay mechanisms for $0\nu\beta\beta$. *Top Left:* A scalar triplet model with a doubly charged field, Δ^{--}, this is the Type II Seesaw mechanism and requires no explicit Majorana neutrino mass. This model is well constrained by present limits, except in the case of left-right symmetric models in which case the W^- are right-handed, and the decay is suppressed by $M_W^4/M_{W_R}^4$. *Top Right:* $0\nu\beta\beta$ decay with the emission of a light or massless Majoron. This decay leads to a modification of the $2\nu\beta\beta$ decay spectrum rather than an excess of events around the endpoint. *Right:* $0\nu\beta\beta$ via an R-parity violating supersymmetric model ($\not{R}_P$ MSSM), moderated by gluino exchange [57].

What makes this model interesting is that it achieves $0\nu\beta\beta$ decay *without* directly invoking any neutrino in the decay. In this model, our limits on the $0\nu\beta\beta$ decay rate translate into *lower* limits on effective mass $m_\Delta^2/(h_{ee}v_L)$. In its basic form, this model has been all but ruled out by present experimental limits, but there is ongoing interest in L-R symmetric models, in which we add a right-handed weak interaction. In this model, $f_{0\nu}$ picks up an extra suppression of $M_W^4/M_{W_R}^4$ [59].

It is common to think of $0\nu\beta\beta$ experiments as trying to measure the very *small* neutrino masses, but they can also be reframed as tests of new physics at very *large* mass scales.

$0\nu\beta\beta$ With Majoron Emission

The last $0\nu\beta\beta$ mechanism we mention is the possibility of decay with emission of one or two light or massless bosons, χ^0, which couple to neutrinos and are called Majorons. This decay channel is depicted in Fig. 2.8, and is given by

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- + \chi^0 \qquad \text{(Single Majoron Emission)} \qquad (2.29)$$
$$(Z, A) \rightarrow (Z + 2, A) + 2e^- + 2\chi^0 \qquad \text{(Double Majoron Emission)} \qquad (2.30)$$

There are many Majoron models with various generating mechanisms. Some models introduce χ^0 as a Nambu-Goldstone boson for a spontaneously broken lepton symmetry [60, 61], others *conserve* lepton number by assigning χ^0 a leptonic charge and still other theories introduce Majorons as supersymmetric partners to SM particles [62, 63].

From an experimental perspective, the Majoron itself will escape any detector unnoticed and appear as missing energy — similar to the signal of $2\nu\beta\beta$. Because of this, the experimental signature of $0\nu\beta\beta$ +Majoron emission is no longer a narrow line at the endpoint of the $2\nu\beta\beta$ spectrum but instead is a tiny modification to the $2\nu\beta\beta$ spectrum itself.

2.5 Experimental Sensitivity

The basic approach to a $0\nu\beta\beta$ experiment is to measure the decay spectrum of the candidate isotope and search for an excess of events around the endpoint of the $2\nu\beta\beta$ spectrum (see Fig. 2.3). We can think of this as a single-bin counting experiment where we count the number of events that occur in a window of width ΔE around the Q-value. In general we will also have some background above which we would hope to see our signal.

In the limit where backgrounds come from an external source and the statistics are gaussian, we can write our 1σ half-life sensitivity as

$$T_{1/2}^{0\nu} = \ln(2)\frac{\varepsilon a_I N_A \eta}{W}\sqrt{\frac{Mt}{b\Delta E}} \qquad (2.31)$$

where ε is the signal detection efficiency, a_I is isotopic abundance, M is the total active mass of the detector, t is the livetime, b is the background rate per unit energy, per unit detector mass, W is the molar mass of the molecule under study, and η is the number of nuclei of interest per molecule. The derivation of this equation and a full poissonian treatment of the sensitivity can be found in [64]. But in essence, Eqn. 2.31 relates the physical parameters of our experiment into a number of candidate nuclei, $N = \frac{a_I N_A \eta}{W}M$, and then compares the number of signal events expected in a given time, $S = \ln(2)\varepsilon Nt/T_{1/2}^{0\nu}$, to the number of background events expected, $B = bMt\Delta E$, and considers the half-life which would correspond to a 1σ fluctuation in the background.

We will return to this equation later in this thesis, but the relation to keep in mind is

$$\boxed{\text{Sensitivity} \propto a_I\sqrt{\frac{Mt}{b\Delta E}}} \qquad (2.32)$$

This relation gives a rough guideline for building a $0\nu\beta\beta$ detector. We aim for a large mass in order to study as many candidate nuclei as possible. We need a good energy resolution and as low a background as possible in order not to drown out the signal. As we will see in the next chapter, these aspects are the central focus of our experimental effort.

2.6   Present Experimental Searches for $0\nu\beta\beta$

The discovery of $0\nu\beta\beta$ decay would have rippling implications for particle physics at both low and high energy scales, would give new insight into the fundamental symmetries of the Universe and could provide compelling evidence for the role of neutrinos in the production of the baryon asymmetry of the Universe. Because of this potential for such fundamental importance, many recent experimental efforts have focused on discovering $0\nu\beta\beta$ decay, but at present $0\nu\beta\beta$ decay has not been convincingly observed. Current limits constrain the half-lives in the isotopes studied to be longer than 10^{24}–10^{25} yr.

There are currently several large experimental efforts to measure $0\nu\beta\beta$ decay in several isotopes, as well as multiple R&D efforts for next-generation experiments. The present best limits on $m_{\beta\beta}$ are shown in Fig. 2.9 and come from Gerda (studying ^{76}Ge), Cuoricino (studying ^{130}Te), and EXO-200 and KamLAND-Zen (studying ^{136}Xe) [65–68]. Together these experiments form a powerful search approach, by covering multiple isotopes with a variety of detector technologies possessing different strengths and systematics uncertainties.

The generation of experiments which are beginning to take data in the next year or two aim to reach as far down as $m_{\beta\beta} \sim 100$ meV by 2020 timescale. And there is a rapidly growing R&D effort for the next generation of detectors which aim to be sensitive to $m_{\beta\beta} \sim 10$ meV. This milestone would either discover $0\nu\beta\beta$ or rule out the inverted hierarchy in the basic light Majorana neutrino exchange model.

This effort is not only on the experimental side; there is also significant effort to improve the estimates of the nuclear matrix elements, $|\mathcal{M}_{0\nu}|^2$. And, as mentioned earlier, there is current significant theoretical interest in the literature in better understanding the quenching of the axial coupling, g_A. As seen in [69], this can significantly impact our calculation of $m_{\beta\beta}$ from a half-life limit. Often this makes our limit worse than previously thought, but work is still ongoing in this area.

Other experimental efforts can place other limits on $m_{\beta\beta}$ from a different (though model-dependent) direction. A measurement of the neutrino mass scale would significantly reduce the space of possible $m_{\beta\beta}$. One possible approach to this is to measure the distortion at the high energy end of a β-decay spectrum caused by the non-zero neutrino mass. This technique is sensitive to the effective electron neutrino mass:

$$m_{\nu_e}^2 \equiv \sum_i |U_{ei}|^2 m_i^2 \qquad \text{Electron Neutrino Mass} \tag{2.33}$$

A measurement of this quantity would help constrain the possible values of $m_{\beta\beta}$ — though in a model-dependent way. Present limits place $m_{\nu_e} < 2.05$ eV (95% C.L.) [71]. Future
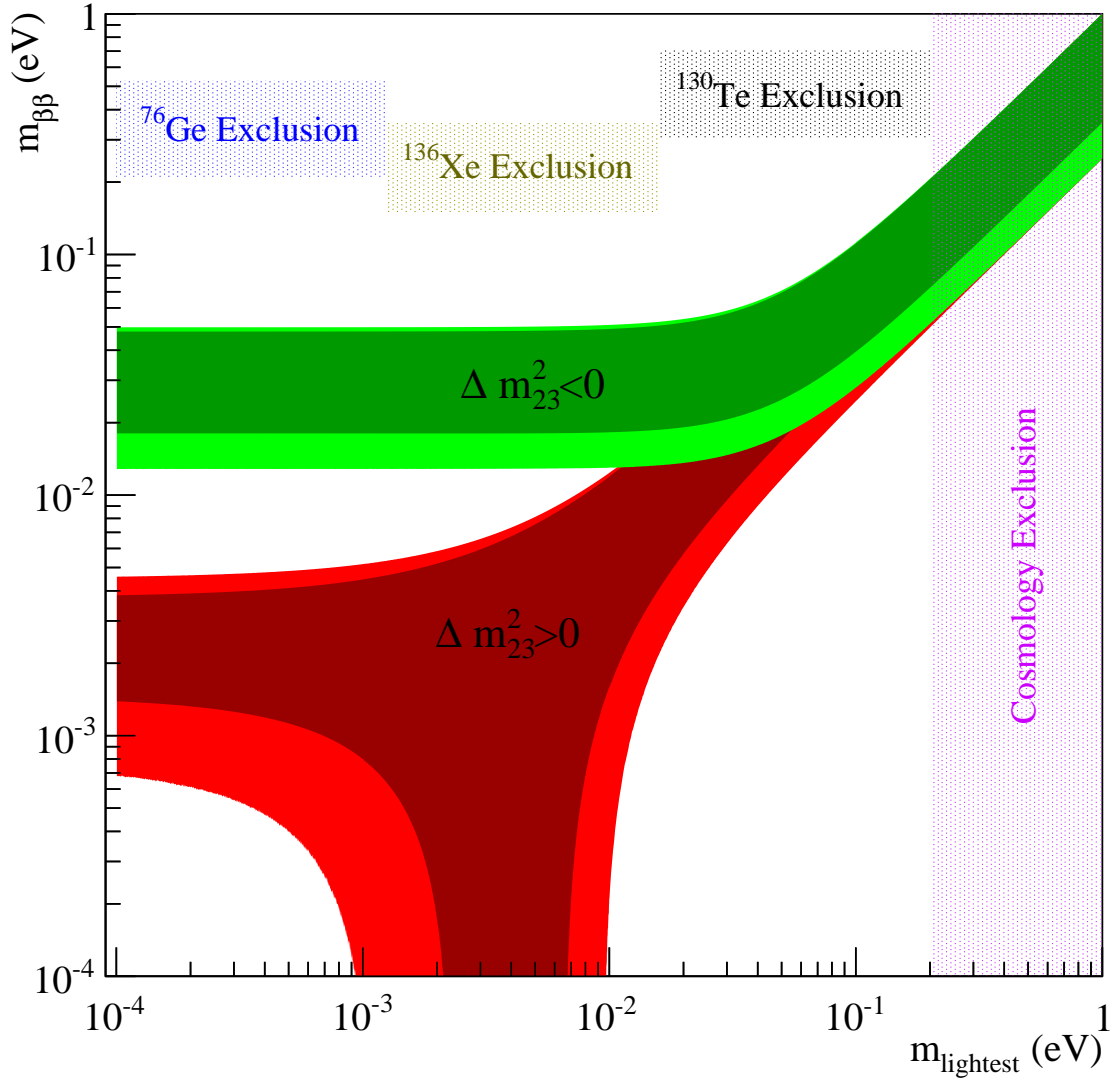
Figure 2.9: Plot of allowed regions for effective Majorana mass, $m_{\beta\beta}$ vs the lightest neutrino mass, m_{Lightest}. The plot includes limits placed before the inclusion of the data presented in this thesis. The shaded horizontal regions are *upper bounds* for a range of nuclear matrix elements. The combined ^{76}Ge limit comes from Gerda + Heidelberg Moscow + IGEX data [65]. The combined ^{136}Xe data comes from EXO-200 + KamLAND-Zen [66, 67]. The ^{130}Te limit comes from Cuoricino [68]. The cosmology limit is a true exclusion ($m_{\text{lightest}} < 0.2$ eV at 95% C.L.) and comes from Planck 2015 [70].

experiments like KATRIN and Project-8 will seek to improve this by an order of magnitude or more [72, 73].

Other experimental efforts [74–76], seek to measure both the hierarchy and the CP violating phase, δ. A definitive measurement of the neutrino mass hierarchy would significantly reduce the space of possible $m_{\beta\beta}$ values in Fig. 2.9.

Still another possibility is the recent advances in precision cosmology, which have been able to place very impressive constraints on the sum of the neutrino masses,

$$m_{\text{Tot}} \equiv \sum_i m_i \tag{2.34}$$

Present limits on this quantity place $m_{\text{Tot}} < 0.63$ eV (95% C.L.) [70].

While these experiments all help constrain the possible values of $m_{\beta\beta}$, none of them is sensitive to the Majorana/Dirac nature of the neutrino. Conversely, a set of searches which is sensitive to this and is often overlooked in the literature is accelerator-based approaches. These are commonly overlooked because in the standard light Majorana neutrino exchange mechanism they have very little sensitivity compared with $0\nu\beta\beta$ searches. However, some of the more exotic models which predict $0\nu\beta\beta$ can be probed at high energy, and indeed many models have been strongly constrained [77–79].

Chapter 3

The CUORE and CUORE-0 Experiments

The Cryogenic Underground Observatory for Rare Events (CUORE) experiment is a ton scale search for $0\nu\beta\beta$ decay of ^{130}Te. It is currently under construction at the Laboratori Nazionali del Gran Sasso (LNGS) in Italy and plans to begin taking data in 2015. The first phase of the experiment, called CUORE-0, ran from March 2013 to February 2015 and forms the basis for this dissertation.

The basic approach to any $0\nu\beta\beta$ search is to measure the decay spectrum of the isotope of interest and look for an excess of events around the Q-Value. There are three qualities that a $0\nu\beta\beta$ search must try to optimize. First, since the decay is so rare, an experiment must observe a very large number of nuclei. Functionally, this means the detector must have a large mass — the next generation of experiments aim to achieve masses in the one to several ton range. Second, an experiment must have a very low background rate, since too high a background rate can bury the $0\nu\beta\beta$ signal in statistical fluctuations. Finally, an experiment needs to have very good energy resolution.

In CUORE we have chosen the candidate isotope ^{130}Te because its high natural isotopic abundance avoids the need for a costly and time-consuming isotopic enrichment. Its high Q-value also places the region of interest (ROI) above the majority of the naturally occurring environmental background lines, in a relatively quiet region of the energy spectrum.

As with all $0\nu\beta\beta$ experiments, CUORE must run underground to get away from the bombardment of cosmic rays at the surface. CUORE will operate in Hall A of the LNGS lab. This provides the detector with an overburden of 1400 m of rock (3600 meters of water equivalent) which reduces the cosmic ray rate by 6 orders of magnitude relative to the surface. The measured residual muon rate is roughly $(3.2 \pm 0.2) \times 10^{-8}$ $\mu/\text{cm}^2/\text{s}$ [80], which corresponds to roughly one muon per hour in CUORE.

CUORE employs a bolometric energy detection method, and operates at a temperature of ${\sim}10$ mK. This method yields an extremely good energy resolution, which results in a narrower ROI and therefore fewer background events to contend with. But even in the limit of no external background an experiment still needs a good energy resolution to separate
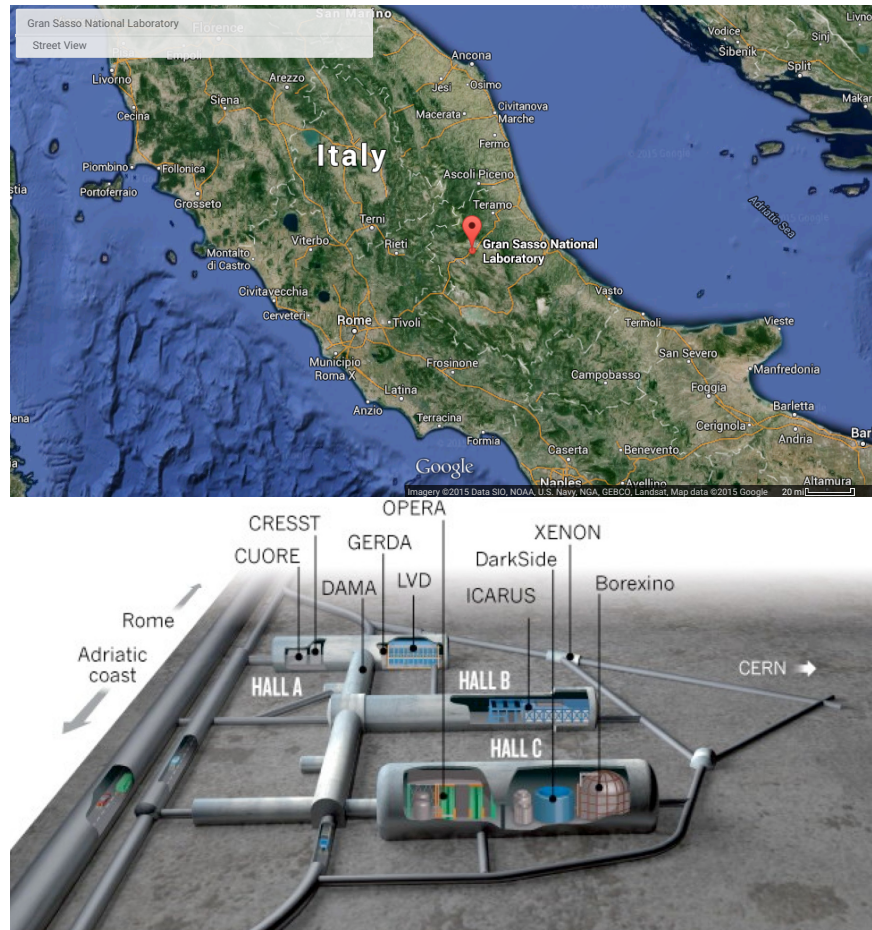
Figure 3.1: *Top:* Map with the location of LNGS marked with an arrow. *Bottom:* Schematic of the LNGS laboratory. The lab is built off of two tunnels that run underneath the highest mountain in the Apenninian range, Gran Sasso.

the $0\nu\beta\beta$ signal from the tail of the $2\nu\beta\beta$ distribution. The payoff of the bolometric method is made clear in Fig. 3.2, but the main drawback is the the technical difficulty of cooling a large mass to such a low temperature. We will spend the rest of this chapter describing the CUORE detector and the bolometric technique.

3.1   A Bolometric Detector

Bolometric detectors are based on the principle of converting the energy of a nuclear decay into heat and thus a change in temperature of the detector. At its most basic level, a bolometer is just an energy absorber and a thermometer. A deposition of energy in the absorber causes an increase in its temperature, which is read out by a thermometer. The
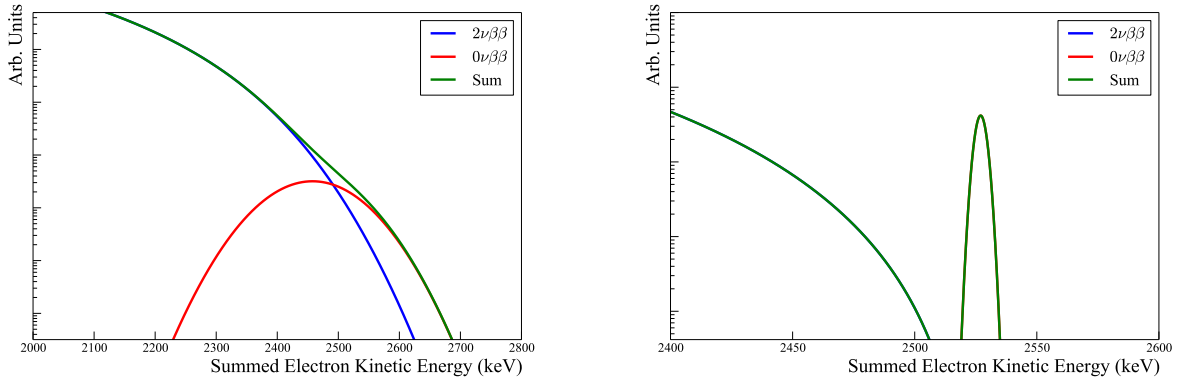
Figure 3.2: Hypothetical double beta decay spectra in the limit of no background, but finite detector resolution and assuming a $0\nu\beta\beta$ branching ratio $\Gamma_{0\nu\beta\beta}/\Gamma_{2\nu\beta\beta} = 10^{-6}$. *Left:* The endpoint of the spectrum of a hypothetical ^{136}Xe experiment with 5.9% energy resolution (FWHM). *Right:* The endpoint of the spectrum for a CUORE style experiment with 5 keV energy resolution (FWHM).

size of the change in temperature corresponds to the amount of energy deposited. The excess heat slowly flows out into a heat bath through a weak thermal link allowing the temperature to gradually return to where it started (see Fig. 3.3).

Since the amount of energy released by a single nuclear decay is extremely small, ~ 0.16 pJ, the change in temperature is correspondingly small. To make this signal measurable, we cool our bolometers down to an operating temperature of $\sim$10 mK. This has two effects: first, at lower temperature the random shot noise fluctuations are much smaller than the size of our signal. Second, the size of the change in temperature, which is given by $\Delta T = E/C(T)$, can be made much larger at lower temperature where the heat capacity, $C(T)$, is significantly smaller.

The heat capacity of the CUORE crystal roughly follows the Debye law [81],

$$C(T) \propto k_B \left(\frac{T}{\Theta_D} \right)^3 \tag{3.1}$$

with a $\Theta_D \approx 232$ K [82]. At the operating temperature of 10 mK, our heat capacity corresponds to a $\Delta T \sim 100$ μK for a 1 MeV energy deposit.

In thermal equilibrium (when $T = T_0$ in Fig 3.3), the uncertainty on the temperature is dominated by thermal fluctuations across the weak thermal link, R. This places a limit on the precision with which we can measure our energy resolution. This can be calculated directly from statistical mechanics to be $\langle \Delta E^2 \rangle \sim k_B T^2 C(T)$. So the thermodynamic limit
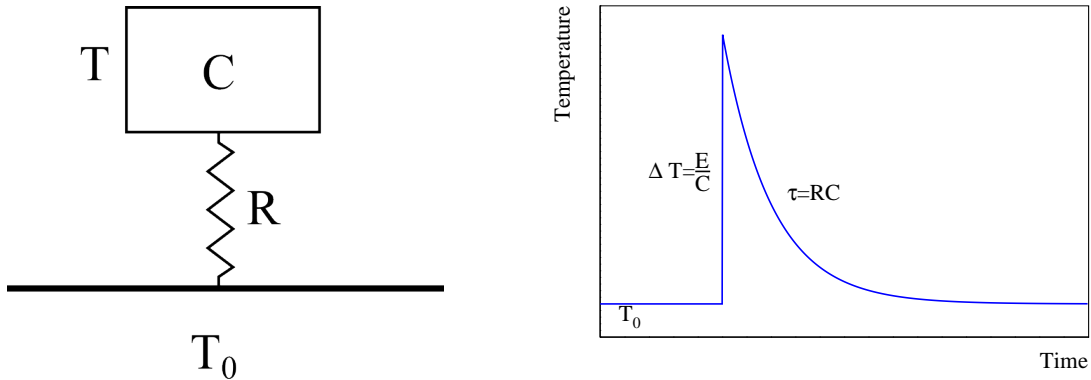
Figure 3.3: *Left*: An idealized bolometer with heat capacity C, at temperature T, connected through a weak thermal link R to a heatbath at temperature T_0. *Right:* An idealized bolometric response to a sudden deposition of energy, E. The temperature slowly falls back to T_0 as the energy escapes through the weak thermal link with time constant RC.

on our energy resolution should go as

$$\sqrt{\langle \Delta E^2 \rangle} \propto k_B T \left(\frac{T}{\Theta_D} \right)^{3/2} \tag{3.2}$$

This strong dependence of resolution on temperature demonstrates why we are so interested in cooling our detectors to such low temperatures.

The power of the bolometric method is really based on the fact that the phonon excitation energy, ϵ_{ph}, is very small and thus the number of phonons excited by a nuclear decay is extremely large. For typical CUORE numbers, we have $\epsilon_{\mathrm{ph}} \approx 1$ μeV, and $C \approx 1$ MeV/100 μK. This gives us a thermodynamic limit on our energy resolution of $\Delta E \sim 10$ eV. However, in practice, our bolometers have other sources of noise, both internal [83] and external, which limit our energy resolution. We typically achieve resolutions on the order of $\Delta E \sim 5$ keV full-width at half-max (FWHM) at 2.5 MeV. This is several orders of magnitude above the thermodynamic limit, but still competitive with the best detector technologies at this energy.

CUORE Energy Absorber

In CUORE the energy absorber is a TeO_2 crystal, which also acts as the source of decays of interest. By building our detector out of our source we gain two advantages: first, we can reduce the amount of unnecessary material near our detector, thus reducing our potential backgrounds; and second we have a very high detection efficiency since the decays in which we are interested occur already inside our detector. This configuration is referred to as "source = detector" and is common among $0\nu\beta\beta$ experiments.

TeO_2 is a convenient material for building a large-scale detector for several reasons. Its thermal contraction properties are similar to those of the copper that composes the

frame that provides structural support to the detector (TeO$_2$ contraction between room temperature and operating temperature is $\Delta l/l \approx 0.27\%$ while copper is $\Delta l/l \approx 0.33\%$ [84]). The crystals are robust to thermal cycling with a low failure rate. The TeO$_2$ crystal has a high Debye temperature, and therefore a small heat capacity at low temperature. And, most importantly, unlike pure Te, TeO$_2$ is far more pleasant to work with and will not make a handler's breath smell like rotten garlic.

The crystals in CUORE are $5 \times 5 \times 5$ cm^3 and weigh an average of 750 g. They are made with ^{nat}Te and so each crystal contains $\sim$208 g of ^{130}Te. The crystals are grown by The Shanghai Institute Of Ceramics, Chinese Academy of Sciences (SICCAS) [85], following a strict growing and handling protocol that is described later in this chapter.

Bolometer Instrumentation

We instrument each crystal absorber with a Neutron Transmutation Doped (NTD) germanium sensor. Ordinarily, germanium is a semiconductor whose conductivity goes to zero at 0 K. The process of neutron transmutation doping exposes pure germanium to a neutron flux. The neutrons capturing on Ge produce Ga, As and Se, doping the semi-conductor material to just below the transition to a metal. Beyond this transition, the material behaves like a metal whose conductivity becomes infinite at 0 K; but just below this transition, the conductivity drops to zero very sharply as the temperature approaches zero.

At a microscopic level, the doping creates a dispersion of "conduction centers" — the Ga, As and Se atoms — randomly spread throughout the bulk of the non-conducting germanium. At relatively high temperatures, the energy of the electrons is high enough that their wavefunction may overlap multiple conduction centers and conduction is mediated by moving between neighboring conduction centers. In this regime, the doped germanium has fairly high conductivity.

As the temperature drops, the wavefunctions of the conduction electrons become more localized. As their energies approach the Fermi energy, the conducting electrons enter a regime called "variable length hopping". Here, the electron wavefunction becomes localized enough to only occupy a single conduction center at a time. In this regime, electrical conduction is mediated by absorbing or emitting phonons to jump from one conduction center to another at a different potential energy. This is not unlike the way an electron absorbs or emits a photon to jump between energy levels in an atom, except here the levels are also separated spatially throughout the material. In this way a conduction electron "hops" from one site to another through the material. In this regime, the resistivity follows a steep exponential temperature dependence [86, 87],

$$\rho(T) = \rho_0 e^{\sqrt{T_0/T}} \tag{3.3}$$

This relation is parameterized by ρ_0 and T_0, which are determined by the doping level. Typically, we determine these numbers empirically and for CUORE they are $\rho_0 \sim 0.1$ $\Omega\cdot$cm and $T_0 \sim 4$ K.
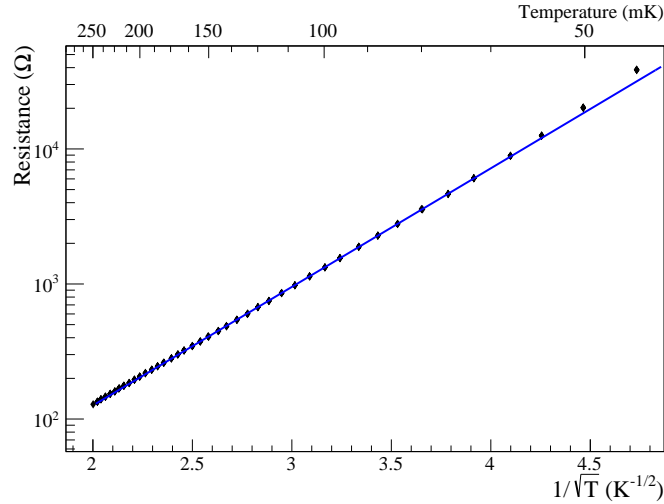
Figure 3.4: Resistance Curve for a CUORE-style NTD with $\rho_0 \approx 0.22$ Ω·cm and $T_0 = 4.1$ K. (From CUORE batch NTD-40A.) The fit is performed over the temperature range 70-250 mK. Below 60 mK the R-T curve displays a divergence from the behavior in Eqn. 3.3 which is common for NTD thermistors.
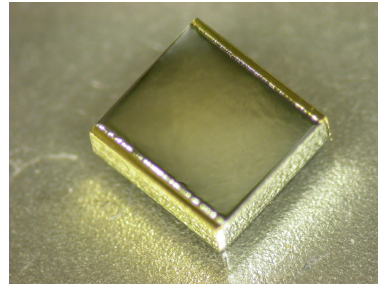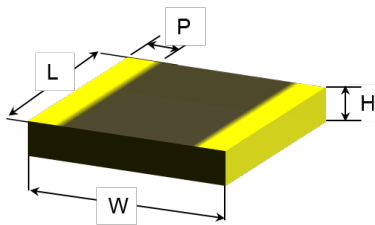
Figure 3.5: *Left:* Sketch of a CUORE style wrap-around NTD, the dimensions are $L = 3.0$ mm, $W = 2.9$ mm, $H = 0.9$ mm, $P = 0.2$ mm. *Right:* A CUORE style NTD. Images taken from [88].

The steep dependence of resistivity on temperature in Eqn. 3.3 makes this material into a very sensitive thermometer. To give a sense of the numbers, a 1 MeV decay in a CUORE crystal at 10 mK results in a 1% increase in temperature which causes a ∼5–10% *decrease* in resistivity. By continually measuring the resistance of an NTD chip, we can very sensitively detect sudden changes in the temperature of the absorber.

To produce the NTD sensors for CUORE, we took large wafers of ultra-pure germanium and irradiated them at the MIT Nuclear Reactor Laboratory. Because the neutron capture cross section is low, the neutrons are able to penetrate the bulk of the Ge wafers and provide a uniform flux to the full volume. This produces a very homogeneous and reproducible doping across the entire wafer [89]. After the short-lived isotopes had decayed away — typically a

waiting period of ∼6 months — we diced the wafers into $3 \times 3 \times 1$ mm^3 chips and gold plated electrodes onto the sides. The electrodes wrap slightly around to the top of the NTD for vertical bonding (see Fig. 3.5). This wrap-around style allows the NTDs to be bonded in-situ after being attached to the bolometer. The typical resistance of these NTDs at ∼10 mK is ∼100 MΩ.

We also instrument each crystal absorber with a Joule heater. This is a simple high-purity Si semiconductor chip with a typical resistance of 300 kΩ. By pulsing these with a fixed voltage we can impart a known amount of energy into our bolometers and monitor the gain. We return to this in Chapter 4.

A CUORE Tower

Once the crystals have been instrumented, we assemble them into a tower of 52 crystals in a $2 \times 2 \times 13$ configuration. The frame of the tower is made from an Electrolytic Tough Pitch (ETP1) copper called "NOSV" copper that was specially chosen for its low hydrogen content, low bulk radioactivity and high thermal conductivity. The copper frame also serves as the heat sink for the bolometers. The crystals are held in the frames with Polytetrafluoroethylene (PTFE) spacers, which provide a weak thermal link to the heat bath.

The electronic readout is carried by flexible polyethylene naphthalate PCB strips with high-density copper wire traces, (referred to as PENCu cables). These strips are attached to two sides of each of the towers and have leaves that fan out on every floor with pads for wire bonding the NTDs and Joule heater. Each PENCu cable instruments 3-4 floors and a total of 9 cables are needed for each tower. The PENCu cables continue past the top of the tower and are plugged in to the NbTi cryostat ribbon cables which carry the signals the rest of the way to the room temperature electronics on top of the cryostat.

The full CUORE detector will consist of 19 of these towers tightly packed into an array (see Fig. 3.7). In total, CUORE will have 988 independent bolometers for a total active mass of 741 kg or 206 kg of ^{130}Te.

Electronic Readout

The CUORE electronics chain (shown in Fig. 3.8) provides both the current bias for the NTDs and the readout for the bolometer signals. The entire chain sits at room temperature and consists of the front-end boards which power the NTDs and act as a first-stage amplifier, the Bessel boards which act as an anti-aliasing filter, and DAQ boards which digitize the signal.

The front-end boards were specially designed for CUORE [90, 91] and serve two purposes. First, they are an extremely stable DC current source that provides low-noise DC bias to the NTDs. The typical value of the current supplied is in the range of $I_{\text{Bias}} \sim 100$ pA, which limits the Joule heating across the sensors to the range of 1 pW. The front-end board also performs the readout of the voltage across the NTD, $V_{\text{NTD}}(T) = I_{\text{Bias}} R_{\text{NTD}}(T)$, and passes the voltage signal through a two-stage amplifier to increase the voltage signal from
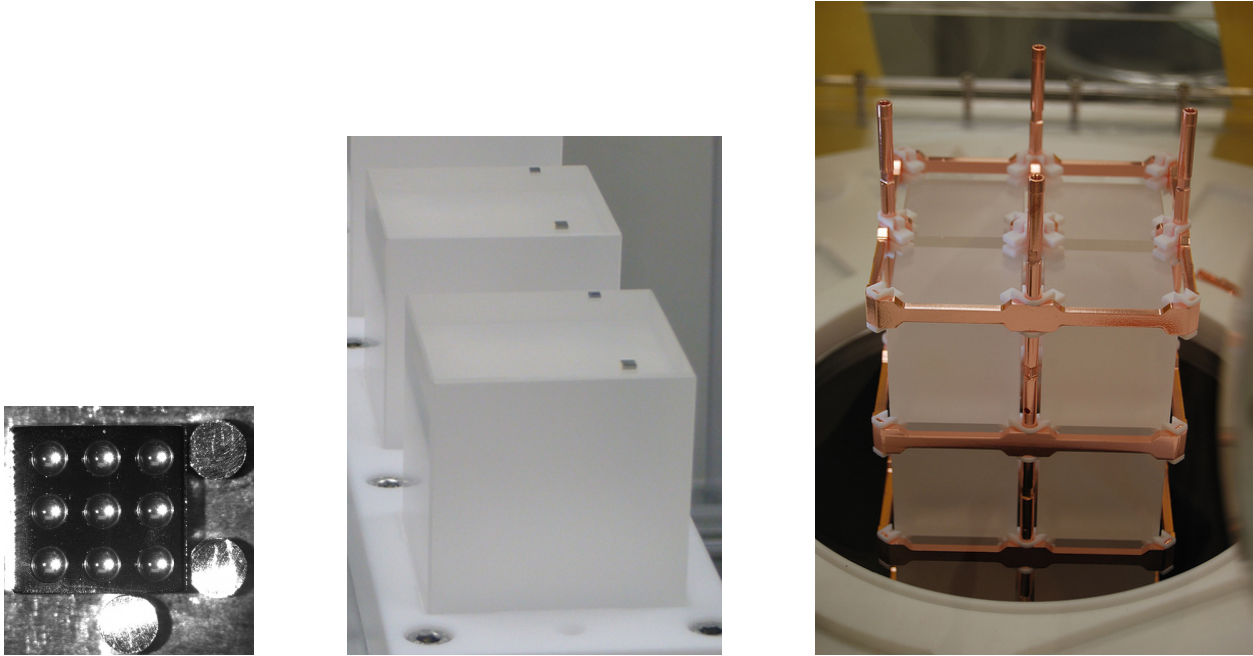
Figure 3.6: *Left:* Underside of an NTD with glue spot matrix before the crystal is lowered on top of it. *Middle:* Example of two CUORE bolometers with NTD (near) and Joule heater (far) attached. These crystals are sitting in the gluing glove box after having been glued. *Right:* CUORE tower assembly in progress. The crystals are arranged in a 2×2 floor and stacked 13 floors tall. The frame is NOSV copper and the crystals are held off with PTFE spacers.

~ 100 μV/MeV to $\sim$1 V/MeV. It also inverts the polarity of the signal so that a positive voltage change corresponds to higher temperature.

After the amplifiers, the voltage signal passes through a 6 pole anti-aliasing Bessel filter whose rolloff frequency can be set between 20 and 120 Hz, above the signal band, which peaks around 1-3 Hz. The Bessel filter also adds a constant offset voltage to bring the signal voltage into range for the National Instruments NI-PXI-6284 high precision ADC boards. The DAQ boards digitize the signal in the range -10.5 V to 10.5 V with 18-bit precision. This yields a voltage precision of 0.08 mV, which is well within the requirement for CUORE. The digitizers sample the signal at 125–1000 Hz and all channels are read in parallel by the APOLLO software.

3.2 Predecessor To CUORE: Cuoricino

CUORE builds on the experience of Cuoricino, which ran in Hall A of LNGS from 2003-2008. Cuoricino was a single tower with a similar design to the CUORE style tower, but with a slightly larger active mass of 40.7 kg. Cuoricino employed a mix of natural abundance TeO_2
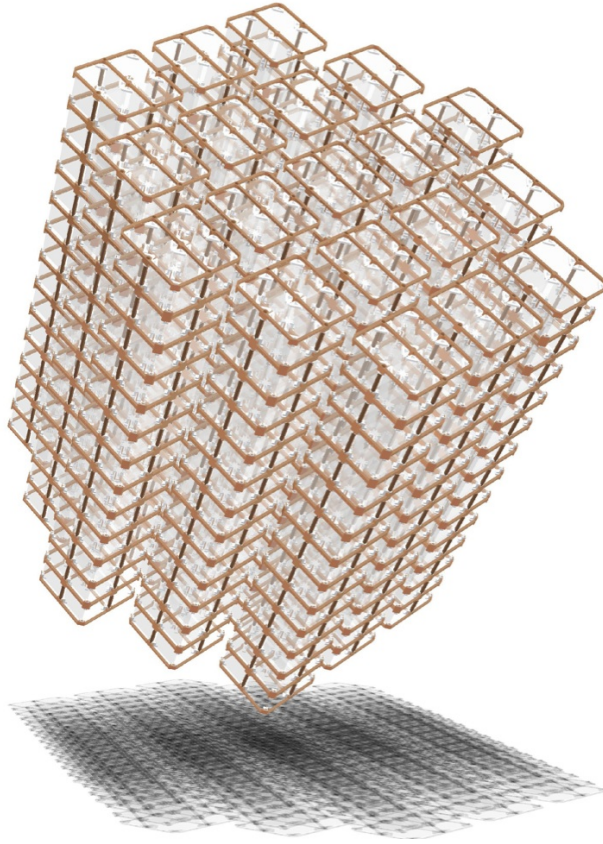
Figure 3.7: Drawing of the CUORE detector. The detector is composed of 19 identical towers, for a total of 988 independent bolometers.
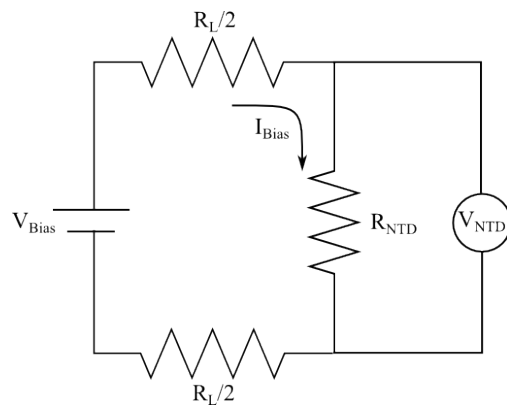
Figure 3.8: Schematic of the bolometer bias and readout circuit. Typical values for CUORE are $R_L \sim 60$ GΩ, $R_{\mathrm{NTD}} \sim 100$ MΩ & $V_{\mathrm{Bias}} \sim 1$ V.
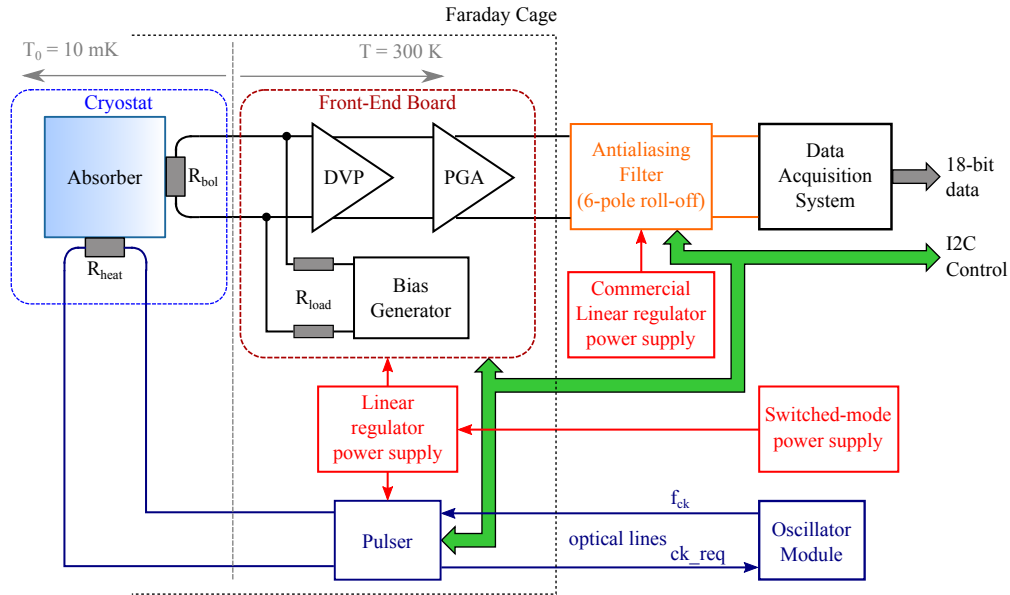
Figure 3.9: Schematic of the CUORE bolometer electronics chain. From [88].

and enriched TeO$_2$ crystals for a total ^{130}Te mass of ~11 kg. It was operated as a single tower (see Fig. 3.10) in a setup essentially identical to CUORE-0, which is described in section 3.5.

Over the 5 year run time it collected 19.75 kg·yr of ^{130}Te exposure and, until the results presented here, it held the best experimental limit on the $0\nu\beta\beta$ half-life of ^{130}Te at $T_{1/2}^{0\nu\beta\beta} > 2.8 \times 10^{24}$ yr (90% C.L.) [68]. The $0\nu\beta\beta$ ROI and full spectrum are shown in Fig. 3.10. Moving forward towards CUORE, there are three design aspects of Cuoricino that we worked to improve upon: the background in the ROI, the energy resolution, and the detector live time.

The main source of down time for Cuoricino— and many cryogenic experiments — comes from maintenance to the cryostat that houses and cools it. In the case of Cuoricino, data collection had to be stopped for ~3–4 hours every 48 hours to refill the liquid helium bath which provides the first stage of cooling for the detector. In CUORE we would like to eliminate this down time by building a new cryogen-free cryostat that is large enough to house the CUORE detector (described in section 3.4).

The Cuoricino tower held a mix of sizes of crystals including smaller $3 \times 3 \times 6$ cm^3 crystals. But among the CUORE-sized, $5 \times 5 \times 5$ cm^3 crystals, it achieved a mean energy resolution at 2615 keV of 6.3 keV. Moving toward CUORE, we sought to improve the energy resolution to 5 keV at 2615 keV. Our approach was a combination of improving the reliability of the thermistor gluing and the robustness of the electrical wiring (described in section 3.3), as well as gaining a better understanding of the noise and detector line shape (described in chapter 5 and section 6.3).

Cuoricino achieved a background in the ROI of 0.169 ± 0.006 cnts/keV/kg/yr, of which 0.110 ± 0.001 cnts/keV/kg/yr came form degraded α decays [92, 93]. These are α decays that
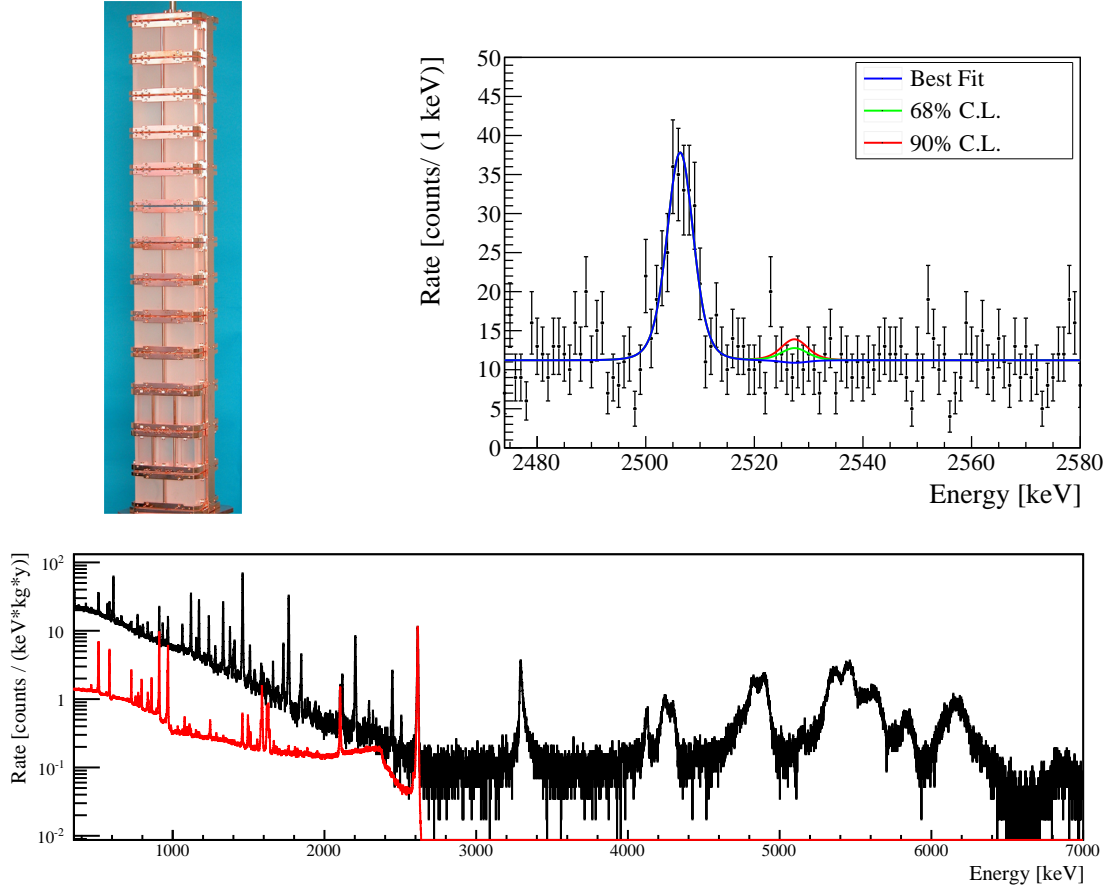
Figure 3.10: *Left:* The Cuoricino tower is similar in design to the CUORE-style tower. It was operated from 2003–2008 in Hall A of LNGS and set the world's best limit on the $0\nu\beta\beta$ half-life of ^{130}Te. *Right:* The final Cuoricino spectrum in the ROI with the $0\nu\beta\beta$ decay fits. *Bottom:* The full Cuoricino background spectrum with the calibration spectrum overlaid in red. The calibration spectrum has been normalized to the background 2615 keV rate. The ROI at 2525 has significant background contribution from degraded α decays.

occur on or near the surface of a material and deposit a random fraction of their energy into a passive material and the rest of their total energy into an active material. These decays form the long tails in Fig. 3.10 that reach from $E > 3$ MeV down through the ROI at 2.5 MeV. Because of the very short range of α particles, these decays necessarily come from contamination on or near the detectors themselves. The remaining 0.059 ± 0.006 cnts/keV/kg/yr come from γ-rays which originate in the cryostat materials and Compton scatter in a passive material before being absorbed by one of the crystals. In CUORE, we sought to reduce the total background in the ROI by a factor of $\sim$17 to 0.01 cnts/keV/kg/yr by developing rigorous surface treatment and material selection and handling procedures, described in the next section.
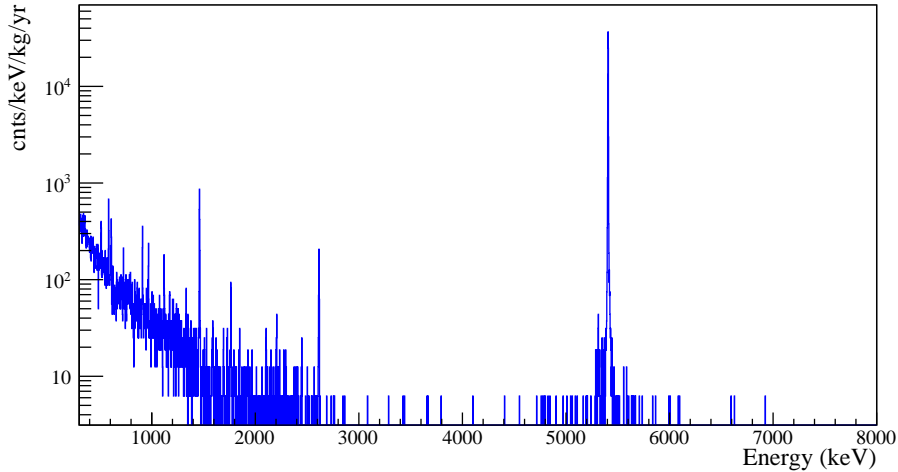
Figure 3.11: Full energy spectrum from 3 crystals in CCVR8. The CCVR runs are performed in the Hall C R&D cryostat, which has a higher γ-ray rate than the Hall A cryostat. Compared to Fig. 3.10, the ^{210}Po peak at 5.4 MeV is much narrower because in the CCVR runs it is a bulk contamination from the crystal growth process. ^{210}Po has a half-life of 138 days, so this contamination is expected to be significantly reduced by the time CUORE starts taking data.

3.3 CUORE Construction and Assembly Line

In moving from Cuoricino to CUORE, we focused much effort on reducing the backgrounds in the ROI. This was accomplished through a combination of careful material selection to keep contaminants out of the detector material bulk, surface cleaning to eliminate surface contamination and parts handling to prevent recontamination after cleaning.

Crystal Growth And Validation

We worked closely with the crystal manufacturer, SICCAS, to develop a strict set of crystal growth and polishing protocols to keep the bulk and surface contamination to a minimum. These protocols stretched all the way from ensuring the radiopurity of the metallic Te, to the growth and shaping process, to the final polishing and shipping protocols.

We grow the raw crystals to roughly $5 \times 5 \times 15$ cm^3, before cutting out and polishing the middle third. The crystallization process pushes most impurities out the edges where they are chopped off and discarded. Once produced, the crystals were triple bagged in cleaned plastic vacuum packaging. In fact, after they are produced, the crystals are kept under vacuum or nitrogen flux throughout the entire CUORE assembly, and will only briefly come into contact with air when they are finally mounted in the cryostat.

The crystals were shipped from Shanghai, China to LNGS, Italy by boat. We kept the crystals at sea level to avoid the cosmogenic activation incurred by air travel [94]. Upon arrival in LNGS, we randomly chose four crystals from each shipment to be tested cryogenically for quality assurance. These tests were called the CUORE Crystal Validation Runs (CCVR). The spectrum for CCVR-8 is shown in Fig. 3.11. The bulk contamination limits for the crystals were $< 3 \times 10^{-13}$ g/g for both ^{238}U and ^{232}Th. During the CCVR runs we placed 90% upper limits on this contamination that were often almost an order of magnitude better than the requirement [95].

During the CCVR runs, we typically did see one significant contaminant in the bulk of the crystals, ^{210}Po. Polonium and tellurium are chemically very similar, which makes Po difficult to separate from the Te powder. However, ^{210}Po decays via a 5.4 MeV α with a half-life of 138 days. So by the time that CUORE begins data taking, the background from this isotope will have decayed significantly.

Parts Cleaning

The CUORE towers and all the copper structures near the CUORE detector were built out of specifically chosen NOSV copper produced by Norddeutche Affinerie. The NOSV copper was designed to have very high thermal conductivity at low temperatures, low hydrogen content and also happened to have very low radioactivity. Moving from Cuoricino to CUORE, we redesigned the frame of the tower to reduce the copper needed by a factor of $\sim$2.5, reducing the background from the bulk of the copper by a similar amount. We also benefit from the smaller surface to volume ratio for the copper radiation shielding reducing the background from surface contamination.

Except for the $\lesssim$4 months required to machine and clean it, the copper spent all of its time underground. This was to minimize cosmogenic activation and specifically minimize the production of ^{60}Co.

Though the bulk contamination of the copper was well within the acceptable limits for CUORE, the handling and machining of the copper introduced contaminants on the surface layer. To address this, we developed a rigorous cleaning procedure that was a combination of tumbling, electropolishing, chemical etching and magnetron plasma etching (TECM). We verified the efficacy of this process using a dedicated experiment setup called the Three Towers Test (TTT) [96]. After cleaning, the copper parts are stored in plastic bags that have been cleaned and flushed with nitrogen. They are transported underground and stored under nitrogen flux to prevent recontamination from radon.

All other parts that make up the tower — PTFE spacers, copper screws, NTDs, heaters, etc — as well as all tools used to build the tower, also undergo aggressive cleaning and are stored under nitrogen flux.

Tower Assembly Line

The entire detector construction took place in a class 1000 clean room in the CUORE hut in Hall A. We designed a tower assembly line to build the towers consisting of three steps: crystal gluing, tower assembly, and tower wiring (see figures 3.6 and 3.12). Each step of the process was performed in a specially designed glove box that kept the detector parts under constant nitrogen flux and out of contact with any of the rest of the environment of the clean room.

We instrument each crystal with one NTD and one Si Joule heater, which are attached using Araldite two component epoxy. We chose this particular epoxy because of its resilience to thermal cycling, rapid curing and its low radioactivity. We glue the NTD with a 3×3 matrix of glue spots. While the epoxy was curing, we maintained a distance between the NTD and the crystal of ~ 50 μm. This gap set after the epoxy dried. This standoff from the crystal and dot matrix absorb the stresses caused by the differential contractions of the crystal, NTD and epoxy during the cooldown. We found that a solid film of epoxy would often lead to too much differential stresses and could result in the NTD popping off or cracking the crystal. The Joule heater was attached in a similar manner, but with a 5-dot matrix, since it was slightly smaller.

To keep this process as uniform and reproducible as possible, we designed a robotic gluing system to move and align the crystals and hold them steady while the epoxy dried. The gluing platform was designed to hold the gap between the bottom of the NTD and the surface of the crystal using vacuum suction. This system ensured the tight constraints and reproducibility required for attaching NTDs to over 1000 CUORE crystals.

We assemble each tower on a specially designed Universal Working Plane (UWP), which can have one of several different glove boxes mounted for each stage of the tower assembly. When the tower is not being worked on, it is lowered into a garage below the UWP and fluxed with nitrogen for storage.

The actual tower assembly was done by hand inside the glove boxes, under constant nitrogen flux. The glove boxes had gloves attached that could be reached into the space to assemble the tower without exposing the detector or any parts to the air of the clean room.

Once the tower has been assembled and the PENCu wire strips have been attached, we mount the wire bonding glove box on the UWP for wiring. The tower is wire bonded in situ with a vertically mounted wire bonder. Each NTD and Joule heater was wire bonded to the PENCu cable pads by 25 μm gold wire, with two wires running from each electrode to the pads for redundancy.

After wire bonding, we loaded each tower into an acrylic storage container and placed it in the corner of the clean room under nitrogen flux for storage. All told, the CUORE detector construction took about 18 months to complete.

Figure 3.12: The CUORE assembly line. *Left:* The gluing robot attaches the NTDs and Joule heaters to the crystals under constant N_2 flux. *Middle:* All construction takes place in specially designed glove boxes under N_2 flux. The tower is stored in a garage immediately below the construction plane. *Right:* Wire bonding is done in situ with a specially mounted wire bonder.

3.4 The CUORE Cryostat

The cryostat being built to hold and cool the CUORE detector will be one of the largest and most powerful dilution refrigerators in the world (see Fig. 3.13). Unlike the cryostat that housed Cuoricino, the CUORE refrigerator will be cryogen-free — meaning that the first stage of cooling is provided by pulse tube refrigerators. This provides continuous cooling without the need to stop data-taking to refill Helium baths like in a cryogen refrigerator.

The stages of the cryostat as well as all thermal shields are made out of NOSV copper to minimize γ backgrounds from the cryostat itself. The towers will be surrounded by an internal and an external lead shield. A 6-cm-thick internal lateral lead shield that surrounds the towers is made from lead recovered from a sunken Roman galleon that had been sitting on the sea floor for ~2000 years. This gave any radioactive isotopes of lead in the bulk many half-lives to decay away. The residual radioactivity of this Roman lead is < 0.004 Bq/kg [97]. Above the detector, between it and the rest of the cryostat will be a 24 cm plate of modern low background lead (16 ± 4 Bq/kg) as well as 6 cm of clean NOSV copper to shield the towers from γs from above. All of this internal shielding will be cooled to < 4 K. When in operation, an external modern lead (150 ± 20 Bq/kg) shield will be raised up around the cryostat. In this configuration the detector will have at least 30 cm of lead shielding in every direction. All told, the cryostat will cool 1.5 tonnes of material down to an operating temperature of 10 mK; and it will have a total mass of ~10 tonnes below 1 K.

The construction of the cryostat is underway and in the first test runs it reached a base temperature of 5.9 mK with no load. Because of its large volume and very low temperature
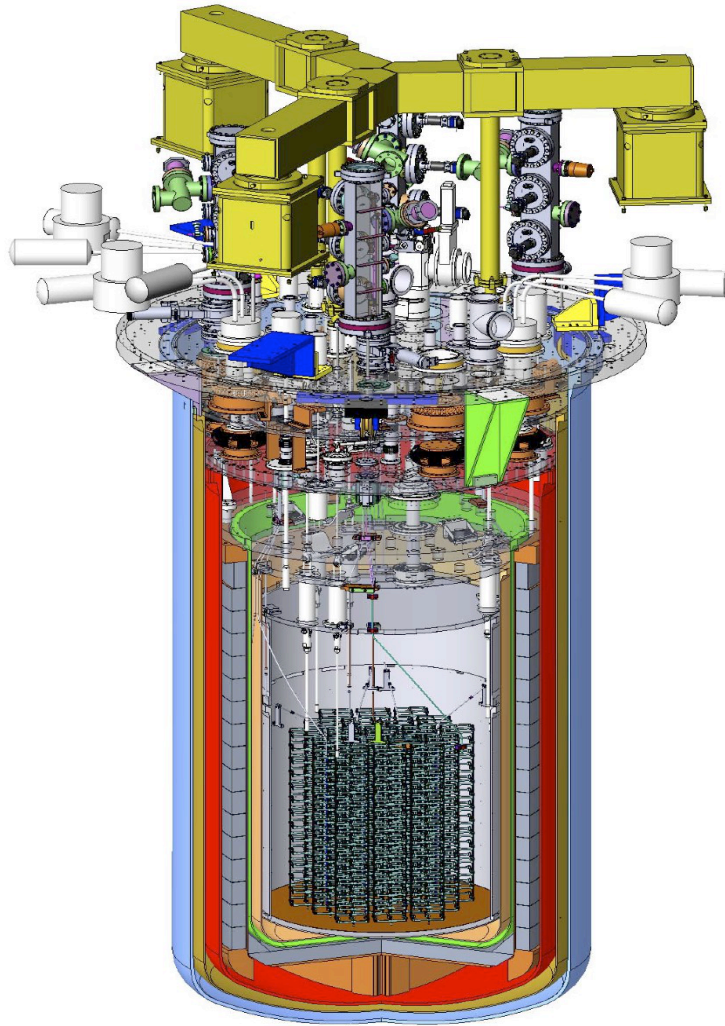
Figure 3.13: Schematic of the CUORE cryostat. The vertical scale is ~4 m.

it was actually the Coldest Cubic Meter in the Known Universe (see Appendix D or [98]).

3.5 The CUORE-0 Experiment

The first tower built on the CUORE assembly line was called CUORE-0 (see Fig. 3.14). This tower was constructed using the same procedures and materials as was used for the other 19. In fact, the only thing that made this tower distinct was that it was operated separately as a single-tower experiment. It was originally conceived as a full-scale commissioning of the CUORE assembly line and verification of the CUORE background reduction techniques, but is also a competitive $0\nu\beta\beta$ search in its own right.

The detector was first assembled in late 2011 and had to be reassembled several times to debug problems with the copper parts and the assembly line. After several failed attempts, we corrected these problems and fully assembled the CUORE-0 detector in early 2012. All told, it took roughly 6 months to assemble CUORE-0, but building off of this experience we built all 19 towers for CUORE in about 18 months. So CUORE-0 was extremely useful as a debugging of the assembly procedure.

The CUORE-0 detector has 39 kg of active mass, or about 10.8 kg of ^{130}Te. Unlike the CUORE towers, the CUORE-0 tower is surrounded by its own lateral copper shield (pictured in Fig. 3.14) and is operated in the same cryostat that housed Cuoricino. The tower is shielded by an inner layer of lead 1.2 cm thick on the sides with a 7.5 cm thick plate above the tower and a 10 cm thick plate below the tower. Both plates are 15 cm in diameter. All internal lead is low-background Roman lead (< 4 mBq/kg). The tower is further shielded by an external 10 cm layer of modern low background (16 ± 4 Bq/kg) inside a 10 cm layer of modern lead (150 ± 20 Bq/kg). This entire setup is surrounded by a 10 cm-thick layer of borated polyethylene serving as a neutron shield and a Faraday cage to shield from electromagnetic noise. The space between the external lead and the cryostat is lined with acoustic dampers to prevent acoustic pick-up. To prevent radon buildup in this space we continually flush it with nitrogen during data taking.

We transported the CUORE-0 detector from the CUORE hut to the Cuoricino hut and stored it in the old Cuoricino cleanroom while preparing the crysotat. While in storage, the detector was under constant nitrogen flux. We installed the detector in late 2012 and began taking data in March 2013.

The electronic readout for CUORE-0 is the same as used in Cuoricino. This is similar to the CUORE electronics, but at a much smaller scale. The only significant differences are that the rolloff frequency of the Bessel boards was set to 12 Hz and the sampling rate of the DAQ was set at 125 Hz.
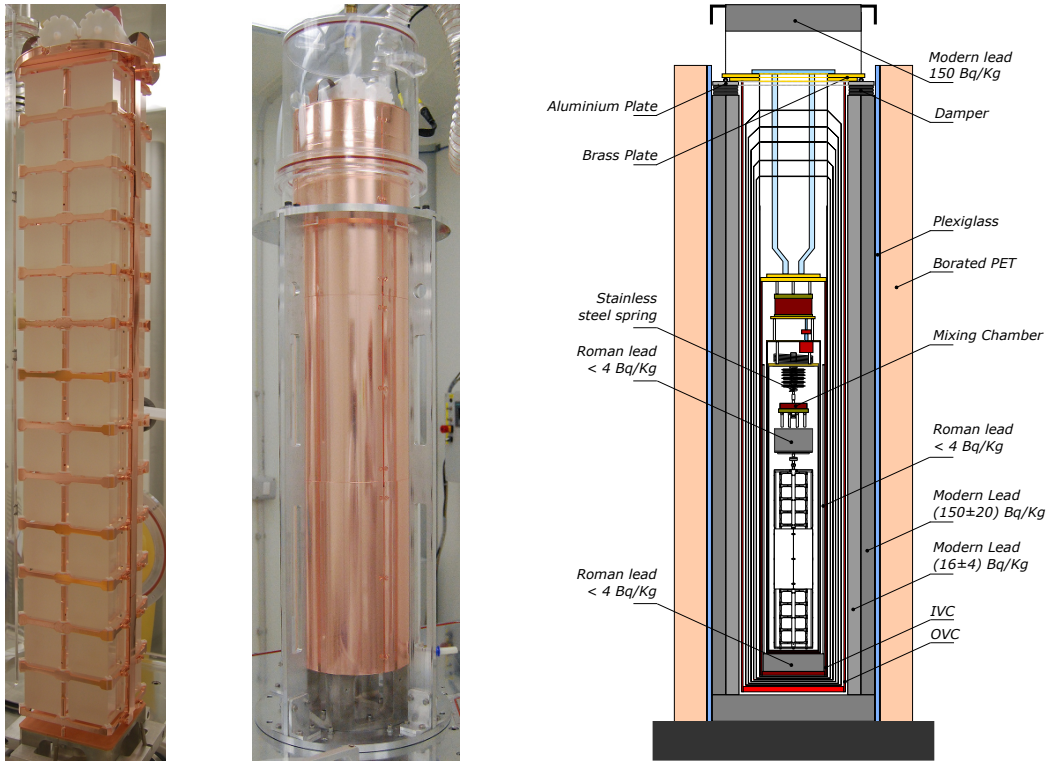
Figure 3.14: *Left:* The CUORE-0 tower raised above the storage garage inside the construction glove box. *Middle:* The CUORE-0 tower enclosed inside its lateral shields and inside the acrylic transport/storage container. *Right:* A schematic of the Cuoricino cryostat that houses the CUORE-0 detector.

Chapter 4

CUORE-0 Data Collection and Processing

4.1 CUORE-0 Detector Setup and Data Taking

Setting the Detector Operating Condition

The first step to operating the CUORE-0 detector is to optimize the detector's working conditions. Most obviously, this entails physically aligning (or verticalizing) the detector so that neither it, nor any of the wires connecting it to the outside are vibrating against the surrounding shield. This is done by adjusting the air pressure of the vibration dampeners and is basically a process of guess and check where we adjust the detector position and check the noise level.

The next step is settings the detector 'working point'. This entails setting the NTD bias, gain and offset for each bolometer. The aim here is to set the bias to maximize the sensitivity of the detector, adjust the gain so as to have a reasonable signal while maintaining a large enough dynamic range, and finally adjust the offset so that the signal lies within an acceptable range with respect to the ADC voltage window.

The bolometer itself acts like a closed loop amplifier whose gain depends on the nominal bias. At small biases, the circuit is linear and the NTD voltage is directly proportional to the bias. However, at large biases, electrothermal feedback kicks in and joule heating of the NTD increasingly warms the NTD — thereby decreasing its resistance — and the NTD voltage falls with increasing bias. At some point between these two extremes, the pulse amplitude is maximized and this is where we want to operate our bolometers.

The gain as a function of bias is mapped by scanning the available frontend biases and measuring the gain using the pulser heater and building this into a Load Curve. Figure 4.1 shows the resulting curve for a CUORE-0 channel.

An alternative approach is not to maximize the pulse amplitude, but rather to maximize the signal to noise ratio (SNR) ratio. This requires measuring not only the pulse height vs input bias but also the average noise power spectrum at each input bias. From this, we build

an Optimum Filter spectrum (see section 4.2), and calculate the SNR for each input bias and choose the maximum. In theory, this process produces a better working point and so was used as the preferred approach. However, it was much more time consuming because it has to integrate an average noise power spectrum. Since the noise was often variable, the resulting SNR vs bias curve was not always smooth enough to find a robust maximum. In these cases the working point was set where the pulse amplitude was maximized.

Once the NTD bias is set, we select a gain such that a 2615 keV pulse has an amplitude of $\sim$1–2 V. Finally, we set an offset voltage, V_{offset}, so that the baseline voltage lies around -4 to -7 V. This gives the electronics enough dynamic range to fully capture pulses up to a few 10s of MeV without saturating the DAQ, but also enough room for the baseline to drift downward without saturating low.
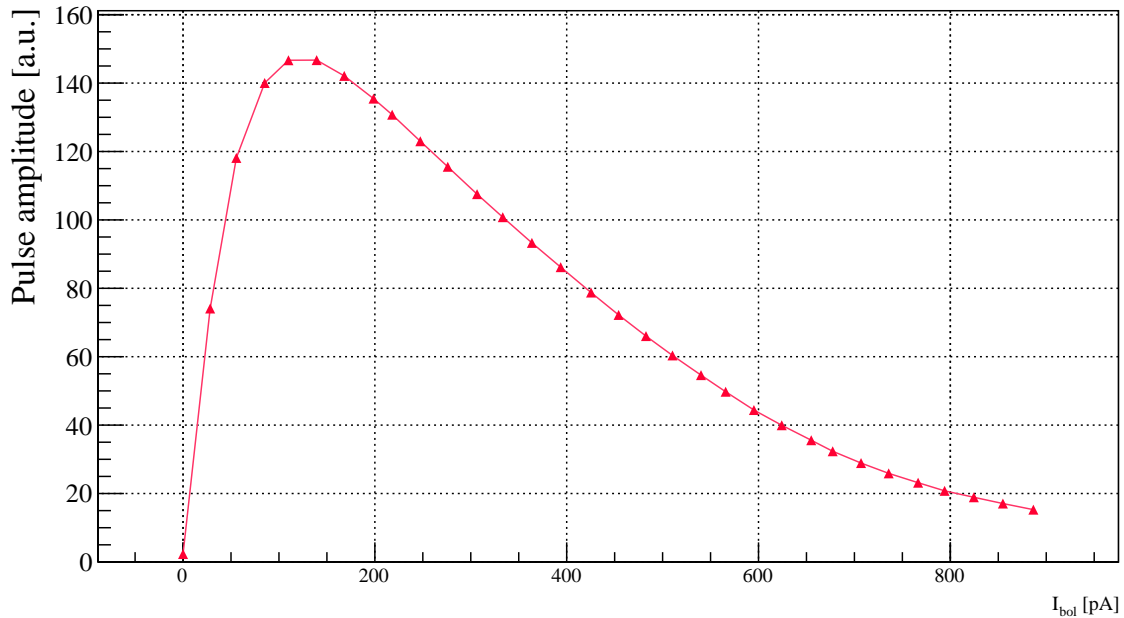
Figure 4.1: Pulser amplitude dependence on NTD bias current for Channel 28 taken before dataset 2079. The working bias is chosen to maximize the pulse amplitude.

Apollo Software

The CUORE-0 data was collected using the APOLLO software package, which custom designed by the CUORE collaboration for data collection. APOLLO handles everything from electronics setup and monitoring, to reading data from the digitizer and event triggering.

During each run, the software reads the continuous data stream from the DAQ and writes to a continuous data file which stores all of the waveform data. It also triggers on the

waveforms and writes the triggered events to `QRaw` files using a separate running instance of the DIANA software (described later in this chapter).

A CUORE-0 Event

When triggered, APOLLO stores a waveform spanning from one second before to four seconds after the trigger to the triggered data file. The one second before the pulse is referred to as the "pretrigger". We will analyze the pretrigger data to understand the behavior of the bolometer before the event occurred. In total, the triggered waveform has 626 ADC samples, collected at 125 Hz. The waveform is stored in the triggered data file with other important event data such as the channel, trigger time, event number and trigger type. APOLLO triggers and stores three types of events: signal events, heater events and noise events. Examples are shown in Fig. 4.2.
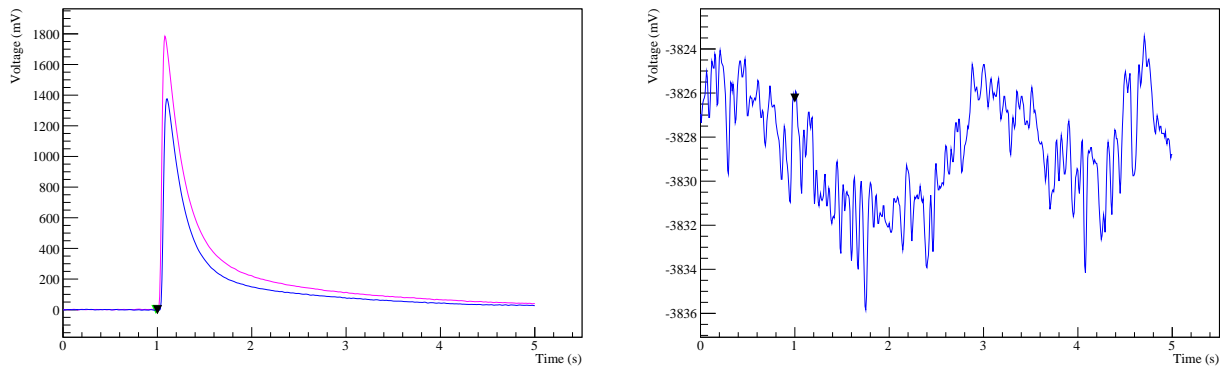
Figure 4.2: *Left:* A CUORE-0 signal event of energy ∼2615 keV (Run 201011, Event Number 4961) with a larger heater pulse of energy ∼3370 keV (Run: 201011, Event Number: 3813). The baselines have been subtracted. The triangular markers indicate where the triggers occurred. The one second of data before the trigger is used to characterize the behavior of the bolometer before the event. The pulser events typically display a faster decay time compared to the signal events. *Right:* A typical noise event (Run 200990, Event Number 815).

Signal Events

The APOLLO software monitors the data streams and triggers a signal event collection when the derivative of the signal exceeds a preset threshold. The thresholds were set for each channel as low as possible without making the trigger rate unacceptably high. The thresholds ranged from ∼20 keV on the best performing channels to ∼40 keV on the noisiest but could vary over the course of data taking. Better thresholds and signal efficiency were achieved

with the new Optimum Trigger which was not used in this analysis but is currently being implemented for CUORE [99].

When a signal event is triggered, APOLLO writes it to the data file along with a set of 'side pulses'. In CUORE-0 these side pulses were defined as the geometric nearest neighbors, so in addition to the waveform that was actually triggered, the waveforms from the crystals immediately surrounding the triggered crystal are also stored. These waveforms are stored as part of the primary event and are assigned the same event number. We use these side pulses later in the Coincidence and Decorrelation analyses.

After a signal trigger, APOLLO places a 1 second dead window where it will not trigger any other signal event on that channel within 1 s of the previous signal trigger. However, a larger dead window will be placed later in the analysis.

Pulser Heater Events

To stabilize the gain of the bolometers against temperatures drifts, we map the gain dependence using a pulser heater which fires every 300 seconds.

The heaters are pulsed with a 100 μs voltage pulse, which causes Joule heating in the chip and heats the crystal suddenly — similar to a particle event. The voltage is set so that the heater pulses reconstruct slightly higher than 2615 keV — the highest energy background γ-ray — typically around 3.2–3.5 MeV. Starting with dataset 2067, we dedicated 10% of the heater pulses to a lower energy pulse, $\sim$1.6 MeV, and 10% to a higher energy pulse, $\sim$7 MeV. We used these pulses to help understand the linearity of the detector response.

Because the heaters are wired in four parallel groups, an entire column of the tower fires their heaters together in unison. The four column pulsers are delayed by 10 seconds with respect to each other, so channels 1–13 fire together first, 10 seconds later 14–26 fire, etc.

Of the 51 crystals with working NTDs, 49 also had functioning heaters; channel 10 lost its heater during tower construction and 1 lost its heater sometime during the initial cool down. For these channels other stabilization techniques were developed, as described in section 4.2.

APOLLO triggers on heater pulses automatically and places a 100 ms dead window around the event, so any trigger that occurs within the 100 ms following a pulser trigger is recorded as a secondary trigger within the pulser event but does not generate its own event and event number. Functionally, this means that every pulser trigger is stored as its own event and the derivative trigger that fires because of the pulser is stored as a secondary trigger in the pulser event. We take advantage of this fact later to measure the efficiency of the derivative trigger.

Noise Events

In order to understand the noise behavior of the detector, APOLLO also collects 'random' triggers. These triggers are only random in the sense that the data that they capture is intended to be random noise, but in fact the triggers are fired at evenly spaced intervals

of 200 seconds. The noise data is collected from all channels in unison so correlated noise information can be derived as well.

We can understand our "baseline resolution" — which is defined as the energy resolution at 0 keV — by measuring the energy resolution of these noise events. Since the triggers are random, we should be measuring no signal, so the baseline resolution determines how well we can measure 0 keV.

Runs, Datasets, Working Points and Phases

The CUORE-0 data is collected in runs, which are grouped into datasets, which are then grouped into working point sets for processing. The datasets are also grouped into data taking phases, which are used in the analysis to partition that data into smaller pieces. We also defined two data taking 'campaigns', which correspond to data taken before and after a major maintenance to the cryostat, which resulted in a significant improvement in the noise performance (see Table A.1).

The majority of the CUORE-0 run time was dedicated to physics runs, which are either "Calibration" runs used to calibrate the detector or "Background" runs in which we perform the $0\nu\beta\beta$ search. The other types of runs are typically short diagnostic runs used either to set up the detector or measure its performance. Each calibration or background run typically lasts 1–2 days. Data taking must be stopped every $\sim$48 hours for $\sim$3 hours, in order to refill the cryostat's helium bath.

A single dataset is comprised of a set of initial calibration runs, a set of background runs, and a set of closing calibration runs. Datasets that are collected in immediate succession usually share a set of calibration runs. During some datasets, the background data taking was interrupted due to some unforeseen circumstance, forcing an early closing of the dataset without any final calibration runs.

A group of datasets that are similar enough in detector behavior are grouped into a single collection that is characterized by the detector's working point. Specifically, in Section 4.2 we will discuss a stabilization algorithm that relies on sets of calibration runs which span several datasets and map a constant (in time) gain dependence on temperature. The number of datasets that make up a working point varies depending on how long the behavior of the detector remained stable; the shortest working point is a single dataset, and the longest is seven datasets.

The datasets are also grouped into data taking "phases". These phases are somewhat arbitrarily defined and were originally tagged in order to correspond with conference presentations. Unfortunately, the boundaries of the working points don't always line up with the boundaries of the data taking phases.

Background

The majority of the CUORE-0 run time was devoted to background runs, which comprise the $0\nu\beta\beta$ decay search. During background data taking the CUORE-0 hut is generally kept

empty to avoid environmental noise. The Faraday cage is kept closed and is flushed with boil-off N_2 to displace any radon gas in the room. During occasions when the cage was not flushed with N_2, we saw a small but significant increase in the rate of ^{214}Bi γ lines.

The background data taking usually lasted about 3–4 weeks, but could be terminated early if working conditions began to change (e.g. accidental warm-ups, equipment malfunction or fire[1]).

Calibration

The detector is calibrated using source wires inserted on opposite sides of the detector. These are thoriated tungsten wires that have a total activity of $\sim$50 Bq. The strings are inserted into tubes that run into the outer vacuum chamber (OVC), inside the outer lead shield but outside the inner lead shield.

The calibration is done using the γ lines from the decay chain of ^{232}Th. Because of the low counting rate, we require at least ~ 60 hours of calibration data to collect enough statistics, so most datasets are flanked by three to four calibration runs on either side. Several datasets had significantly more opening or closing calibration runs to characterize the behavior of the detector.

During this time, the conditions in the CUORE-0 hut are kept similar to that of the background data collection.

Working Point Measurements

We performed test runs between every physics run to monitor the detector working condition. These runs usually take about 5-10 minutes and characterize the working point of the bolometers. This measurement involves decreasing the pre-amp gain, zeroing the voltage offset, and alternating the polarity of the voltage measurement to determine the NTD resistance. The result of this measurement is the NTD resistance, which is used later as part of the stabilization without heater algorithm (see section 4.2.)

CUORE-0 Data Taking Summary

The CUORE-0 $0\nu\beta\beta$ data collection began in March 2013 and finished in February 2015, amassing a total of 35.21 kg·yr of exposure. The data collection was performed in four data taking Phases. The first ran from March – August 2013 and was stopped to warm up the detector for a two-month long maintenance.

The second data taking Phase began in November 2013 and saw an immediate improvement in detector performance. The noise levels were lower than in Phase I and as a result the detector up-time could be increased. This Phase ran through May 2014, and was closed for a data release at Neutrino 2014. Phase III began immediately after and in the same

---

[1]No comment.

working conditions — in fact, Phase II and Phase III share calibration runs. Phase III ran through October 2014 and was closed for another few weeks of detector maintenance.

Phase IV began in November 2014 and ran through the end of February 2015. Phase IV was special in that it wasn't included in the CUORE-0 analysis decision making. All the cuts and techniques were tuned on the first three Phases and the last Phase was added separately.

The breakdown of the final CUORE-0 data history plus run statistics can be found in Appendix A.

4.2   First-Level Data Processing

The first-level data production includes all the steps that take the data from the triggered files through the data blinding and NTuple production for physics analysis. The CUORE-0 data production is done using the Diana v02.30 software developed for CUORE. Diana is a modular software that loops through all the triggered events and processes each in turn through a series of production steps.

In this section, we walk through the first-level production in the order that it is run. At each stage, we can reference measured quantities whose origin we have already described. However, in this order, it is easy to lose the forest for the trees. So we briefly outline the goal of the first-level production.

The first-level production can be very briefly summarized in the following major steps: amplitude evaluation, gain stabilization, energy calibration, coincidence rejection, and data blinding. For each event, we measure the amplitude of each event using a waveform filter that is "optimized" for our signals. We then stabilize each event, by regressing out the gain dependence on temperature. From here, we convert each event's amplitude to an energy using a known calibration spectrum. Since we expect that roughly 88% of $0\nu\beta\beta$ events will be contained in a single crystal we reject events that occur simultaneously in different crystals. These coincidence events are primarily α decays or scattered γ-rays. Finally, we blind our data in the ROI to prevent any subconscious bias in the second-level analysis.

Several of the steps in this process require other bits of information before they can run. For example, the optimized waveform filter requires an estimate of the average signal response and the expected noise, which must be calculated before hand. The gain stabilization requires an estimate of the temperature before the event, for which we use the 1 second pretrigger that we collect with each event.

The entire production is complicated by the fact that we actually have three possible optimized filters and two gain stabilization algorithms. At the end of the first-level data production, we will have six energy estimators for each event, that we will sort out in the second-level analysis.

For each event we also calculate several pulse shape parameters that we will later use for rejecting spurious events (e.g. pile-up events, electronic noise, etc.). We find that all of these pulse shape parameters are energy dependent, which makes it difficult to place cuts

on them without introducing an energy dependence to our signal efficiency. To counter this, we include a pulse shape normalization as part of the first-level production.

Throughout this chapter and the rest of this thesis, we refer to bolometers by their channel number. The mapping between channel number and position in the tower can be seen in Fig. A.1.

Setting Bad Intervals

The first-level data production actually begins with the setting of "bad intervals" in the database. These bad intervals remove periods of time when the detector — or an individual channel — is misbehaving. These intervals can reject events that occur on individual channels or the entire detector, depending on the cause of the bad interval. The majority of the bad intervals set in the data production are due to either noisy intervals with an elevated baseline RMS or periods of unstable baseline oscillations. Figure 4.3 shows a channel that had a period of time with elevated baseline noise removed. Bad intervals can be caused by things like equipment malfunction, DAQ saturation, and of course earthquakes.
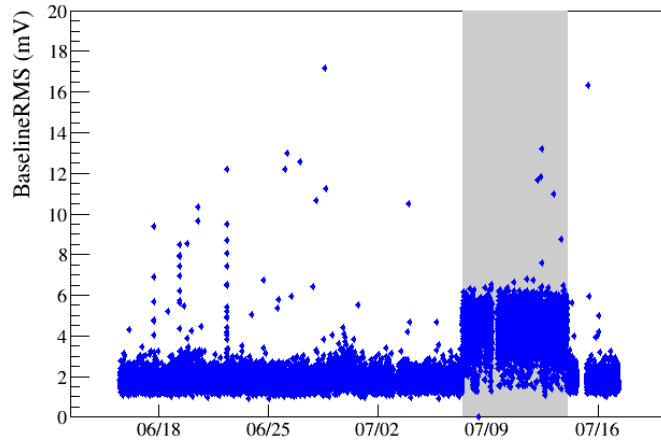
Figure 4.3: Baseline RMS vs time for channel 8 in dataset 2109. The interval in grey is a noisy interval that is excluded. The X-axis spans the length of the background runs in dataset 2109; several runs had to be excluded for this channel when the noise spontaneously increased significantly. Leaving these runs in would adversely affect the average noise power spectrum, decreasing the efficacy of the filter for the entire dataset.

Preprocessing

The preprocess consists of two main steps: measuring the baseline and counting pulses. In the first step, DIANA measures the pretrigger data to determine behavior of the detector

before the signal event occurred. The module measures the baseline voltage — defined as the mean voltage before pulse begins. DIANA then measures the slope of the pretrigger data by fitting a line to the first 3/4 s of data. The baseline slope is used later in the pulse shape analysis. We also measure the "Baseline RMS", which is actually the RMS of the residuals about this line of best fit over the first 3/4 s of the waveform. The Baseline RMS is useful for monitoring the detector performance.

The second important step of the preprocessing is to count the number of pulses in the event. This process begins by differentiating the pulse and determining the RMS of the derivative. Next, peaks in the derivative waveform are determined by looking for local maxima. Anywhere the derivative exceeds five times the RMS is counted as a pulse. The resulting number of pulses is stored. It is worth emphasizing that only positive pulses are counted and stored.

Evaluating the Detector Response & Detector Noise

The waveform filtering, described in the following sections, requires prior knowledge of the detector's response to a signal event and the noise behavior while no signal is present. These are approximated by using averages measured over each dataset.

Building The Average Pulse

The average pulse is measured for each channel in each dataset using the ^{208}Tl 2615 keV events in the calibration runs[2]. The implicit assumption here is that an individual channel's signal response does not vary significantly over the dataset, and this has been shown to be a safe assumption.

The averaging process involves two steps — pulse alignment and averaging — and thus requires two passes through the data. During the first pass, DIANA measures the peak position (in time) of each pulse and determines an average peak time. The peak position is determined by differentiating the pulse and interpolating where the derivative crosses from positive to negative. During the second pass, DIANA aligns all the pulses so that the maximum occurs at the same time. The pulses are aligned by shifting the samples in the relevant direction in memory, dropping samples that are shifted out of range, and approximating new samples shifted into range with a linear extrapolation. Once the pulses are aligned, they are all linearly averaged together with equal weighting. Typically, an average pulse will be built out of a few hundred pulses. An example average pulse is shown in Fig. 4.4.

---

[2]It is worth pointing out that using events that lie in the 2615 keV line requires prior knowledge of the event energy. For this reason, the entire data production process from Preprocess to energy evaluation is repeated twice: the first time using the pulser heater events as a stand-in for the 2615 keV line, and the second time with the actual physics events.
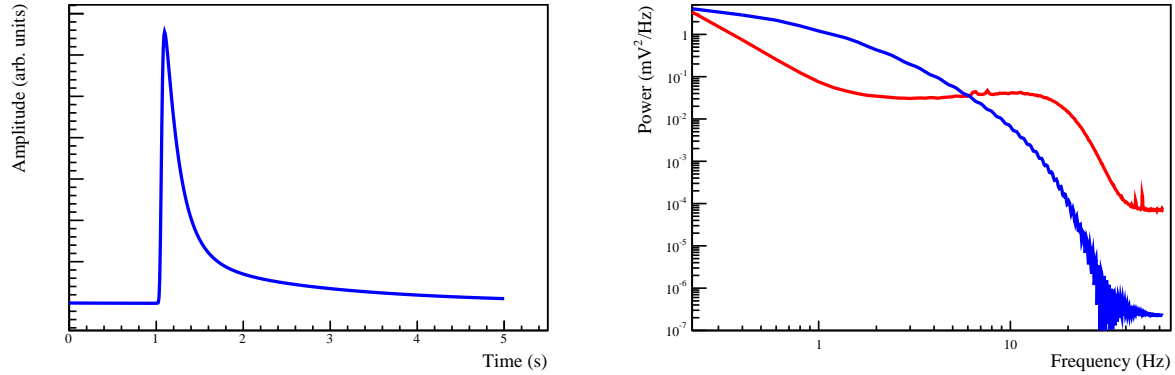
Figure 4.4: *Left:* Example detector response template, $s_{18}(t)$, for channel 18 in dataset 2073. The template is built by averaging many 2615 keV pulses from the dataset. *Right:* In red is an example noise power spectrum, $N_{18}(\omega)$, for channel 18 in dataset 2073. The noise spectrum is built by averaging the power spectra of many noise pulses from the dataset. The signal response frequency spectrum is overlaid in blue. The signal has been normalized to have the same power at low frequency. Note that the noise floor of the signal is much lower than the true noise floor in red, because the signal response is an average over many pulses.

Building The Average Noise Covariance Matrix

The detector noise behavior is understood by studying the noise waveforms and building an estimate of the noise power spectrum and noise covariance matrix. This is done in frequency space, so all noise events are first Fourier-transformed into frequency space using Welch windowing.

The CUORE-0 channel-channel covariance matrix is defined as

$$C_{ij}(\omega) = \langle n_i(\omega) n_j^*(\omega) \rangle \tag{4.1}$$

and describes the covariance between the noise n on channel i at frequency ω and channel j at frequency ω. (This matrix is actually an array of matrices, one for each frequency bin. This is covered in greater detail in Chapter 5.) This can be estimated by evaluating the sum over random noise triggers of the detector,

$$C_{ij}(\omega_p, \omega_q) = \frac{1}{N_{ij}} \sum_{\text{events}} n_i(\omega_p) n_j^*(\omega_q) \tag{4.2}$$

Here the sum is over triggered events of the entire tower, not just of a single channel.

The noise power spectra, can be pulled from the on-diagonal elements of this matrix or calculated explicitly as

$$N_i(\omega_p) \equiv C_{ii}(\omega_p) = \frac{1}{N_{ii}} \sum_{\text{events}} |n_i(\omega_p)|^2$$

An example noise power spectrum is shown in Fig. 4.4. To exclude non-noise events from this sum, events are preselected through a set of event filters, so different channel pairs will generally have a different number of events contributing. This means that while the estimate of the covariance matrix, $C_{ij}(\omega_p)$, is guaranteed to be Hermitian, it is not guaranteed to be rank 1.

Waveform Filtering

The data production uses three parallel filtering techniques to during the amplitude evaluation. The first uses the same "Optimum Filter" technique that was used in Cuoricino, which we denote OF, as well as two new filters which have been implemented in DIANA. One is a re-implementation of the old optimum filter and is denoted NewOF, the other is a generalization of the "Optimum Filter" that accounts for correlations amongst the channels, called DecorrOF. These filters are explained in detail in Chapter 5.

The filtering process is the same for all filters and takes place in Fourier space, by comparing an event's spectral shape to the expected spectral shape of the average detector response, and accounting for the expected noise on the detector. The DecorrOF filter further includes the neighboring channels in the process, by subtracting the correlated components of the neighboring channels from the signal before applying the optimum filter.

It's worth pointing out that the normalization for the old optimum filter and the new filters are slightly different. The new filters are normalized such that the filtered Fourier components of the average pulse are on average unity. In effect, this means that the RMS of the unfiltered noise and the filtered noise can be directly compared. On the other hand, the standard optimum filter has a normalization such that the *measured* gain of the average pulse is unity — and this depends on the amplitude evaluation technique described in the next section. In practice, these two procedures are only different by O(0.1%), but this means that the NewOF pulses and the standard optimum filter needed to be calibrated independently.

More details about the filtering process can be found in Section 5.4 and Appendix B.

Amplitude Evaluation

DIANA evaluates the pulse amplitude on the filtered waveforms in two steps: first locate the pulse peak and interpolate to measure the amplitude. This process is identical for all waveform filters (OF, NewOF, DecorrOF).

The peak identification algorithm searches for the first local maximum within the waveform window after the trigger. The algorithm searches cyclically, (i.e. if the first local maximum appears before the trigger, it will still locate it). Once the peak position has been located, the interpolation takes the data sample identified as the local maximum and its two immediate neighboring samples and interpolates a parabola through them. The amplitude is defined as the global maximum of this parabola, and the time of the peak position is defined as the time of this maximum.

This amplitude evaluation algorithm is performed on each of the three filtered waveforms (Optimum Filter, NewOF, and DecorrOF) and the amplitudes of each is stored. From here on, there are three different amplitudes that need to be considered at each step.

Summary Of Variables After Amplitude Evaluation

At the end of the amplitude evaluation modules, we will have three separate estimates of the pulse height — one for the old optimum filter, the new optimum filter, and the decorrelating filter. We denote these A_{OF}, A_{NewOF} and $A_{DecorrOF}$, respectively. From this point on in the analysis, each variable will follow parallel production paths.

Pulse Amplitude Stabilization

Since the gain of the bolometer depends on its operating temperature, we need to stabilize the gain to counter any time variation in the operating temperature of the detector. In the final CUORE-0 processing, this is done using two different algorithms running in parallel, but both are based on the same principle.

We use a constant energy event to trace the dependence of the bolometer gain on the baseline. Notice that we are NOT measuring the time variation of the gain and interpolating in time. Each event has a measurement of the baseline, and our stabilization algorithm derives a gain function that takes this baseline as input. Thus each event can be stabilized using its own measured baseline. No interpolation in time is involved.
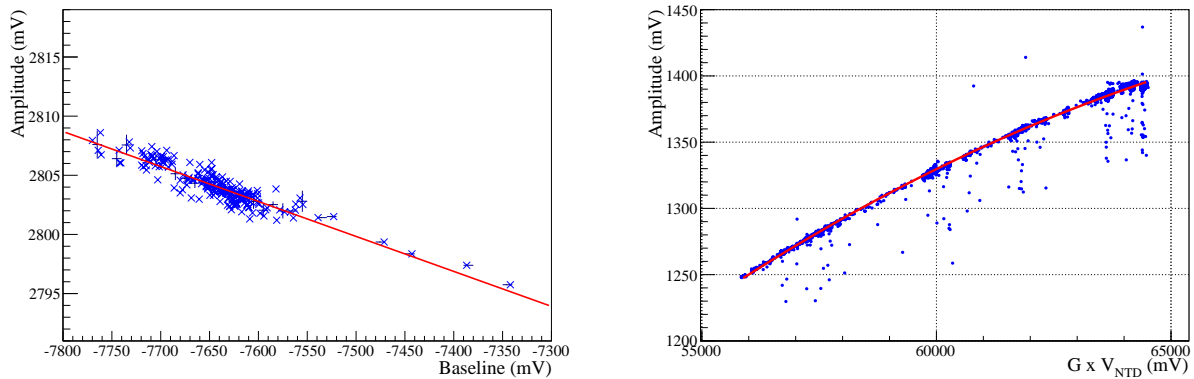
Figure 4.5: *Left:* Pulser amplitude dependence on temperature for a sample of heater pulses for channel 18 from run 201367. *Right:* 2615 keV pulse amplitude dependence on absolute baseline for channel 18 on working point 3. The measured baseline and V_{NTD} are related by $V_{NTD} = V_{offset} - V_{Baseline}$, hence the relative negative sign between the amplitude dependence of these two variables.

Stabilization With Heater

The first stabilization algorithm uses the constant energy heater pulses as a tracer of the gain dependence. The heater pulsing technique is described in section 4.1.

DIANA measures the gain dependence on temperature by a linear regression to the amplitude of the stabilization pulses versus the value of the baseline just before the pulse, which acts as a proxy for the temperature. For most channels, this gain dependence is close to linear over the typical range of temperature variation.

Once the amplitude vs baseline dependence of the stabilization heater is determined for each channel, the amplitudes of all pulses are corrected using the formula:

$$\text{StabAmplitude} = 5000 \times \frac{\text{Amplitude}}{\text{HeaterAmplitude(Baseline)}} \tag{4.3}$$

This correction maps the stabilization heater pulses to the arbitrarily chosen value of 5000 (arb. units).

Stabilization Without Heater

The stabilization without the pulser heater algorithm is based on the same idea as the stabilization with heater, however, instead of using the pulser heater to trace the temperature gain dependence, this algorithm uses the ^{208}Tl events in the calibration runs.

Unlike the stabilization with heater algorithm, the stabilization without heater algorithm determines a single baseline trend for an entire working point — which may span multiple datasets. This algorithm requires a careful accounting of the voltage offset and gain — which are measured in the WorkingPoint runs — to determine the absolute baseline voltage. It then fits a trend of the pulse amplitude vs absolute baseline voltage. Since the stabilization spans multiple runs, and thus a much larger spread in the baseline, the amplitude vs baseline fit is quadratic to account for non-linearity in the temperature gain curve.

The correction equation is the same as the stabilization with heater, except since the stabilization pulses have a known energy — namely the 2615 ^{208}Tl line — they are mapped to an amplitude of 2615 rather than 5000.

More information about this algorithm can be found in [100].

Summary of Variables After Amplitude Stabilization

At this stage in the processing, there are three amplitude variables and two stabilization algorithms, so we can calculate 6 stabilized variables: S_{OF}, S_{NewOF}, S_{DecorrOF}, S_{WoH}, S_{NewOFWoH}, & $S_{\text{DecorrOFWoH}}$.

However, there is one subtlety that's worth pointing out. Since we have 3 amplitude estimators and 2 stabilization algorithms, we *could* derive 6 stabilization trends. That is, we could determine the gain dependence on baseline for each of the estimators in each of the two ways. However, since the three estimators are only different in noise fluctuations and *overall* gain, but *not* in gain dependence on baseline, we simplify the book-keeping by

only calculating *two* stabilization trends — one for each stabilization algorithm on the OF amplitude estimator. Then the 6 stabilized amplitudes are produced by plugging each of the three amplitude estimators into each of the two stabilization trends.

Calibration

The next step after stabilizing the pulse amplitudes is to calibrate the stabilized pulse amplitudes to energies using the calibration runs taken at the beginning and end of each dataset. For each variable on each channel in each dataset, the calibration procedure derives a mapping from stabilized amplitude to energy. In this data production, there are 6 energy variables that need to be calibrated independently, but the procedure is nearly identical for each.

The calibration algorithm can be broken into two parts: peak finding, and fitting a calibration trend. In the first part, DIANA attempts to find seven peaks from the thorium spectrum; these peaks and their origins are detailed in Table 4.1. The algorithm considers the four strongest of these peaks to be "primary" peaks and the other three to be "secondary". If the peak search does not locate at least three of the four primary peaks, it asks for user intervention. If it still can not locate at least three of the four primary peaks, the calibration is considered unsuccessful.

Once enough calibration peaks are located, they are fit with a single Gaussian plus a linear background. Two of the peaks are actually double peaks and are fit with two Gaussians of the same width, and fixed distance between.

The mean of these Gaussians, in units of stabilized amplitude, are used to regress the calibration function. In this production, the calibration trend is determined by regressing the peak means with a second degree polynomial which passes through the origin:

$$E = aS + bS^2$$

Here S is the stabilized amplitude (in arbitrary units), and E is the known peak energy. The uncertainty on the mean from the Gaussian fit to the peak is used as the uncertainty in the S value in the fit. An example is shown in Fig. 4.6. Typically, successful fits have χ^2/d.o.f. of $\lesssim 10$, while fits that have failed completely have χ^2/d.o.f. of $\gtrsim 50$ and need to be redone or excluded.

Summary of Energy Variables

At this point in the analysis, we have produced 6 different energy estimators. We will denote these as E_{OF}, E_{NewOF}, E_{DecorrOF}, E_{WoH}, E_{NewOFWoH}, & $E_{\mathrm{DecorrOFWoH}}$. Later we will compare the performance of these variables to choose a final energy estimator for our analysis.

Pulse Shape Analysis

One of the tools we have for rejecting non-signal events is to analyze the pulse shape of the event. The bolometric response in CUORE-0 is too slow to permit any kind of particle

Table 4.1: γ peaks from thorium daughters used by the calibration algorithm.

Energy (keV)	Source	Peak Type
511	e^+e^-	Primary
583.191	^{208}Tl	Secondary
911.204	^{228}Ac	Primary
964.766, 968.971	^{228}Ac	Primary
1588.19, 1592.533	^{228}Ac, ^{208}Tl Double-Escape	Secondary
2103.533	^{208}Tl Single-Escape	Secondary
2614.511	^{208}Tl	Primary
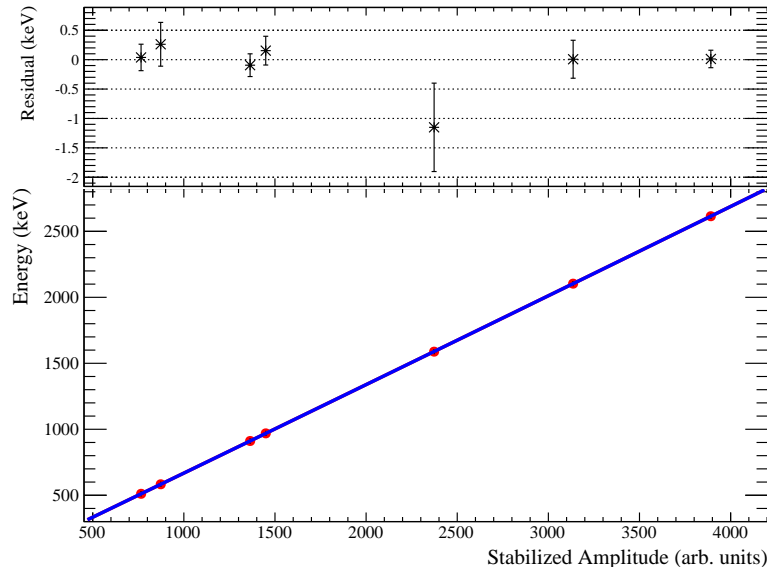
Figure 4.6: Example of a fitted calibration trend (Channel 18 from Dataset 2070), as well as the peak residuals in keV. The uncertainties on Energy, are actually uncertainties in the peak means which have been converted to vertical error bars by dividing by the slope. The large residual at the 1593 keV double escape peak is typical on all channels and likely due to the fact that the peak is overlapping with an ^{228}Ac peak at 1588.2 keV. This may lead to a systematic parabolic calibration bias that we discuss further in Chapter 6.

identification, or bulk vs surface discrimination but we can reject things like noise spikes, pile-up or events that for some reason may not have their energy measured properly.

We measure 6 pulse shape properties:

- The pulse rise time.

  This is the amount of time in ms it takes for the pulse to rise from 10% to 90% of its total height.

- The pulse decay time.

  This is the amount of time in ms it take for the pulse to fall from 90% to 30% of its total height.

- The baseline slope.

  This is the best fit slope of the first 3/4 s of the pretrigger and is measured in mV/S.

- The peak delay.

  This is the time of the measured peak with respect to the beginning of the pulse waveform measured in ms. Typically the signals peak around 1100–1400 ms after the start of the window, and events that deviate significantly from that typically have poorly measured amplitudes.

- TVL and TVR

  These two variables are acronyms for 'Test Variable Left' and 'Test Variable Right' and are basically a χ^2 statistic for how much the filtered pulse looks like the filtered template pulse on the left side and the right side of the peak, respectively.

Typically, the distributions of these variables vary from channel to channel and even from dataset to dataset. Further complicating the matter, the distributions also have a very strong energy dependence. As such, cuts on these variables would need to be tuned for each channel, for each dataset, and carefully as a function of energy. However, instead of doing this we normalize these variables to create a set of Normalized Pulse Shape Variables that have distributions which are — at least in principle — independent of energy, channel and dataset [101].

For each channel all of the runs in a particular dataset are compiled (both Background and Calibration runs) and for each pulse shape variable the distribution is plotted as a function of energy for several peaks in the spectrum: 145 keV, 511 keV, 583 keV, 911 keV, 969 keV, 1588 keV, 2104 keV, 2615 keV, 5450 keV. At each energy peak, the median and median absolute deviation (MAD) is determined and the energy dependence of these median values are fit with a function that interpolates the median value of the shape parameter between the peak values. A similar procedure is used to interpolate the MAD as a function of energy.

Once the median and MAD have been interpolated as a function of energy for each pulse shape variable on each channel in the dataset, all events are assigned their normalized pulse shape variables. These are calculated using the formula

$$\text{NormPSA} = \frac{\text{PSA} - \text{Median}(E)}{\text{MAD}(E)}$$

Which gives the number of median absolute deviations from the median value a particular event lies — with both the median and the MAD now an interpolated function of energy.

Geometric Coincidences

Since the bolometric response time is very slow compared to the light travel time from one crystal to another (or even the traverse time of an α between crystals), a physical event which deposits energy in more than one crystal simultaneously is expected to have the same time, up to a measurement uncertainty. Since our $0\nu\beta\beta$ signal is expected to be primarily contained in a single crystal, we search for coincidences between multiple crystals and reject events with a multiplicity greater than one.

Event A is said to be in coincidence with event B if event B is a signal event with energy over 10 keV ($E_{OF} > 10$ keV) and both have their measured peak times within a narrow window ΔT. No geometric cut is placed on the channels. Any two channels can be in coincidence, even if they are on opposite sides of the tower. This is discussed later in this section. If event B satisfies all of these requirements, then event A is said to be in coincidence with event B. It's worth mentioning, that if event A does not satisfy those requirements, then B need not be considered to be in coincidence with A.

We use the requirement that $E_{OF} > 10$ keV, but in practice this requirement is rarely actually enforced because the low energy threshold is actually dominated by the channel trigger thresholds.

Correcting For Peak Time Jitter

Two events that are truly in coincidence can nevertheless have a relatively large delay between their *measured* times. The primary source of the spread in time comes from the difference in response of different crystals (e.g. the difference in rise time between two bolometers). See Fig. 4.7. To account for this, we run a DIANA sequence to measure the "jitter" between detectors.

DIANA collects a list of 'true coincidences' in the calibration data — i.e. events that occur in two crystals and have a total energy around 2615 keV and are thus likely to be a Compton scatter ^{208}Tl event. These pairs are used to determine the constant time offset between each pair of detectors' response times — the channels' jitter. These constant time delays between detector responses are then accounted for in the coincidence module to achieve a better estimate of the time between events on different channels. This process significantly reduces the size of the time window needed to identify a coincidence. When the jitter is not taken into account, we require a window of $\Delta T = 200$ ms (100 ms on each side), but after the jitter is removed, the window is reduced to ΔT of 10 ms (5 ms on either side). More information on this jitter subtraction process can be found in [102].

For each coincidence event DIANA stores the multiplicity, which is the total number of channels found in coincidence with each other, as well as the total energy, which is just the sum of energies from these channels. It is worth pointing out that for simplicity the coincidence analysis uses only the E_{OF} energy variable, not any of the new energy estimators.
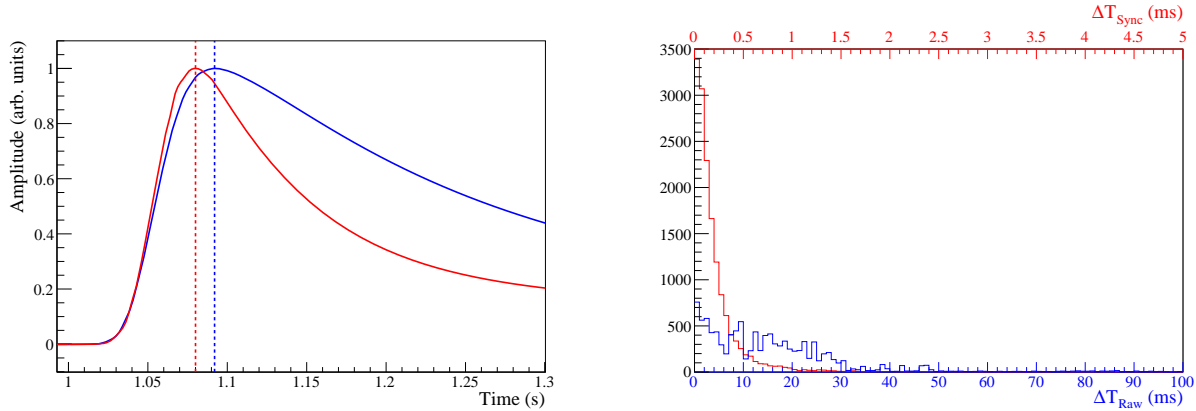
Figure 4.7: *Left:* Average pulses for channels 18 and 20 from dataset 2073. The time delay between the pulse peaks is a consequence of the different responses of the two bolometers. The constant time between the peaks is referred to as the "jitter" and increases the spread in measured time between two events that are a true coincidence. *Right:* Distribution of time delays between pairs of events determined to be in coincidence in dataset 2085 calibration data with total energy around 2615 keV. These events are almost entirely true coincidences. Without accounting for the peak jitter (*blue*) the necessary coincidence window size is 200 ms wide (100 ms on either side); after accounting for the peak jitter (*red*) it improves to 10 ms wide (5 ms on either side).

Spatially Correlated Coincidences

One avenue we investigated briefly was placing a spatial cut on the coincidence events. This is based on the fact that it is highly unlikely a photon would Compton scatter on one side of the detector, traverse all the crystals, and be absorbed on the other end of the detector. We speculated that perhaps we could improve the coincidence efficiency by including a spatial cut, requiring that the two coincident channels must be in close enough proximity to have a coincidence. This is demonstrated in Fig. 4.8, which shows that the $\sim 98\%$ of multiplicity 2 events occur within one floor of each other.

This may prove to be a powerful technique to reject accidental coincidences in CUORE, however in CUORE-0 the rate of accidental coincidences in the background data is too low to warrant a spatial cut. This is because the coincidence jitter subtraction has narrowed the coincidence window to the point that accidental coincidences are not an issue. Thus a spatial coincidence cut is not implemented in CUORE-0.

Data Blinding

The data blinding procedure we employ is a form of data salting, where we randomly move a random fraction of events from the ROI to the ^{208}Tl line and vice-versa. The effect is
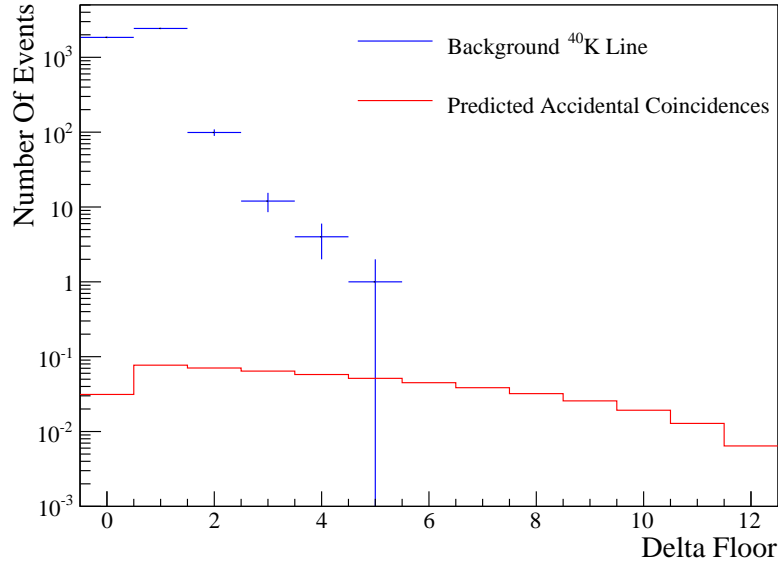
Figure 4.8: Distribution of the distance between floors for multiplicity 2 events with total energy around 1460 keV (*blue*) and the predicted spectrum of accidental coincidences with ^{40}K events.

to create a false peak in the region of interest, masking any possible peak from $0\nu\beta\beta$ or statistical fluctuations.

The data blinding procedure is applied only to the background runs and only to multiplicity one events. The process takes as input a password, which is used to generate a deterministic random seed. From this seed, a random fraction is generated between 1% and 3% — the same fraction for every run in every dataset. For each event between 2527.518 ± 10 keV, a random number is generated — again from a predetermined seed — which will determine whether the event gets shifted. If shifted, the event has a constant 86.993 keV added to its energy variables. Similarly, every event between 2614.533 ± 10 keV has the same probability of being shifted down 86.993 keV.

The process is deterministic, though the exact process is impossible for a human to predict. As such, it is reversible by rerunning the same procedure. The true energy is stored as well but is encrypted using an RSA key pair, with the private key known only to one person.

4.3   CUORE-0 Second-Level Data Processing

The second-level data analysis begins where the first-level production leaves off. At the end of the first-level production, we have six energy estimators for each event. As we will
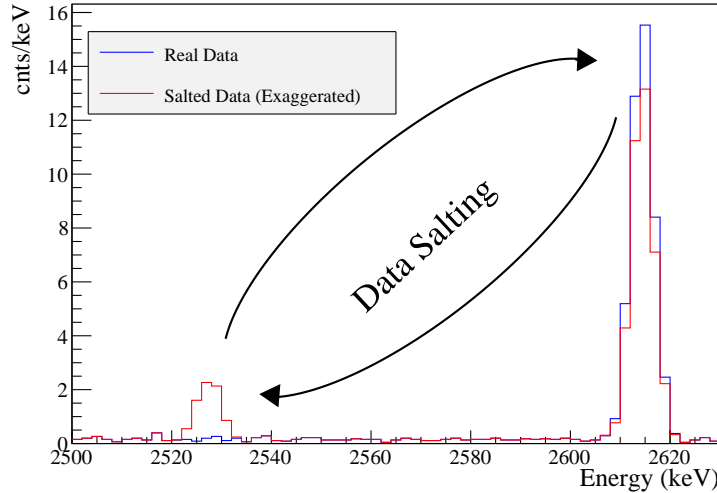
Figure 4.9: A fraction of events from the ^{208}Tl photo peak are shifted down into the ROI, while a fraction of events from the ROI are shifted up into the ^{208}Tl photopeak. Since there are more events in the photopeak, this produces a false peak in the ROI. The fraction of events shifted was blinded, but was chosen to be large enough that the resulting peak in the ROI was unphysically large.

show, occasionally some of these energy estimators fail, occasionally all of them fail. The second-level analysis begins by determining which of these estimators are usable. We phrase this in terms of setting a second set of "Bad Intervals" — similar to where we started in the first-level production — because when an energy estimator fails it fails for an entire channel-dataset pair. We consider like a bad interval of time for an energy variable, and this means that our exposure depends on our choice of energy variables.

The majority of the second-level analysis concerns selecting a set of energy estimators from which we will form our final spectrum. Rather than picking one energy estimator for of all the data, we choose to mix them to take advantage of their relative strengths. The majority of this section focuses on choosing these energy estimators and testing our selection procedure. We also describe the method for evaluating the final CUORE-0 exposure — which becomes entangled in the process of choosing a set of energy estimators.

All of the analysis tuning and decision making were performed on Phases I-III, and Phase IV data was added later. For this reason, many of the results in this section deal with and present only numbers for the first three Phases, and Phase IV is presented separately. All of the final numbers for the $0\nu\beta\beta$ are presented in both Chapter 6 and Appendix A.

Unblinding Procedure

All data selection and analysis occurs on blinded data only. The cuts and fitting technique were tuned on datasets 2049–2124, at which point the analysis procedure was frozen and the data were unblinded.

While the analysis was being finalized and the final paper being written, the CUORE-0 detector continued to collect usable data. Before unblinding in February 2015, we decided that all data collected between October 2014 (the close of dataset 2124) and March 1, 2015 would be added to the final data release. All data collected in these last 5 months would go through both the first-level data production — including blinding — and through the usual data selection procedure, but the event cuts and fitting procedure are not changed after being frozen at the initial unblinding.

Bad Intervals In the Second-Level Analysis

After the initial data production, the data quality is checked and a second set of bad intervals are set. We exclude intervals of time where the stabilization failed, or intervals where the calibration drifted between the initial calibration to the final calibration, or intervals where the filter produced a poor energy resolution. These bad intervals are generally set over an entire channel-dataset pair for a particular analysis approach; the bad interval indicates that that particular analysis approach (i.e. energy estimator) failed for that channel on that dataset.

The important thing to note, and what differentiates the second set of bad intervals from the first, is that by the end of the first-level data production we have several parallel analysis tracks. Since the data are treated differently in each track, we have the situation where a particular interval of time may be bad in one analysis track but good in another. For this reason, bad intervals in the second level analysis are tabulated in a separate DB table called `bad_for_second_level` that includes a bit-flag field specifying for which analysis tracks the interval is bad for.

In CUORE-0 we have two relevant options for flagging bad intervals in the second-level analysis: 'Bad for Standard Stabilization' and 'Bad for WoH Stabilization'. These categories specify that either the standard stabilization routine failed and E_{OF}, E_{NewOF} and E_{Decorr} are unusable, or that the Without Heater analysis failed and E_{WoH}, E_{NewOFWoH} and $E_{\text{DecorrOFWoH}}$ are unusable. Of course, it was also possible that all 6 variables failed and nothing could be used.

Shifting Calibrations in the Standard Analysis

In the standard analysis (OF filter + heater stabilization) we saw that in the case of several channels the baseline dependence of the heater events did not parallel the baseline dependence of the signal events and thus the calibration would drift over the course of a dataset. This is shown in Fig. 4.10. If the center of the ^{208}Tl peaks (as determined by a Gaussian fit)

drifted by more that than 1 FWHM (also determined by a Gaussian fit) between the initial and final calibrations, then the entire dataset was flagged 'Bad for the Standard Stabilization' on that channel. We refer to these channels as shifting channels.

At some point during the CUORE-0 data taking this condition for being flagged as shifted changed from 1 FWHM to $\frac{1}{2}$ FWHM, and this change was not backwards propagated. Needless to say, the inconsistent bad intervals setting was a source of confusion that we hope to improve in CUORE.

These shifting calibration peaks were a reoccurring problem in the CUORE-0 data, and accounted for a loss of 2.7 kg·yr of exposure in the standard analysis variables ($\sim 7\%$). As we will see later, the WoH stabilization routine was very effective in recovering much of this lost data.

One thing we noticed, was that the even-numbered channels were far more likely to shift than the odd-numbered channels. We do not have a definitive explanation for this. The only thing consistently different between even and odd-numbered channels is the chirality of the gluing platforms that attached the NTDs and heaters, but it is unclear how this might lead to calibration shifts in some channels but not others.
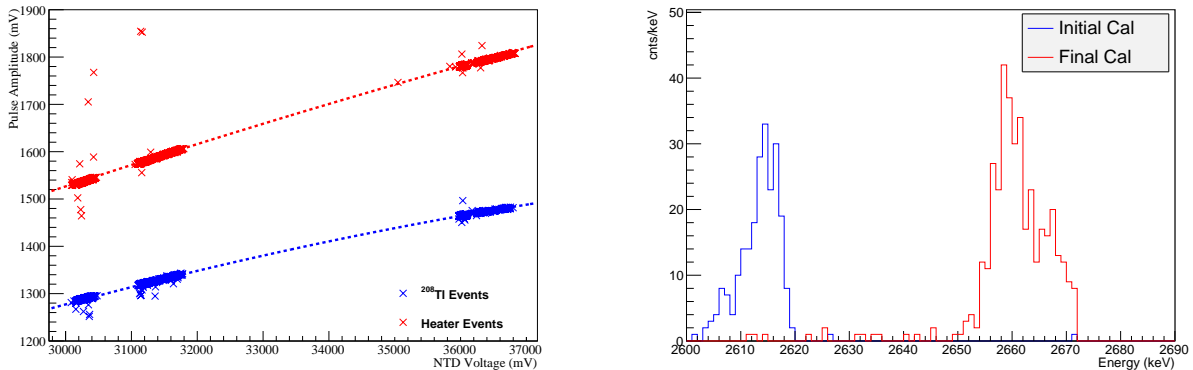
Figure 4.10: Channel 48 in dataset 2049. *Left:* Comparison of the gain dependence on Total NTD voltage for the 2615 keV events and the pulser heater. The relative difference between the two trends results in a shift in the calibration. *Right:* As a result, this channel's calibration shifted by far more than 1 FWHM over the course of a single dataset. This dataset must be removed from the standard analysis for this channel, but as we will see later, this data can be recovered via the Without Heater analysis.

Failed Without Heater Stabilization

The WoH stabilization algorithm depends on accurately knowing the absolute voltage of the baseline and thus the working point of the run. Occasionally, the working point measurement could fail, thus creating a false measurement of the absolute baseline. Alternatively, the

channel setup may be changed in the middle of the dataset, and the baseline gain dependence could change entirely. This issue will be corrected in CUORE, but in CUORE-0 these runs were flagged and are removed from the WoH analysis.

Exposure Evaluation

The background exposure is calculated over all background runs that contribute data to the final result. Typically, this is all background runs that did not crash and have not been manually excluded from the data.

The calculation sums the total runtime for each channel in all background runs, and removes the intervals that have been flagged as bad. All intervals that have been flagged as bad for the first-level analysis are removed from the exposure.

Later, we will combine our six energy estimators into a single spectrum, and this means removing second-level bad intervals in accordance with the energy estimator chosen for channel and dataset. In other words, since the energy estimator may be different for each channel and dataset, we must be careful to properly remove bad intervals in the second-level analysis.

In addition to the flagged bad intervals that are removed from the exposure calculation, the one second and last four seconds of every run are subtracted by default.

Table 4.2: CUORE-0 detector exposures. Phases I-III were used for tuning the final CUORE-0 analysis. Phase IV was added after unblinding, but no decisions were made using this data.

	Exposure (kg·yr)
Phase I	8.1
Phase II	11.7
Phase III	9.2
Phase IV	6.2
Final CUORE-0 Exposure	**35.2**

Calibration Exposure

The calculation of the calibration exposure parallels that of the background exposure, with one exception. Since the calibration runs are often shared between two adjacent datasets, the exposure cannot simply be the sum of the exposures for each dataset. And since the calibration runs may have bad intervals which apply to one dataset but not another, the exposure cannot simply be the sum of the individual calibration runs. Consequently, the calibration exposure is only calculated on a per-dataset basis.

Analysis of the Energy Estimators

No one energy estimator performed consistently better than the others in every situation; each had its respective strengths and weaknesses. Overall, the standard energy estimator

E_{OF} and E_{NewOF} had the fewest major failures resulting in worse energy resolution. The E_{WoH} variable failed completely due to a software bug and thus had to be excluded from the rest of the analysis. A detailed discussion of the decorrelated variables can be found in Chapter 5.

Performance of the New Optimum Filter

The new optimum filter energy estimator, E_{NewOF}, performed extremely well. The only major difference between the new and old optimum filters is a slight difference in amplitude gain of up to $\sim 0.01\%$, which varied over each channel-dataset pair. But within a single dataset, the ratio $E_{\mathrm{NewOF}}/E_{\mathrm{OF}}$ had a typical spread of $\sim 0.001\%$. This is likely due to slight differences in the way each filter calculates its noise spectrum (see Appendix B), but does not rise to the level of concern. This analysis shows that the new optimum filter behaves well enough to replace the old one.
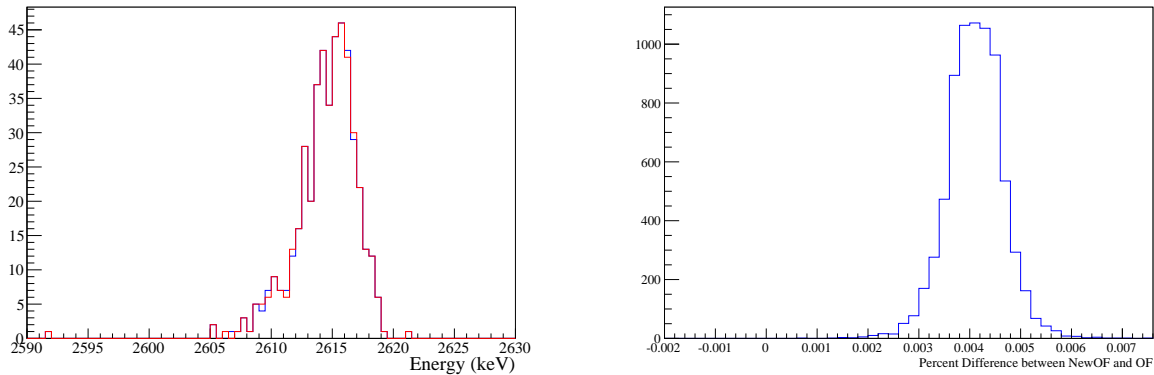
Figure 4.11: Dataset 2073 Channel 18 Calibration Data. *Left:* Comparison of the ^{208}Tl spectra built with the old (*blue*) vs new (*red*) optimum filter. *Right:* Distribution of the percent difference between amplitudes measured with the new and old optimum filters over the whole spectrum. This channel had a difference in gain of $\sim 0.004\%$ with a spread of $\sim 0.001\%$.

Recovering Shifting Channels with the WoH Analyses

We developed the WoH stabilization to solve two problems: first, channels 1 and 10 did not have functioning heaters and thus could not be stabilized with the heater based algorithm, and second, as we have seen above, the standard stabilization algorithm seemed to fail on some channels resulting in calibration shifts.

To these ends, the WoH stabilization was extremely successful. It recovered $\sim 80\%$ of the livetime on channels 1 & 10, as well as many of the shifting channels. Altogether, this resulted in an overall recovery of 12% of the total exposure. On top of this, the stabilization

algorithm improved the resolution on many channels that had shifted slightly, but not enough to be flagged 'shifted'.

Despite its many successes, the algorithm did have a few drawbacks. For non-shifting channels, the energy resolutions produced with the WoH stabilization algorithm were often not as good as those produced with the standard heater based algorithm. There were also a handful of cases where the algorithm simply failed, usually due to unstable or changing working points or software bugs with the new algorithm.
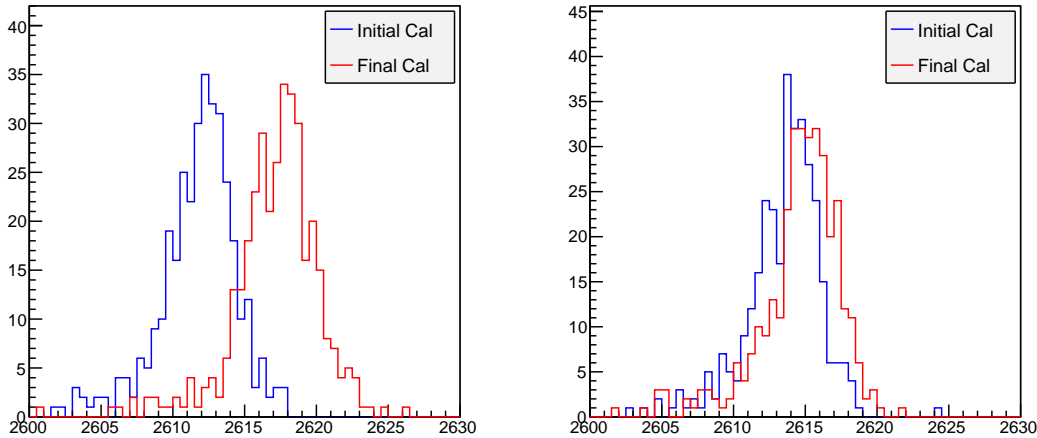
Figure 4.12: Channel 50 initial and final calibrations in dataset 2073 with the standard stabilization (*left*) and the WoH stabilization (*right*).

Combining Energy Estimators

There were situations for each energy variable where it outperformed the others, but there were also situations where each energy variable underperformed. These successes and failures can be directly attributed to different behaviors of the detector i.e. large thermal drifts, correlated noise, etc. Instead of restricting our analysis to a single energy estimator, we chose to mix and match them and thereby take advantage of their respective strengths.

An energy estimator was chosen for each dataset and channel, and the final CUORE-0 spectrum was built by summing these. We decided on three possible combinations that increasingly optimized the analysis while becoming increasingly reliant on the new analysis variables.

- **Ultra-Conservative Approach:** The ultra-conservative approach used none of the new analysis techniques. This is essentially the same approach that was used for Cuoricino. This approach uses the standard energy estimator, E_{OF}. This "tried and true"

approach has the disadvantage of having significantly lower exposure and potentially worse energy resolution.

- **Moderately Conservative Approach:** This approach uses the ultra-conservative approach as a starting point, but incorporates the new energy estimator E_{NewOFWoH} (as sparingly as possible) to achieve the maximum possible exposure. In other words, it uses the new without heater stabilization technique to fill in lost exposure where the standard stabilization technique failed. This recovers channels 1 and 10, which lacked stabilization heaters, and recovered most of the channels whose calibrations had shifted.

- **Aggressive Approach:** This approach utilizes all energy estimators in order to optimize the sensitivity. For each channel and dataset, the "best" estimator is chosen and the final spectrum is built by combining these into a single spectrum.

Creating an Optimized Energy Estimator Combination for the Moderately Conservative Approach

The moderately conservative approach uses the E_{OF} estimator whenever possible, and uses E_{NewOFWoH} for channels 1 and 10 and whenever a channel-dataset has been flagged as "shifted". This was deemed the "conservative" thing to do in that it made as little use of the new variables as possible. However, in practice this approach had many subtle problems.

The definition of "shifted channel" was any channel whose calibration shifted by more than 1 FWHM between its initial calibration run and a final calibration run. The FWHM was measured using a single gaussian fit, which was later determined to be a poor fit to the detector response line shape (we discuss this further in Chapter 6). Because of this, the measured FWHM and mean had a typical statistical uncertainty of 10–20%. This caused our decision of whether a channel shifted to be heavily dependent on statistical fluctuations. At some point during the CUORE-0 data production, the threshold for declaring a channel "shifted" went from 1 FWHM to 0.5 FWHM, and this new threshold was not backward propagated.

Since this inconsistent procedure determined which channels were flagged "shifted". For the Moderately Conservative approach, this directly determined whether the selected energy estimator was E_{OF} or E_{NewOFWoH}. Thus, the moderately conservative approach was left with an inconsistent and unreproducible selection of energy estimators which were heavily influenced by statistical fluctuations.

So while the Moderately Conservative approach maximized the exposure without making too much use of the new variables, there were many nonsensical situations (like those shown in Fig. 4.13) that worsened the resolution, and the approach was difficult to defend.

Table 4.3: Energy estimator breakdown for the moderately conservative approach. (Phases I-III.)

	Number of Channel,Dataset pairs	Exposure	Fractional Exposure
E_{OF}	788	26.62	92.1%
E_{NewOFWoH}	67	2.29	7.9%
Total	855	28.90	
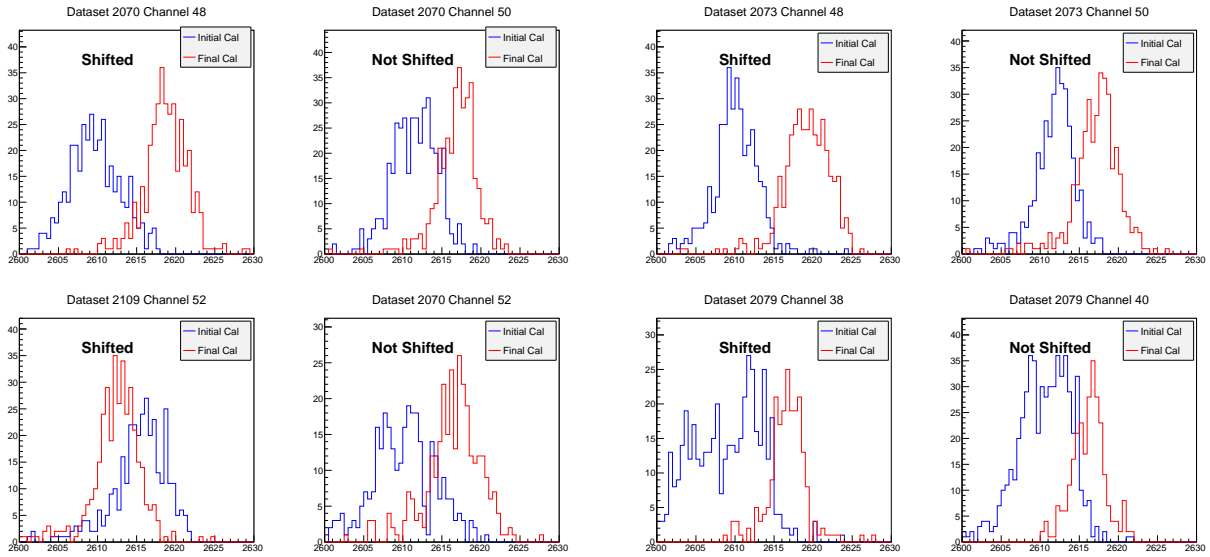
Figure 4.13: The moderately conservative approach was based on a poorly defined and inconsistent definition of "shifting channels."

Creating an Optimized Energy Estimator Combination for the Aggressive Approach

In order to find a combination of energy estimators that optimizes the sensitivity, the "best" energy estimator has to be chosen for each channel in each dataset. We chose the figure of merit for determining "best" variable to be the ratio of the background exposure squared to the variance of the ^{208}Tl line, $T_{\text{Bkg}}^2/\sigma_{\text{Tl}}^2$. The background exposure needs to be included in this calculation because even over the same data, two analyses could have different bad intervals and thus different exposures. Thus it is possible that even if the resolution gets worse it may be advantageous to choose the approach that maximizes the exposure. This combination is proportional to the fourth power of the half-life sensitivity and is easy to calculate.

For each dataset-channel, there are four possible energy estimators: E_{OF}, E_{DecorrOF}, E_{NewOFWoH}, and $E_{\text{DecorrOFWoH}}$. (E_{NewOF} was not included as it was functionally equivalent to the standard energy variable and E_{WoH} was excluded because of a bug in the data processing.)
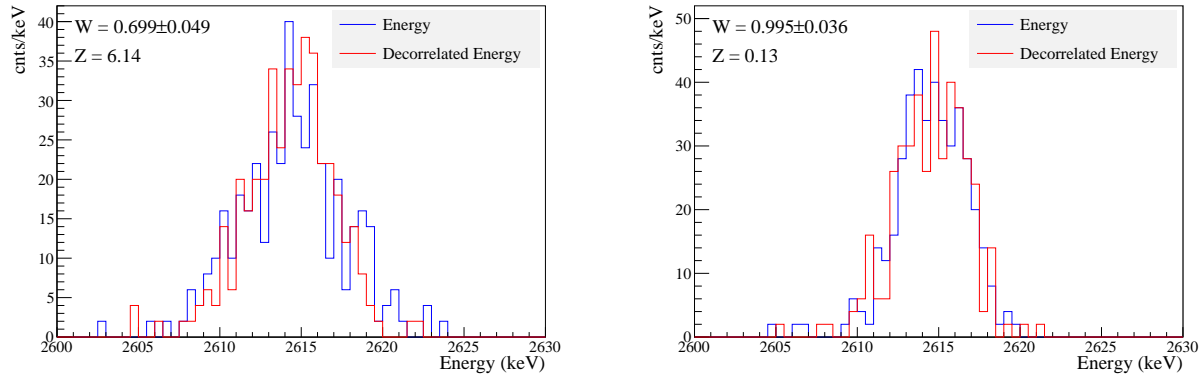
Figure 4.14: Comparison between the standard energy variable and the decorrelated energy variable for dataset 2070 on channels 13 (*left*) and 18 (*right*). Channel 13 saw a statistically significant improvement in the calculated variance. Channel 18 saw a small improvement but not statistically significant. We chose to used the decorrelated energy estimator for channel 13 and the standard energy estimator for channel 18.

The approach to choosing the "best" variable was a single elimination tournament. First we compared the figure of merits for E_{OF} and E_{DecorrOF}. The ratio of the figures of merit was determined,

$$W \equiv \frac{\sigma^2_{\mathrm{Tl}}(E_{\mathrm{DecorrOF}})}{T^2_{\mathrm{Bkg}}(E_{\mathrm{DecorrOF}})} \bigg/ \frac{\sigma^2_{\mathrm{Tl}}(E_{\mathrm{OF}})}{T^2_{\mathrm{Bkg}}(E_{\mathrm{OF}})} \tag{4.4}$$

Its uncertainty, σ_W, was calculated from direct propagation of errors and the number of standard deviations from the null hypothesis (that $W = 1$) was calculated according to

$$Z \equiv \frac{1 - W}{\sigma_W} \tag{4.5}$$

If the figure of merit was improved by at least 2% ($W < .98$), and the number of standard deviations from 1 was more than 1.5 ($Z > 1.5$), then E_{DecorrOF} was chosen, if not E_{OF} was chosen. In other words, if E_{DecorrOF} was at least a 2% improvement *and* the hypothesis that the two approaches are equivalent can be rejected at $\sim$93% C.L., then E_{DecorrOF} was chosen. If either of these was not satisfied, then E_{OF} was chosen. See Fig. 4.14.

This process was repeated between E_{NewOFWoH} and $E_{\mathrm{DecorrOFWoH}}$, and repeated again between the two surviving variables to select the "best" variable. If no winner can be satisfactorily determined, the algorithm defaults to E_{OF} if it has any background exposure or E_{NewOFWoH} otherwise.

This approach is not a statistically robust comparison amongst 4 variances, and the thresholds were chosen somewhat arbitrarily, but this method was fast to implement and worked well enough. The Aggressive distribution of variables is shown in Fig. 4.15.

To be clear, this approach is not guaranteed to be the optimal approach; it is only shown to be better than the Moderately Conservative and Ultra-Conservative approaches. One downside to this approach, is that it implicitly assumes the variances being compared are a good estimate of the resolution of the ^{208}Tl line. However, because of the sub-structure present in the line as well as the Compton continuum, this assumption is not fully true — i.e. events that occur in its Compton continuum can act like extreme outliers and skew the statistics. Thus the comparison, particularly the Z parameter, was still somewhat sensitive to statistical fluctuations in the Compton continuum. This is therefore an area for future improvement.
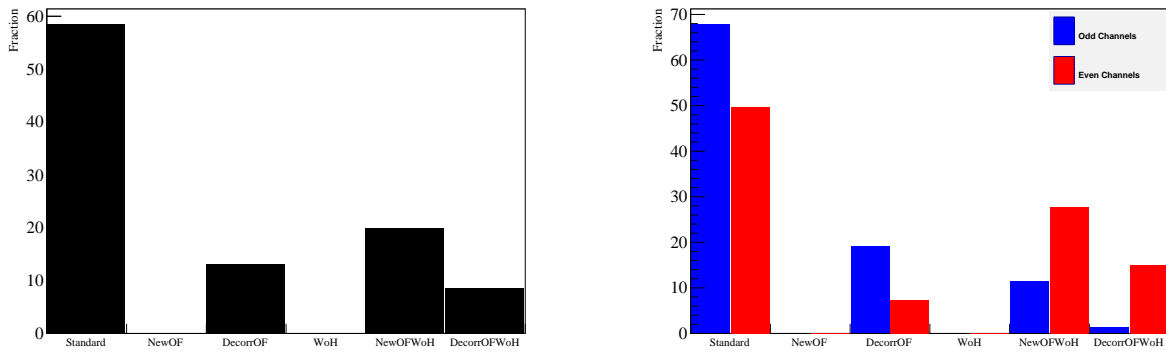
Figure 4.15: *Left:* The distribution of energy estimators chosen by the optimization routine. The standard analysis (OF + heater stabilization) is heavily favored, as it is the default approach. *Right:* The same distribution, but subdivided by even and odd channels. The WoH variables are favored in the even channels as they are more likely to shift — an effect that is still not fully understood.

Table 4.4: Energy estimator breakdown for the aggressive approach. The data included here are for Phases I, II and III, which were studied before unblinding. The full statistics are shown in table A.3.

	Number of Channel,Dataset pairs	Exposure	Fractional Exposure
E_{OF}	496	17.41	60.0%
E_{DecorrOF}	120	3.80	13.1%
E_{NewOFWoH}	171	5.52	19.0%
$E_{\mathrm{DecorrOFWoH}}$	68	2.32	8.0%
Total	855	29.04	

To ensure that this technique was not tuning on statistical fluctuations, we compared the binned resolutions of other prominent peaks: the background ^{208}Tl peak, background ^{40}K peak, and the calibration single and double-escape peaks. Each summed line was fit

with a channel and dataset-independent model PDF built from a sum of two gaussians with identical means — this fit is an empirical fit which produces reasonable results. The FWHM of the peak is determined from the resulting fit. To understand the uncertainties on the FWHM, we determined the errors by bootstrapping. The empirical distribution was randomly resampled and refit O(1000) times and the resulting distributions were compared. These distributions, shown in Fig. 4.17, indicate that the Aggressive approach produces a better FWHM than the Moderately Conservative approach the majority of the time. This lends confidence to our energy variable selection procedure.
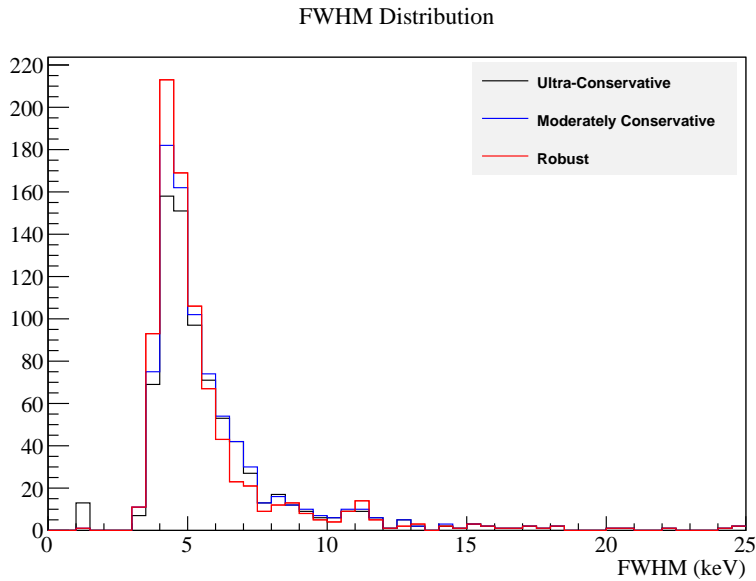
Figure 4.16: Distribution of FWHM values for each channel-dataset pair derived from binned fits for each of the three proposed data combinations.

Summarizing the Three Energy Estimator Approaches

The first point for comparison of the three approaches is the exposure. Since some energy estimators failed on some of the channel-datasets, the ultimate exposure of an approach depends on the particular combination of energy estimators. The exposures for the three approaches are given in table 4.5. Since the Ultra-Conservative approach had a significantly lower exposure, it was quickly disfavored as a candidate for the final CUORE-0 data.

We measured the FWHM of the calibration ^{208}Tl with simple binned fits and on average the Aggressive approach performed $\sim$5% better than the Moderately Conservative at the ^{208}Tl line. The results of these comparisons are detailed in Table 4.5 and Fig. 4.16.

The aggressive approach yielded a $\sim$3% better sensitivity over the moderately conservative approach. Ultimately we chose to use the aggressive energy variable combination approach, and the rest of the analysis proceeded using its variables as our energy estimators.
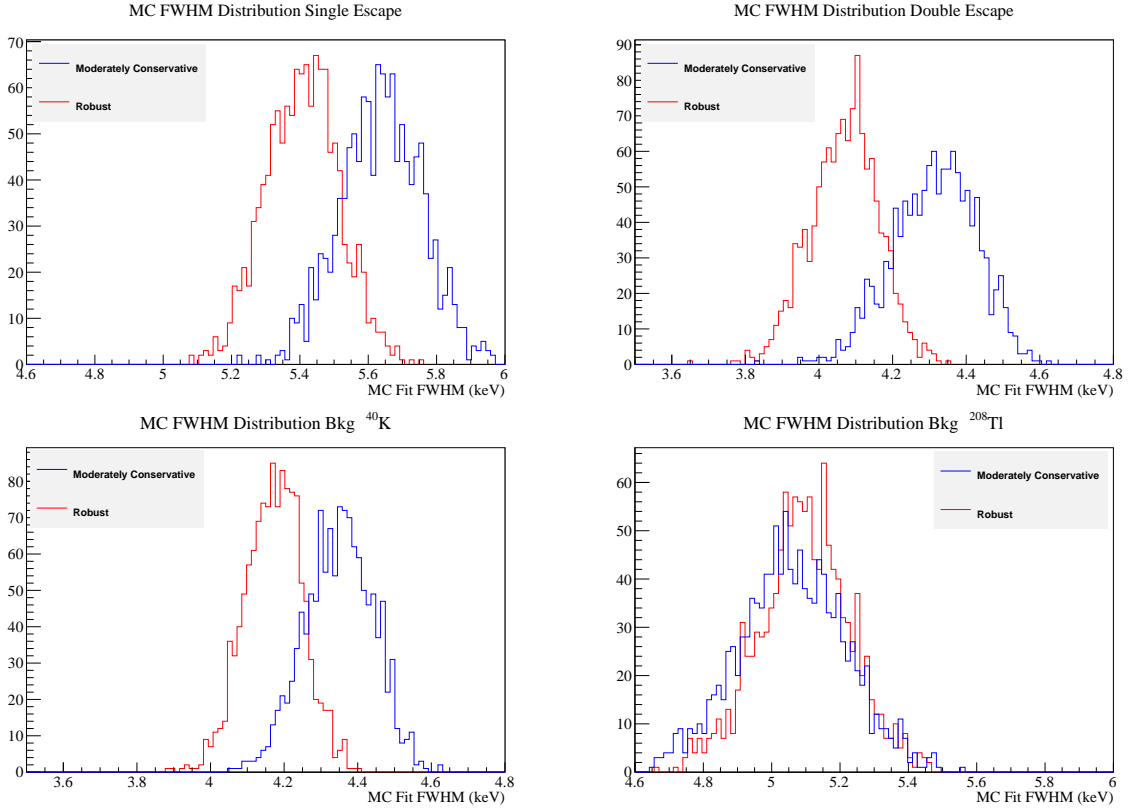
Figure 4.17: Bootstrapped distributions of FWHM distributions for various lines using a double gaussian fit. In blue are the distributions for the Moderately Conservative approach, in red are the distributions for the aggressive approach.

Table 4.5: Comparison of the three energy estimator combinations for Phases i–iii. The FWHM numbers are from binned fits to the ^{208}Tl line. These numbers are less accurate than those from the approach presented in Chapter 6 and are NOT the official FWHM numbers. The "Effective FWHM" is described in Chapter 6.

Approach	Exposure (kg·yr)	FWHM (keV)	Effective FWHM (keV)
Ultra-Conservative	26.8	4.88	5.12
Moderately Conservative	28.9	4.83	5.09
Aggressive	29.0	4.71	4.92

Cuts Tuning

The cuts used for the final $0\nu\beta\beta$ analysis are a combination of basic event cuts, event timing cuts, pulse shape cuts and multiplicity cuts. It is worth pointing out that while none of the cuts depend explicitly on the energy estimators, the PSA variables, the PSA normalization and the multiplicity evaluation all depend on the E_{OF} energy estimator only.

The standard cuts require that an event has been flagged by ABOLLO as a signal event, that ABOLLO triggered no other signal events within the event window, that there only be a single peak within the event and finally that the event not occur during a bad interval.

The event timing cuts create a 7.1 second dead window around a pulse and are designed to prevent overlap between two consecutive events on the same channel. Since a pulse typically takes ∼3 seconds to return to the baseline, we require that any previous event on the same channel have not occurred within a window of 3.1 s before the event. This ensures that the previous event has had sufficient time to return to the original baseline so that the event in question can achieve a reasonable measurement of the baseline. We also require that any subsequent event on the same channel not occur within a window of 4 s after the event. This gives the event in question sufficient time to return to baseline so that we can measure its properties accurately. There is a fair amount of redundancy between these dead window cuts and the cut on the number of pulses within the event window.

We also impose a set of asymmetric pulse shape cuts on the normalized pulse shape variables. These were tuned by maximizing the signal efficiency over the square root of the background efficiency, $\varepsilon_{\mathrm{Signal}}/\sqrt{\varepsilon_{\mathrm{Bkg}}}$. We used several peaks in the background spectrum to stand in as our signal sample and the flat regions around the peak to stand in as our background sample. To avoid statistical bias, these cuts were tuned on Phases I–III using only the half of the data with even-numbered time stamps. In the chapter 6, we evaluate the efficiency on the half of the data with odd-numbered time stamps. The signal efficiency of the resulting cuts was roughly on all peaks from 511 keV to 5400 keV. For more on this, see [101].

The applied Normalized Pulse Shape cuts were

$$-4 < \text{Normalized Baseline Slope} < 4.8$$
$$-4 < \text{Normalized Delay} < 4.8$$
$$-4 < \text{Normalized Rise Time} < 4.8$$
$$-4 < \text{Normalized Decay Time} < 4.8$$
$$-6 < \text{Normalized TVL} < 5.3$$
$$-6 < \text{Normalized TVR} < 5.3$$

The pulse shape normalization failed on a few datasets for a few channels. For these channels the signal efficiency was extremely low, and so the offending cut (usually only one of the PSA cuts) was relaxed to increase the signal efficiency at a small cost to the background efficiency.

Finally, we impose a multiplicity 1 cut to eliminate any events that occurred in multiple crystals.

Chapter 5

CUORE-0 Noise Analysis And Decorrelation

During CUORE-0 we did an extensive study into the sources and behavior of the observed noise and how it fed into the detector resolution. This was originally undertaken in the development of a decorrelating filter, but led to several useful realizations about the noise sources and improvements for CUORE. In this chapter, we dive into deeper detail about the decorrelating filter that was discussed in the previous chapter. We begin by discussing a general analysis of correlated noise in the detector, before focusing on several specific cases of correlated noise and the subsequent insights into weaknesses in our cryogenic setup. We then describe the new filtering module that was developed for CUORE-0 and CUORE to remove the correlated noise and how it was implemented in the DIANA v02.30 processing. Finally, we discuss how it performed on the CUORE-0 data, its weaknesses, and directions for future improvement.

5.1   The Full Noise Covariance Matrix and Subsets

The discussion about the behavior of the noise in CUORE-0 begins with our most general description of the noise: the covariance matrix. The full covariance matrix is defined as

$$C_{ij}(\omega, \omega') = \langle n_i(\omega) n_j^\dagger(\omega') \rangle \tag{5.1}$$

and describes the covariance between the noise, n, on channel i at frequency ω and channel j at frequency ω'. This captures the entire two-point correlation information of the detector. This matrix is both complex and hermitian. Each diagonal element is real and represents the variance of a particular channel at a particular frequency.

The noise of the CUORE-0 detector is characterized by studying the noise triggers. These triggers are fired every 200 s and collected on all channels simultaneously. Because all channels are collected together, we get a snapshot of how the entire tower is behaving for a short period of time. Using this, we can study the correlated noise on the detector and build

an estimate of the covariance matrix. We evaluate the average:

$$C_{ij}(\omega_p, \omega_q) = \frac{1}{N_{ij}} \sum_{\text{Snapshots}} n_i(\omega_p) n_j^\dagger(\omega_q) \tag{5.2}$$

This covariance matrix can be parameterized as an $N_{\text{Channel}} \times N_{\text{Channel}} \times M_{\text{Freq}} \times M_{\text{Freq}}$ matrix. It turns out, that this is too time consuming to compute in full. We therefore present here only a portion of this matrix in as the *correlation* matrix (Fig. 5.1), which is defined as

$$\rho_{ij}(\omega_p, \omega_q) = \frac{|C_{ij}(\omega_p, \omega_q)|}{\sqrt{C_{ii}(\omega_p, \omega_p) C_{jj}(\omega_q, \omega_q)}} \tag{5.3}$$

The most obvious thing to point out about the correlation matrix is that the majority of it is consistent with zero. Specifically, the detector displays very little cross-frequency correlations:

$$C_{ij}(\omega_p, \omega_q) \approx 0 \text{ for } \omega_p \neq \omega_q$$

In other words, the matrix in Fig. 5.1 is pretty close to blocked diagonalized[1]. This is actually expected since we expect cross frequencies to average to zero (i.e. $\langle \sin(\omega t) \sin(\omega' t) \rangle = 0$ for $\omega \neq \omega'$), but mostly this tells us that we do not see a lot of noise harmonics or inter-modulation, which is convenient. The second thing to point out is that the majority of the correlation appears at low frequency, $\lesssim 1$ Hz, with spikes at higher frequencies (here we see the lowest one at ~ 7 Hz). We discuss this low and high-frequency noise separately in more detail in the following sections.

Channel-Channel Covariance

Because the majority of elements in the covariance matrix are negligible — and because the full covariance matrix is too unwieldy to be useful — we consider only a small subset of the covariance matrix by keeping only the channel-channel covariances. In terms of the full covariance matrix, the channel-channel covariance matrix is defined as

$$C_{ij}(\omega_p) \equiv C_{ij}(\omega_p, \omega_p) \tag{5.4}$$

This matrix corresponds to the M_{Freq} $N_{\text{Channel}} \times N_{\text{Channel}}$ sub-matrices along the diagonal of Fig. 5.1, one channel-channel covariance matrix at each frequency.

This covariance matrix is smaller in size by a factor of M_{Freq} and thus much more manageable. Because of its smaller size and the fact that it still contains the majority of the correlation information of the detector, we will largely be working with this subset of the covariance matrix in CUORE-0. Indeed, throughout the rest of this thesis, when we refer to the "covariance matrix", we are referring to this channel-channel covariance unless specifically stated otherwise.

---

[1]The frequency bins that are one off from diagonal (i.e. $(\omega_p, \omega_{p\pm1})$) seem to show a little bit of correlation, but this is actually just an artifact of the FFT and the fact that the waveforms have finite length.
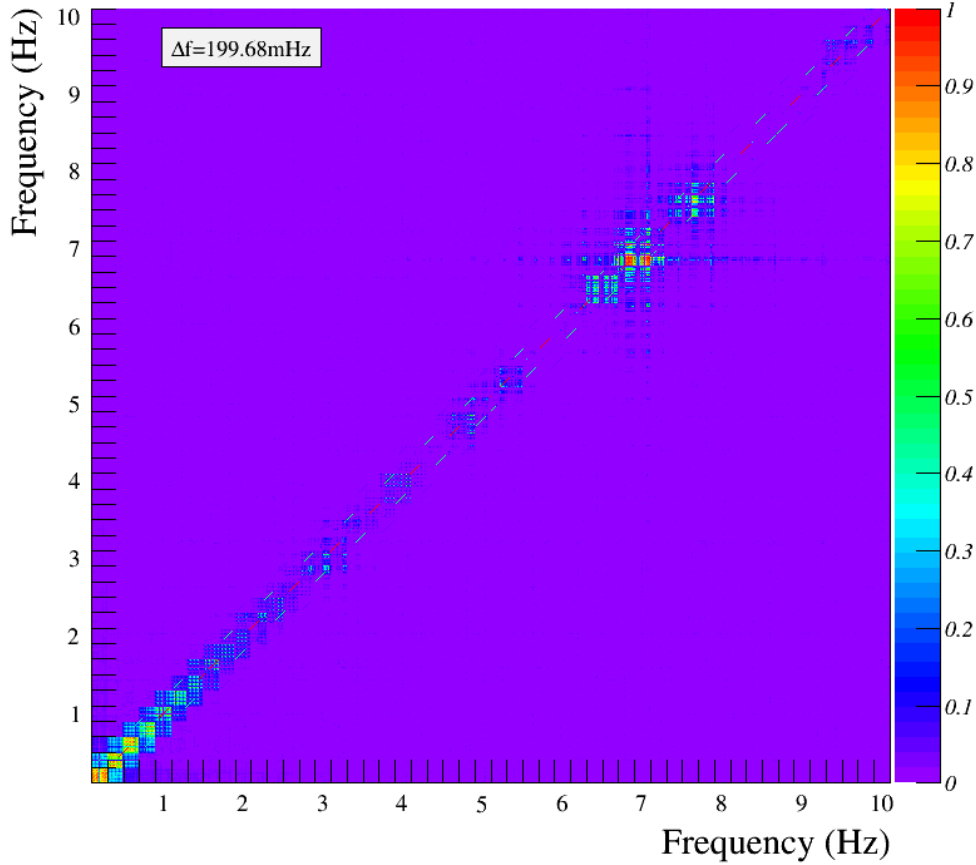
Figure 5.1: The full CUORE-0 correlation matrix, from ds2073 over the range 0–10.2 Hz. Each tick on the X and Y axes represents a frequency bin that is 0.2 Hz wide, and within each frequency box is a 51×51 matrix representing each channel-channel correlation in the CUORE-0 tower. This matrix contains 6.5 million complex entries, and represents only $\sim2.6\%$ of the full covariance matrix.

To be consistent with other works [103], we plot the channel-channel covariance matrix averaged over frequency ranges in the following manner:

$$\bar{R}_{ij} = \sqrt{\frac{\sum_{\omega_p} |\rho_{ij}(\omega_p)|^2 C_{ii}(\omega_p)}{\sum_{\omega_p} C_{ii}(\omega_p)}} \tag{5.5}$$

where

$$\rho_{ij}(\omega_p) \equiv \rho_{ij}(\omega_p, \omega_p)$$

$\bar{R}_{ij}$ can be understood as the average correlation between channel i and j, with the average
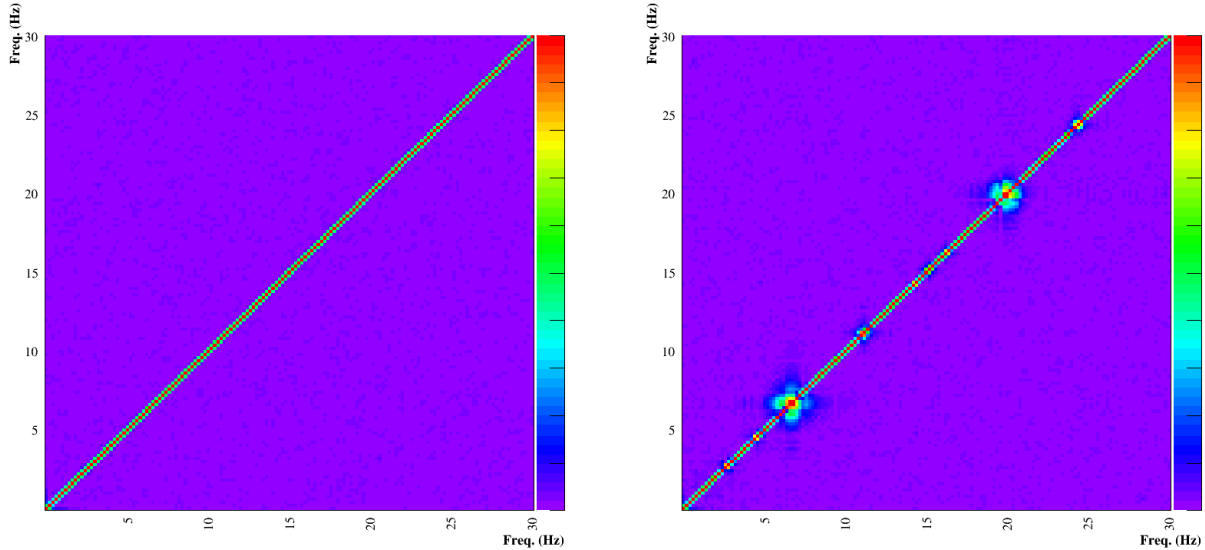
Figure 5.2: Frequency-frequency covariance matrices from ds2073 for channel 18 (*left*) and 43 (*right*). The majority of the frequency-frequency covariance matrices showed similar characteristics to the one on the left, with little or no cross frequency noise, however, a few channels demonstrated a few harmonic peaks like the matrix on the right.

being weighted by the total noise on channel i. As such, $\bar{R}_{ij}$ is not symmetric. Examples of this can be seen in Figs. 5.4 and 5.9 for the low frequency noise and the noise spike at 7 Hz.

Frequency-Frequency Covariance

Another reasonable subset of the covariance matrix is to look within a single channel at the correlations between frequencies. This frequency-frequency covariance matrix is described for each channel in terms of the full covariance matrix as

$$C_i(\omega_p, \omega_q) \equiv C_{ii}(\omega_p, \omega_q) \tag{5.6}$$

This subset forms the basis of a generalization of the optimum filter found in [92], upon which this approach *further* generalizes. The approach in [92] used the frequency-frequency covariance matrix to remove any cross-frequency correlations within a single channel. However, as seen in the full covariance matrix, CUORE-0 exhibited very little cross-frequency correlation (though there are a few channels that are exceptions to this).

In CUORE-0, these frequency-frequency covariance matrices were not used and are only included here for future reference.

Average Noise Power Spectrum

The final subset of the covariance matrix that must be mentioned is the Average Noise Power Spectrum. This is mentioned frequently throughout this thesis because it is the input to the old optimum filter. We denote this spectrum as $N_i(\omega_p)$ and it can be derived from the covariance matrix as

$$N_i(\omega_p) \equiv \frac{2}{f_{\text{Samp}} M_{\text{Freq}}} C_{ii}(\omega_p, \omega_p) \tag{5.7}$$

$N_i(\omega_p)$ is, by definition, real. It is standard practice to normalize it so that if C is in units of mV^2, then $N_i(\omega_p)$ has units of mV^2/Hz and can be readily converted to the amount of *power* per unit frequency. Here f_{Samp} is the sampling frequency. Further, with this normalization it has the property that the mean RMS of the noise on channel i is given by $\sqrt{\sum_{\omega_p} N_i(\omega_p)}$. Ultimately, this normalization is a matter of convenience and has no effect on our final result.

5.2 Correlated Noise and Crosstalk

We will discuss three sources of correlated noise in CUORE-0: high-frequency correlated noise, low-frequency correlated noise, and signal crosstalk. Since the low frequency correlations require a separate investigation we will handle them in the next section.

High-Frequency Correlated Noise

The correlated noise in CUORE-0 was first noticed as noise spikes on many channels in the frequency range 7–20 Hz. This noise seems most correlated between channels that are on the same PENCu readout strips, and has been attributed to microphonics or capacitive pickup on these strips.

A similar correlated noise was seen in Cuoricino, though at a lower level and between fewer channels [103]. However, the Cuoricino electronic readouts did not use PENCu strips and thus were not coupled together as they are in CUORE-0. So it's possible that the correlated noise seen in CUORE-0 has a different origin than pickup on the PENCu cables.

These noise spikes are very narrow in frequency, and tend to lie above the signal frequency band. As such, they actually do not contribute too much to the overall resolution; instead the majority of the noise was induced in the low frequency part of the spectrum.

Signal Crosstalk

So far we have been focusing on correlated noise, but another type of correlation between bolometers is *signal* crosstalk. This is qualitatively different from our correlated noise, in that our noise is common-mode noise that occurs when multiple bolometers pick up noise from a common source, whereas signal crosstalk is when one of the bolometers *is* that source
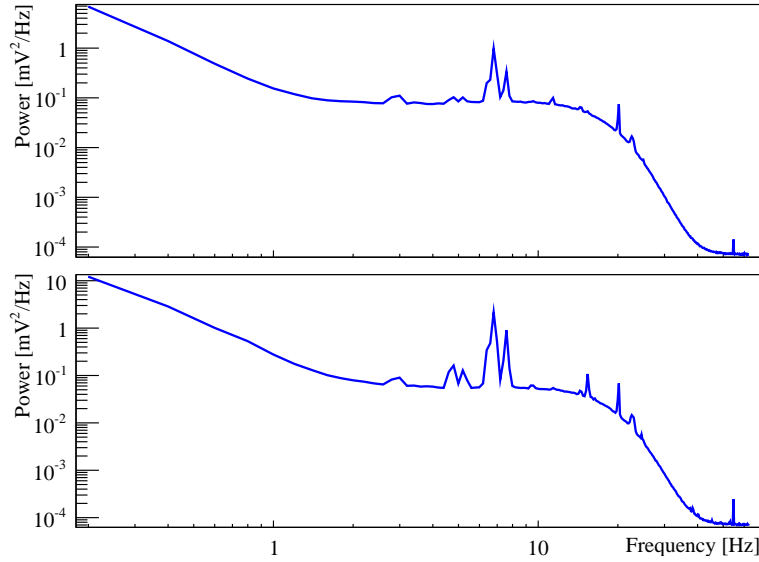
Figure 5.3: Spikes in the average noise power spectra for channels 32 and 34 in ds2073. These peaks appear on many of the channels all over the detector and suggest a correlated source.

of noise. Ultimately, the question is whether a signal is shared between channels, or just the noise.

To test this, we averaged the side pulses of heater events. Specifically, we used the same technique described in section 4.2 to create average pulses for each channel, but instead of triggering on 2615 keV events, we triggered on heater events in channels 1–13. The average pulses of channels 1–13 showed the average of the heater pulses, as expected. But the average pulses of channels 14–52 showed the average behavior of the neighboring channels during a heater event.

The result was a clear temperature rise and fall with slow time constants on almost all neighboring channels. These types of signals are consistent with a gradual heating of the tower from an event on a crystal (i.e. signal crosstalk). However, the typical amount of energy that leaked to the neighboring channels was of order a few keV compared to a signal of $\sim$3 MeV on 13 channels simultaneously. This corresponds to a level of signal cross talk of less than $\lesssim$.1%. It's also worth noting that this approach actually gives an overly conservative estimate of the size of the crosstalk. Because it averages over heater events, and the heater fires in parallel on 13 channels simultaneously, we are actually looking at the crosstalk from a major event on an entire column of the tower. Particle events are typically constrained to one or two crystals so the heating of the neighboring crystals would be smaller. Because of this, we feel confident in assuming that the crosstalk between channels is negligibly small in the signals that we are interested in.
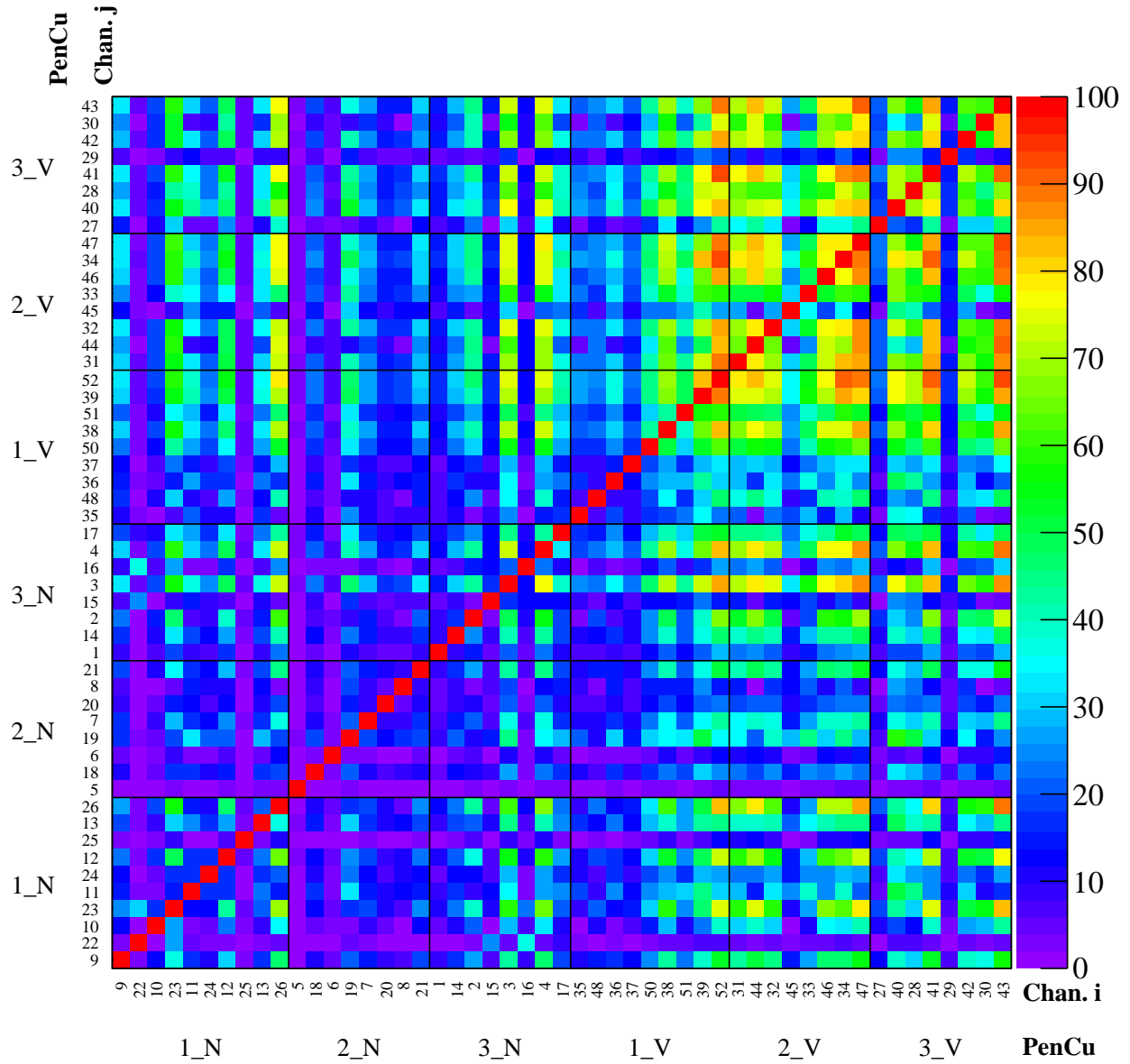
Figure 5.4: The averaged channel-channel correlation matrix for ds2073, averaged over the frequency range 6–8 Hz. The channels have been grouped by PENCu readout cable to emphasize the correlations.
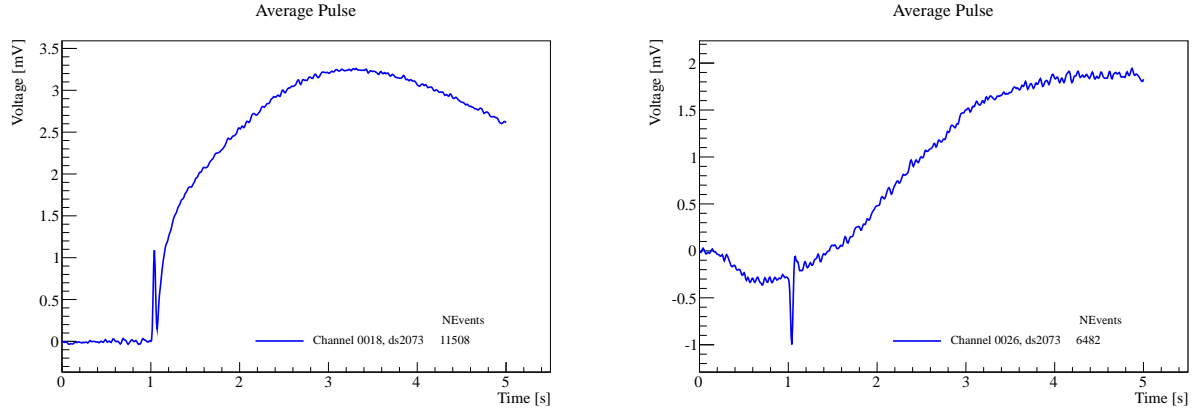
Figure 5.5: Average pulse on channel 18 (*left*) and channel 26 (*right*) that was built out of events triggered when the heater fired on channels 1–13. This shows an obvious temperature rise on the neighboring channels. The amplitude of this pulse corresponds to a few keV, which corresponds to a crosstalk of order $\sim$0.1%. The spike at $t = 1$ corresponds to electronic pickup when the pulser fired.

5.3 Low-Frequency Correlated Noise In CUORE-0

In CUORE-0, we collect and store the continuous waveform data on all channels for possible later use. In order to study the low-frequency noise, we ran the continuous waveform data of dataset 2073 back through APOLLO and retriggered it as if it were new data. We modified the trigger to collect much more noise data by storing 20-second-long waveforms with a 10 s deadtime between triggers[2]. This resulted in a very high statistics sample of noise data with longer waveforms that gave us better frequency resolution. Like the noise triggers in the standard data collection, all channels (including thermometers) were collected together for correlated noise analysis. We processed these retriggered waveforms through the standard DIANA preprocess and average noise modules.

The most immediately obvious observation in these noise spectra was the difference in low-frequency noise between the bottom of the tower and the top. All channels demonstrated a $1/f$ noise spectrum at low frequencies, but the channels at the top showed excess noise above the channels at the bottom by up to a factor of 10. This is seen in Fig. 5.7. The correlation matrix built from this re-triggered dataset (see Fig. 5.9) indicates that not only is there more low-frequency noise at the top of the tower, it is also more *correlated*. This suggests that within this frequency range, all channels are picking up noise from a common source that is stronger at the top of the tower than the bottom. The most obvious candidate is thermal fluctuations from the mixing chamber propagating down the tower.

---

[2]The intention here was to have a 10 second dead time between windows; however, an incorrect software setting led to a 10 second dead time between *triggers*. This meant that every event is overlapped with the previous events by exactly the same amount. However, other than edge cases, this just has the effect of
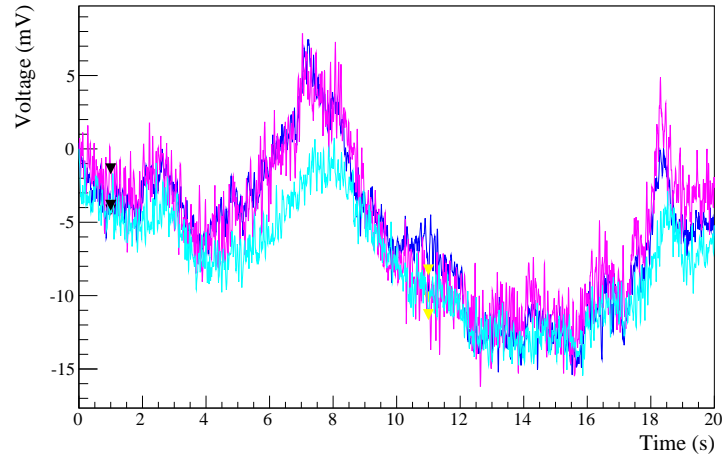
Figure 5.6: Retriggered 20-second-long waveforms collected from dataset 2073 on channels 13 (*blue*), 26 (*red*) and 23 (*cyan*). The channels display a visible amount of correlated noise, with channel 23 (on floor 10) lagging slightly behind the other two channels (top floor).
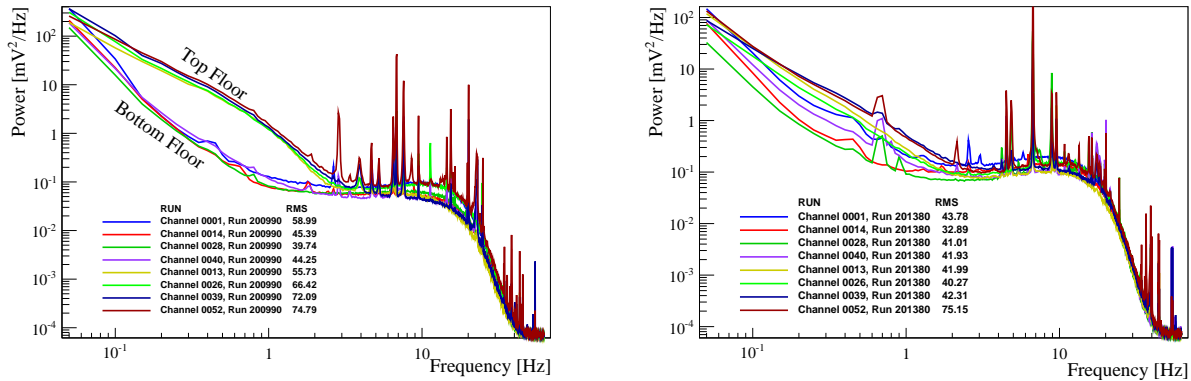
Figure 5.7: Average noise power spectra from the top of the tower vs the bottom of the tower for the retriggered run 200990 from dataset 2073 (*left*) and the retriggered run 201380 from dataset 2085 (*right*). The bottom floor consists of channels 1, 14, 28 & 40 and the top floors are channels 13, 26, 39 & 52.

Low-Frequency Thermal Propagation

Aside from just looking at the magnitude of the correlations, we can actually learn a lot by studying the phase evolution of the correlated noise. Unlike the correlation matrix depicted here, the covariance matrix itself is complex and contains information about the average

---

doubling the data and does not change the conclusions quoted here.

phases between correlated noise on two channels. This can be thought of as the time delay between two channels seeing a common signal.

The phase evolution for channels 14-26 is depicted in Fig. 5.8 and displays a clear trend: the noise on the lower floors lags behind the upper floors and the lag grows faster with increasing frequency. We compare this trend to the phase of the noise at high frequency which appears to display more or less a single common phase. The noise at high frequency is consistent with capacitive pickup, which affects all channels in phase; but the low frequency noise behaves differently.

The leading contender for the source of this behavior is thermal fluctuations propagating down from the mixing chamber through the top of the tower and down to the bottom of the tower. Naively, we can model each floor of the tower as a low pass filter, with some thermal resistance to the next floor, R, and its own heat capacity, C. In this *very* simplified model we can measure the RC time constant, by fitting the phase delay between floors.

$$\Delta \tan(\phi)_{ij} = \Delta N_{ij} RC\omega \tag{5.8}$$

Here, we are fitting the average phase delay between two channel, i and j, as a function of the number of floors between them, ΔN_{ij} and the frequency, ω. Fitting each column of the tower separately, we measure the values listed in Table 5.1. The numbers from each column are in good agreement.

We can perform an *extremely* rough estimate of the RC time constants assuming C to be the heat capacity of four crystals, and the thermal resistance, R, comes from the four copper posts of the frame, yields an expected RC time constant of 15 ms. Though, it should be emphasized that this is an extremely rough model, which could vary by up to an order of magnitude depending on the specifics of the copper and heat conduction.

With the present data, it is impossible to tell if the thermal propagation is proceeding through the copper frame and to the bolometers through the PTFE or through the copper wire strips and directly to the NTDs. There are two points that are worth mentioning. First, though this low-frequency noise was never explicitly discussed in Cuoricino, tests of a decorrelating filter significantly reduced the noise at frequencies below 1 Hz [103], which implies the presence of correlated low-frequency noise. Cuoricino did not use the copper wire strips but rather used Constantine wires which have much worse thermal conductivity than copper. This might imply that the thermal propagation in CUORE-0 proceeds through the frames, as it may have in Cuoricino. Second, as we will discuss later in this chapter, many of the thermometers for monitoring the tower temperature on the CUORE-0 tower did not work or were too noisy to be correlated with the bolometers; but of all the thermometers, the one that correlated most strongly with the thermal fluctuations on the tower was thermally anchored to the mixing chamber rather than the tower. This might imply that the thermal propagation is missing the other thermometers and thus proceeding through the wire strips and directly to the NTDs on the bolometers. These two facts point toward opposite propagation paths for the heat, but neither is conclusive. We may be able to study this further in CUORE or dedicated R&D runs.

Table 5.1: The RC time constants for modeling the CUORE-0 tower as a series of thermal low pass filters. Measured from the covariance matrix in dataset 2073.

Channels	RC (ms)
1-13	38
14-26	29
27-39	30
40-52	31
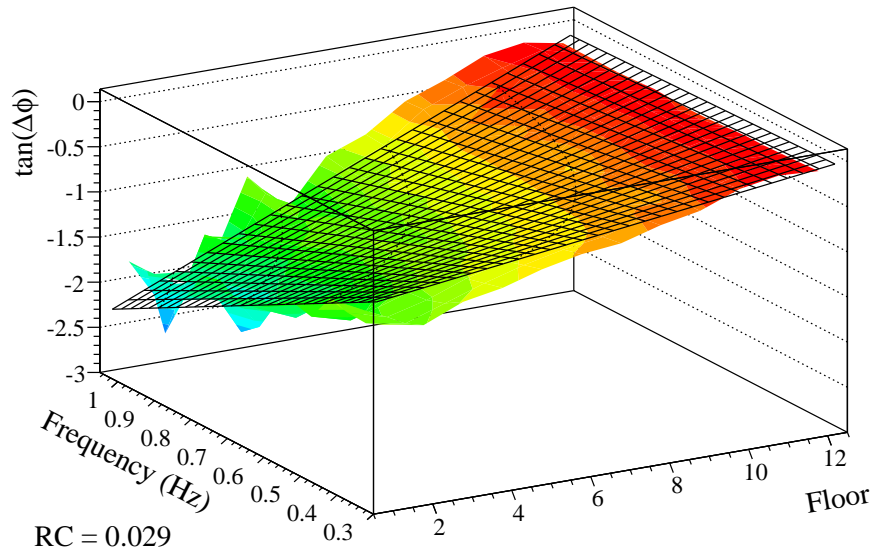
Figure 5.8: Average phase delay for channels 14-26 in dataset 2073. Z-axis shows relative phase between the top floor of the tower (channel 26) and channels in the same column plotted against floor and frequency. Negative values indicate phase delay. We see that channels lag farther behind the top floor the lower they are in the tower. The phase lag grows faster at higher frequency. The contour represents the best fit curve $\tan(\Delta\phi) = \Delta N RC \omega$.
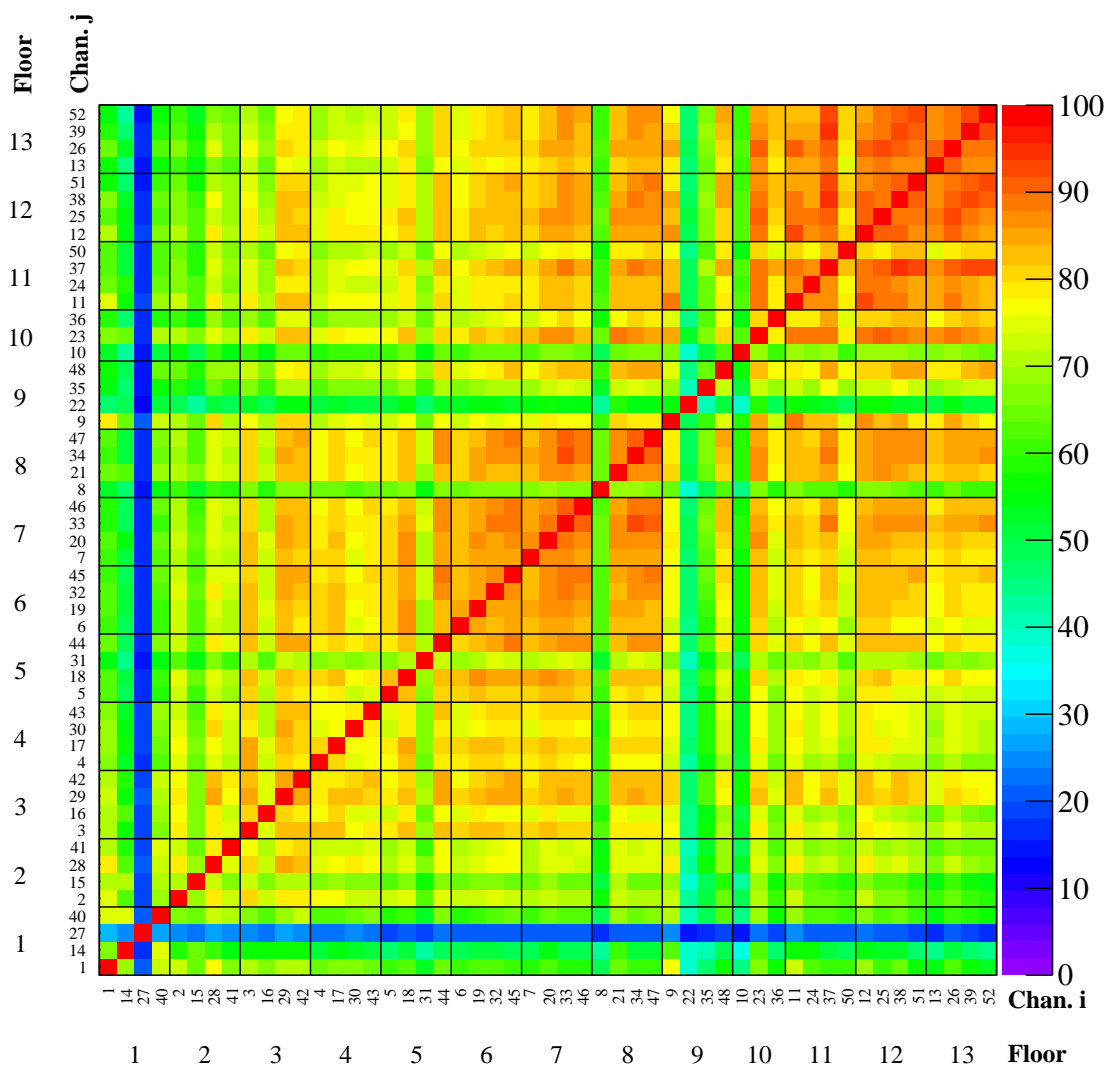
Figure 5.9: Channel-channel correlation matrix for ds2073 retriggered waveforms averaged over the frequency range 0.3–2 Hz. Strong correlations between channels are present at the top of the tower, getting weaker down towards the bottom floors. Channels 22, 27 & 35 show lower correlation than the surrounding channels because of a higher uncorrelated noise level.

Low-Frequency Noise During Campaign I and II

Starting in dataset 2079, we began stabilizing the temperature of the mixing chamber in addition to the tower itself and this lead to a significant improvement in the detector stability and a decrease in the amount of correlated noise. In Fig A.2, we see a significant improvement in the detector behavior after the cryostat maintenance between datasets 2076 and 2079 — but this could also have been caused by the maintenance to the cryostat as well. To examine the behavior of the correlated noise in Campaign II, we analyzed dataset 2085 which — in terms of resolution — was the best dataset in the CUORE-0 data taking, and compared this to dataset 2073 which was among the worst.

The channel-channel correlation matrix from dataset 2085 is shown in Fig. 5.10 and displays a significantly smaller degree of correlation among the channels. Still, the channels at the top of the tower are more correlated than those at the bottom, but to a much smaller extent than in dataset 2073. This is not simply indicative of a lower overall noise level in 2085 — but rather a decrease *especially* in the correlated noise.
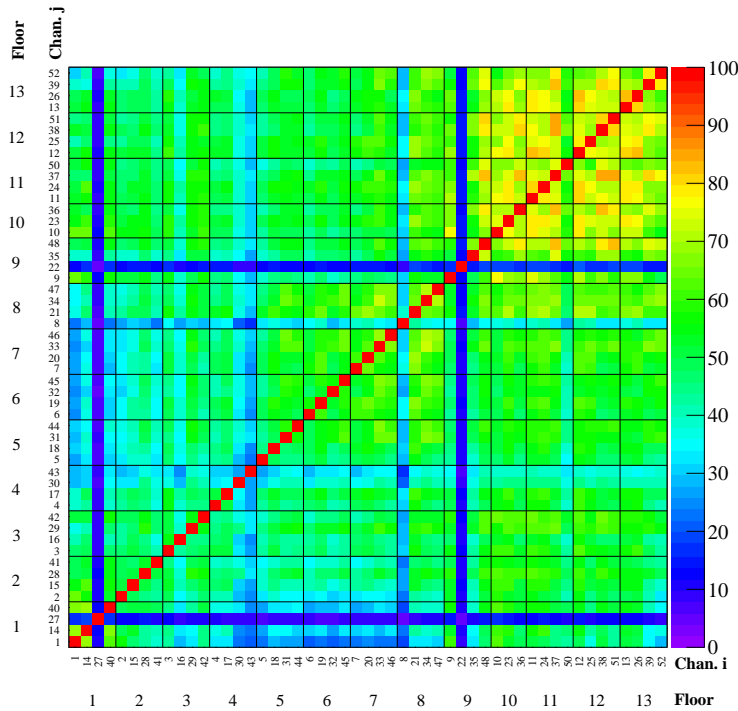
Figure 5.10: Channel-channel correlation matrix for ds2085 retriggered waveforms averaged over the frequency range 0.3–2 Hz. The channel correlations in dataset 2085 display the same pattern as dataset 2073 but at a lower level. See Fig. 5.9.

Temperature Stabilization in CUORE

In CUORE-0, we stabilized the detector temperature using a proportional feedback loop. We fed the output of one of the NTD thermometers through an amplifier and directly back into a heater attached to the tower. The main difficulty with this approach was that the majority of the thermometers that were mounted for temperature monitoring either didn't work, or were too noisy to use for stabilization. As a result, we had to stabilize the temperature of the CUORE-0 tower using a sub-optimal feedback arrangement. In this arrangement, our stabilization heater was on the top of the tower, close to the thermal anchor with the mixing chamber, but our stabilization thermometer was on the bottom of the tower. Due to the phase delay discussed in the previous section, our stabilization thermometer lagged about half a second behind the noise, and was not very useful for stabilizing temperature fluctuations that occurred within the signal band. Though, this feedback system was effective in stabilizing longer term temperature trends (i.e. of order tens of seconds to minutes).
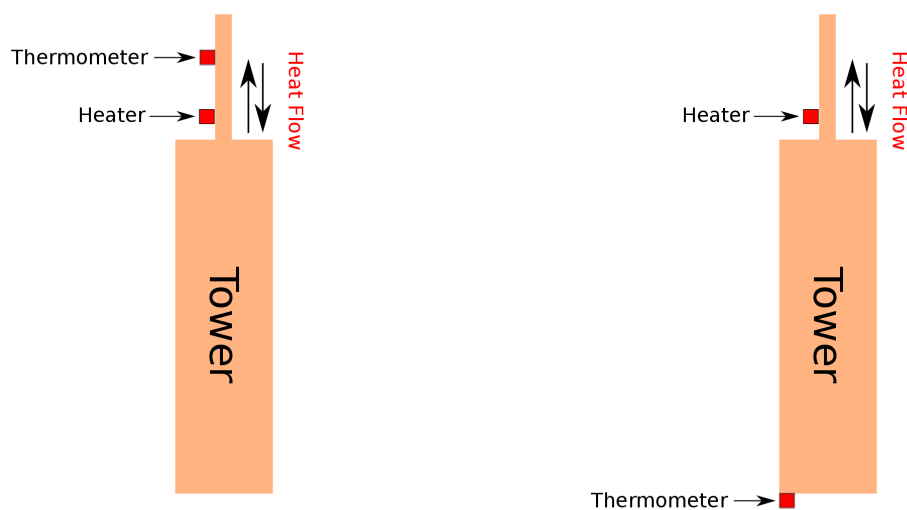
Figure 5.11: Possible temperature stabilization configurations for either CUORE-0 or CUORE. The majority of the heat flow comes in and out through a bottle-neck at the top of the tower(s). *Left:* The optimal feedback configuration has the temperature sensor before the heater so as to measure temperature fluctuations *before* they reach the tower. *Right:* The configuration used in CUORE-0, the temperature sensor did not detect temperature fluctuations until *after* they had already propagated through the tower, and is constantly trying to play catch up.

For CUORE, we are developing a new Proportional-Integral-Derivative (PID) feedback system to stabilize the temperature of the towers against fluctuations in the signal band. Aside from upgrading it to a PID system (as opposed to proportional only), the major focus during the construction phase of CUORE is on ensuring an optimal configuration of the feedback system as well as the quality (and redundancy) of the temperature sensors.

Pulse Tubes

A possible new source of correlated noise in CUORE that wasn't present in CUORE-0 is the Pulse Tube coolers. CUORE will have typically three but up to five pulse tubes running during data taking. These replace the liquid helium bath and 1 k pot in a wet fridge and provide the first stage cooling for the cryostat. This has the potential to be much quieter and stable than a 1 K pot (which is constantly boiling liquid helium and can lead to noise) — however, it also has the potential to induce either thermal or pick-up noise from the operation of the pulse tubes. In the construction of CUORE, we have taken steps to mitigate this noise source by placing as much of the moving parts of the pulse tubes far away from the cryostat.

The pulse tubes typically pulse at a frequency of ∼1.4 Hz which lies right within our signal band. At this point, it's not yet clear what effect this will have on the bolometers, or indeed if this will be a problem at all. But if it does turn out to induce noise, it's likely that the effect will be correlated between all channels. To address this, we could use the decorrelation algorithm described later in this chapter but more likely since the frequency of the pulse tubes is extremely regular, a better option may be to use a very narrow software based notch filter. This will be studied in the first test runs of the CUORE cryostat.

5.4   Towards A Decorrelating Filter

For CUORE-0, we developed a DIANA module that generalizes upon the approaches taken in [92, 104], to create general filtering framework. Here, we describe very broad outline for this filter, but include a more detailed description in Appendix B.

Any linear filter of time series data can be expressed as a filter in frequency space and on an even sampled time series can be expressed in the following matrix equation

$$\boldsymbol{Y}^f = \boldsymbol{W}^\dagger \boldsymbol{Y} \tag{5.9}$$

where $\boldsymbol{Y}$ is a column vector of the unfiltered data, $\boldsymbol{Y}^f$ is a column vector of the filtered data, and $\boldsymbol{W}$ is a matrix of weights[3]. This form can be used to filter data over a single channel's time series or over multiple channel's time series depending on the layout of the data vector and weight matrix. For instance a single channel may be laid out as

$$\boldsymbol{Y} = \begin{pmatrix} y(\omega_0) \\ y(\omega_1) \\ \vdots \\ y(\omega_M) \end{pmatrix} \tag{5.10}$$

---

[3]We express Eqn. 5.9 in terms of the hermitian conjugate of the weights, $\boldsymbol{W}^\dagger$. We do this for later convenience, but the choice of definition is arbitrary.

whereas a group of N channels may be laid out in terms of their single channel layouts as

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_0 \\ \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_N \end{pmatrix}$$

Specifying a filter now amounts to choosing the matrix of weights, $\boldsymbol{W}$ and the process of filtering amounts to solving the matrix multiplication in Eqn. 5.9.

We have implemented this formalism in a new DIANA module. Unlike previous filters in DIANA, which implement specific filters one at a time, this new module is general and evaluates Eqn. 5.9 for an arbitrary set of weights. We have also implemented this class with a set of three filters, the Optimum Filter, a Decorrelating filter, and a Decorrelating Optimum Filter. In this section, we describe the Optimum Filter and Decorrelating Optimum Filter, which were both used in the CUORE-0 data production.

The NewOF and DecorrOF Modules

In the DIANA 2.30 data processing, we used two implementations of this new filtering module, which we refer to as NewOF and DecorrOF. The former is a re-implementation of the same Optimum Filter as was used in Cuoricino [92, 93, 104], and the latter is a generalization of the Optimum Filter which takes into account correlations between channels. For these two implementations, equation 5.9 can be rewritten in component form as

$$y_i^f(\omega_p) \;\; = \;\; A \frac{s_i^\dagger(\omega_p)}{C_{ii}(\omega_p)} y_i(\omega) \tag{5.11}$$

$$y_i^f(\omega_p) \;\; = \;\; B \sum_j s_i(\omega_p)^\dagger C_{ij}^{-1}(\omega_p) y(\omega_p) \tag{5.12}$$

These filters take advantage of prior knowledge of the detector response function, $s_i(\omega_p)$, described in section 4.2, and the channel-channel noise covariance matrix, $C_{ij}(\omega_p)$, described earlier in this chapter. A and B are normalization constants that force the filters to have on average unity gain[4]

The derivation of this choice of weights — and the justification for the name 'Optimum Filter' — is described in detail in appendix B. But functionally, these filters act similarly to a bandpass filter, with a transfer function peaking around 1-2 Hz.

To Differentiate or Not to Differentiate?

One problem with frequency based filtering is that the Fourier transform implicitly assumes that the signal is periodic — with period equal to the size of our window, 5 s. Since our

---

[4]This is different from the old Optimum Filter module (OF) which normalizes the filter so that the *measured* gain is unity — a constraint that depends on the amplitude evaluation technique. But typically, the difference in gain is $\lesssim .1\%$.
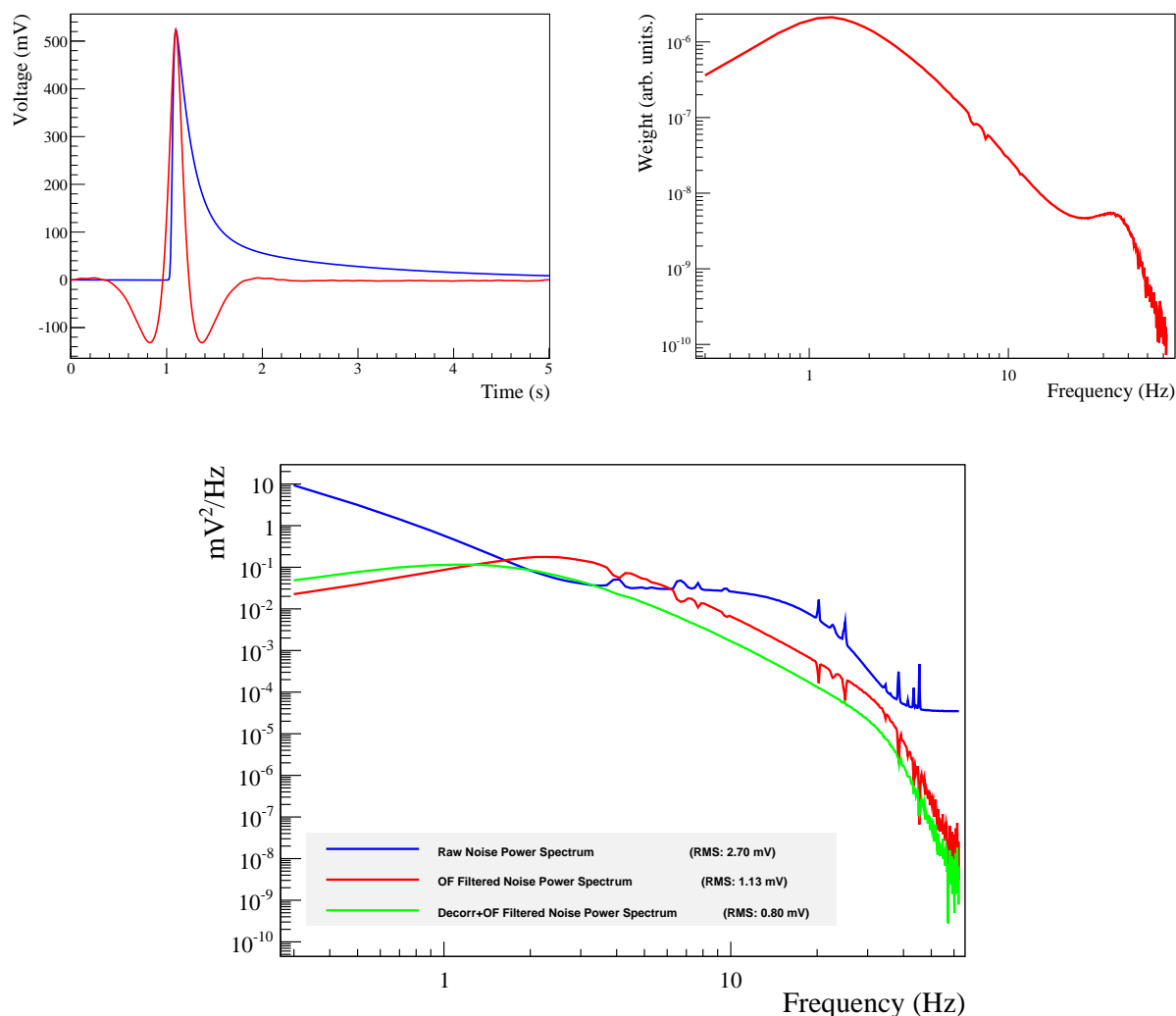
Figure 5.12: *Top Left:* An example of an `NewOF` filtered template pulse for dataset 2073, channel 18. The pulse has been cyclically shifted to align with the unfiltered peak (an artifact of the cyclic nature of the FFT). *Top Right:* Magnitude of the transfer function for the `NewOF` filter for dataset 2073, channel 18. *Bottom:* A comparison of the average noise power spectrum for the unfiltered samples (*blue*) and the predicted filtered noise power spectra for the `NewOF` filtered (*red*) and the `DecorrOF` filtered (*green*) for dataset 2073, channel 13. Channel 13 was strongly correlated with other channels, and so is predicted to be significantly improved by using the decorrelation.

filtering procedure is essentially a convolution, we can run into problems with wrap-around. Here the behavior at the end of the unfiltered signal wraps around and affects the shape at the beginning of the filtered signal. This common problem is often dealt with by zero

padding either the beginning or end of the pulse.

In CUORE-0, we do not zero pad. Wrap-around can become a problem for events with non-zero baseline. In terms of the Fourier transforms, this is akin to adding a signal on top of a saw-tooth function. When passed through a bandpass filter, a saw-tooth function oscillates at the beginning and end of the window, and this deteriorates our energy resolution. This is most easily demonstrated in Fig. 5.13.

This problem can be significantly reduced by filtering the derivative of the pulse rather than the pulse itself. This has two effects, first, any non-zero baseline slope is now reduced to a DC offset. Second, it makes the effective length of the signal significantly shorter; this is because the derivative returns to zero significantly faster than the signal itself. This second feature also allows us to avoid the need to zero pad, as the derivative is already effectively zero padded.

To do this, first we can trivially differentiate the average pulse

$$s'(t_{p+1}) = s(t_{p+1}) - s(t_p)$$

however, we must also transform the covariance matrix (or average noise power spectrum) appropriately to account for the differentiation. We use the property of the DFT that

$$\mathcal{F}\left(n(t_{p+1}) - n(t_p)\right) = \mathcal{F}\left(n(t_p)\right)\left(e^{\frac{2\pi i}{M_{\text{Freq}}}p} - 1\right)$$

to show that the covariance matrix is transformed to

$$C_{ij}(\omega_p, \omega_q) \rightarrow 2\left(1 - \cos\left(\frac{2\pi}{M_{\text{Freq}}}p\right)\right)C_{ij}(\omega_p, \omega_q)$$

5.5 The Decorrelation Procedure in the Diana v02.30 Production

The decorrelation procedure has been designed to be incorporated into the second iteration of the standard CUORE-0 data processing. This means that rough estimates of the stabilization and calibration parameters were produced previously and are accessible.

Grouping Channels for the Decorrelation Algorithm

To remove the common mode noise from a particular signal event, the decorrelation algorithm needs several samples of the common mode noise in addition to the signal waveform to be decorrelated. In principle, we could input all of the other channels in the tower for maximum noise removal. However, decorrelating each event against all channels in the tower can be very computationally intensive. In CUORE-0, we reduce the size of the problem by decorrelating with only the waveforms of the geometric neighbors which were collected along with every
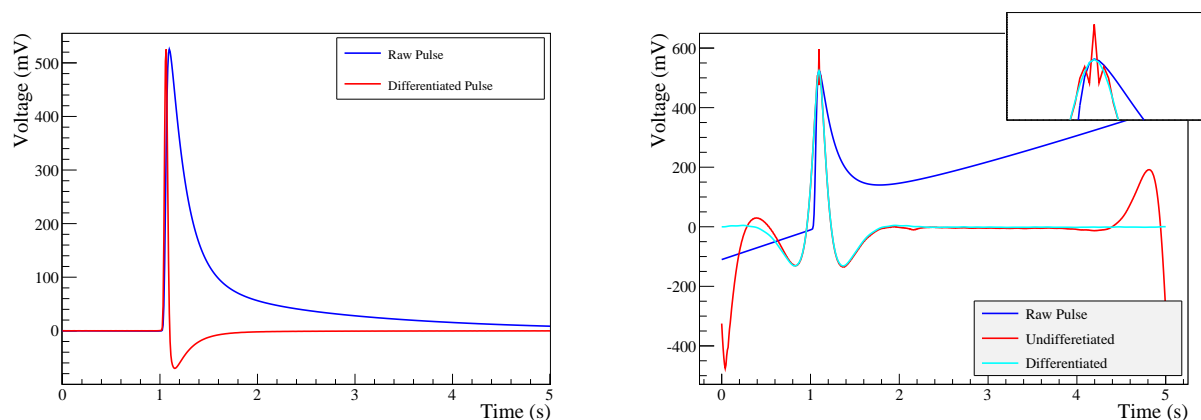
Figure 5.13: *Left:* A pulse vs a differentiated pulse, normalized to the same amplitude. *Right:* A pulse with an exaggerated slope that has been passed through the Optimum Filter. The same pulse is also passed through after having been differentiated. The inset zooms in on the top of the pulse.

signal event. These 'side pulses' were originally collected for the coincidence analysis, but are also useful for decorrelation.

For a channel in the middle of the tower the geometric neighbors were the other three channels on the same floor, the four channels on the floor above, and the four channels on the floor below. Channels on the top or bottom floor of the tower simply had 4 fewer geometric neighbors. This defines 13 overlapping groups of crystals that are all decorrelated together — 11 groups of 12 channels and 2 groups of 8 channels. From the point of view of matrix inversion, this grouping allows us to reduce one large matrix, into 13 smaller matrices by ignoring correlations between channels that are more than one floor apart.

Since the majority of the correlated noise comes from thermal fluctuations, this particular grouping of channels is nice, in that it groups channels that are strongly thermally linked. However, there is one caveat. The decorrelation assumes that the side channels contain samples of noise and only noise. If a side channel instead contains a particle event, it must be excluded from the decorrelation. Unfortunately, the choice to use geometric neighbors maximizes the probability that a side channel will have a coincident signal event. This is an unavoidable consequence of this grouping.

An Alternate Possible Grouping

An alternative grouping is to group channels by the PENCu readout cables rather than geometric nearest neighbors. Typically, a PENCu cable joins about three to five floors on one side of the tower. This grouping would be more ideal for removing the noise induced by the capacitive pickup on these cable strips.

However, this pickup results in narrow peaks around 7 Hz - 10 Hz. This is above the

signal band and so does not contribute as much to the noise as the thermal fluctuations at lower frequency. Further, when the data was collected, the geometric neighbors were stored in the triggered data files, so grouping by PENCu cable would require re-triggering the data from the continuous waveforms. Because of this we decided to use the nearest neighbors as the grouping for decorrelation.

The Decorrelation Module

The decorrelation itself is done in with a new DIANA module. As input, this module takes the average pulses from each channel, the channel covariance matrix, and a file which specifies the above grouping of channels for decorrelation. It then breaks up the covariance matrix into smaller sub-matrices and inverts them to build the weights that appear in the sum in Eqn. 5.12. To save time, these weights are built once before the module begins processing the list of events. Once the weights have been evaluated, DIANA begins to process the list of triggered events in order. For each triggered event, the side pulses that APOLLO wrote with the event, must be 'unrolled' and processed before the decorrelation sum can take place.

Here, we encounter the most complicating step in the decorrelation process. The algorithm explicitly assumes that the side channels are accurate snapshots of the *noise* on the tower at the time of the trigger. If instead a side pulse contains its own signal event, then the decorrelation would smear it into the primary channel being decorrelated thus worsening the resolution.

To prevent this, we must exclude all side channels with any signal event *anywhere* within the 5 second window. This is not just an anti-coincidence cut, where we exclude any associated signals within a small window of the primary trigger. This is a much more aggressive cut that removes any side channels with even an *accidental* coincidence *anywhere* within the 5 s window.

In the DIANA 2.30 data production, we used the requirement that the side channels have no pulses anywhere within the window with |Energy| > 12 keV. By placing a cut on the calibrated *filtered* amplitude of a waveform, we are specifically limiting the amount of energy that can be smeared between two channels *after* filtering. In other words, this limits not just the highest sample or derivative in the window, but the amplitude of the full detector response shape, $s_i(t)$. Though, this has the undesirable overheard of needing to filter every side pulse. In CUORE, we intend to try to improve on this cut.

The decorrelation itself proceeds in four steps: Differentiate the signal waveform and all relevant side channel waveforms, transform into frequency space, evaluate the sum in Eqn. 5.12, transform back into time domain. This process is fairly straight forward, though somewhat computationally expensive.

The only subtlety here is how to treat side channels that have been excluded from the decorrelation. The problem is that with channels missing, the sum in Eqn. 5.12 can not be completed. The correct way to handle this would be to return to Eqn. 5.12 and re-evaluate the weights starting from a subset of the covariance matrix where the missing channels have been removed. However, this potentially requires re-inverting a covariance sub-matrix for

a significant fraction of the event. This would be *significantly more* computationally time consuming. Alternatively, we could try to precalculate all possible combinations of sub-matrices, but this is impractical. One possible future avenue, which has not been tried yet, is to derive the inverse of the covariance sub-matrix from the inverted covariance-matrix itself, which can be done much faster.

In the DIANA v02.30 processing we simply evaluate the sum *without* the missing channel. This is the fastest solution, but comes at a cost to the final energy resolution. We discuss the consequences of this choice later in this chapter.

Amplitude Evaluation, Stabilization & Calibration

The amplitude evaluation proceeds just like the standard OF described in the previous chapter. The decorrelated amplitudes are stabilized using the parameters that are calculated on the standard OF amplitude. This is done for two reasons: first, since the decorrelated filter and the optimum filter provide unbiased estimates of the true pulse amplitude, they should have the exact same gain dependence on temperature. Second, since the heater pulses are fired in unison for an entire column of the tower and in reasonably quick succession, the decorrelation of the heater pulses is ALWAYS missing two channels, and is thus intrinsically different from the signals.

The decorrelating filter displayed a slightly different gain from both the OF. This was not unexpected, but meant that the calibration coefficients needed to be calculated separately for each of the filtering techniques.

5.6 Decorrelating Filter Performance

On whole, the decorrelation procedure worked very well in the DIANA data production. However, at the end of the data production, the results were somewhat underwhelming. Several things became clear. First, the correlated noise is not the primary driver of the energy resolution at 2615 keV. Second, our handling of pulses on neighboring channels was insufficient, and this has a major impact on the performance of the filter. Ultimately, the decorrelation filter had only a marginal effect on the energy resolution on the 2615 keV ^{208}Tl peak in the calibration data, but did significantly improve the noise level in the background data. On some channels, decorrelation improved the energy resolution at 0 keV up to ~25%. This could be useful for improving the low energy thresholds for anti-coincidence analyses or even Dark Matter searches.

Performance on the Noise During Background Runs

Focusing on the noise data collected during background runs, we see that as expected, the decorrelation has the most significant impact on the channels at the top of the tower. Some channels at the top of the tower had their baseline noise improved by ~15-20%. This is
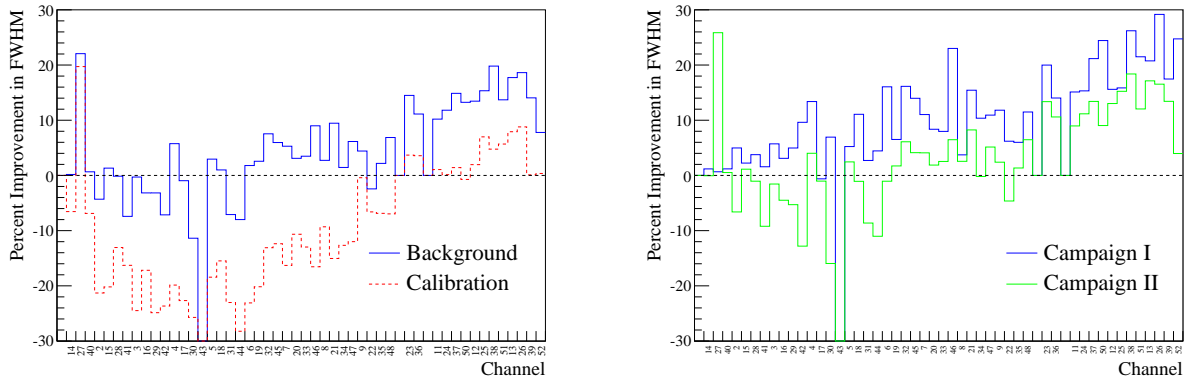
Figure 5.14: Percentage improvement of the decorrelating filter over the standard optimum filter as measured by the FWHM at 0 keV. Channels are ordered from the bottom of the tower on the left to the top of the tower on the right. The decorrelation filter performed significantly better on the top of the tower compared to the bottom. Overall about 2/3 of the channels saw improvement in the background data, but most got worse in the calibration data (*left*). In the background data, we saw more improvement in Campaign I than II (*right*).

consistent with what we expected since the channels at the top of the tower are the ones most directly affected by the thermal fluctuations.

Comparing the performance between the two campaigns, we see that Campaign I showed significantly more improvement from the decorrelation, with some channels improving by as much as $\sim 30\%$. Again, this is as expected. Starting at the begin of Campaign II, we began stabilizing the mixing chamber temperature with its own feed back loop and the amount of correlated noise decreased.

In Fig. 5.15, we show a comparison between the expected filtered noise power spectrum and the measured noise power spectrum after the filter. For the background data, they line up nearly exactly, which gives us confidence that the decorrelating filter is working as it should. We see that the NPS turns up over the predicted curve above a frequency of $\sim 30 - 40$ Hz. We believe that this has to do with the fact that the template pulse (which is built by averaging many pulses) runs into its own noise floor. See Fig. 4.4.

Performance on the Noise During Calibration Runs

Focusing on the noise data collected during calibration runs, we see significantly worse performance. In Fig. 5.15 we show a comparison between the measured filtered NPS and the predicted for the background and calibration runs. While the background runs closely follow the prediction, the calibration runs show excess noise at low frequency, which ruins the energy resolution. There are three possibilities as to why this could be, which we briefly explain:
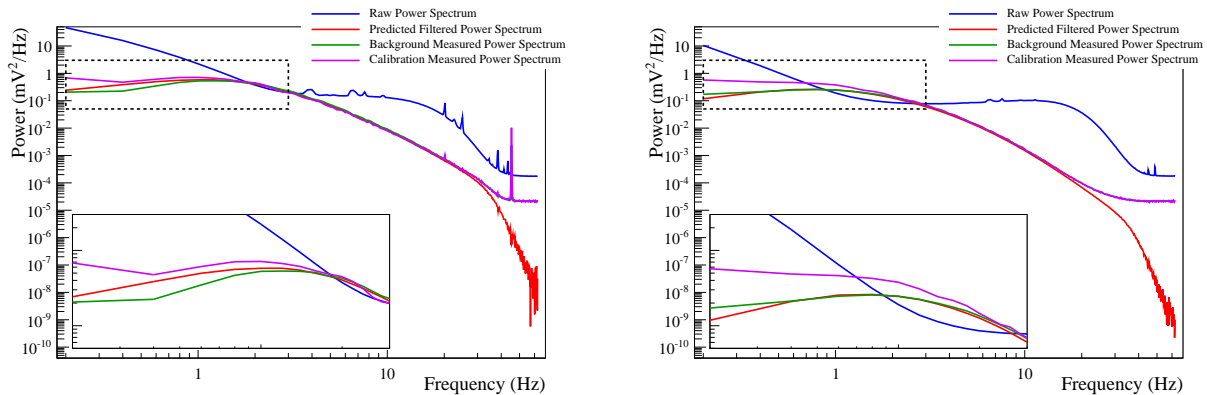
Figure 5.15: Theoretical filtered noise power spectrum vs actual filtered noise power spectrum for channel 13 (*left*) and channel 18 (*right*) on dataset 2073.

First, our measurement of the decorrelation performance on the background data is done using the same noise pulses that were used to build the covariance matrix. In other words, we are testing the filter on the very data that was used to *build* the filter. So, its possible that we are simply over-tuning our filter. However, this is unlikely since we are only building an average covariance matrix which has much fewer degrees of freedom than there are noise events to be decorrelated. Further, we have seen good results using a covariance matrix built on neighboring datasets, so this explanation is unlikely.

An second possibility is that the noise is significantly different in the calibration runs than in the background runs. In other words, perhaps we are using the wrong covariance matrix. However, this too does not pan out. We see no obvious differences between the noise power spectra produced on calibration runs vs background runs.

This leaves the event rate and the most problematic part of the decorrelation: finding signal free noise. The calibration runs typically have a ∼100 times higher event rate than the background runs. The first assumption of the decorrelation procedure is that the side channels being used to decorrelated a signal pulse contain only noise information and no signals of their own. If this assumption breaks down, then our decorrelation algorithm stops decorrelating noise and begins correlating signals. To counter this, we implemented a cut on the side channel waveforms before they could be used in a decorrelation. To give a rough idea of the numbers, the typical event rate for a channel during a calibration run is ∼60 mHz, which means that probability of no events in any 5 second window is roughly ∼75% — but the probability of no events in any 5 second window *on all 11 side channels* around any given channel is ∼ 4% (for background data, this number is closer to 70%).

There are two possible scenarios where a high event rate could cause this our decorrelation to fail. In each decorrelation, either we are successfully removing all the side channel waveforms that contain signals and what remains is too few side channels to reliably decorrelate or we are not successfully removing all of the side channel waveforms with signals and

these signals are worsening our resolution. The troubling aspect of this, is that the correction for either of the two scenarios will lead us into the other scenario.

In Fig. 5.16, we show the distribution of number of channels used per noise event for channel 18 in dataset 2073. Each noise event for channel 18 is decorrelated using up to 12 channels (itself included), but when a side channel waveform has its own signal that waveform must be excluded and the number of channels used in the decorrelation decreases. We see that for background data, most events were decorrelated with all 12 available channels — and those that weren't were typically missing only one or two side channels. However, in the calibration data, it was far more typical for an event to be decorrelated with only 8 or 9 channels and not uncommon for an event to be missing 6 or 7 side channels. This was typical of other channels as well.

What's more telling is that the energy resolution does not appear to degrade with decreasing number of pulses used. This is indicative of the second scenario, where we are not fully removing side channels with signal events.

It's possible that the conclusion of this study is that the calibration runs may just have too high of an event rate to accurately run the decorrelation procedure — or perhaps we need to think of a more clever way to deal with signals on side channels.
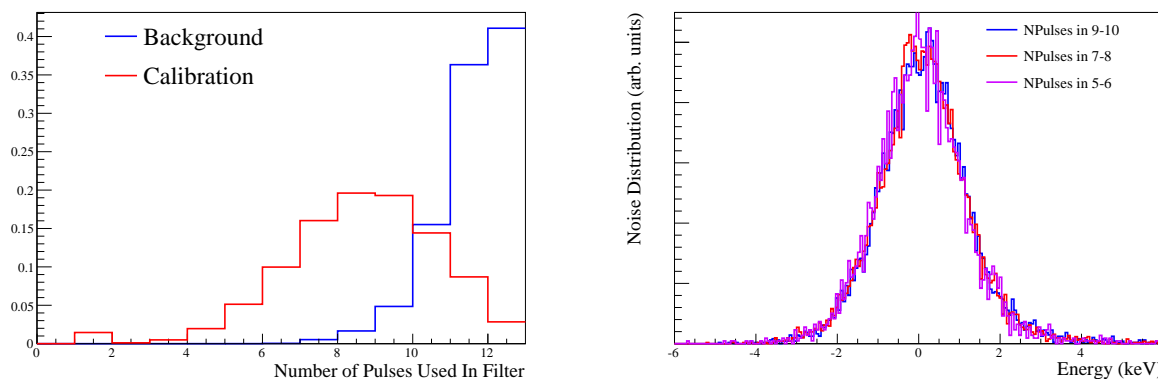
Figure 5.16: *Left*: Distribution of the total number of pulses used in the decorrelation procedure for channel 18 over all datasets for background noise (*blue*) and calibration noise (*red*). For channel 18, the maximum number of channels used in the decorrelation is 12 (including channel 18). We see that in the background data, most events used 11 or 12 channels in the decorrelation. In the calibration runs, the filter usually excluded a significant number of side pulses from each event. *Right:* Energy spectra for noise events which had a different number of side pulses excluded. We see (contrary to expectation) that the fewer side pulses used does *not* lead to worse energy resolution.

Decorrelation in ^{208}Tl Resolution

The typical energy resolution at 2615 keV is about 5-6 keV FWHM, while at 0 keV, it is closer to 2-3 keV FWHM. If we assume that the noise adds in quadrature, this means that the baseline noise accounts for roughly 10% of the resolution. (In Chapter 6, we make this estimate more concrete.) So at best, our improvements in the noise at 0 keV will be reduced by $\sim$10% at 2615 keV.

Despite the poor performance on the calibration data, the decorrelation performed quite well on the background data. We expect that this could translate into improvements in the energy resolution at 2615 keV. But in order to measure the energy resolution at 2615 keV, we must perform a calibration run to build enough statistics to measure the energy resolution at 2615 keV — but of course, this ruins the improvement we are trying to measure. A rather beautifully ironic paradox: a measurement ruins the measurement.

In the final CUORE-0 dataset, the decorrelation is used on $\sim$20% of the data. The algorithm for deciding whether or not to use the decorrelated variable is described in section 4.3. Mostly, the decorrelation was useful on the top three floors of the tower.

Decorrelation Outlook for CUORE

The decorrelation algorithm performed well in the background noise data and so can be useful as-is for low threshold analyses like anti-coincidence and Dark Matter searches. However, for a $0\nu\beta\beta$ search the discrepancy between the filter performance in background and calibration data needs to be addressed.

We have concluded here that the degradation of the filter performance on calibration data is caused by the high rate of accidental coincidences on side channels that goes along with the increased event rate. This causes a breakdown of the assumptions that go into the decorrelating filter (see Appendix B), and thus worse resolution than the non-decorrelating filter. There are several possible avenues to address this in CUORE.

The first approach is to better handle excluded side channels with accidental coincidences. Here, we have chosen to just exclude them from the decorrelation, which degrades the resolution. This required us to be very permissive in letting accidental coincidences into the decorrelation, which also degrades our resolution. One possibility is to actively search for channels in the tower without accidental coincidences rather than using a predefined set of side-channels. Computationally, this is a difficult and time consuming proposition. This could require evaluating pulses on many potential side channels until an adequate number are located, and reinverting covariance matrices on the fly. There are numerical tricks to make this faster, so perhaps this possibility is worth exploring.

Alternatively, since accidental coincidences are uniformly distributed in time, they can be thought of as a kind of 'noise' on our amplitude evaluation and folded into the covariance matrix. This would add an additional source of channel-uncorrelated 'noise' which would be rate dependent — the more accidental coincidences we expect the more channel-uncorrelated

'noise' is in our covariance matrix[5]. The addition of this uncorrelated noise would decrease the expected performance of our decorrelating filter; but by accounting for the additional 'noise' source in the calibration data it could bring our filter performance in line with our expectations. The limiting case of this is when the covariance matrix is completely dominated by uncorrelated noise and our decorrelating filter becomes identical to our non-decorrelating Optimum Filter. (This limit is demonstrated in Appendix B.)

This raises another possibility. We could also use a different filter for the calibration data and the background data. Since we know that the decorrelating filter seems to perform well on the background data but not the calibration data, why not use the Optimum Filter on calibration data, and the decorrelating filter on background data? The only constraint is to require that the two filters have the same gain at all energies in order to calibrate the background data without bias. But this also leads to the problem that we mentioned earlier: How do we measure the detector resolution in the background data at the $0\nu\beta\beta$ Q-value? One possibility is to split the detector resolution into an energy dependent component and a constant offset

$$\sigma^2(E) = f(E) + \sigma^2(0 \text{ keV}) \tag{5.13}$$

However, this requires a more detailed understanding of the sources of noise in the detector than we presently have. As we will see in Chapter 6, our ignorance of this dependence is already becoming a problem in the present analysis.

---

[5]It is worth noting that this is slightly complicated by the fact that the accidental coincidence 'noise' would be frequency-correlated.

Chapter 6

CUORE-0 Analysis and $0\nu\beta\beta$ Fit

This chapter focuses on fitting the ROI and extracting a limit on the $0\nu\beta\beta$ decay rate, $\Gamma_{0\nu\beta\beta}$. Our general approach is to perform a maximum likelihood fit of a detector response function, $f_{\mathrm{Det}}(E)$, at the position we believe a $0\nu\beta\beta$ signal would appear. From the resulting fit signal amplitude, we are able to extract a decay rate. The key ingredients to this approach are understanding the detector response function, $f_{\mathrm{Det}}(E)$, the expected location of the $0\nu\beta\beta$ signal, and the signal detection efficiency, $\varepsilon_{0\nu\beta\beta}$.

This chapter begins by presenting and discussing the final unblinded CUORE-0 spectra. We then use the background data to estimate the signal detection efficiency, $\varepsilon_{0\nu\beta\beta}$. We then switch tracks and discuss the detector response function, $f_{\mathrm{Det}}(E)$. Our approach to estimating this is to use the high-statistics ^{208}Tl line in the calibration data to build a detector response for *each* channel-dataset pair.

Before we use $f_{\mathrm{Det}}(E)$ to fit our ROI, we will first use it to fit other lines in the background spectrum. Partly this allows us to validate our fitting technique, but it also gives us insight into a residual calibration bias, $\Delta\mu(E)$, that exists as a result of our energy reconstruction procedure, as well as an energy dependent component of our detector energy resolution. This residual bias slightly modifies where we search for our $0\nu\beta\beta$ peak, and the scaling of the energy resolution with energy slightly modifies how wide we expect our signal to be. With these modifications in hand, we fit our ROI using an unbinned extended maximum-likelihood (UEML) fit and obtain a limit on $\Gamma_{0\nu\beta\beta}$ that includes statistical uncertainties alone.

The latter part of this chapter is devoted to understanding our systematic uncertainties and how these affect our limit. Specifically, these include our uncertainties on all of the above steps: our choice of $f_{\mathrm{Det}}(E)$, our uncertainty on $\varepsilon_{0\nu\beta\beta}$, our uncertainty on the position of the $0\nu\beta\beta$ line and its resolution, as well as any bias introduced by our UEML fitting technique. We quantify the effect that most of these have on our final result using toy Monte Carlo experiments, which we describe later in this chapter.

This chapter concludes by presenting a limit on $\Gamma_{0\nu\beta\beta}$ including both statistical and systematic uncertainties. We compare the limit set here to the one set by Cuoricino and TTT, and combine them to form a single limit on the decay rate of ^{130}Te. Finally, we use a

range of matrix elements to convert this to a range of limits on $m_{\beta\beta}$.

So hold on tight, because here we go.

6.1 Final Physics Spectra

We produce our final physics spectra by summing over all datasets. For the rest of this thesis, we will refer generically to the set of Aggressive energy estimators that we defined in Chapter 4 as the "Energy".

Creating a Summed Calibration Spectrum

To build a final calibration spectrum that is representative of the background data, we chose a scheme that weights calibration data by the amount of background exposure.

We begin by normalizing the histogram (or fit), $H(D, C; E)dE$, from dataset D on channel C, to unity over the range 2560–2650,

$$R_{\mathrm{Cal}}(D, C; E) = \frac{H(D, C; E)dE}{\int_E H(D, C; E)dE}$$

As long as the number of events in the range is large, this is a good approximation to dividing by the *expected* number of events, which cancels out variations in the detector counting rate between channels and datasets.

From here, we combine these normalized PDFs over channels and over datasets by weighting by the *background* exposure, $T_{\mathrm{Bkg}}(D, C)$. For instance, the total weighted calibration spectrum is given by

$$R_{\mathrm{Cal}}(E)dE = \frac{\sum_{D,C} T_{\mathrm{Bkg}}(D, C) R_{\mathrm{Cal}}(D, C; E)}{\sum_{D,C} T_{\mathrm{Bkg}}(D, C)} dE \tag{6.1}$$

Keep in mind, that since the calibration rate depends on the placement of the source strings — which can vary from dataset to dataset — the overall normalization of this calibration spectrum is arbitrary. For presentation, we normalize the resulting spectrum so that the integral of the ^{208}Tl line matches that of the background spectrum. This is shown, plotted against the final CUORE-0 background spectrum in Fig. 6.1.

CUORE-0 vs Cuoricino in the α Region

We can compare the background spectrum of CUORE-0 to that of Cuoricino, to measure the efficacy of the background reduction procedures. We compare the full range and the α region alone in Fig. 6.2.

We measure the expected α-continuum rate in the ROI by measuring the continuum in the region 2.7–3.9 MeV, but excluding the ^{190}Pt line between 3.2–3.4 MeV. ^{190}Pt is a naturally
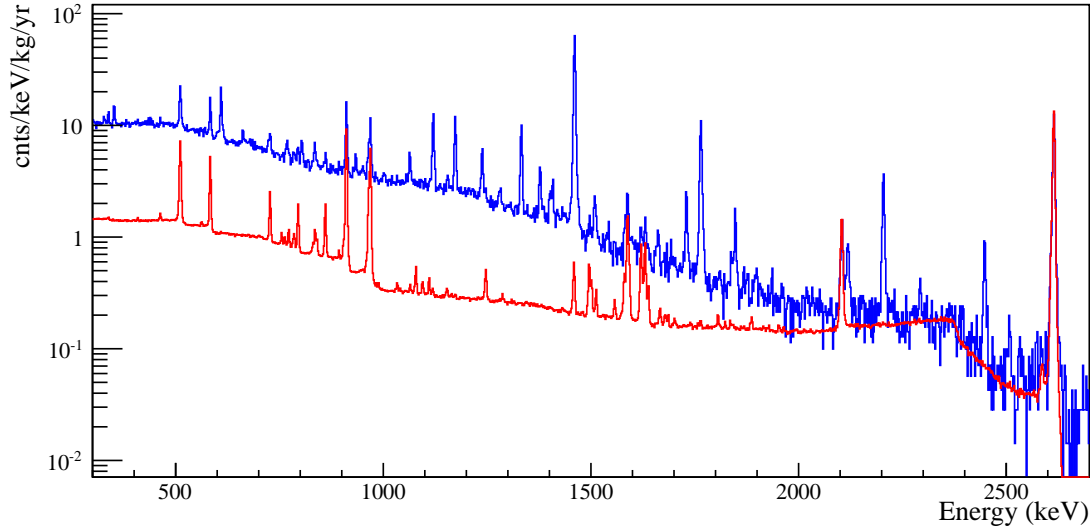
Figure 6.1: The final CUORE-0 background spectrum (*blue*) and the final weighted calibration spectrum (*red*). The weighted calibration spectrum has been normalized so that the rate at the ^{208}Tl peak matches that in the background.

occurring isotope of Pt. The contamination is incurred during the crystal production which uses thin Pt bags to hold the crystals as they grow. Typically, this contamination manifests as tiny Pt inclusions within the crystal. But because of the extremely short range of α particles, we get little to no degraded α spectrum from ^{190}Pt — in other words, all ^{190}Pt decays are reconstructed in the ^{190}Pt peak. So this line does not present a background for our $0\nu\beta\beta$ analysis. It is worth noting that these inclusions do affect the crystal's heat capacity, causing a change in the pulse shape (see Fig. A.4).

The rate between 2.7–3.9 MeV excluding 3.2–3.4 MeV is measured to be

$$b_\alpha = 0.0162 \pm 0.0007 \text{ cnts/keV/kg/yr} \tag{6.2}$$

We compare this to the rate measured for Cuoricino of $b_\alpha = 0.110 \pm 0.001$ cnts/keV/kg/yr [68]. This is a factor of almost 7 reduction in background in the α region. This is consistent with expectation and demonstrates the efficacy of our background reduction program for CUORE.
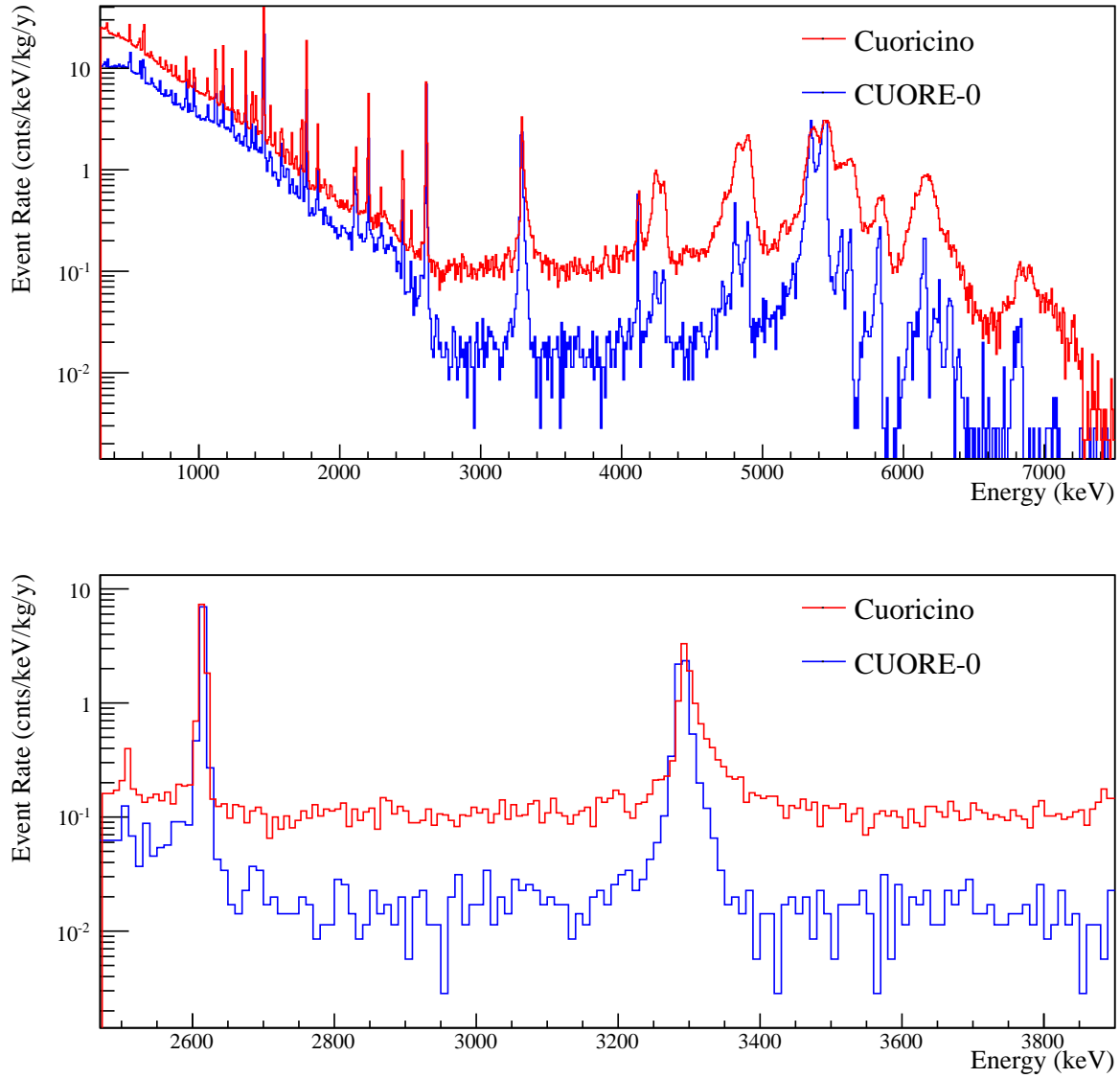
Figure 6.2: The final spectra for CUORE-0 and Cuoricino over 0–10 MeV (*top*) and the α region alone (*bottom*). The α background rate is measured over the range $[2.7, 3.2] \cup [3.4, 3.9]$ MeV. The line at 3.3 MeV is a ^{190}Pt line that comes from bulk contamination during the crystal growth process.

6.2 Efficiency Evaluation

The signal efficiency is calculated as the product of the containment efficiency, the trigger and reconstruction efficiency, and the cut efficiency. These are, in order, the probability a $0\nu\beta\beta$ event deposits its full energy into a single crystal, the probability that we correctly trigger on it and reconstruct its amplitude correctly, and the probability that the event passes all of the signal cuts. A detailed description of the efficiency evaluation can be found in [105].

The containment efficiency is estimated by Monte Carlo. We generated 5×10^6 $0\nu\beta\beta$ events uniformly throughout the volume of the crystals. We simulate detector effects like resolutions, channel thresholds, event pile-up and dead channels that have all been measured from the CUORE-0 detector. The containment efficiency is calculated as the number of events that lie within the $0\nu\beta\beta$ peak compared to the number of events generated.

$$\varepsilon_{\text{Containment}} = (88.345 \pm 0.04(\text{stat}) \pm 0.075(\text{syst}))\%$$

The trigger efficiency and energy reconstruction efficiency are calculated together using the pulser events. APOLLO automatically flags and records each heater event with a flag indicating that it is associated with the pulser. But if the event passes the signal trigger threshold, then the signal trigger is recorded in the event as a secondary trigger. The trigger efficiency is calculated as the fraction of heater events that also have a signal event as a secondary trigger.

The energy reconstruction efficiency is calculated by taking the fraction of pulser heater events that occur within $\pm 3\sigma$ of the mean heater energy for each channel in each run and dividing by .997.

The total trigger and reconstruction efficiency is the product of these two numbers and the uncertainty comes from counting statistics.

$$\varepsilon_{\text{Trigger}} = (98.529 \pm 0.004)\%$$

This efficiency cannot be calculated on channels without functioning heaters (i.e. channels 1 and 10), so for these channels, we just assume the average efficiency of the other channels. For $\gtrsim 500$ keV events, this is generally a safe assumption since the efficiency is driven by pile-up and noise scattering events, rather than anything having to do with the heaters.

We split the cut efficiency into two parts: the pile-up and pulse-shape cut efficiency, and the multiplicity cut efficiency. Since the pulse-shape cuts were tuned on the half of the events with even numbered timestamps, we performed the efficiency calculations on the half of events with odd numbered timestamps.

We calculate the pile-up and pulse-shape efficiency by measuring the fraction of events in the 2615 keV ^{208}Tl background peak that survive the pile-up and pulse-shape cuts.

$$\varepsilon_{\text{Pile-Up \& PSA}} = (93.7 \pm 0.7)\%$$

The efficiency due to accidental pile-up (which is the dominant component of the pile-up efficiency) is independent of energy as is the efficiency of the PSA cuts (see section 4.3).

Since a fraction of the ^{208}Tl 2615 keV events are expected to occur in coincidence with a 538 keV γ, we calculate the multiplicity cut efficiency on the ^{40}K 1460 keV line, which is expected to be a true multiplicity-one event. We take the pile-up efficiency to be the fraction of events in the ^{40}K peak which survive the multiplicity cut, which yields

$$\varepsilon_{\text{Multiplicity}} = (99.6 \pm 0.1)\%$$

This efficiency represents the probability of an accidental coincidence between two crystals, and is therefore independent of energy.

The uncertainties on each of these efficiencies were estimated using a Clopper-Pearson interval [106]. The total cut efficiency is given by

$$\varepsilon_{\text{Cut}} = (93.3 \pm 0.7)\%$$

The final total $0\nu\beta\beta$ signal efficiency is given by

$$\varepsilon_{0\nu\beta\beta} = (81.3 \pm 0.6)\%$$

This is on par with the efficiency seen in Cuoricino, $82.8 \pm 1.1\%$.

Table 6.1: Summary of $0\nu\beta\beta$ signal efficiencies.

Containment Efficiency	$88.345 \pm 0.04 \pm 0.075\%$
Trigger Efficiency	$98.529 \pm 0.004\%$
Cut Efficiency	$93.3 \pm 0.7\%$
Total Efficiency	$81.3 \pm 0.6\%$

6.3 ^{208}Tl Line Shape From Calibration

We model the detector response to $0\nu\beta\beta$ by fitting the shape of the ^{208}Tl 2615 keV calibration line and scaling some of the shape parameters down to the ROI. We use this shape because the 2615 keV line is only 87 keV above the expected $0\nu\beta\beta$ decay value, so any energy-dependent detector effects should be small. However, this approach is not without its caveats which we will discuss in the systematics section.

Line Shape Model

We fit the calibration 2615 keV line in ROOFIT with a large simultaneous fit over all channels and datasets, with each channel-dataset (ChDs) pair having its own set of mean and resolution parameters while sharing some common shape parameters. The final CUORE-0

dataset contained 1008 ChDs pairs, so this fit is actually 1008 simultaneous fits with some 3129 free parameters over 542826 events. The free parameters are listed in Table 6.2.

The fit was performed over the range 2560–2650 keV. For CUORE-0 we used an empirical model for each ChDs pair that can be written as

$$
\begin{aligned}
f_{\text{Cal}}(D,C;E) \;=\; & f^C_{\text{Det}}\left(E_0(D,C),\sigma_0(D,C);E\right) \\
& + \alpha_{30\text{keV}} f_{30\text{keV}}\left(E_0(D,C),\sigma_0(D,C);E\right) \\
& + \frac{\alpha_{\text{Compt}}}{2}\text{Erfc}\left(\frac{E_0(D,C)-E}{\sqrt{2}\sigma_0(D,C)}\right) \\
& + \alpha_{\text{Bkg}}
\end{aligned}
\tag{6.3}
$$

This model describes a primary detector response with a shape given by $f^C_{\text{Det}}(\mu,\sigma;E)$, centered at $E_0(D,C)$, with a resolution of $\sigma_0(D,C)$. The central energy, $E_0(D,C)$, and resolution, $\sigma_0(D,C)$, are both free to float for each ChDs. We have several models for the detector response function, $f_{\text{Det}}(D,C;E)$, that we will describe later in this section; the fitting procedure is identical for each response function.

The second element of the calibration shape describes an X-ray escape peak that lies $\sim$30 keV below the main peak at $E_0(D,C)$. The amplitude of the escape peak is suppressed by a fraction $\alpha_{30\text{keV}}$. This fraction is a global parameter common to all channels and datasets but is free to float in the fit.

The next part is a gaussian-smeared "step function" that mimics the spectrum of γ-rays that scatter and lose energy before interacting with the bolometer. The number of events in the Compton shoulder is also pegged to the number of events in the photopeak and is characterized by the ratio α_{Compt}. This ratio is also common to all channels and datasets but is a free parameter in the fit.

The final element is a flat background that fits the continuum at energies above the primary peak. These events are mostly from accidental coincidences within a single crystal and poorly reconstructed events and therefore this background should be proportional to the signal rate in the crystal. Thus the amplitude of the flat background is also expressed as a fraction of the main peak, α_{Bkg}, which is also common to all channels and datasets.

The channels are summed together to form a total PDF,

$$
f_{\text{Cal}}(E) = \sum_{\text{ChDs}} N_{\text{Sig}}(D,C) f_{\text{Cal}}(D,C;E)
\tag{6.4}
$$

Here $N_{\text{Sig}}(D,C)$ is a parameter that encompasses the number of events in each ChDs's main peak, and is also free to float.

Detector Response

In Cuoricino, we modeled the 2615 keV γ line with a simple gaussian, but the improved energy resolution in CUORE-0 reveals a more complicated substructure just below the primary photopeak. We have several empirical models for the detector response to a γ-ray,

Table 6.2: Floating parameters in the simultaneous fit, their limits, and their domains of applicability. The detector response, f_{Det}, is characterized by one or two parameters, here we quote the parameters for the standard approach: r_{Sub} and α_{Sub}.

Parameter	Domain	Range	N_{Par}
R_{Sig}	ChDs	10^3 - 10^5	1008
$E_0(D,C)$	ChDs	2612-2618	1008
$\sigma_0(D,C)$	ChDs	0.8-15	1008
$r_{\mathrm{Sub}}(C)$	Channel	.997-.999	51
$\alpha_{\mathrm{Sub}}(C)$	Channel	0.01-0.5	51
α_{30keV}	Global	10^{-4} - 10^{-2}	1
α_{Compt}	Global	0.01-0.05	1
α_{Bkg}	Global		1
Total Parameters			3129

$f_{\mathrm{Det}}(\mu, \sigma; , E)$, but our primary approach models the peak as a sum of two gaussians, one for the main photopeak and a second sub-peak for the substructure.

$$f_{\mathrm{Det}}^C(\mu, \sigma; E) = (1 - \alpha_{\mathrm{Sub}}(C))\mathrm{Gauss}(E; \mu, \sigma) + \alpha_{\mathrm{Sub}}(C)\mathrm{Gauss}(E; r_{\mathrm{Sub}}(C)\mu, \sigma) \qquad (6.5)$$

The shape of the detector response varies by channel and is characterized by two parameters: $\alpha_{\mathrm{Sub}}(C)$ and $r_{\mathrm{Sub}}(C)$, which determine the amplitude and the position of the sub-peak gaussian. Both gaussians share the same σ.

Some of the better channels in Cuoricino showed hints of this substructure, but it was really the improved resolution of the CUORE-0 detector that has made this substructure clear. The gaussian substructure tends to correspond to a ~0.25% energy loss — or alternatively, we can look at it as a peak shifted down from the main peak by ~6.5 keV. The amplitude of this substructure is typically $\sim 5\%$ of the main photopeak. Because we saw hints of this effect in Cuoricino, as well as the ubiquity of this substructure among the peaks in the CUORE-0 data as well as the absence of the substructure in the reconstructed pulser heater line, we believe this is a detector effect and not a result of our energy reconstruction. At present, however, we have no convincing physical explanation for its origin.

For these reasons, we are inclined to believe that the substructure should also be present in the $0\nu\beta\beta$ signal shape. However, as we don't yet understand its physical origin, we accept the possibility that this may not be the case. Our primary approach includes the substructure as part of the $0\nu\beta\beta$ line shape, but because of our ignorance of it origin we will study the effect of the lineshape as part of our systematic uncertainty study.

30 keV X-ray Escape

When a γ-ray scatters off an electron in the crystal lattice, it generally ionizes the atom to which that electron was bound. If the ejected electron comes from one of the inner atomic shells, then atom will often release X-rays as the electrons rearrange themselves. If one of

these X-rays escapes the crystal, this results in a slightly lower detected energy. Tellurium has a set of characteristic X-rays between 27–31 keV whose escape peak we refer to generically as the 30 keV X-ray escape peak.

This peak is an interesting case unto itself. It is in many ways a toy version of the actual $0\nu\beta\beta$ fit. The 30 keV peak is a small ($\sim$.1%) peak that lies $\sim$30 keV below a much larger peak; a $0\nu\beta\beta$ signal would create a small peak that lies $\sim$20 keV *above* a (presumably) larger peak, the ^{60}Co sum peak at 2506 keV. So this line provides a good demonstration of the resolving power of our bolometric technique in regards to our $0\nu\beta\beta$ search.

The 30 keV line is not actually a single line, but is instead a sum of $\sim$6 lines. (Actually, there are more than 6 lines, but here we use the strongest 6.) We model this line as a constrained sum of 6 separate lines that are pinned to the main photopeak energy, $E_0(D, C)$:

$$f_{30\mathrm{keV}}(D, C; E) = \sum_{i=1}^{6} \alpha_i \mathrm{Gauss}\left(E; E_0(D, C) - \delta E_i, \sigma_0(D, C)\right) \tag{6.6}$$

The relative intensities and energies can be looked up in X-ray data tables [107], but here we use values taken from Monte Carlo data which account for the relative probabilities of escaping the crystal. They are listed in Table 6.3.

Table 6.3: Relative intensities and energies for the 30 keV X-ray lines. Numbers are taken from the CUORE Monte Carlo.

Line	δE_i (keV)	α_i (%)
1	27.20	26.5
2	27.47	50.4
3	30.96	21.6
4	31.22	0.49
5	31.65	0.50
6	31.80	0.42

A Tail of Two Gaussians

We also investigated another model for the calibration detector response. This model replaced the single-gaussian substructure shape with two sets of X-ray escape peaks similar to the 30 keV structure. The model is based on the observation that Te has three lower energy X-rays at $\sim$3–4 keV in addition to the 30 keV lines. However, the line shape is not modeled well with just these X-ray escape lines, but rather we must include a double-escape line where two X-rays escaped. In this model we parameterize the substructure as

$$
\begin{aligned}
f_{\mathrm{Det}}^{C}(\mu, \sigma; E) &= (1 - \alpha_{\mathrm{Single}}(C) - \alpha_{\mathrm{Double}}(C))\mathrm{Gauss}(\mu, \sigma; E) \\
&\quad + \alpha_{\mathrm{Single}}(C) \sum_{i=1}^{3} \alpha_i \mathrm{Gauss}\left(\mu - \delta E_i, \sigma; E\right) \\
&\quad + \alpha_{\mathrm{Double}}(C) \sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_i \alpha_j \mathrm{Gauss}\left(\mu - \delta E_i - \delta E_j, \sigma\right)
\end{aligned} \tag{6.7}
$$

Here, the sums run over the three Te X-rays with energies around 3–4 keV; δE_i and α_i are their energies and relative intensities. These energies and intensities are listed in Table 6.4. The parameters $\alpha_{\text{Single}}(C)$ and $\alpha_{\text{Double}}(C)$ are the ratio of the number of events in the single or double-X-ray escape peaks to the main photopeak; these parameters are floated in the fit and are only channel dependent.

This model looks far more complicated then the standard approach. However, it has the *same number of free parameters* as the fit in the previous section. The fit with the single gaussian substructure had a floating mean and amplitude; this model has two floating amplitudes but fixed means.

Table 6.4: Parameters for the Te X-rays around 3–4 keV. Numbers are taken from the CUORE Monte Carlo.

Line	δE_i (keV)	α_i (%)
1	3.77	35.5
2	4.03	52.1
3	4.25	12.4

This model produced similar residuals to the model with a single-gaussian substructure, produced a comparable χ^2 and negative log likelihood. However, this model implies an unphysically large number of X-ray escapes from the crystal. Because the attenuation length of 4 keV X-rays in TeO_2 is so short, we expect all X-ray escapes to come from interactions in a very shallow surface layer, and thus to be suppressed by ~ 5 orders of magnitude relative to the main peak. But the number of events observed in the substructure is only suppressed by a factor of ~ 20 to the main peak. As a result, we have a hard time attributing the observed substructure to X-rays. For now, we leave this interpretation as an interesting possibility that could be explored further in future investigations.

Calibration ^{208}Tl Fit

The final fit to the calibration ^{208}Tl 2615 keV line using the detector response in Eqn. 6.5 is shown in Fig. 6.3. Overall, the model does a good job describing most of the features in the lineshape over ~ 3 orders of magnitude. However, the residuals display some clear structure around the main photopeak of the line, possibly indicating that we are overlooking or oversimplifying some small shape effects. The fit also reproduces the shape of 30 keV escape peak, which we are modeling for the first time as a sum of 6 individual gaussians.

Figure 6.4 shows the distribution of the ^{208}Tl photopeak means. There are two points to take away from this plot. First, the distribution is centered at ~ 2614.7 keV, instead of the true energy, 2614.511 keV. This 0.2 keV shift stems from the fact that the first-level production calibration module fits the energy peaks with simple gaussian line shapes, and the substructure causes these lines to be slightly biased downward. The second point to take away from the distribution of means is that the RMS of means is ~ 0.2 keV, which is the
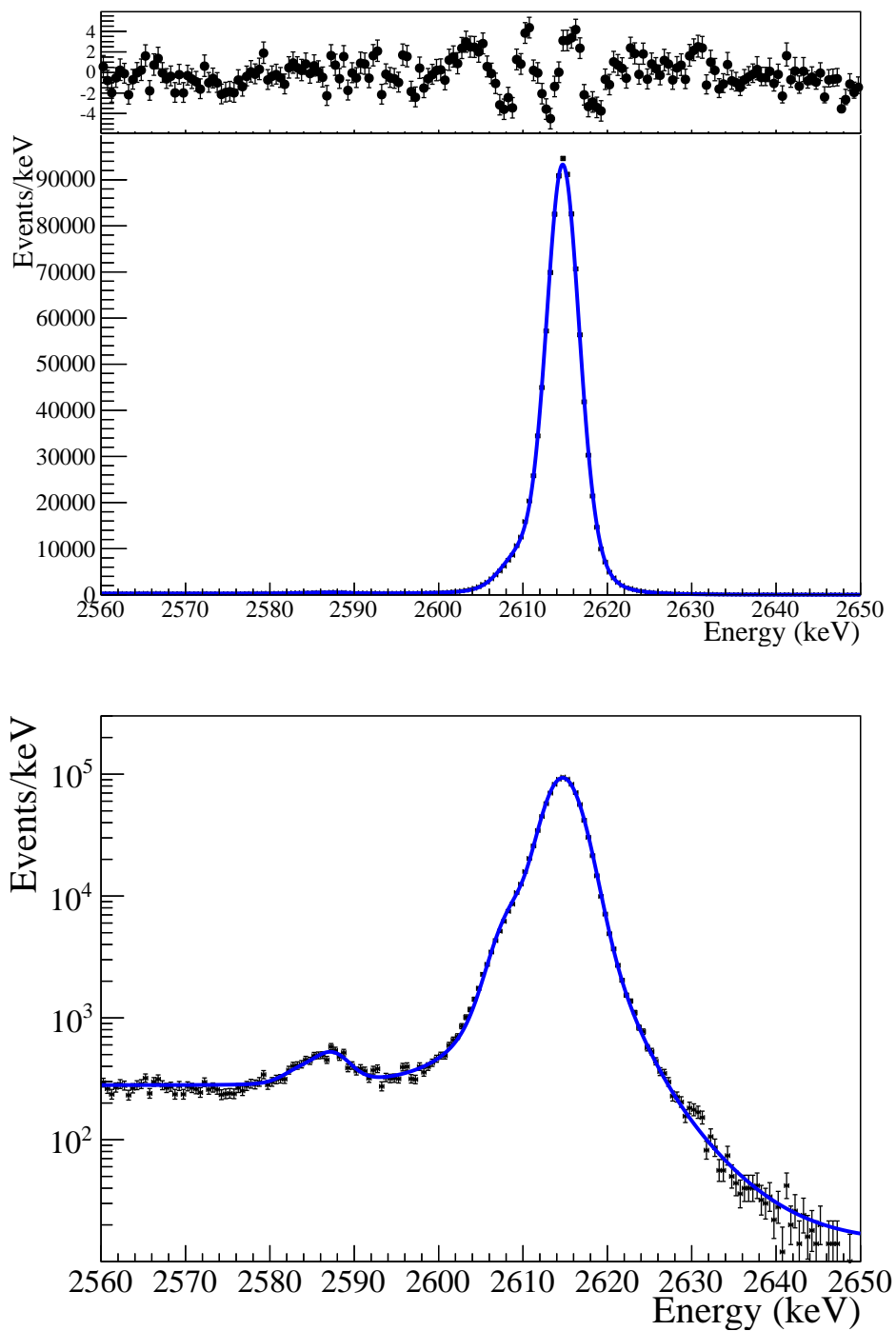
Figure 6.3: Fit to the calibration ^{208}Tl 2615 keV line with normalized residuals in linear and log scales. Here, we use the standard approach of a gaussian photopeak with a single sub-gaussian to model the substructure. The bump around 2585 keV is the 30 keV X-ray Te escape line modeled with a sum of 6 individual gaussian lines.

residual shift after the calibration fit. We return to this issue and how we handle it in the final $0\nu\beta\beta$ fit later in this chapter.
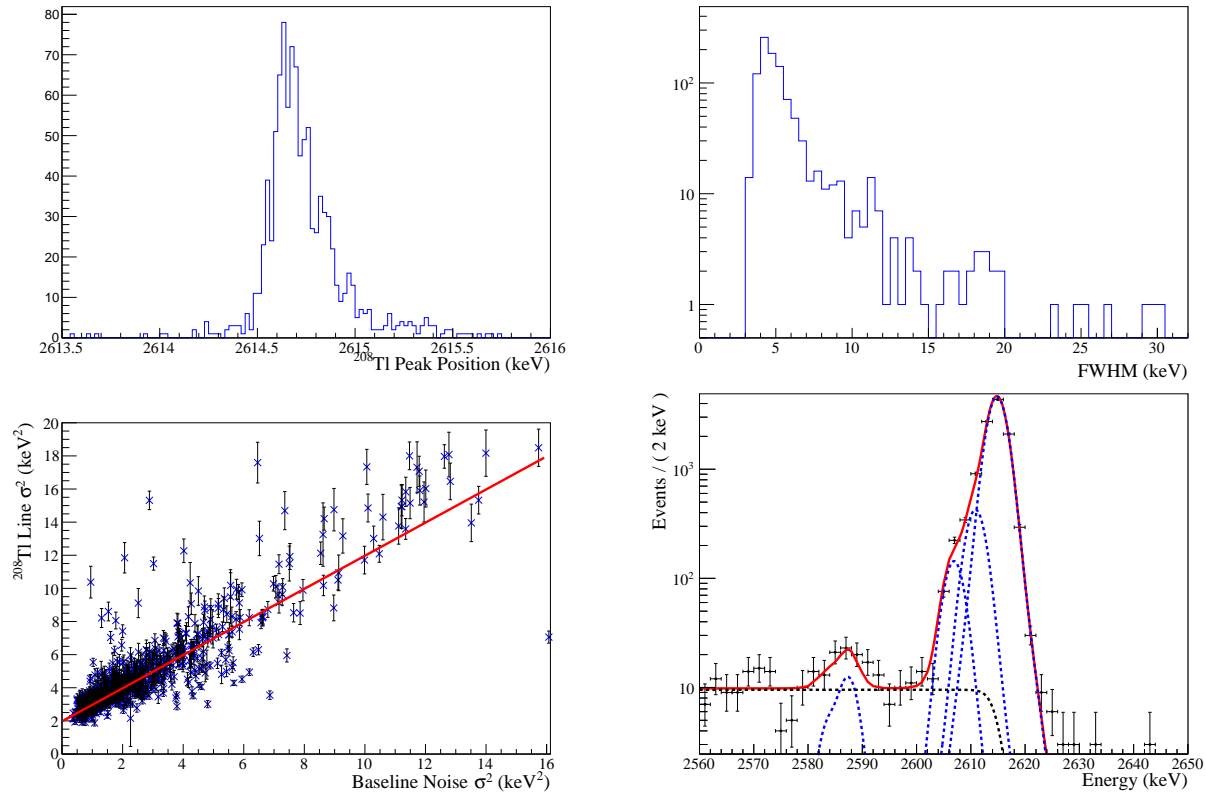
Figure 6.4: *Top Left:* Distribution of photopeak means from the line shape fit. *Top Right:* Distribution of measured FWHM values from lineshape fit. *Bottom Left:* Scatter plot of ^{208}Tl 2615 keV gaussian variance vs calibration baseline noise variance with best-fit line. *Bottom Right:* Gauss + X-ray detector response fit to ^{208}Tl calibration data on channel 18. In blue, the Gaussian lines for the main peak, 4 keV escape, double escape, and 30 keV X-ray escape. In black is the step function.

Figure 6.4 also shows the distribution of the fit variance at 2615 keV for each ChDs plotted against the variance of the baseline noise for that dataset. Unsurprisingly, we see strong correlation between the two variables, indicating that higher baseline noise leads to higher noise level at 2615 keV. But, surprisingly, we see an intercept of $\sigma^2 = 1.97$ keV^2 indicating that at 2615 keV we have an excess variance of ~ 2 keV^2 (FWHM of $\sim$3.3 keV) on top of any baseline noise. This is an order of magnitude larger than the RMS of the photopeak means and indicates some other source of noise above the baseline and calibration shifting.

Presently, we do not understand the source of this excess noise, but we can say several

things about it. First, the pulser heater — which typically reconstructs around 3.5 MeV — does not show this excess variance, but has roughly the same energy resolution as the baseline noise (∼2-4 keV FWHM). We thus believe that this excess noise over the baseline seen in the 2615 keV peak is intrinsic to particle interactions. We could imagine several sources for this, such as a position-dependent gain within the absorber, the down-conversion of athermal phonons to thermal phonons, trapped electron-hole pairs within the crystal lattice or something else completely. In the future, we would like to better understand this through a combination of R&D runs and absorber modeling.

The second thing that we can say about the excess noise, is that it is energy dependent — but we do not know the form of that energy dependence. Later in this chapter, we will fold this into our $0\nu\beta\beta$ fit as a systematic uncertainties — in fact this is one of our largest sources of systematic uncertainty.

Detector Performance Figure Of Merit

When quoting the energy resolution of our detector, it is common to report the FWHM at the 2615 keV line because of its proximity to the $0\nu\beta\beta$ Q-Value. However, as seen here, we actually have 1008 FWHM values which all contribute to the final physics result. During CUORE-0, we reported two values which were meant to be figures of merit: the summed FWHM and the *effective* FWHM.

The summed FWHM is obtained by summing the calibration data and plotting a summed line shape (as in Fig. 6.3) and measuring the FWHM of the summed line. This has the advantage of being true to the plot — if someone were to examine the plot, this is the FWHM they would measure. However, this number has a few disadvantages. First, it contains all of the inter-dataset and channel shifting shown in Fig. 6.4 (which is not part of our detector resolution when doing a ChDs dependent fit). Second, it bears little resemblance to the actual $0\nu\beta\beta$, since it is heavily weighted by channels with higher event rate during calibration.

An alternative number to quote is the effective FWHM. This is defined as

$$\text{Effective FWHM} \equiv \left(\sum_i T_{\text{Bkg},i} \right) \bigg/ \left(\sum_i \frac{T_{\text{Bkg},i}}{\text{FWHM}_i} \right) \tag{6.8}$$

This is the harmonic mean of the ChDs FWHM values *weighted* by the background exposure, T_{Bkg}. This is inspired by Eqn. 2.31, and is more akin to weighting the FWHM by each ChDs's contributions to the sensitivity.

The reported calibration FWHM values are shown in Table. 6.5 and are representative of our detector performance. Note, however, that these numbers do not enter into our $0\nu\beta\beta$ fit.

Table 6.5: Summed and effective FWHM at 2615 keV values for CUORE-0, by phase. These numbers are meant to be indicative of the detector performance, but do not appear anywhere in the actual $0\nu\beta\beta$ analysis.

	Summed FWHM (keV)	Effective FWHM (keV)
Phase I	5.31	5.65
Phase II	4.56	4.77
Phase III	4.63	4.81
Phase IV	4.44	4.65
Total	**4.81**	**4.93**

6.4   Projecting the Detector Response

Our approach to fitting the peaks in the ROI is to use the detector response function described in Eqn. 6.5, fix the shape of the substructure to the best-fit from the high statistics calibration fit, and shift the whole line down to the ROI, leaving as few free parameters as possible. We begin by validating this technique on other known lines in the background spectrum to understand a few systematics of the technique. In this section, we use these fits to first understand a residual bias in the reconstructed energy estimators, and then to understand the detector resolution. Then, in the next section, we will apply this fit to the ROI and measure the peaks in the ROI.

We fit each peak with a fit function that depends on channel and dataset and is given by

$$f_{\mathrm{Peak}}(D, C; E) = f_{\mathrm{Det}}^{C}\left(\mu, r_\sigma \sigma_0(D, C); E\right) + p_1 E + p_0 \qquad (6.9)$$

Here, we project the best-fit detector response from the calibration ^{208}Tl down to fit a peak centered at μ (note that in this projection, μ is a global parameter for all channel-dataset pairs). We fit the background with a polynomial; in most cases a linear background produces a good enough fit. The $\sigma_0(D, C)$ are the best-fit gaussian resolutions from the calibration ^{208}Tl fit. All ChDs pairs are scaled together by a global scaling factor, r_σ[1]. One important aspect of this approach is that we assume the substructure sits at a constant *fraction* of the main peak energy. In doing this, we make an implicit assumption about the origin of the substructure — namely that it scales with energy. This is a purely phenomenological choice justified by the quality of the resulting fits.

The ChDs pairs are summed together slightly differently for calibration and background data. For calibration data, we assume that the number of events in each peak in the spectrum is proportional to the number of events in the ^{208}Tl peak. We thus sum together the ChDs

---

[1]We also briefly investigated an alternate scaling of the gaussian resolution. Instead of scaling the full resolution, $\sigma(D, C)$, we scaled only the variance in excess of the baseline variance. We wrote $\sigma^2(D, C) = r_\sigma^2(\sigma_0^2(D, C) - \sigma_{0\,\mathrm{keV}}^2(D, C)) + \sigma_{0\,\mathrm{keV}}^2(D, C)$. This approach assumes that the noise is partly a constant from the baseline noise plus another source which is energy-dependent. This approach produced promising results, but we did not pursue it due to time constraints.

pairs as

$$f_{\text{Peak}}(E) = \sum_{\text{ChDs}} \alpha_{\text{Peak}} N_{\text{Sig}}(D, C) f_{\text{Peak}}(D, C; E) \tag{6.10}$$

where N_{Sig} is the best-fit number of events in the calibration ^{208}Tl fit and α_{Peak} is a global parameter that determines each peak's intensity relative to ^{208}Tl. For the background lines, we assume that all channel-datasets have the *same* event rate. In this case, the number of peak events for each channel-dataset is proportional to its exposure times the global event rate. We know that this assumption is not completely correct and that the background event rates have a non-zero channel dependence (see Fig. A.3), but we find that accounting for this has little effect on our results. Thus, in our projected fit there are 5 floating parameters: the peak position, μ, the energy resolution scaling, r_σ, two parameters for the background, and the global event rate (for the background peaks) or the ratio of peak intensity to ^{208}Tl (for calibration peaks). We perform this fit on 33 lines in the background data and 7 lines in the calibration data. Several of these projections are shown in Fig. 6.5. The distribution of peak residuals is shown in Fig. 6.6; this is the $\mu_{\text{Measured}} - \mu_{\text{Expected}}$.

The 2506 keV ^{60}Co sum peak is a significant outlier from the rest of the distribution, and possibly the ^{208}Tl single escape peak at 2103 keV as well. A similar effect was seen in Cuoricino where the ^{60}Co peak was centered at 2506.5 ± 0.32 keV (gaussian fit). These peaks have fundamentally different topologies from the rest of the peaks in that they are, by definition, a coincidence of more than one high energy γ in a single crystal. The ^{60}Co peak is a coincidence of two γ-rays at 1.17 MeV and 1.33 MeV, while the single escape peak is a 2615-keV γ that pair-produces within the crystal and creates two 511-keV γ-rays, one of which is absorbed in the same crystal. These two types of events are fundamentally different from most other peaks, either in that they are more spatially distributed across the crystal or in that they produce a different spectra of athermal phonons. Though we do not yet fully understand the microscopic mechanism for why these peaks are shifted, preliminary Monte Carlo modeling indicates that the spatial extent of the energy deposition for these events is indeed different. Combining this suspicious feature with the fact that that other 31 peaks in the background spectrum are very well reproduced, we do *not* consider the shift in the ^{60}Co peak to be indicative of a problem with our energy calibration. It is just an unfortunate coincidence that this peak happens to lie in our ROI.

The average of the peak residuals in Fig. 6.6 (weighted by the inverse square error) is 0.11 keV. We attribute this residual to the fact that we calibrate using gaussian line shapes, which slightly underestimates the peak position because of the substructure that became evident later in the analysis. The residuals also display a small parabolic energy dependence, which we believe is also a relic of the calibration process — namely, that the calibration point at 1592 keV is actually a sum of two lines and is systematically underestimated. We plan to address these issues in CUORE.

For CUORE-0, we fit these peak residuals to understand our calibration bias as a function of energy, $\Delta\mu(E)$. This is the amount we expect a peak to be reconstructed off from its true energy as a function of the true energy. In the next section, we will peg our fit in the ROI
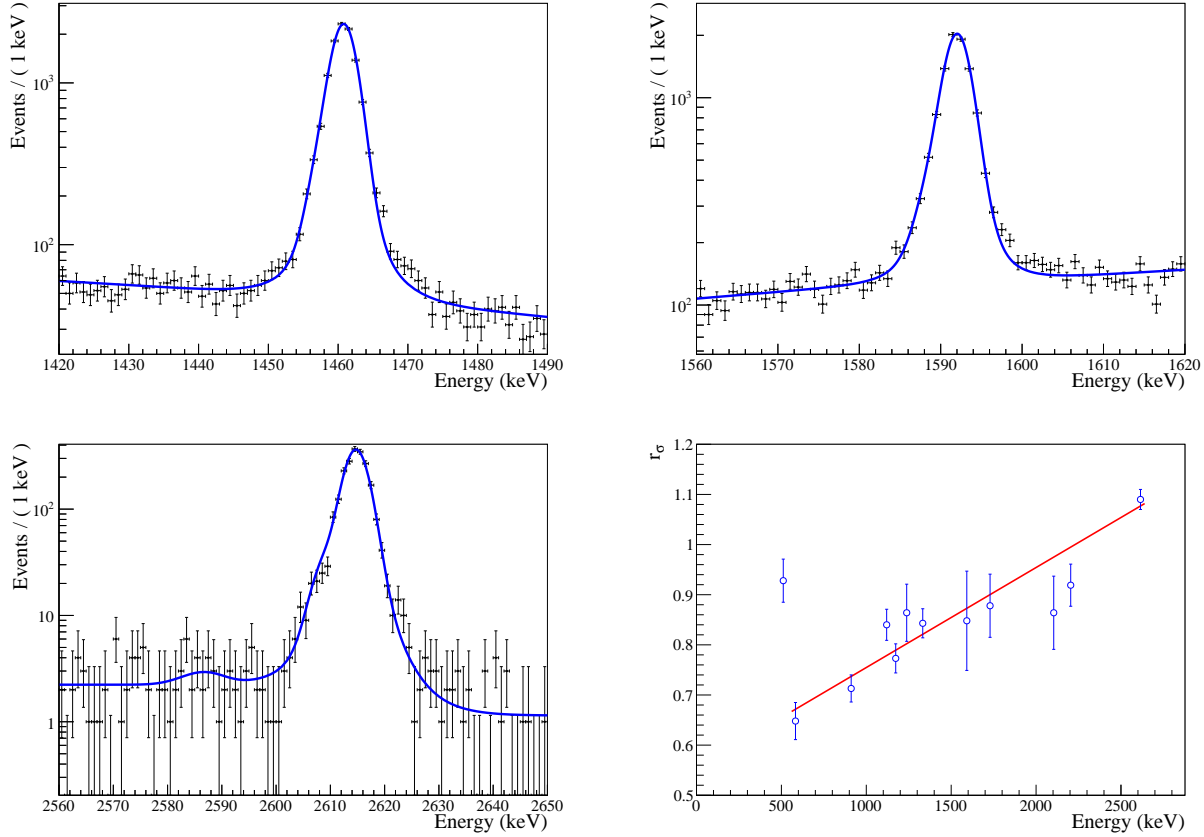
Figure 6.5: Projection of the lineshape model onto the background 1460 keV line (*top left*), calibration ^{208}Tl double-escape peak at 1592 keV (*top right*), and the background ^{208}Tl 2615 keV peak (*bottom left*). *Bottom Right:* The resolution scaling r_σ of the fits to the background peaks as a function of peak energy. The error bars are statistical only. Some peaks' resolutions are broadened by other effects. For example, the 511 keV peak is broadened by a 510.77 keV ^{208}Tl line as well as positron decay in flight and is not used in this fit.

to the ^{208}Tl calibration value and shift our lines down by an amount close to $Q_{\mathrm{Tl}} - Q_{0\nu\beta\beta}$. Because of this, we are not interested in the *absolute* bias at the $0\nu\beta\beta$ energy, but rather on the *relative* bias between the calibration ^{208}Tl peak and the $0\nu\beta\beta$ position. To reflect this, we consider the calibration ^{208}Tl line to be a fixed point and constrain $\Delta\mu(E)$ to pass through the ^{208}Tl calibration peak residual, 0.21 keV. The fit is performed over the background lines, and the ^{60}Co sum peak and the ^{208}Tl single escape are excluded. This result, shown in Fig. 6.6, yields a reasonably good fit with a $\chi^2 = 64.3$ for 29 degrees of freedom:

$$\Delta\mu(E) = (1.78\pm0.21)\times10^{-7}(E-2614.511)^2+(3.94\pm0.36)\times10^{-4}(E-2614.511)+0.21 \quad (6.11)$$

Evaluating the difference between expected bias at the true ^{130}Te Q-value and the ^{208}Tl
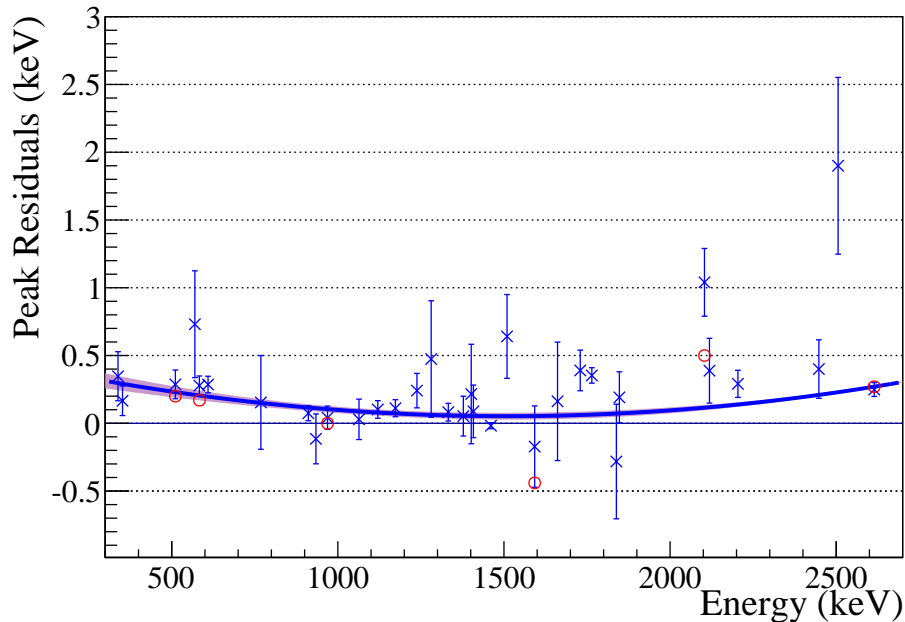
Figure 6.6: The differences between best-fit peak position and true value as a function of true peak value for the calibration peaks (red circles) and background peaks (blue crosses). The error bars for the calibration lines are smaller than the points themselves. The two outliers are the 2103 keV ^{208}Tl single escape peak and the 2506 keV ^{60}Co sum peak. The calibration residual at 1592 keV is the ^{208}Tl double-escape peak. The best-fit line does not include either the single escape or ^{60}Co peak and is constrained to pass through the ^{208}Tl calibration residual.

photopeak gives us our expected excess energy calibration residual of

$$\Delta_{0\nu\beta\beta} \equiv \Delta\mu(2527.513) - \Delta\mu(2614.511) = -0.033 \pm 0.120 \pm 0.005 \text{ keV}$$

In other words, because of our non-constant calibration bias, we expect the ^{130}Te Q-Value to be 0.033 keV *farther* from the ^{208}Tl 2615 keV line than it would otherwise be. The uncertainty from the fit itself (0.005 keV) is negligibly small, owing to the fact that the ROI is so close to the constraint at the ^{208}Tl calibration peak. Instead, we consider the primary uncertainty on our peak position to be given by the weighted RMS of the residual points about this line. (Again, neither the single escape nor the ^{60}Co sum peak are included.) This weighted RMS evaluates to 0.12 keV and is the 0.12 uncertainty in the equation above. Later in this chapter, we fold the 0.033 keV bias into our fit as a small correction to the position of our $0\nu\beta\beta$ line fit, and take 0.12 keV to be the systematic uncertainty on the peak position.

Resolution in the Background Spectrum

In addition to just measuring the peak residuals, we can also examine the resolution ratio scaling, r_σ, as a function of energy. Our central question here is if and how we need to scale the resolution from our ^{208}Tl fit to our $0\nu\beta\beta$ ROI. This will turn out to be our largest source of uncertainty, and it stems from the fact that we do not yet understand what drives our energy-dependent resolution.

There are two differences to account for when projecting our detector response to the $0\nu\beta\beta$ ROI. First, the Q-Value is at a lower energy — about 96.6% the energy of the ^{208}Tl peak position. Second, the ROI fit is done in the background data, which may have a different resolution than the calibration data. Comparing the ^{208}Tl peak fits to the background vs calibration data, we can get a direct comparison between background resolution and calibration resolutions. Here, we find that the resolution in the background data is worse by 9% ($r_\sigma = 1.089 \pm 0.019$) — at least for this one peak. While the statistics in the background data are too low to pinpoint exactly the cause of the worsened resolution, it is likely due to unresolved calibration shifting in the background data. We remove much of this calibration shifting with the WoH stabilization algorithm, but since the background data collection is longer than the calibration, the background data have more opportunity to shift. Though, at this point, this is just speculation.

In Fig. 6.5, we show the distribution of r_σ vs energy from the background peak fits. Because of the low statistics in the background data, the error bars are quite large and we cannot say with any great certainty what energy dependence the resolution scaling follows. By fitting this distribution with various functions, and picking and choosing the higher statistics peaks that seem to be fit "best", we can tune our predicted resolution scaling to just about any value between 1.0 and 1.09. In order to reflect this ignorance, we place a large uncertainty on our resolution scaling parameter at the ROI, $r_\sigma = 1.05 \pm 0.05$. This conservative approach gives us a $\sim$60% chance that we lie between 1.0 and 1.09 as we believe we do, but also a not insignificant chance that we lie outside this range — this roughly reflects our belief of r_σ. We hope to improve upon this in CUORE with dedicated R&D runs to better understand the resolution behavior.

6.5 $0\nu\beta\beta$ ROI Fit Technique

The $0\nu\beta\beta$ region of interest was defined as the range 2470–2570 keV. We produced the final unblinded CUORE-0 spectrum by removing the salted peak. This left 233 events in the $0\nu\beta\beta$ ROI. The $0\nu\beta\beta$ fit is performed as a simultaneous fit over all channel-dataset pairs, but with most of the parameters scaled from the values fit on the calibration 2615 keV peak and fixed. The model for each channel can be parameterized as

$$
\begin{aligned}
f_{\text{ROI}}(E) \;=\; \sum_{C,D} \;\; & T_{\text{Bkg}}(D,C) R_{0\nu\beta\beta} f_{\text{Det}}^{C}\left(E_{0\nu\beta\beta}(D,C), r_\sigma \sigma_0(D,C); E\right) \\
& + T_{\text{Bkg}}(D,C) R_{\text{Co}}(D) f_{\text{Det}}^{C}\left(E_{\text{Co}}(D,C), r_\sigma \sigma_0(D,C); E\right) \\
& + T_{\text{Bkg}}(D,C)(E_{\text{Max}} - E_{\text{Min}}) b_{\text{Bkg}}
\end{aligned} \tag{6.12}
$$
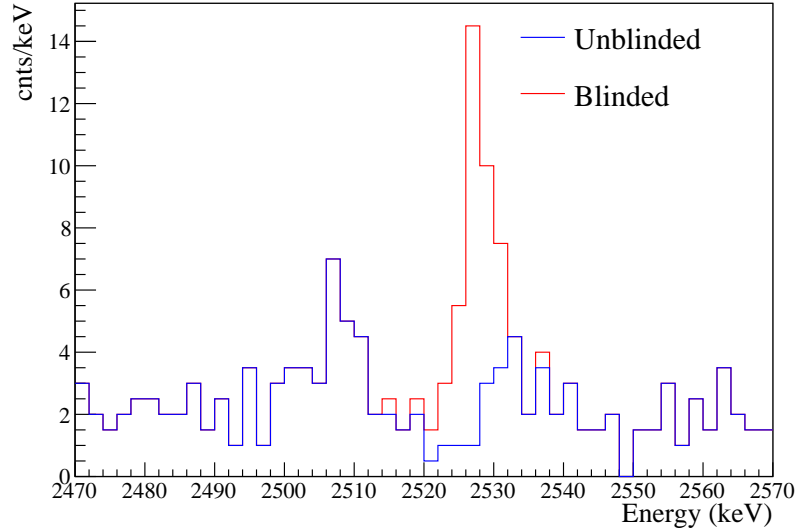
Figure 6.7: The final blinded and unblinded CUORE-0 spectra in the ROI. The peak around 2507 is the sum peak from the coincidence of the two ^{60}Co γ-rays.

This model describes two peaks, each modeled after Eqn. 6.5, which sit on top of a flat background continuum. Unlike the fits to the background peaks in the previous section, the locations of the peaks in ROI fit are set relative to the location of the ^{208}Tl peak. That is, instead of fixing all the lines to peak at $Q_{0\nu\beta\beta} = 2527.513$, the $0\nu\beta\beta$ peak position is fixed at a ChDs-dependent energy of

$$E_{0\nu\beta\beta}(D,C) = \frac{2527.513 + \Delta_{0\nu\beta\beta}}{2614.511} E_0(D,C) \tag{6.13}$$

where $\Delta_{0\nu\beta\beta} = -0.033$ keV is the small correction for our parabolic calibration bias and $E_0(D,C)$ is the fit peak location for that ChDs calibration ^{208}Tl line. By pegging the fit positions to the measured ^{208}Tl value, we accomodate the residual calibration drifts from ChDs to ChDs on top of $\Delta\mu(E)$. In essence, we are saying that for each ChDs, we expect the $0\nu\beta\beta$ peak to occur at a fraction $2527.513/2614.511$ of the *measured* ^{208}Tl peak, plus a small correction for the *average* bias of the peak positions, $\Delta_{0\nu\beta\beta}/2614.511$. This approach attempts to account for the residual shifts between the ChDs pairs (i.e. $E_0(D,C) - 2614.511 - \Delta\mu(2614.511)$) by scaling the values from the ^{208}Tl calibration line down to the ROI. We did not scale these values for the other peak fits because we cannot assume that the ChDs residuals about the ^{208}Tl peak are relevant too far away in another region of the spectrum — or to say this differently, we have 1008 residuals about the reconstructed peak position at $\approx$2614.7 keV, and we do not expect these residuals to be correlated to the residuals at 1460 keV (for example); we can only assume this in the ROI because of the close proximity to the calibration line. Anyway, this choice is a small effect and is encompassed in our systematic uncertainties.

The ^{60}Co peak position is similarly scaled down from the ChDs-dependent ^{208}Tl peak location, but the scaling r_{Co} is left floating, to account for the miscalibration of the ^{60}Co sum peak.

The parameters $R_{0\nu\beta\beta}$ and R_{Co} describe detector-wide rates for $0\nu\beta\beta$ and the ^{60}Co sum peak. Both of these rates are in units of cnts/kg/yr. The background rate, b_{bkg}, is in units of cnts/keV/kg/yr.

The $0\nu\beta\beta$ decay rate, $\Gamma_{0\nu\beta\beta}$, can be derived from $R_{0\nu\beta\beta}$ via the relationship

$$R_{0\nu\beta\beta} \equiv \varepsilon_{0\nu\beta\beta} \frac{a_I N_A}{W} \Gamma_{0\nu\beta\beta} \tag{6.14}$$

where a_I is the isotopic abundance, W is the molar mass of TeO$_2$, and N_A is Avagadro's number.

We account for the ^{60}Co lifetime by calculating the rate at the middle of each dataset

$$R_{\mathrm{Co}}(D) = R_{\mathrm{Co}} e^{-t(D)/\tau_{\mathrm{Co}}} \tag{6.15}$$

where $t(D)$ is the time of the middle of the dataset relative to the start of CUORE-0 data-taking, τ_{Co} is the ^{60}Co lifetime, and R_{Co} is the decay rate at $t = 0$, (Mar 16, 2013).

The background rate, b_{Bkg}, is in units of cnts/keV/kg/yr, and represents the time and channel-averaged total background rate between 2470 and 2570 keV.

We perform our fit using an unbinned extended maximum likelihood (UEML) method. We define our likelihood function as

$$\mathcal{L}(\mathrm{Data}|\boldsymbol{\theta}) = \frac{\lambda^N}{N!} e^{-\lambda} \prod_i^N f_{\mathrm{ROI}}(D_i, C_i, E_i; \Gamma_{0\nu\beta\beta}, \boldsymbol{\theta}) \tag{6.16}$$

where $\Gamma_{0\nu\beta\beta}$ is the $0\nu\beta\beta$ decay rate and $\boldsymbol{\theta}$ represents the three remaining free parameters of the ROI fit: the integrated ^{60}Co sum rate, R_{Co}, the ^{60}Co sum position, and the total background rate, b_{Bkg}. N is the number of events in the ROI, and λ is the expected number of events in the ROI given $\Gamma_{0\nu\beta\beta}$ and $\boldsymbol{\theta}$.

We maximize the likelihood function by using RooFit to numerically solve the system of equations

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_i} &= 0 \\ \frac{\partial \mathcal{L}}{\partial \Gamma_{0\nu\beta\beta}} &= 0 \end{aligned} \tag{6.17}$$

The best-fit values are listed in Table 6.6.

The fitted background rate in the ROI is

$$b_{\mathrm{Bkg}} = 0.059 \pm 0.005 \ \mathrm{cnts/keV/kg/yr} \tag{6.18}$$

Comparing this value to the rate measured in Cuoricino, $b_{\mathrm{Bkg}} = 0.169 \pm 0.006$ cnts/keV/kg/yr, we see that the CUORE-0 background rate at the ROI is a factor of ~ 2.5 better.

It is worth pointing out that the improvement in background appears to have come entirely from the reduction in the α background. The excess backgrounds in the ROI over

Table 6.6: The best-fit results to the ROI fit. All errors are statistical.

Parameter	Best-fit value
R_{Co}	0.87 ± 0.21 cnts/kg/yr
^{60}Co Sum Position	2507.62 ± 0.65 keV
b_{Bkg}	0.059 ± 0.005 cnts/keV/kg/yr
$\hat{\Gamma}_{0\nu\beta\beta}$	$0.00^{+0.13}_{-0.11} \times 10^{-24}$ yr
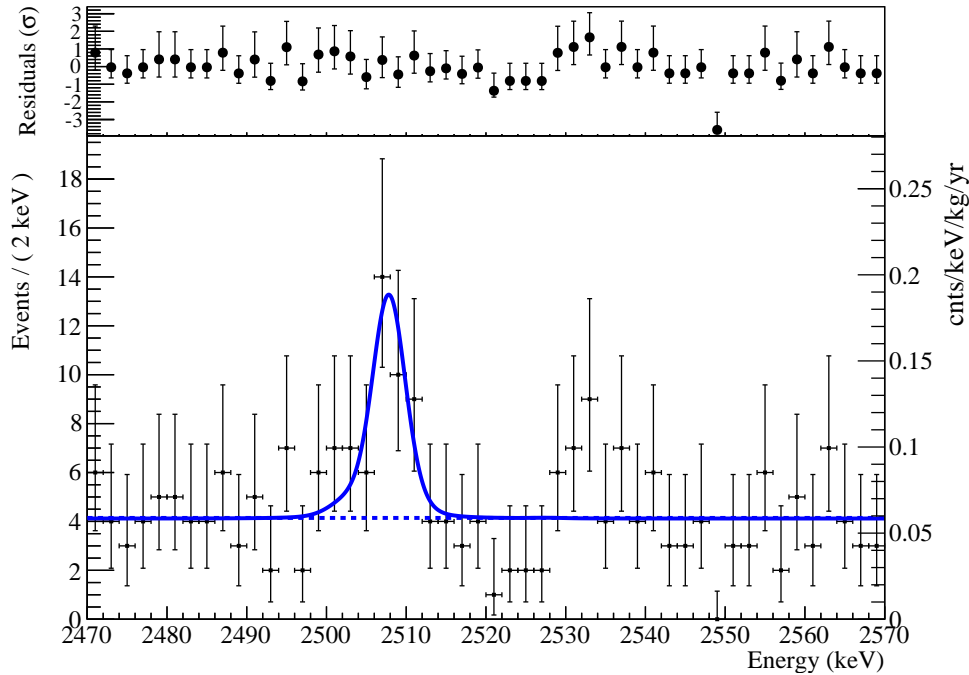
Figure 6.8: The best-fit in the ROI and the normalized residuals. The error bars are Poissonian 68% C.L. intervals.

the extrapolated α backgrounds are $b_{\mathrm{Bkg}} - b_\alpha = 0.043 \pm 0.005$ cnts/keV/kg/yr for CUORE-0 and 0.059 ± 0.006 cnts/keV/kg/yr for Cuoricino. This is consistent with the hypothesis that the excess background in both CUORE-0 and Cuoricino comes from γ contamination within the cryostat materials. This can also be seen in Fig. 6.1, which shows that the background spectrum around 2525 keV follows the tail of the calibration ^{208}Tl line. This background forms an irreducible background for this cryostat but will not be an issue for the CUORE cryostat, largely due to careful material selection, increased shielding and better self-shielding.

It is also worth pointing out that the observed ^{60}Co rate is consistent with the same rate seen in Cuoricino after accounting for the decay of ^{60}Co. This *could* imply that the majority of our ^{60}Co contamination is in the cryostat shields and not our frames. This is unusual,
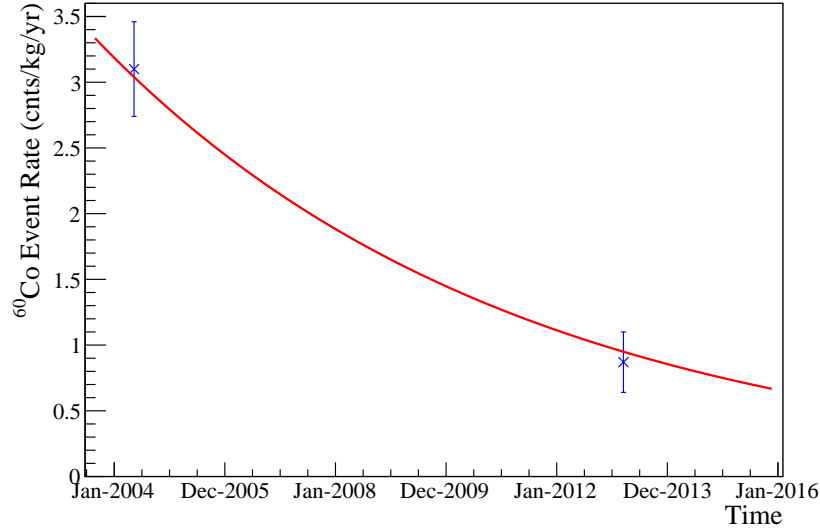
Figure 6.9: Best-fit event rate in the ^{60}Co sum peak as a function of time with its expected natural decay rate. The reduction in ^{60} rate measured between Cuoricino and CUORE-0 is *consistent* with what we expected from natural decay alone. But this claim is not conclusive, since there are other factors in play as well.

since the ^{60}Co sum peak is a coincidence of two γ-rays, it is suppressed by $\sim 1/r^4$, so the source needs to be very close to the detector. If the ^{60}Co is entirely from the cryostat, it would imply that we have no little to no cosmogenic activation in our copper frames, which would be odd, though not impossible. However, the decrease in ^{60}Co rate could also be attributed to the fact that we reduced the amount of copper in our frame by a factor of 2.3. So at this stage, this result is inconclusive.

Limit Setting Approach

Since we see no evidence for a nonzero $0\nu\beta\beta$ signal, we place a limit on the $0\nu\beta\beta$ decay rate by building a posterior Bayesian probability distribution given our data,

$$\mathcal{P}_{\text{stat}}^{\text{post}}(\Gamma_{0\nu\beta\beta}|\text{Data}) = \mathcal{P}_{\text{stat}}(\text{Data}|\Gamma_{0\nu\beta\beta})\pi(\Gamma_{0\nu\beta\beta}) \qquad (6.19)$$

We build our profile likelihood function, $\mathcal{P}_{\text{stat}}(\text{Data}|\Gamma_{0\nu\beta\beta})$, by evaluating

$$-\log\mathcal{P}_{\text{stat}}(\text{Data}|\Gamma_{0\nu\beta\beta}) \equiv -\min_{\boldsymbol{\theta}}\left\{\log\mathcal{L}(\text{Data}|\Gamma_{0\nu\beta\beta},\boldsymbol{\theta})\right\} + \log\mathcal{L}(\text{Data}|\hat{\Gamma}_{0\nu\beta\beta},\hat{\boldsymbol{\theta}}) \qquad (6.20)$$

where $\boldsymbol{\theta}$ and $\hat{\Gamma}_{0\nu\beta\beta}$ are the best-fit parameters from Eqn. 6.17 and thus yield the global minimum of the likelihood function[2]. This function is shown in Fig. 6.14.

[2]We use the symbol $\mathcal{P}$ to denote that the function is not a true probability distribution as it is not properly normalized. We will normalize it later to obtain proper probabilities.
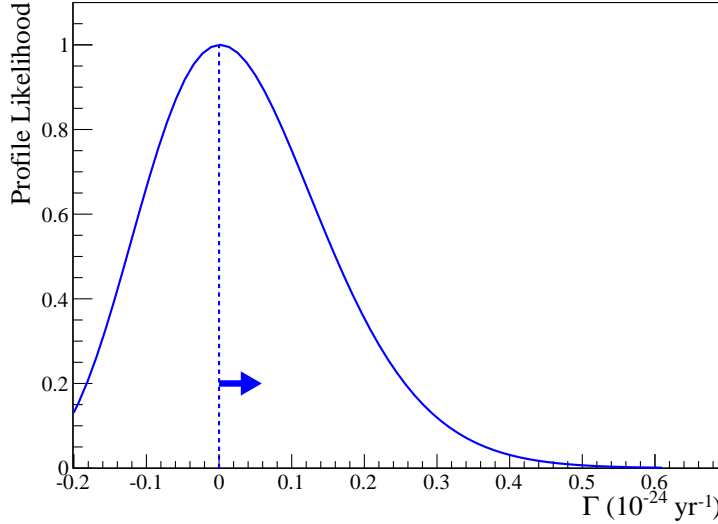
Figure 6.10: Profile likelihood curve for CUORE-0. Both the curves with and without systematics are plotted in this figure, but they cannot be discerned on a linear y-scale.

We use a flat prior, subject to the physical requirement that the decay rate $\Gamma_{0\nu\beta\beta}$ be positive. We write this as

$$\pi(\Gamma_{0\nu\beta\beta}) = \begin{cases} 1, & \Gamma_{0\nu\beta\beta} > 0 \\ 0, & \text{Otherwise} \end{cases} \tag{6.21}$$

We evaluate our limit on $\Gamma_{0\nu\beta\beta}$ by integrating our posterior likelihood function up to the required confidence level $\alpha_{\text{C.L.}}$, remembering to normalize our posterior probability to unity:

$$\alpha_{\text{C.L.}} = \int_{-\infty}^{\Gamma_{0\nu\beta\beta}^{\text{Limit}}} P_{\text{stat}}^{\text{post}}(\Gamma_{0\nu\beta\beta}|\text{Data}) \, d\Gamma_{0\nu\beta\beta} \tag{6.22}$$

$$= \frac{\int_{0}^{\Gamma_{0\nu\beta\beta}^{\text{Limit}}} \mathcal{P}(\text{Data}|\Gamma_{0\nu\beta\beta}) \, d\Gamma_{0\nu\beta\beta}}{\int_{0}^{\infty} \mathcal{P}(\text{Data}|\Gamma_{0\nu\beta\beta}) \, d\Gamma_{0\nu\beta\beta}} \tag{6.23}$$

This results in a 90% C.L. upper limit on $\Gamma_{0\nu\beta\beta}$ (accounting only for statistical uncertainty), of

$$\boxed{\Gamma_{0\nu\beta\beta} < 0.248 \times 10^{-24} \text{ yr}^{-1} \quad (90\% \text{ C.L. Stat. Only})} \tag{6.24}$$

or a half-life limit of

$$\boxed{T_{0\nu\beta\beta}^{1/2} > 2.80 \times 10^{24} \text{ yr} \quad (90\% \text{ C.L. Stat. Only})} \tag{6.25}$$

Based on statistical uncertainties only, we compare this to a limit set using a similar approach on the Cuoricino profile likelihood [93]:

$$\boxed{\Gamma_{0\nu\beta\beta} < 0.26 \times 10^{-24} \text{ yr}^{-1} \quad (\text{Cuoricino } 90\% \text{ C.L. Stat. Only})} \tag{6.26}$$
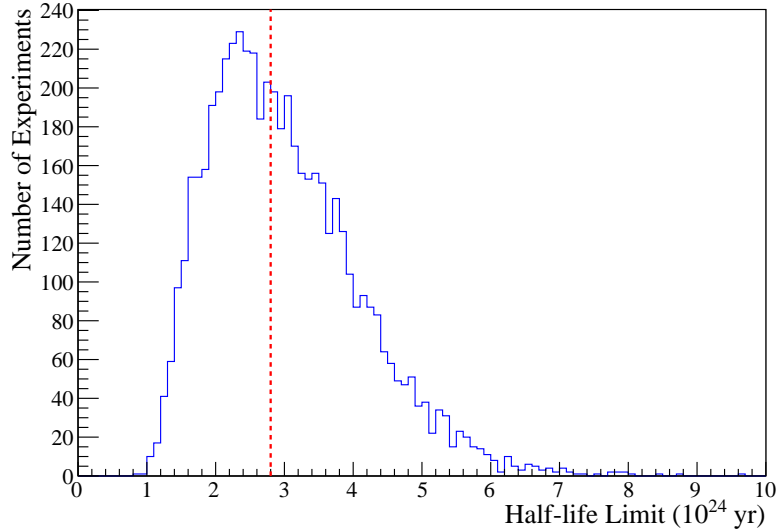
Figure 6.11: Distribution of 5485 half-life 90% Bayesian confidence limits from toy Monte Carlo experiments assuming the best-fit CUORE-0 parameters. Roughly 51% of toy Monte Carlo experiments set a stronger limit than we find in CUORE-0 (dashed red line).

or a half-life limit of

$$\boxed{T^{1/2}_{0\nu\beta\beta} > 2.7 \times 10^{24} \text{ yr} \quad (\text{Cuoricino 90\% C.L. Stat. Only})} \quad (6.27)$$

We can also compare this to the expected sensitivity given our best-fit parameters. To do this, we resample our best-fit function, with $\Gamma_{0\nu\beta\beta} = 0$ and plot the distribution of resulting 90% limits (see Fig. 6.11). We see that our result is pretty close to the median limit. We expect that about 51% of experiments would set a stronger limit.

Consistency Of the Data With The Model

Qualitatively examining the 233 events in the ROI region, we see no obvious correlation between events in time or channel. It's worth noting that we saw two events right at the Q-Value in the datasets added after the initial unblinding. This was unfortunate in that — assuming they are background events — they worsened our final result significantly.

One prominent feature of the final spectrum is the structure on either side of the $0\nu\beta\beta$ Q-value — a dip right below and a peak right above. We show in this section that these structures are consistent with statistical fluctuations of the background. As a first rough approach, we look at the region immediately above the $0\nu\beta\beta$ position, which shows a slight excess of events and ask what the probability of such an excess is. First, we note that we see no such structure in the Cuoricino final spectrum, (see Fig. 3.10) — if anything, we see

a slight deficit of events. We can treat this as a simple counting experiment; the number of events in this excess between 2531–2535 keV is 16 events, compared to the expected number of 8.3 from the best-fit flat line. The probability of an upward fluctuation of this size or larger is $p = \sum_{k=16}^{\infty} \text{Pois}(8.3; k) \approx 0.5\%$. This makes the fluctuation seem statistically significant, but this calculation singles out one 4 keV bin in a 100 keV wide window with 25 such bins. The probability that at least one 4 keV bin fluctuating up by this amount or more is $1 - (1-p)^{25} \approx 12\%$. This is well within a reasonable probability of fluctuation — this is referred to as the "look elsewhere effect".

To make the above argument more rigorous, we perform two non-parametric tests between our data and our model: a Kolmogorov-Smirnov (KS) test and an Anderson-Darling (AD) test. Both tests compare the observed cumulative distribution function (CDF), $F_n(E)$, to the predicted CDF, $F(E)$.

The KS test statistic is given by finding the maximum absolute deviation between the two CDFs,

$$D = \sup|F_n(E) - F(E)|$$

For our data, this evaluates to $D = 0.046$ which corresponds to a fluctuation probability of $P(D > .046) \approx 69\%$. So despite prominence of the observed fluctuations, we see that it is actually about twice as likely that we should have had more.

This KS test is generally a good test of the bulk of two distributions — i.e. when the two CDFs diverge significantly — but it is not very powerful when the two CDFs don't match, but never get too far from one another — i.e. when the two CDFs oscillate around each other. The AD test statistic is built by looking at the sum square distance between the two CDFs and is more sensitive in the case where the two CDFs never diverge significantly. The AD statistic can be written as

$$A^2 = -N - \sum_i \frac{2i-1}{N} (\ln F(E_i) + \ln F(E_{N+1-i}))$$

Here, the E_i are the measured energy values and have been ordered

$$E_1 < E_2 < ... < E_N$$

For our data, we have an AD statistic of $A^2 = 0.55$, which corresponds to a probability of $P(A^2 > 0.55) \approx 69\%$. Again, we see that our data are on the side of lower fluctuations[3].

6.6   Systematics Accounting

In this section, we consider the systematic uncertainties on our $0\nu\beta\beta$ limit. There are several sources that contribute significantly to our systematic uncertainty on $\hat{\Gamma}_{0\nu\beta\beta}$:

---

[3]Strictly speaking, the probabilities listed for both the KS and AD tests correspond to a comparison to a model with *no* free parameters. Since our model is fit to the data, there are fewer degrees of freedom and the percentages should increase. However, since the values listed do not rule out our model, we can use them as a conservative estimate of the likelihood of our data given the model.
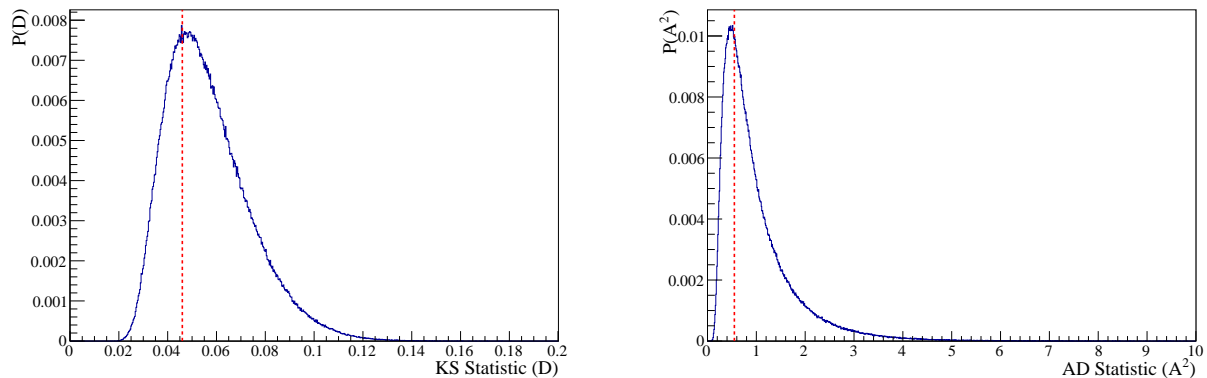
Figure 6.12: Distribution of 10^6 non-parametric statistics from toy Monte Carlos with 233 uniformly distributed points each for Kolmogorov-Smirnov (*left*) and Anderson-Darling (*right*). The value from the true data is indicated by the dashed red line.

- Uncertainty in the $0\nu\beta\beta$ peak location due to the energy reconstruction.

- Uncertainty in the resolution scaling from calibration ^{208}Tl to background $0\nu\beta\beta$.

- Uncertainty in the $0\nu\beta\beta$ signal efficiency.

- Uncertainty in the exact shape of the detector response to an $0\nu\beta\beta$ event.

- Uncertainty in the exact shape of the background.

- Bias introduced by the fitting technique itself.

The effect of these systematic uncertainties on $\hat{\Gamma}_{0\nu\beta\beta}$ is evaluated using a Monte Carlo approach and incorporated into our final limit as a modification to the profile likelihood. We begin by describing our approach to combining systematic uncertainties and then we describe the toy Monte Carlos. Finally, we present the final CUORE-0 limit including both statistical and systematic uncertainties.

Approach to Convolving Systematic Uncertainties

We include our systematic uncertainties as a modification to the previously calculated likelihood profile, $P_{\text{stat}}(\text{Data}|\Gamma_{0\nu\beta\beta})$. We begin by splitting our systematic uncertainties into two categories: absolute uncertainties and uncertainties that scale with the *true* decay rate. For each source of systematic uncertainty we calculate our error as

$$\sigma_{\text{syst,i}}(\Gamma) = \sigma_{\text{abs,i}} + \Gamma\sigma_{\text{rel,i}} \tag{6.28}$$

An example of a relative uncertainty is uncertainty in the $0\nu\beta\beta$ signal efficiency — the more events that we expect to see, the more a systematic error in $\varepsilon_{0\nu\beta\beta}$ will bias our result. We combine the σ_{syst} from multiple sources in quadrature

$$\sigma_{\text{syst}}^2(\Gamma) = (\sigma_{\text{abs},1} + \Gamma\sigma_{\text{rel},1})^2 + (\sigma_{\text{abs},2} + \Gamma\sigma_{\text{rel},2})^2 + \dots \tag{6.29}$$

We define a χ^2 statistic from our profile likelihood:

$$\chi_{\text{stat}}^2(\Gamma_{0\nu\beta\beta}) \equiv -2\log\left[\mathcal{P}_{\text{stat}}(\text{Data}|\Gamma_{0\nu\beta\beta})\right] \tag{6.30}$$

and one for our systematic uncertainties

$$\chi_{\text{syst}}^2(\Gamma_{0\nu\beta\beta}) \equiv \frac{(\Gamma_{0\nu\beta\beta} - \hat{\Gamma}_{0\nu\beta\beta})^2}{\sigma_{\text{syst}}^2(\Gamma_{0\nu\beta\beta})} \tag{6.31}$$

where $\hat{\Gamma}_{0\nu\beta\beta}$ is the best-fit decay rate value from our fit. We combine our statistical and systematic uncertainties by calculating a χ_{total}^2.

$$\frac{1}{\chi_{\text{total}}^2(\Gamma_{0\nu\beta\beta})} = \frac{1}{\chi_{\text{stat}}^2(\Gamma_{0\nu\beta\beta})} + \frac{1}{\chi_{\text{syst}}^2(\Gamma_{0\nu\beta\beta})} \tag{6.32}$$

From χ_{total}^2, we calculate our new profile likelihood, $\mathcal{P}_{\text{total}}(\text{Data}|\Gamma_{0\nu\beta\beta})$, from which we can calculate a limit that includes both statistical and systematic errors.

This approach is actually the gaussian limit of a more general approach and this equation is derived in Appendix C. The advantage here is that we only need to evaluate the effect of our nuisance parameter at the 1σ value, rather than fully sampling the distribution.

Calculating Systematic Errors on $\Gamma_{0\nu\beta\beta}$ Using Monte Carlo

Most of the systematic uncertainties on our nuisance parameters are difficult to translate into uncertainties on $\Gamma_{0\nu\beta\beta}$, so we do this by Monte Carlo. Our approach is

1. Begin with our best-fit model.

2. Modify some nuisance parameter by 1σ.

3. Generate a sample of Monte Carlo events with a simulated signal in the range $\Gamma_{0\nu\beta\beta} \in [0, 2 \times 10^{-24}]$ yr^{-1}.

4. Fit the simulated data assuming the unmodified nuisance parameter.

5. Repeat. (Typically 1,000–10,000 times depending on convergence.)

The number of events generated in each iteration is given by 233 plus the number of events expected by the assumed decay rate, $\Gamma_{0\nu\beta\beta}$. We then fit the resulting distribution of best fit rates, $\hat{\Gamma}_{0\nu\beta\beta}$, vs true decay rates, $\Gamma_{0\nu\beta\beta}$, to determine σ_{abs} and σ_{rel}.

The $0\nu\beta\beta$ line shape is treated as a discrete nuisance parameter. We tested the systematic errors using the above approach by assuming two line shapes: the model of X-ray escapes and a gaussian-only model where we set $\alpha_{\mathrm{Sub}} = 0$. The X-ray model produced a statistically significant bias, but the bias was extremely small. The gaussian-only model produced much more significant bias and we used this as our primary source of uncertainty. The MC produced a $\sigma_{\mathrm{abs}} = 7 \times 10^{-27}$ yr^{-1} and $\sigma_{\mathrm{rel}} = 2.6\%$. Since this corresponds to the extreme where the $0\nu\beta\beta$ line contains no substructure, we halve this and consider our uncertainty to be $\sigma_{\mathrm{abs}} = 3.5 \times 10^{-27}$ yr^{-1} and $\sigma_{\mathrm{rel}} = 1.3\%$ to encompass the range of possibilities from $0\nu\beta\beta$ does have substructure to $0\nu\beta\beta$ has no substructure.

For the systematic bias of the fit itself, we used the same procedure but we did not modify any nuisance parameters at all. We generated data from the best-fit model and then fit it with the best-fit model. The resulting systematics were consistent with no bias to the level of $< 1.7 \times 10^{-27}$ yr^{-1} absolute bias and $< 0.15\%$ relative bias, which was the statistical limit of the Monte Carlo.

The systematic uncertainties are summarized in Table 6.7.
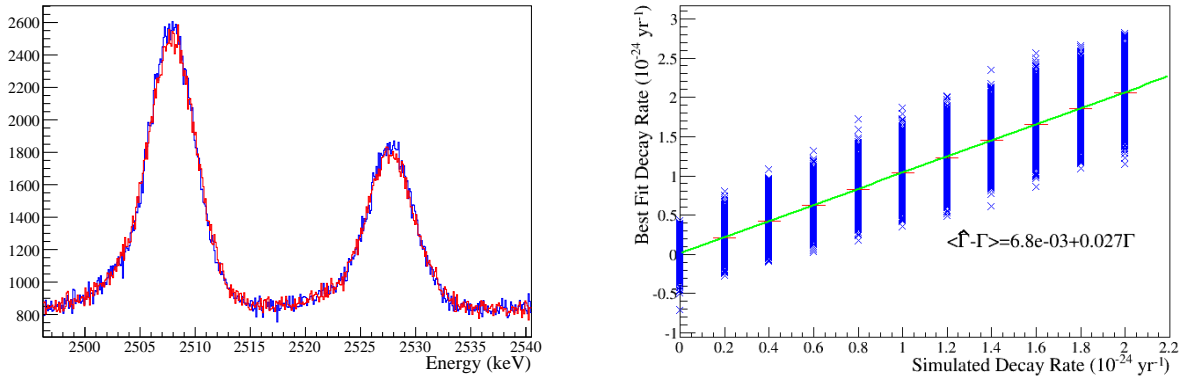
Figure 6.13: *Left:* Distribution of events from a toy MC of 1.2 million events, for $r_\sigma = 1.05$ (*blue*) and $r_\sigma = 1.10$ (*red*) and a simulated decay rate $\Gamma_{0\nu\beta\beta} = 0.4 \times 10^{-24}$ yr^{-1}. *Right:* Results of systematics toy Monte Carlo. Best-fit decay rates, $\hat{\Gamma}_{0\nu\beta\beta}$, vs simulated signal rate, $\Gamma_{0\nu\beta\beta}$ for data generated with no substructure and fit with a lineshape including substructure.

Table 6.7: Summary of systematic uncertainties and their effect on the decay rate.

	Absolute Uncertainty $(10^{-24} \text{ yr}^{-1})$	Fractional Uncertainty
Line Shape	0.004	1.3%
Energy Resolution	—	1.2%
Background Shape	0.004	0.8%
Efficiency	—	0.7%
Energy Calibration	0.005	0.4%
Fit Bias	—	—

6.7 Final CUORE-0 $0\nu\beta\beta$ Limit

In summary, including only statistical uncertainties, we are able to set the following limits on the $0\nu\beta\beta$ decay rate of ^{130}Te

$$\Gamma_{0\nu\beta\beta} < 0.248 \times 10^{-24} \text{ yr}^{-1} \quad (90\% \text{ C.L. Stat. Only}) \tag{6.33}$$

or a half-life limit of

$$T_{1/2}^{0\nu} > 2.80 \times 10^{24} \text{ yr} \quad (90\% \text{ C.L. Stat. Only}) \tag{6.34}$$

When we include systematic uncertainties, the limit on the decay rate becomes

$$\Gamma_{0\nu\beta\beta} < 0.249 \times 10^{-24} \text{ yr}^{-1} \quad (90\% \text{ C.L. Stat.+Syst.}) \tag{6.35}$$

or a half-life limit of

$$T_{1/2}^{0\nu} > 2.79 \times 10^{24} \text{ yr} \quad (90\% \text{ C.L. Stat.+Syst.}) \tag{6.36}$$

Note: This is different from the official CUORE-0 result, which used a slightly different approach and slightly different numbers. Partly the difference comes from the method of calculating and combining systematics; the rest of the difference is rounding errors. We discuss this further at the end of this chapter.

Combination With TTT and Cuoricino

The 35.2 kg·yr exposure from CUORE-0 corresponds to a ^{130}Te exposure of 9.8 kg·yr. We can combine our present result with the results of the previous two ^{130}Te bolometer experiments to form a combined analysis. We include 1.2 kg·yr of ^{130}Te exposure from the TTT experiment [92, 96], which collected data from September 2009 – January 2010 as well as the 19.75 kg·yr of ^{130}Te exposure from Runs I & II of the Cuoricino experiment [68, 93],

which collected data from 2003–2008. Cuoricino observed a typical background in the ROI of 0.169 cnts/keV/kg/yr, while TTT improved on this slightly to 0.129 cnts/keV/kg/yr.

Both Cuoricino and TTT used a natural isotopic abundance of a_I=33.8%, which has since been updated to $a_I = 34.167\%$, which we use here. We correct both the Cuoricino and TTT results to reflect this new number.

Combining the CUORE-0 result with TTT yields a 90% upper limit on the decay rate of

$$\Gamma_{0\nu\beta\beta} < 0.215 \times 10^{-24} \text{ yr}^{-1} \quad \text{(CUORE-0 + TTT)} \tag{6.37}$$

or a half-life limit of

$$T_{1/2}^{0\nu} > 3.23 \times 10^{24} \text{ yr} \quad \text{(CUORE-0 + TTT)} \tag{6.38}$$

Including the Cuoricino result yields a 90% upper limit on the decay rate of

$$\Gamma_{0\nu\beta\beta} < 0.154 \times 10^{-24} \text{ yr}^{-1} \quad \text{(CUORE-0 + TTT + Cuoricino)} \tag{6.39}$$

or a half-life limit of

$$T_{1/2}^{0\nu} > 4.50 \times 10^{24} \text{ yr} \quad \text{(CUORE-0 + TTT + Cuoricino)} \tag{6.40}$$

This is the most stringent limit on the $0\nu\beta\beta$ decay of ^{130}Te to date (see Fig 6.14).

Limit on $m_{\beta\beta}$

Using our limit on $\Gamma_{0\nu\beta\beta}$, we can derive a range of limits on the effective Majorana mass, $m_{\beta\beta}$. We use the most up-to-date calculations of $|\mathcal{M}_{0\nu}|^2$ from a variety of techniques, [48, 52–56].

We quote two limits. One limit is for comparison between similar limits placed by ^{76}Ge and ^{136}Xe, using all the same matrix element calculations. Our official limit also includes a recent ISM calculation for which the analogous calculation in ^{76}Ge is unavailable. So, our official limit is up-to-date with the most recent calculations, but our limit for comparison places all isotopes on the same footing.

Our official limit is

$$m_{\beta\beta} < 250 - 710 \text{ meV} \quad \text{(CUORE-0 + TTT + Cuoricino 90\% C.L.)} \tag{6.41}$$

We compare this to previous experiments and other experiments in Fig. 6.15.

6.8 Differences with Official CUORE-0 Result

The results presented here differ slightly from the published CUORE-0 results the collaboration presented in [108]. Here we briefly outline the differences.
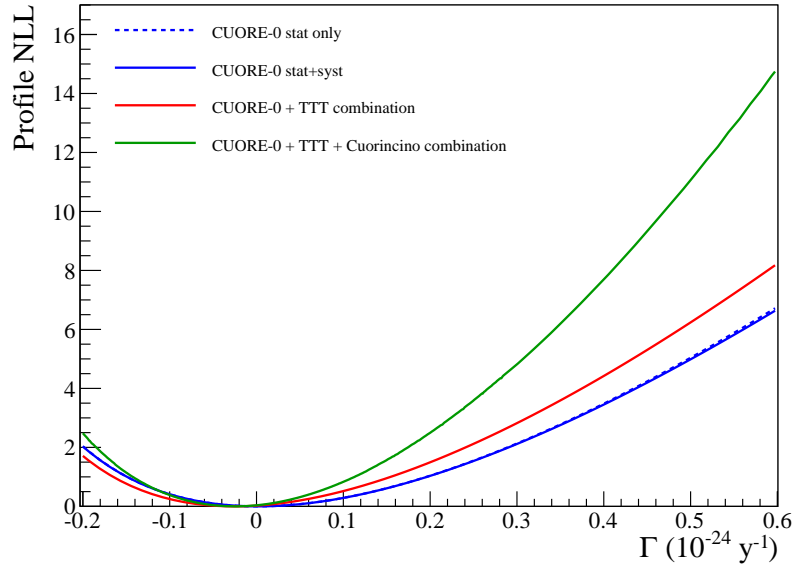
Figure 6.14: Combined negative log likelihood curves for CUORE-0, CUORE-0 + TTT and CUORE-0 + TTT + Cuoricino data.

Table 6.8: Comparison of 90% C.L. on $m_{\beta\beta}$ from various isotopes. The range quoted for ^{130}Te does not include a recent ISM calculation and is thus slightly different from our official range limit.

Isotope	$m_{\beta\beta}$ 90% Upper Limit Range	Source
^{76}Ge	$< 220 - 540$	Gerda + HDM + IGEX
^{136}Xe	$< 200 - 500$	EXO-200 + KamLAND-Zen
^{130}**Te**	$< 250 - 610$	**Cuoricino + TTT + CUORE-0**

There are two subtle differences in the method of fitting the ^{208}Tl lineshape, neither of which has any major effect on the results. As explained in [109], in the [108] result we allow the parameters that characterize the detector response substructure to vary between data-taking Campaigns I and II as well as between channels. In this thesis, the substructure is only allowed to vary by channel and not campaign. In [108], we fit the 30 keV X-ray escape peak below the ^{208}Tl line with a single gaussian with the same σ as the main peak, whereas in this thesis we account for the multiple X-ray escapes by fitting multiple constrained gaussians.

The approaches to evaluating the systematics using Monte Carlo in both this thesis and [108] are identical; however, the resulting numbers in some cases are slightly different. As explained in [110], when combining systematics in [108] we treat the additive and relative systematic errors as uncorrelated:

$$\sigma_{\text{syst}}^2(\Gamma) = \sigma_{\text{abs},1}^2 + \Gamma^2\sigma_{\text{rel},1}^2 + \sigma_{\text{abs},2}^2 + \Gamma^2\sigma_{\text{rel},2}^2 + \dots$$
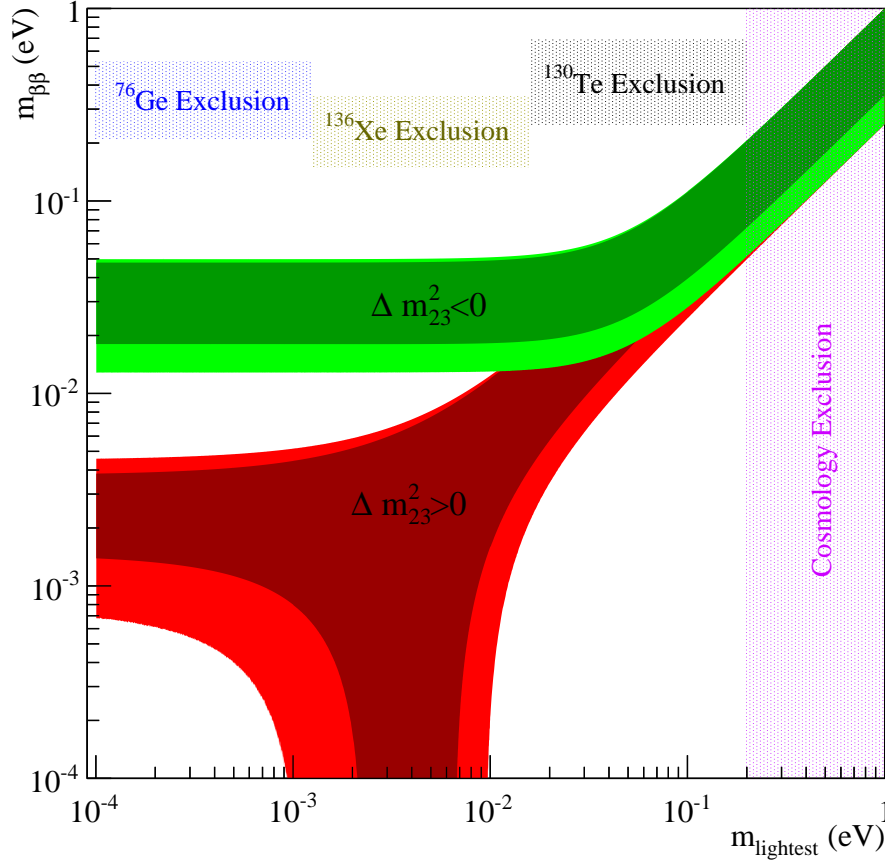
Figure 6.15: The allowed $m_{\beta\beta}$ as a function of the lightest neutrino mass. Allowed regions for normal and inverted hierarchy (assuming the mechanism of $0\nu\beta\beta$ is light Majorana neutrino exchange). The horizontal exclusion regions are an upper bound range, with all parameter space above excluded. The combined ^{76}Ge limit comes from Gerda + Heidelberg Moscow + IGEX data [65]. The combined ^{136}Xe data comes from EXO-200 and KamLAND-Zen [66, 67]. The ^{130}Te limit is the result presented here. The cosmology exclusion ($m_{\text{lightest}} < .2$ eV at 95% C.L.) comes from Plank 2015 [70].

as compared to treating them as correlated as in Eqn. 6.29.

The primary difference between the approach we used in [108] and the one we present here is how each approach handled the calibration bias. Here, we scaled our channel-dataset dependent ^{208}Tl calibration positions down to the $0\nu\beta\beta$ ROI by a scale factor $(2527.513 + \Delta_{0\nu\beta\beta})/2614.511$, where $\Delta_{0\nu\beta\beta}$ is a tiny correction, due to the energy bias, that evaluates to $\Delta_{0\nu\beta\beta} = -0.033$ keV. In [108], we do not scale the peak position down, but rather translate them down by a global energy shift given by $2614.511 - (2527.513 + \Delta_{0\nu\beta\beta})$.

These two procedures by themselves produce compatible results, but the difference is in how we estimate $\Delta_{0\nu\beta\beta}$.

For the results presented in [108], we estimated $\Delta\mu(E)$ by repeating the process of translating the detector response down to measure multiple peaks in the background spectrum, measuring each peak shift *relative* to the calibration ^{208}Tl peak, then fitting an *unconstrained* parabola to the residuals (see [111]). Notice that this function has a different meaning in [108] than it does in this thesis. In this thesis, this function is the expected total deviation between the true energy and the reconstructed energy, whereas in [108] it is the expected *relative* deviation from the expected difference between the reconstructed energy and the calibration ^{208}Tl reconstructed energy.

In [108], we estimate $\Delta_{0\nu\beta\beta}$ by evaluating $\Delta\mu(E)$ at 2527.513 keV, whereas here we take the *difference* in $\Delta\mu(E)$ at 2527.513 compared to 2614.511 keV. In [108], this yielded $\Delta_{0\nu\beta\beta} = 0.05$ keV — meaning that the $0\nu\beta\beta$ ROI moved *closer* to the calibration ^{208}Tl peak by 0.05 keV. The distance between the ROI position used in this thesis and the one used in [108] is 0.08 keV, which is smaller than the systematic uncertainty on this value of 0.12 keV.

Fundamentally, these two approaches make different assumptions about what dominates the bias $\Delta_{0\nu\beta\beta}$. In [108], we assume that the bias is dominated by a relative shift between the calibration and background data — hence why we do not constrain the fit at the calibration ^{208}Tl peak. However, in order to measure this bias, we must assume that we can measure a ≈ 50 eV effect in peaks that are over > 1 MeV away from our ROI. In this thesis, we are slightly more conservative on this point and assume that we cannot measure such an effect, and thus we must be slightly less conservative than [108] and assume that the calibration energy provides an unbiased estimate of background events.

As it happened, the approach in [108] moved the fitting region into a slight excess in the spectrum, while the approach here moved it into a slight dip. Ultimately, this question of an 80 eV discrepancy is only an issue because the difference is exacerbated by Poissonian fluctuations in the CUORE-0 background. For CUORE, we intend to address and hopefully resolve the issue of calibration bias in the data production. But even so, a 120 eV uncertainty on the $0\nu\beta\beta$ position is a sub-leading source of systematic uncertainty and will be less of a concern with the higher statistics of CUORE.

Chapter 7

CUORE and Beyond

The CUORE detector is scheduled to turn on and begin taking data in late 2015. In this chapter, we build upon the experience that we have gained with CUORE-0 to understand the outlook for CUORE, detail problems that need to be addressed for CUORE, and finally outline the key issues that an experiment beyond CUORE will need to address.

7.1 Outlook for CUORE

The first extrapolation from CUORE-0 that we can make is in regards to the expected background for CUORE. With the new cleaning and storage protocols that we developed, our goal for the CUORE background in the ROI is 0.01 cnts/keV/kg/yr. We expect that this number will be dominated by the residual α contamination originating from the surfaces of either the copper frame or PTFE supports. In CUORE-0, we measured the α component of the background to be 0.0162 ± 0.0007 cnts/keV/kg/yr. This is consistent with the prediction from our models. Specifically, if the residual α background in CUORE-0 is entirely due to surface contamination of copper, then we expect to achieve our background goal in CUORE of 0.01 cnts/keV/kg/yr (due to the smaller ratio of copper to active mass in CUORE than in CUORE-0). If, however, the α contamination is entirely from contamination on PTFE, then we expect our background to remain at $\sim$0.016 cnts/keV/kg/yr. Comparing with simulations, our data are reasonably consistent with both hypotheses, but favor the former hypothesis at $\sim 3\sigma$ level.

We measured a total background in the ROI in CUORE-0 of 0.059$\pm$0.005 cnts/keV/kg/yr. This number is dominated by γ-contamination originating in the old Cuoricino cryostat. The γ-background is expected to be significantly lower in CUORE due to more careful screening of materials, better shielding, and active anti-coincidence rejection.

Overall, we feel CUORE-0 has achieved its background goal and gives us confidence that the CUORE goal of 0.01 cnts/keV/kg/yr is within reach.

CUORE-0 also achieved its resolution goal of 5 keV FWHM at 2615 keV. This was achieved through a combination of better reproducibility in the NTD gluing as well as better

understanding of the detector response shape. Both of these aspects will carry over to CUORE. So, while the noise environment of the new cryostat is still an open question, we are confident CUORE will be able to maintain this excellent energy resolution.

Open Questions for CUORE

The analysis presented in this thesis has left open several questions and issues that should be answered or addressed for CUORE. These can be seen by simply looking through the systematic uncertainties listed in Table 6.7.

$0\nu\beta\beta$ Signal Shape and Location

The largest source of systematic uncertainty is the detector lineshape — namely, the substructure seen in the ^{208}Tl peak. We do not know the origin of this substructure and we do not know if a potential $0\nu\beta\beta$ signal would show it as well. In order to better understand the details of our detector and analysis chain, we should try to better understand the cause of this substructure through dedicated tests and bolometric modeling.

The second question which raises concern is the reconstructed position of the ^{60}Co sum peak. Presently, we do not have a concrete explanation for why this peak reconstructs 2 keV too high. We have argued in this thesis that it has to do with the specific topology of a multi-photon event, but we do not have a microscopic or bolometric model for how this occurs. So while we are confident that a $0\nu\beta\beta$ event will not share the same topology as a ^{60}Co sum event, this question opens up the possibility that a $0\nu\beta\beta$ event may not reconstruct where we think it should.

Both of these issues can actually be directly addressed without the need for modeling. There is currently a proposal to perform several studies on CUORE-0 with a ^{56}Co source. ^{56}Co has a γ line at 3549 keV, which places a double escape peak at *exactly* the $0\nu\beta\beta$ ROI. A ^{56}Co double escape mimics the exact same event topology as a $0\nu\beta\beta$ event at exactly the energy we might expect to see it. By tagging on one or both of the 511 keV annihilation photons we can reduce the background and perform a very clean measurement of both the $0\nu\beta\beta$ reconstruction position and the $0\nu\beta\beta$ line shape. We hope to have answers to both of these questions before CUORE turns on.

Bias Between Calibration and Background

The other major source of systematic uncertainty is in the validity of using the calibration data to model the background data. These issues are slightly harder to address directly. First, we believe that the parabolic shape of the peak residuals in Fig. 6.6 is due to biases in the calibration step in the data production. Specifically, we feel that modeling the peak shape with gaussians rather than the more complicated lineshape and failing to account for double lines in the calibration spectrum is systematically pulling some of the peak positions. We plan to address both of these issues in the CUORE analysis software. Further, we expect

our calibration spectrum in CUORE to be quite a bit cleaner than in CUORE-0, due to the fact that CUORE will not have shielding between the crystals and the calibration sources.

A second question that we encountered during the analysis presented in this thesis is whether the energy estimators from the calibration runs provided an unbiased estimate of the energy of background events. Measuring a bias between the calibration and background is difficult, mainly due to the low event rate in the background data. So, arguably the best way to address this issue is to remove the things that could cause a bias. Mainly this is the calibration drifts that we saw on some of the channels — particularly channel 48. It is unclear what caused some channels' calibrations to drift more than others, but it appears the best way to prevent this drift is to keep the detector and mixing chamber temperature extremely stable. This leads back to the development of the PID feedback system that we discussed in Chapter 5. We hope that this will not be an issue in CUORE, but it remains to be seen.

The last concern about using the calibration data to model the background data is in the detector energy resolution. In Chapters 5 and 6 we saw systematic differences between the calibration and background resolutions both at 2615 keV and in the filter performance on the noise at 0 keV. We currently use an empirical argument to scale the detector energy resolution from the calibration line at 2615 keV to the $0\nu\beta\beta$ ROI, but this scaling is fairly subjective and has a large error bar. In the future we would like to better understand the energy resolution behavior as a function of energy. Specifically, we would like to build a model of energy resolution of the form

$$\sigma^2(E) = f(E) + \sigma^2(0 \text{ keV})$$

where $\sigma(0 \text{ keV})$ is the RMS at 0 keV and $f(E)$ is some, currently unknown, function of energy. This would greatly help in extrapolating energy resolutions between calibration and background and between different energies.

7.2 Beyond CUORE

Looking beyond CUORE, the next generation of experiments seeks to fully explore the inverted hierarchy down to $m_{\beta\beta} \sim 10$ meV. We briefly outline the issues that a future generation $0\nu\beta\beta$ search will need to surmount in order to reach this goal.

It is possible that CUORE has reached the point of diminishing returns in terms of how clean a detector can be fabricated and assembled. It is unclear at this point if any more sensitivity can be gained out of increased radiopurity procedures. The remaining background will likely need to be removed through "active background rejection."

Particle Identification

The main drawback to bolometric detectors as they are implemented in CUORE is that they lack the ability to discriminate by particle type. The primary background for CUORE

comes from degraded α-particles from passive surfaces, whereas the signals in which we are interested are β/γ-type events. Presently, our detectors lack a way to discriminate between these two particle types. A major focus for a future detector upgrade is to develop a method for particle identification.

The energy deposited in any particle interaction is typically split between heat, light and ionization of the surrounding material. In a good bolometer, the majority (if not all) of the energy is converted into heat — this is what yields such excellent energy resolution. However, the exact split between light, heat and ionization is determined both by the properties of the detector material and the nature of the interaction itself. It is this latter fact that we would like to leverage to discriminate α-type events from β/γ-type.

Several R&D projects have shown very promising results in this direction. For example, the LUCIFER R&D project used a second bolometer made from a semi-conductor material as a light collector next to a scintillating primary bolometer. By measuring both the heat signature on the primary bolometer and the heat signature on the light collector (called the light yield), they were able to perform very powerful background rejection [112]. Several other projects have shown similarly promising capabilities, (see [113] for a summary).
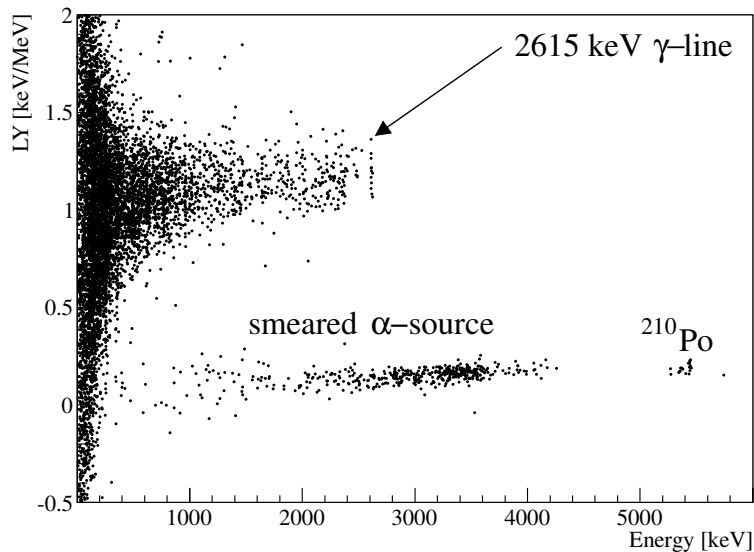
Figure 7.1: An example of particle type discrimination using a scintillating $ZnMoO_4$ crystal. The plot shows energy deposited as heat (X-axis) vs Light Yield on the light collector (Y-axis). There is a clear separation between α type events and β/γ type events. (Plot taken from [112].)

This approach is difficult (or impossible) in TeO_2, because it is such a poor scintillator. But another possibility along similar lines, is to try to detect the Cherenkov light that would be emitted with β particles, but not α particles. Since the amount of energy released in Cherenkov light is small ($\sim$200 eV) we would require a light collector with an extremely low

threshold. One of the promising technologies towards this goal is Transition Edge Sensors (TES). These are superconductors that are operated in the middle of their transition from normal to super conducting. These can be tuned to have a very narrow transition, making their resistances a very steep function of temperature. Thus, these make very good temperature sensors. Often these sensors have low electrical impedance which gives them several advantages over CUORE-style NTDs: they have larger bandwidth, can handle faster signals and they can be multiplexed in large arrays. This technology has been well established in the fields of X-ray astronomy and Cosmic Microwave Background (CMB) cosmology.

One difficulty with TES sensors is that they are only useful over a very narrow range of temperatures around their transition temperature, T_c. There are very few known materials which have a transition temperature around ∼10–15 mK. One technology which shows promise is bilayer superconductors. Here, we prepare a thin film composed of one superconducting material and one non-superconducting material. By tuning the ratio of the thicknesses, we can tune the transition temperature. In Fig. 7.2, we show a transition for a Ir/Au bilayer sample which has a $T_c \approx 21.4$ mK.
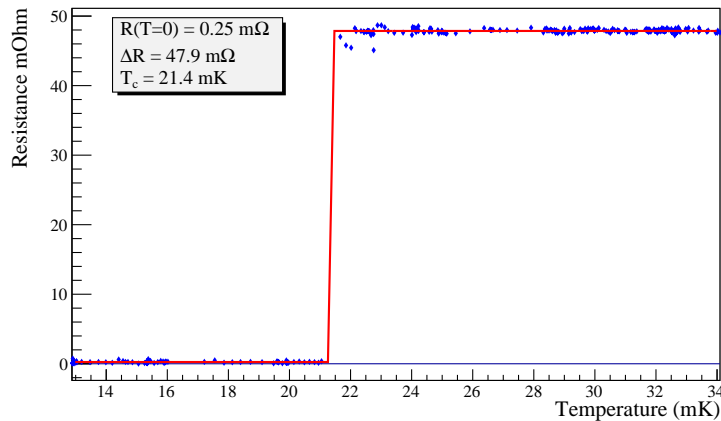
Figure 7.2: Transition curve for a Ir/Au bilayer TES. This sample has a transition temperature around 21 mK. This measurement was performed with a current-biased resistance bridge, so the transition is artificially narrow.

There is currently an intense ongoing research effort into developing particle identification techniques for bolometric detectors. Within CUORE we have formed an interest group to accomplish just this [113, 114]. This new group is called the **CUORE Upgrade with Particle ID**entification — CUPID.

Bibliography

[1]   Aad, G., Abajyan, T., Abbott, B. *et al.* (ATLAS Colaboration). *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.* Phys. Lett. B (2012). **716** (1), 1.

[2]   Chatrchyan, S., Khachatryan, V., Sirunyan, A. *et al.* (CMS Collaboration). *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.* Phys. Lett. B (2012). **716** (1), 30.

[3]   Ade, P. A. R., Aghanim, N., Armitage-Caplan, C. *et al.* (Planck Collaboration). *Planck 2013 results. XVI. Cosmological parameters* (2013). p. 67. `arXiv:1303.5076`.

[4]   Kolb, E. W. and Turner, M. S. *The Early Universe* (Westview Press, 1990).

[5]   Steigman, G. *Observational tests of antimatter cosmologies.* Annu. Rev. Astron. Astrophys. (1976). **14** (1), 339.

[6]   Prokhorov, D. *Upper bounds on matter-antimatter admixture from gamma-ray observations of colliding clusters of galaxies with the Fermi Large Area Telescope.* Phys. Rev. D (2015). **91** (8), 083002.

[7]   Sakharov, A. D. *Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe.* Pisma Zh. Eksp. Teor. Fiz. (1967). **5**, 32.

[8]   Huet, P. *Electroweak baryogenesis and the standard model* (1994). pp. 77–91. `hep-ph/9406301`.

[9]   Weinberg, S. *Cosmological production of baryons.* Phys. Rev. Lett. (1979). **42**, 850.

[10]   Toussaint, D., Treiman, S. B., Wilczek, F. *et al. Matter-antimatter accounting, thermodynamics, and black-hole radiation.* Phys. Rev. D (1979). **19**, 1036.

[11]   Affleck, I. and Dine, M. *A New Mechanism for Baryogenesis.* Nucl. Phys. B (1985). **249**, 361.

[12]   Kuzmin, V. A., Rubakov, V. A., and Shaposhnikov, M. E. *On anomalous electroweak baryon-number non-conservation in the early universe.* Phys. Lett. B (1985). **155**, 36.

[13] Fukugita, M. and Yanagida, T. *Baryogenesis without grand unification.* Phys. Lett. B (1986). **174** (1), 45 .

[14] Cleveland, B. T., Daily, T., Raymond Davis, J. *et al. Measurement of the solar electron neutrino flux with the homestake chlorine detector.* Astrophys. J. (1998). **496** (1), 505.

[15] Fukuda, Y., Hayakawa, T., Ichihara, E. *et al. Measurements of the Solar Neutrino Flux from Super-Kamiokande's First 300 Days.* Phys. Rev. Lett. (1998). **81**, 1158.

[16] Fukuda, Y., Hayakawa, T., Inoue, K. *et al. Solar neutrino data covering solar cycle 22.* Phys. Rev. Lett. (1996). **77**, 1683.

[17] Abdurashitov, J. N., Gavrin, V. N., Gorbachev, V. V. *et al. Measurement of the solar neutrino capture rate with gallium metal. III. Results for the 2002–2007 data-taking period.* Phys. Rev. C (2009). **80**, 015807.

[18] Anselmann, P. *et al.* (GALLEX). *GALLEX solar neutrino observations: The Results from GALLEX-I and early results from GALLEX-II.* Phys. Lett. B (1993). **314**, 445.

[19] Hampel, W. *et al.* (GALLEX). *GALLEX solar neutrino observations: Results for GALLEX IV.* Phys. Lett. B (1999). **447**, 127.

[20] Altmann, M. *et al.* (GNO). *Complete results for five years of GNO solar neutrino observations.* Phys. Lett. B (2005). **616**, 174. `hep-ex/0504037`.

[21] Fukuda, S. *et al.* (Super-Kamiokande). *Determination of solar neutrino oscillation parameters using 1496 days of Super-Kamiokande I data.* Phys. Lett. B (2002). **539**, 179. `hep-ex/0205075`.

[22] Ahmad, Q. R., Allen, R. C., Andersen, T. C. *et al. Measurement of the Rate of $\nu_e + d \to p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory.* Phys. Rev. Lett. (2001). **87**, 071301.

[23] Ahmad, Q. R., Allen, R. C., Andersen, T. C. *et al. Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory.* Phys. Rev. Lett. (2002). **89**, 011301.

[24] Eguchi, K., Enomoto, S., Furuno, K. *et al. First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance.* Phys. Rev. Lett. (2003). **90**, 021802.

[25] Gando, A., Gando, Y., Ichimura, K. *et al. Constraints on θ_{13} from a three-flavor oscillation analysis of reactor antineutrinos at KamLAND.* Phys. Rev. D (2011). **83**, 052002.

[26] K. A. Olive *et al.* (Particle Data Group), Chin. Phys C, 38, 090001 (2014).

[27] Aguilar, A., Auerbach, L. B., Burman, R. L. *et al. Evidence for neutrino oscillations from the observation of $\overline{\nu}_e$ appearance in a $\overline{\nu}_\mu$ beam.* Phys. Rev. D (2001). **64**, 112007.

[28] Aguilar-Arevalo, A. A., Anderson, C. E., Brice, S. J. *et al. Event Excess in the Mini-BooNE Search for $\overline{\nu}_\mu \rightarrow \overline{\nu}_e$ Oscillations.* Phys. Rev. Lett. (2010). **105**, 181801.

[29] Aguilar-Arevalo, A. A., Brown, B. C., Bugel, L. *et al. Improved Search for $\overline{\nu}_\mu \rightarrow \overline{\nu}_e$ Oscillations in the MiniBooNE Experiment.* Phys. Rev. Lett. (2013). **110**, 161801.

[30] Mention, G., Fechner, M., Lasserre, T. *et al. Reactor antineutrino anomaly.* Phys. Rev. D (2011). **83**, 073006.

[31] Mueller, T., Lhuillier, D., Fallot, M. *et al. Improved Predictions of Reactor Antineutrino Spectra.* Phys. Rev. C (2011). **83**, 054615. `arXiv:1101.2663`.

[32] Abdurashitov, J. N., Gavrin, V. N., Girin, S. V. *et al. The Russian-American Gallium Experiment (SAGE) Cr Neutrino Source Measurement.* Phys. Rev. Lett. (1996). **77**, 4708.

[33] Abdurashitov, J. *et al.* (SAGE). *Measurement of the response of the Russian-American gallium experiment to neutrinos from a Cr-51 source.* Phys. Rev. C (1999). **59**, 2246. `hep-ph/9803418`.

[34] Xing, Z.-Z. *Properties of CP violation in neutrino-antineutrino oscillations.* Phys. Rev. D (2013). **87**, 053019.

[35] Šimkovic, F., Bilenky, S. M., Faessler, A. *et al. Possibility of measuring the CP Majorana phases in $0\nu\beta\beta$ decay.* Phys. Rev. D (2013). **87**, 073002.

[36] King, S. F. and Luhn, C. *Neutrino Mass and Mixing with Discrete Symmetry.* Rept. Prog. Phys. (2013). **76**, 056201. `arXiv:1301.1340`.

[37] Capozzi, F., Fogli, L., G. Lisi, E. *et al. Status of three-neutrino oscillation parameters, circa 2013.* Phys. Rev. D (2014). **89**, 093018.

[38] Goeppert-Mayer, M. *Double beta-disintegration.* Phys. Rev. (1935). **48**, 512.

[39] Ekström, L. P. and Firestone, R. B. *WWW Table of radioactive isotopes* (2004). URL `http://ie.lbl.gov/toi/`.

[40] Barabash, A. S. *Precise half-life values for two-neutrino double-β decay.* Phys. Rev. C (2010). **81** (3), 035501. `arXiv:1003.1005`.

[41] Arnold, R., Augier, C., Baker, J. *et al. Measurement of the $\beta\beta$ Decay Half-Life of ^{130}Te with the NEMO-3 Detector.* Phys. Rev. Lett. (2011). **107** (6), 062504. `arXiv:1104.3716`.

[42] Albert, J. B., Auger, M., Auty, D. J. *et al.* (EXO Collaboration). *Improved measurement of the 2νββ half-life of* ^{136}Xe *with the EXO-200 detector.* Phys. Rev. C (2014). **89** (1), 015502. `arXiv:1306.6106`.

[43] Primakoff, H. and Rosen, S. P. *Double beta decay.* Reports on Progress in Physics (1959). **22** (1), 121.

[44] Majorana, E. *Teoria simmetrica dellelettrone e del positrone.* Nuovo Cim. (1937). **14** (4), 171.

[45] Racah, G. *On the symmetry of particle and antiparticle.* Nuovo Cim. (1937). **14**, 322.

[46] Furry, W. H. *On transition probabilities in double beta-disintegration.* Phys. Rev. (1939). **56**, 1184.

[47] Schechter, J. and Valle, J. W. F. *Neutrinoless double-β decay in SU(2)×U(1) theories.* Phys. Rev. D (1982). **25**, 2951.

[48] Barea, J., Kotila, J., and Iachello, F. *Nuclear matrix elements for double-β decay.* Phys. Rev. C (2013). **87** (1), 014315. `arXiv:1301.4203`.

[49] Šimkovic, F., Pantis, G., Vergados, J. D. *et al. Additional nucleon current contributions to neutrinoless double β decay.* Phys. Rev. C (1999). **60** (5), 055502. `hep-ph/9905509`.

[50] Yao, W.-M. *et al. Review of particle physics.* J. Phys. G: Nuclear and Particle Physics (2006). **33** (1), 1.

[51] Kotila, J. and Iachello, F. *Phase-space factors for double-β decay.* Phys. Rev. C (2012). **85**, 034316.

[52] Šimkovic, F., Rodin, V., Faessler, A. *et al. 0νββ and 2νββ nuclear matrix elements, quasiparticle random-phase approximation, and isospin symmetry restoration.* Phys. Rev. C (2013). **87**, 045501.

[53] Hyvärinen, J. and Suhonen, J. *Nuclear matrix elements for 0νββ decays with light or heavy majorana-neutrino exchange.* Phys. Rev. C (2015). **91**, 024613.

[54] Neacsu, A. and Horoi, M. *Shell model studies of the* ^{130}Te *neutrinoless double-β decay.* Phys. Rev. C (2015). **91**, 024309.

[55] Menendez, J., Poves, A., Caurier, E. *et al. Disassembling the Nuclear Matrix Elements of the Neutrinoless beta beta Decay.* Nucl.Phys. (2009). **A818**, 139. `arXiv:0801.3760`.

[56] Rodríguez, T. R. and Martínez-Pinedo, G. *Energy density functional study of nuclear matrix elements for neutrinoless ββ decay.* Phys. Rev. Lett. (2010). **105**, 252503.

[57] Faessler, A., Kovalenko, S., and Šimkovic, F. *Bilinear R-parity violation in neutrinoless double beta decay.* Phys. Rev. D (1998). **58** (5), 055004. `hep-ph/9712535`.

[58] Rodejohann, W. *Neutrino-less double beta decay and particle physics.* Int. J. Mod. Phys. E (2011). **20** (09), 1833. `arXiv:1106.1334`.

[59] Tello, V., Nemevšek, M., Nesti, F. *et al. Left-Right Symmetry: From the LHC to Neutrinoless Double Beta Decay.* Phys. Rev. Lett. (2011). **106** (15), 151801. `arXiv:1011.3522`.

[60] Chikashige, Y., Mohapatra, R. N., and Peccei, R. *Are There Real Goldstone Bosons Associated with Broken Lepton Number?* Phys. Lett. (1981). **B98**, 265.

[61] Chikashige, Y., Mohapatra, R. N., and Peccei, R. D. *Spontaneously broken lepton number and cosmological constraints on the neutrino mass spectrum.* Phys. Rev. Lett. (1980). **45**, 1926.

[62] Aulakh, C. and Mohapatra, R. N. *Neutrino as the Supersymmetric Partner of the Majoron.* Phys.Lett. (1982). **B119**, 136.

[63] Carone, C. D. *Double beta decay with vector majorons.* Phys. Lett. B (1993). **308**, 85. `hep-ph/9302290`.

[64] Alessandria, F., Ardito, R., Artusa, D. R. *et al. Sensitivity and Discovery Potential of CUORE to Neutrinoless Double-Beta Decay* (2011). p. 14. `arXiv:1109.0494`.

[65] Agostini, M., Allardt, M., Andreotti, E. *et al. Results on Neutrinoless Double-β Decay of ^{76}Ge from Phase I of the GERDA Experiment.* Phys. Rev. Lett. (2013). **111** (12), 122503. `arXiv:1307.4720`.

[66] Albert, J. B., Auty, D. J., Barbeau, P. S. *et al.* (EXO Collaboration). *Search for Majorana neutrinos with the first two years of EXO-200 data.* Nature (2014). **510**, 229. `arXiv:1402.6956`.

[67] Gando, A. *et al.* (KamLAND-Zen Collaboration). *Limit on Neutrinoless $\beta\beta$ Decay of ^{136}Xe from the First Phase of KamLAND-Zen and Comparison with the Positive Claim in ^{76}Ge.* Phys. Rev. Lett. (2013). **110** (6), 062502. `arXiv:1211.3863`.

[68] Andreotti, E., Arnaboldi, C., Avignone, F. T. *et al.* ^{130}Te *neutrinoless double-beta decay with CUORICINO.* Astropart. Phys. (2011). **34**, 822. `arXiv:1012.3266`.

[69] Dell'Oro, S., Marcocci, S., and Vissani, F. *New expectations and uncertainties on neutrinoless double beta decay.* Phys. Rev. D (2014). **90** (3), 033005. `arXiv:1404.2616`.

[70] Ade, P. *et al.* (Planck). *Planck 2015 results. XIII. Cosmological parameters* (2015). `arXiv:1502.01589`.

[71] Aseev, V. N., Belesev, A. I., Berlev, A. I. *et al.* (Troitsk Collaboration). *Upper limit on the electron antineutrino mass from the Troitsk experiment.* Phys. Rev. D (2011). **84** (11), 112003. `arXiv:1108.5034`.

[72] Parno, D. S. *The Katrin Experiment: Status and Outlook.* In A. Kostelecky, ed., *CPT and Lorentz Symmetry - Proceedings of the Sixth Meeting* (2014) pp. 77–80. `arXiv:1307.5289`.

[73] Doe, P. J., Kofron, J., McBride, E. L. *et al. Project 8: Determining neutrino mass from tritium beta decay using a frequency-based method* (2013). `arXiv:1309.7093`.

[74] Agarwalla, S. K., Agostino, L., Aittola, M. *et al.* (LAGUNA-LBNO Collaboration). *The mass-hierarchy and CP-violation discovery reach of the LBNO long-baseline neutrino experiment* (2013). `arXiv:1312.6520`.

[75] Patterson, R. B. *The NOνA experiment: status and outlook.* Nucl. Phys. B. (Proc. Suppl.) (2013). **235**, 151. `arXiv:1209.0716`.

[76] LBNE Collaboration, Adams, C., Adams, D. *et al. The Long-Baseline Neutrino Experiment: Exploring Fundamental Symmetries of the Universe* (2013). `arXiv:1307.7335`.

[77] Aad, G., Abbott, B., Abdallah, J. *et al. Search for heavy neutrinos and right-handed W bosons in events with two leptons and jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector.* Eur. Phys. J. C (2012). **72**, 2056. `arXiv:1203.5420`.

[78] *Search for a heavy neutrino and right-handed W of the left-right symmetric model in pp collisions at 8 TeV.* Tech. Rep. CMS-PAS-EXO-12-017, CERN, Geneva (2012).

[79] CMS Collaboration. *Search for heavy Majorana neutrinos in $\mu^{\pm}\mu^{\pm}+$ jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV* (2015). `arXiv:1501.05566`.

[80] Ambrosio, M., Antolini, R., Auriemma, G. *et al. Vertical muon intensity measured with MACRO at the Gran Sasso laboratory.* Phys. Rev. D (1995). **52**, 3793.

[81] Kittel, C. *Introduction to Solid State Physics* (Wiley, 2004), 8 ed.

[82] Barucci, M., Brofferio, C., Giuliani, A. *et al. Measurement of low temperature specific heat of crystalline TeO_2 for the optimization of bolometric detectors.* J. Low Temp. Phys. (2001). **123** (5-6), 303.

[83] McCammon, D. *Thermal Equilibrium Calorimeters – An Introduction.* In C. Enss, ed., *Cryogenic Particle Detection*, p. 32 (Springer; Berlin 2005, 2005). `arXiv:physics/0503045`.

[84] Alessandrello, A., Brofferio, C., Cremonesi, O. *et al. First tests on a large mass, low temperature array detector.* Nucl. Instrum. Meth. (1995). **A360**, 363.

[85] Arnaboldi, C., Brofferio, C., Bryant, A. *et al. Production of high purity TeO$_2$ single crystals for the study of neutrinoless double beta decay.* J. Cryst. Growth (2010). **312**, 2999. `arXiv:1005.3686`.

[86] Mccammon, D. *Semiconductor Thermistors.* In C. Enss, ed., *Cryogenic Particle Detection*, p. 35 (Springer Berlin Heidelberg, 2005).

[87] Efros, A. L. and Shklovskii, B. I. *Coulomb gap and low temperature conductivity of disordered systems.* J. Phys. C: Solid State Physics (1975). **8**, L49.

[88] Alfonso, K. *et al. CUORE-0 detector: design, construction and operation* (In preparation).

[89] Haller, E. E. *Isotopically engineered semiconductors.* J. Appl. Phys. (1995). **77** (7), 2857.

[90] Arnaboldi, C., Liu, X., and Pessina, G. *The preamplifier for cuore, an array of large mass bolometers.* In *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE* (2009) pp. 389–395.

[91] Arnaboldi, C., Bucci, C., Campbell, J. *et al. The programmable front-end system for cuoricino, an array of large-mass bolometers.* Nuclear Science, IEEE Transactions on (2002). **49** (5), 2440.

[92] Bryant, A. D. *A Search for Neutrinoless Double Beta Decay of ^{130}Te.* Ph.D. thesis, University of California, Berkeley (2010).

[93] Carrettoni, M. A. *Data Analysis For Neutrinoless Double Beta Decay.* Ph.D. thesis, Universitá Degli Studi Di Milano Bicocca (2011).

[94] Barghouty, A. F., Brofferio, C., Capelli, S. *et al. Measurements of proton-induced radionuclide production cross sections to evaluate cosmic-ray activation of tellurium* (2010). `arXiv:1010.4066`.

[95] Alessandria, F., Andreotti, E., Ardito, R. *et al. CUORE crystal validation runs: Results on radioactive contamination and extrapolation to CUORE background.* Astropart. Phys. (2012). **35**, 839. `arXiv:1108.4757`.

[96] Alessandria, F., Ardito, R., Artusa, D. R. *et al. Validation of techniques to mitigate copper surface contamination in CUORE.* Astropart. Phys. (2013). **45**, 13. `arXiv:1210.1107`.

[97] Alessandrello, A., Arpesella, C., Brofferio, C. *et al. Measurements of internal radioactive contamination in samples of roman lead to be used in experiments on rare events.* Nucl. Inst. Meth. Phys. Res. B (1998). **142** (12), 163 .

[98] LBL News Center. *Creating the coldest cubic meter in the universe [press release]* (2014). URL `http://newscenter.lbl.gov/2014/10/28/creating-the-coldest-cubic-meter-in-the-universe/`.

[99] Di Domizio, S., Orio, F., and Vignati, M. *Lowering the energy threshold of large-mass bolometric detectors.* J. Instrum. (2011). **6**, 2007. `arXiv:1012.1263`.

[100] Martinez, M. *Stabilization without heater* (2014). CUORE document ID: cuore-doc-26-v1.

[101] Biassoni, M., Gironi, L., and Pozzi, S. *Pulse shape normalization* (2015). CUORE document ID: cuore-doc-8-v5.

[102] Tomei, C. *Coincidences with response synchronization in* Diana (2014). CUORE document ID: cuore-doc-9-v3.

[103] Mancini-Terracciano, C. and Vignati, M. *Noise correlation and decorrelation in arrays of bolometric detectors.* J. Instrum. (2012). **7** (06), P06013.

[104] Gatti, E. and Manfredi, P. *Processing the Signals from Solid-State Detectors in Elementary-Particle Physics.* La Rivista del Nuovo Cimento (1986). **9** (1). URL `http://link.springer.com/article/10.1007/BF02822156`.

[105] Lim, K. *CUORE-0 cut efficiency evaluation* (2015). CUORE document ID: cuore-doc-6-v5.

[106] Clopper, C. J. and Pearson, E. S. *The use of confidence or fiducial limits illustrated in the case of the binomial.* Biometrika (1934). **26** (4), 404.

[107] Thompson, A. C. and Vaughan, D., eds. *X-Ray Data Booklet* (Lawrence Berkeley National Laboratory, 2001), 2 ed.

[108] Alfonso, K. *et al.* (CUORE). *Search for Neutrinoless Double-Beta Decay of ^{130}Te with CUORE-0* (2015). (Submitted to PRL), `arXiv:1504.02454`.

[109] Banks, T. *Evaluation of the CUORE-0 resolution* (2015). CUORE document ID: cuore-doc-7-v4.

[110] Han, K. *CUORE-0 background rates and projected sensitivity* (2014). CUORE document ID: cuore-doc-27-v3.

[111] Cushman, J. *Calibration performance in CUORE-0* (2015). CUORE document ID: cuore-doc-201-v2.

[112] Beeman, J., Bellini, F., Capelli, S. *et al. ZnMoO$_4$: A Promising bolometer for neutrinoless double beta decay searches.* Astropart. Phys. (2012). **35**, 813. `arXiv:1202.0238`.

[113] Group, T. C. I. *R&D towards CUPID (CUORE Upgrade with Particle IDentification)* (2015). `arXiv:1504.03612`.

[114] Group, T. C. I. *CUPID: CUORE (Cryogenic Underground Observatory for Rare Events) Upgrade with Particle IDentification* (2015). `arXiv:1504.03599`.

[115] Fehr, M. A., Rehkmper, M., and Halliday, A. N. *Application of mc-icpms to the precise determination of tellurium isotope compositions in chondrites, iron meteorites and sulfides.* Int. J. Mass Spec. (2004). **232** (1), 83 .

[116] Wieser, M. E., Holden, N., Coplen, T. B. *et al. Atomic weights of the elements.* Pure Appl. Chem. (2013). **85** (5), 1047.

[117] Redshaw, M., Mount, B. J., Myers, E. G. *et al. Masses of ^{130}Te and ^{130}Xe and Double-β-Decay Q Value of Te130.* Phys. Rev. Lett. (2009). **102** (21), 212502. `arXiv:0902.2139`.

[118] Fixsen, D. J. *The Temperature of the Cosmic Microwave Background.* Astrophys. J. (2009). **707**, 916. `arXiv:0911.1955`.

[119] Wegner, G. and Glass, I. S. *A new bipolar nebula in Centaurus.* Mon. Not. R. Astron. Soc. (1979). **188**, 327.

[120] Sahai, R. and Nyman, L.-A. k. *The Boomerang Nebula: The Coldest Region of the Universe?* Astrophys. J. (1997). **487** (2), L155.

[121] Schaeffer, D., Nucciotti, A., Alessandria, F. *et al. The cryostat of the CUORE Project, a 1-ton scale cryogenic experiment for Neutrinoless Double Beta Decay Research.* JPCS (2009). **150** (1), 012042.

[122] Alessandria, F., Biassoni, M., Ceruti, G. *et al. The 4 K outer cryostat for the CUORE experiment: Construction and quality control.* Nucl. Inst. Meth. Phys. Res. A (2013). **727**, 65. `arXiv:1306.1351`.

[123] Nucciotti, A., Schaeffer, D., Alessandria, F. *et al. Design of the Cryogen-Free Cryogenic System for the CUORE Experiment.* J. Low Temp. Phys. (2008). **151** (3-4), 662.

[124] Schaeffer, D., Nucciotti, A., Alessandria, F. *et al. The cryostat of the CUORE Project, a 1-ton scale cryogenic experiment for Neutrinoless Double Beta Decay Research.* JPCS (2009). **150** (1), 012042.

[125] Lang, R. F. *Search for Dark Matter with the CRESST Experiment.* Ph.D. thesis (2008).

[126] *Edelweiss website.* `http://edelweiss.in2p3.fr/Presentation/experiment.php`. Accessed July 29, 2014.

[127] CDMS Collaboration. *CDMS II A Search for Cold Dark Matter with Cryogenic Detectors at the Soudan Mine Proposal.* Tech. Rep. March (1999).

Appendix A

CUORE-0 Dataset Data

Table A.1: Grouping of datasets into working points and data taking phases. The working points are determined by detector performance, while the data taking phases were driven by data processing and conference presentations.

Campaign	Phase	Working Point	Dataset
I	I	1	2049
		2	2061
			2064
		3	2067
			2070
			2073
			2076
II	II	4	2079
		5	2085
			2088
			2091
			2097
			2100
	III		2103
			2109
		6	2118
			2124
	IV	7	2130
			2133
			2139
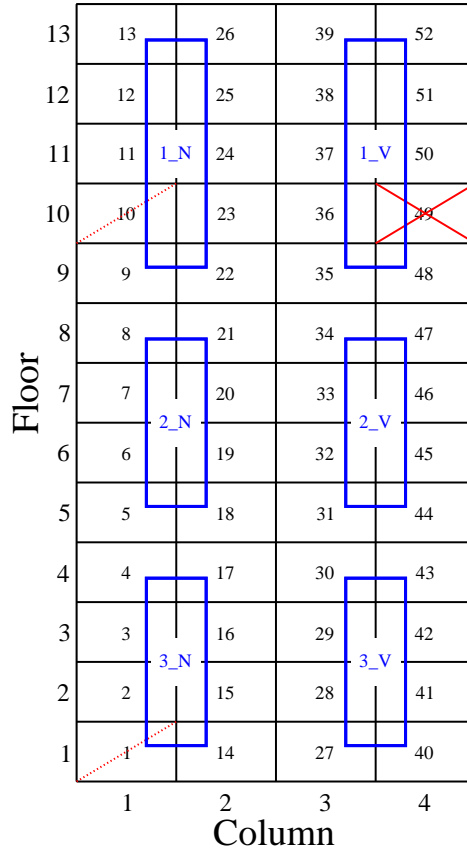
Figure A.1: Mapping of the bolometers in the CUORE-0 tower. Channels 1 & 10 are missing their Si Joule heaters and Channel 49 is not functional at all. The blue labels represent the PENCu cables and read out either 8 or 10 channels. The Si heaters are wired along an entire column.

Table A.2: Values for various physical parameters used in this analysis.

Parameter	Value	Reference
Isotopic Abundance of ^{130}Te	34.167%	[115]
Standard Atomic Weight of Tellurium	127.61 u	[115]
Standard Atomic Weight of Oxygen	16.00 u	[116]
Molar Mass of TeO$_2$	159.61 g/mol	
Q-value of ^{130}Te	2527.518$\pm$0.013 keV	[117]

Table A.3: Energy estimator breakdown for the CUORE-0 data. The data included here are for the full statistics.

	Number of Channel,Dataset pairs	Exposure	Fractional Exposure
Energy	582	20.57	58.4%
EnergyDecorrOF	133	4.32	12.3%
EnergyNewOFWoH	213	7.48	21.2%
EnergyDecorrOFWoH	80	2.84	8.1%
Total	1008	35.21	100.0%

Table A.4: Summary of 90% Bayesian limits on the $0\nu\beta\beta$ decay rate of ^{130}Te, including various experiments.

Experiment	$\Gamma_{0\nu\beta\beta}$ (10^{24} yr^{-1})	$T^{0\nu}_{1/2}$ (10^{24} yr)
CUORE-0 (Stat Only)	< 0.25	> 2.8
CUORE-0	< 0.25	> 2.8
CUORE-0 + TTT	< 0.22	> 3.2
CUORE-0 + TTT + Cuoricino	< 0.15	> 4.5
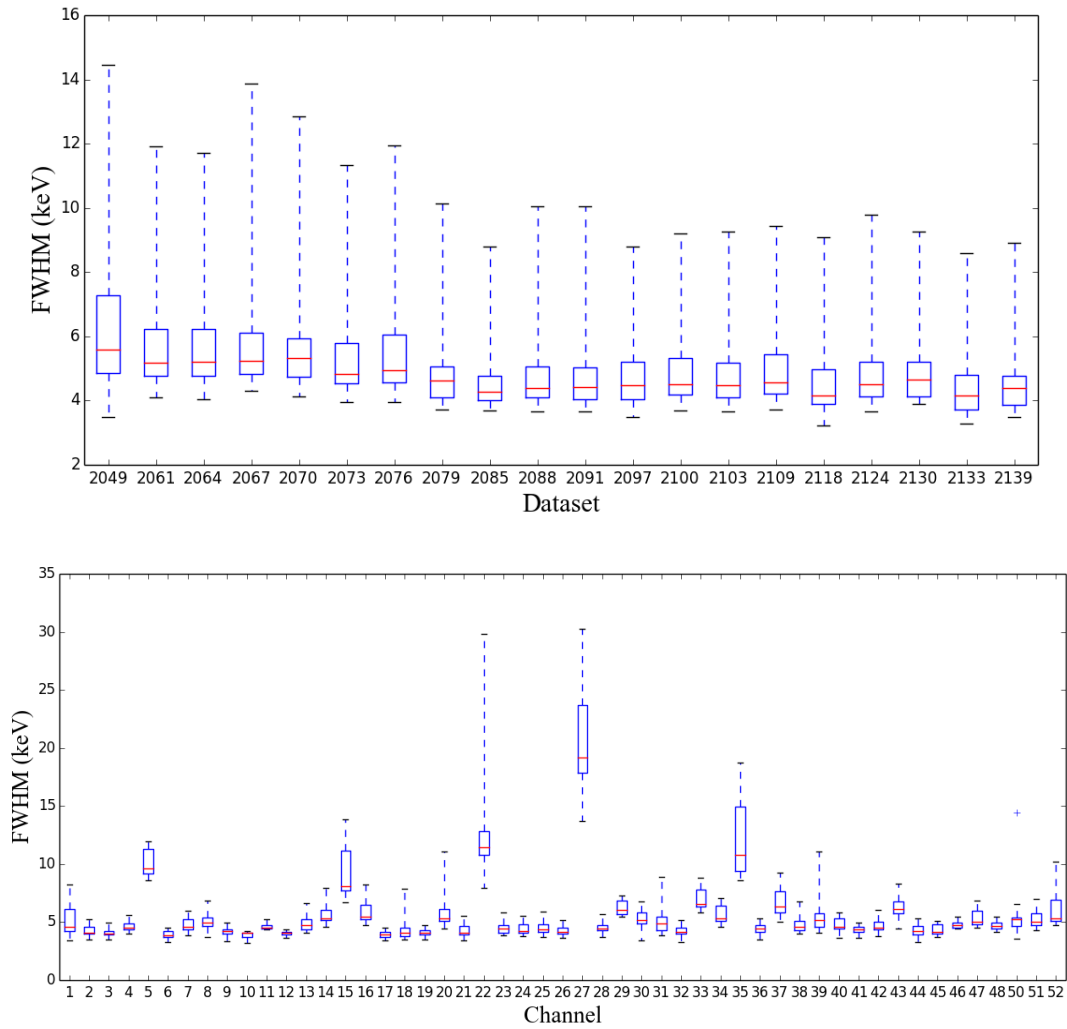
Figure A.2: A candlestick plot of FWHM at 2615 keV split by channel (*top*) and dataset (*bottom*). The box of the plot stretches from the 25th percentile to the 75th. The whiskers stretch from the maximum to the minimum values. Channels 22, 35 and 27 are excluded from the top plot.
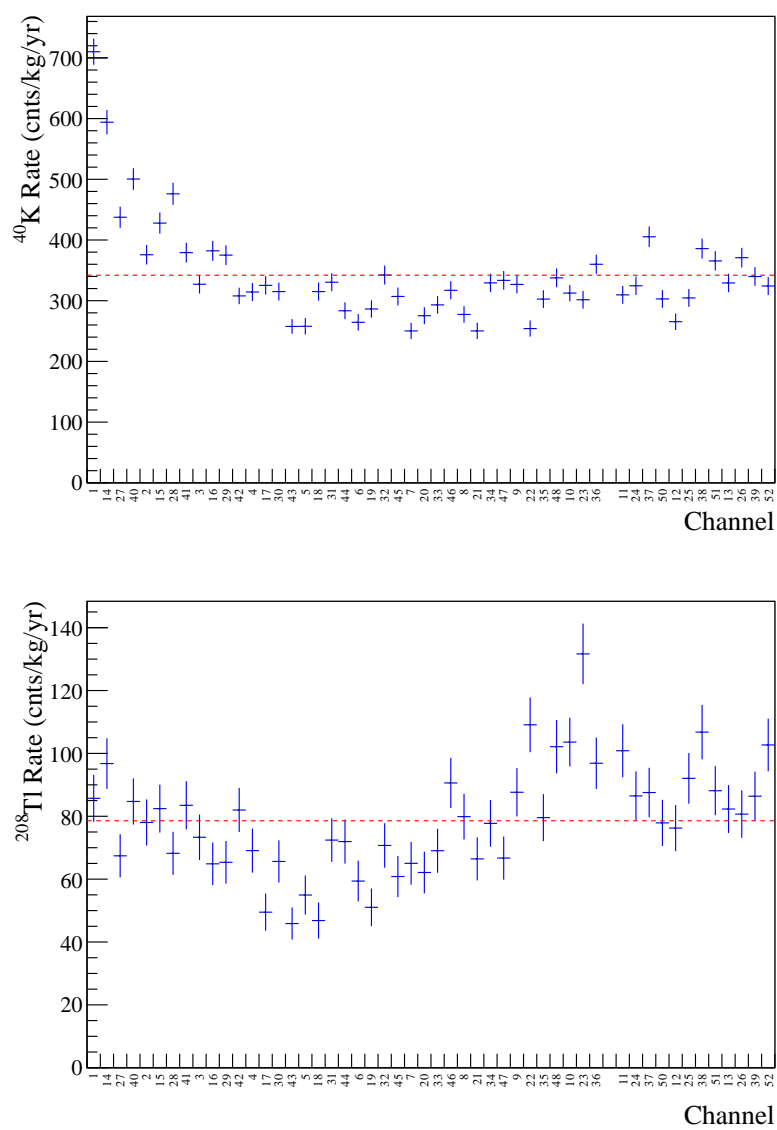
Figure A.3: Background ^{40}K rate (*top*) and ^{208}Tl rate (*bottom*) by channel over the entire CUORE-0 exposure. The channels are ordered by floor from bottom on the left to top on the right.
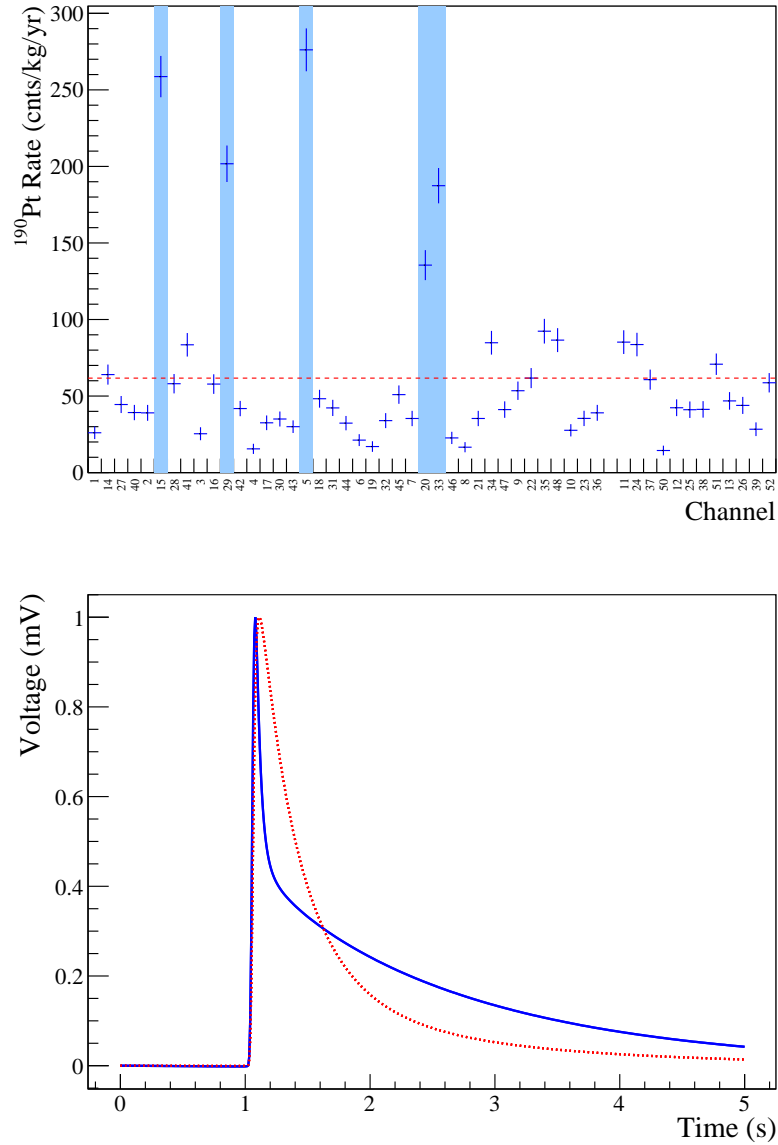
Figure A.4: *Left:* Background ^{190}Pt rate by channel over the entire CUORE-0 exposure. The channels are ordered by floor from bottom on the left to top on the right. The five channels with the highest Pt rates are highlighted, these channels show distinctly different pulse shapes from the other channels. *Right:* The average pulses from channel 5 (*blue*) and channel 6 (*red*) from dataset 2073. A similar sharp spike and slow roll off shape can be seen on channels 15, 20, 29 and 33 as well.
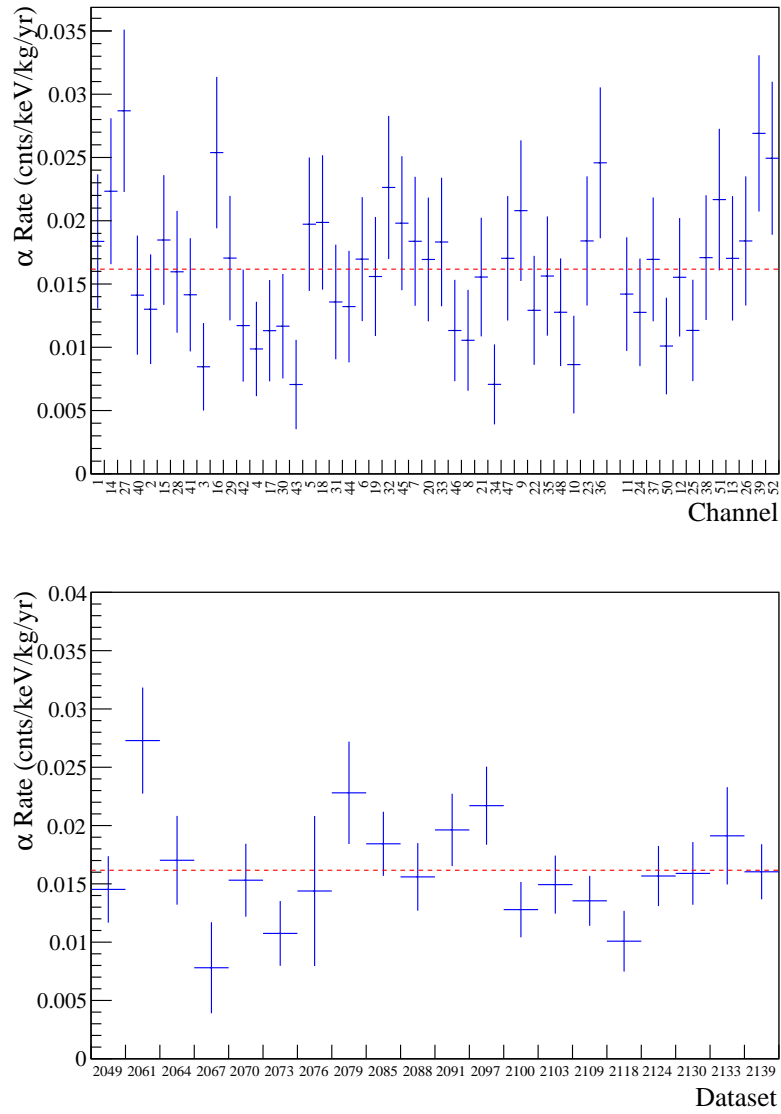
Figure A.5: *Left:* Background α rate (as measured between 2700–3200 $\cup$ 3400–3900 keV) by channel over the entire CUORE-0 exposure. The channels are ordered by floor. *Right:* Background α rate by dataset of the entire CUORE-0 exposure.
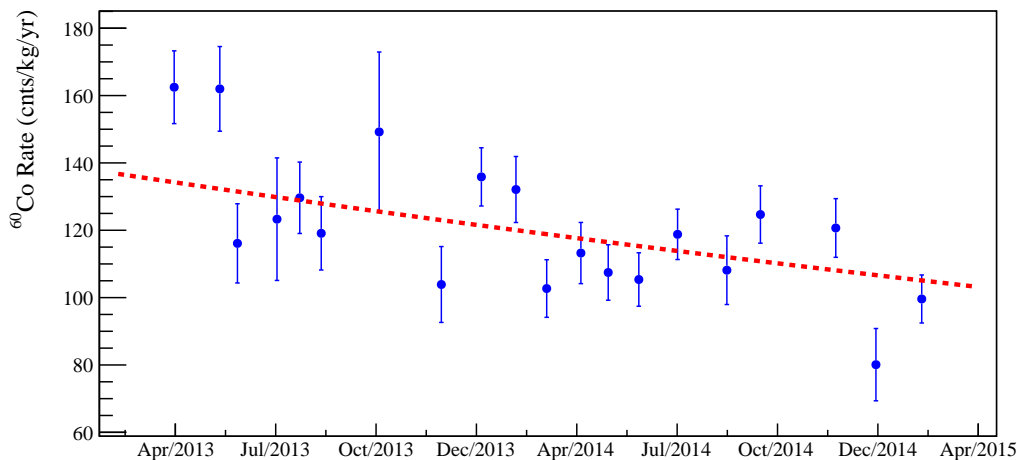
Figure A.6: The sum of the event rates in the 1173 and 1332 keV ^{60}Co peaks as a function of time. The dashed line is the best-fit rate assuming the ^{60}Co lifetime.
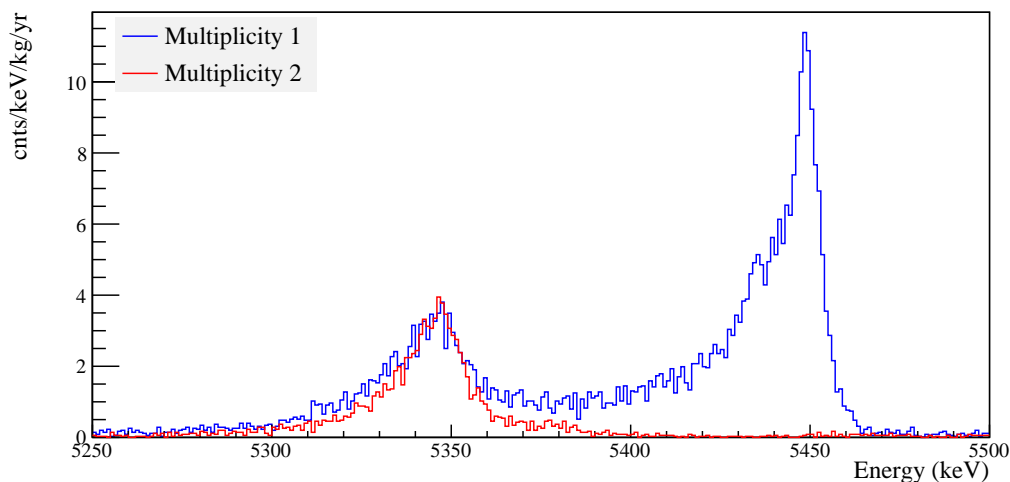
Figure A.7: The ^{210}Po peak for the multiplicity 1 (*blue*) and the multiplicity 2 (*red*) spectrum. The peak to the right is the Q-Peak is is composed of decays where both the α-particle and the nuclear recoil are contained in one crystal. The left peak is the α-Peak where only the α-particle is detected. The multiplicity 2 spectrum contains only the α-Peak. Both peaks are shifted up by $\sim$50 keV from their expected position due to an α quenching.
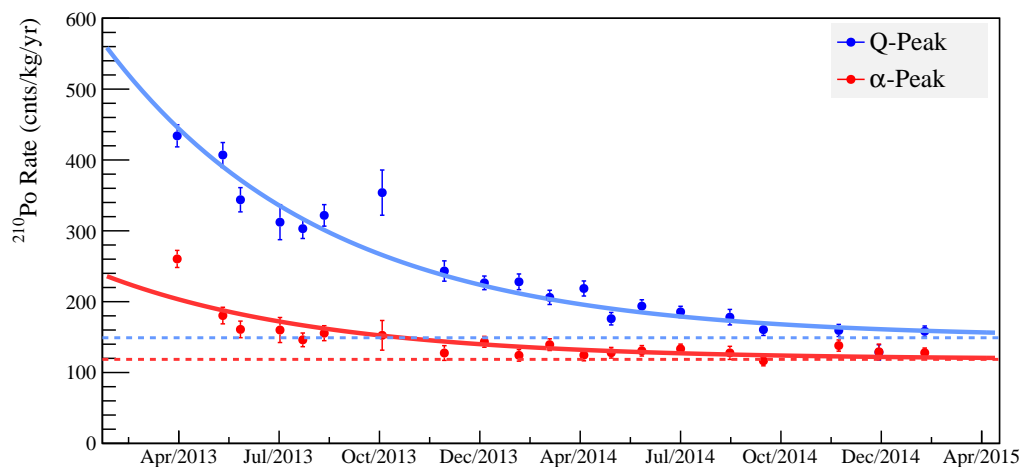
Figure A.8: The multiplicity 1 ^{210}Po rate as a function of time. The Q-Peak (*blue*) is defined as the range 5.4–5.5 MeV, while the α-Peak (*red*) is the range 5.3–5.4 MeV. The Q-Peak is composed of both bulk ^{210}Po, which originates from Po contamination in the TeO$_2$ powder before the crystal growth, decays away with a 138 day halflife, and a surface contamination which is likely in equilibrium with a ^{210}Pb surface contamination from Rn exposure possibly during installation which decays with a 22 yr halflife. The α-peak is primarily due to the surface contamination.
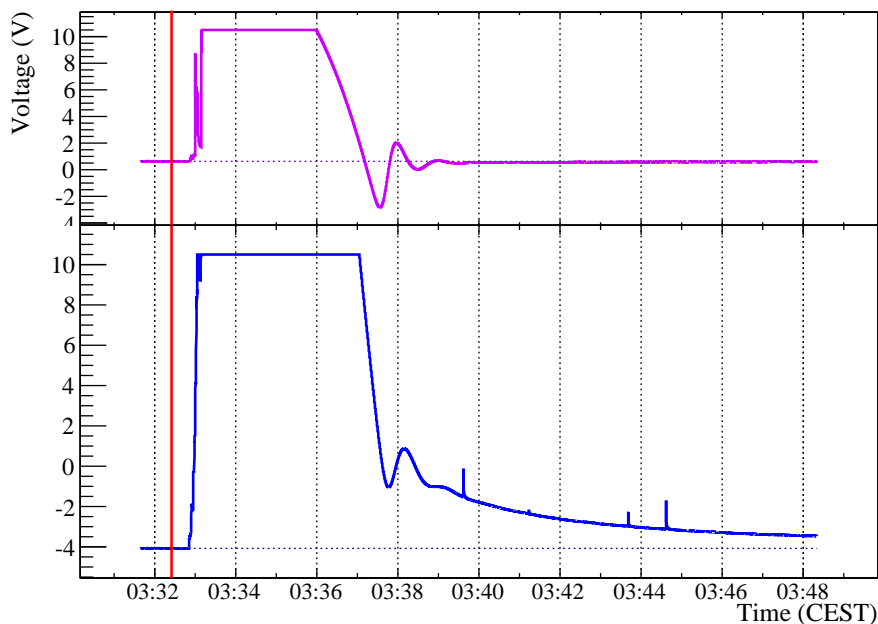
Figure A.9: On July 21, 2013 at 3:32:24 CEST a magnitude 5.4 earthquake struck near Numana, Italy. This earthquake was seen in the CUORE-0 detector as a sudden saturation of all channels. Here, we show channel 18 (*blue*) saturating and slowly returning to the baseline. We also show channel 55 (*violet*) which is the thermometer used for stabilization. Channel 18 begins returning to baseline rather quickly, then experiences a blip, then returns to baseline slowly. This is driven by the behavior of the stabilization thermometer, which initially overshoots its set point (the dashed line), over-heats the tower and eventually settles back to its stabilization value — before the bolometers have cooled back to equilibrium.

Appendix B

Generalized Amplitude Evaluation

This section describes the most generalized amplitude evaluation. It is a generalization of the Optimal Filter commonly used in Cuoricino [68] as well as the algorithm found in [92, Appendix C].

B.1   Derivation of the Generalized Optimum Filter

We begin by describing a collection of waveforms T seconds long, collected from N_C channels. We assume that each channel has its own signal response function, $s_i(t)$, which is known or measured. Each channel also has its own *additive* noise signal $N(t)$, which can have many origins (e.g. electronic noise, microphonics or thermal instabilities). We assume that over time the noise averages to zero, $\langle N(t) \rangle = 0$. More precisely, we assume that the noise and signal are independent, which to very good approximation is true since they are of different origins. Further, we assume that the noise is stationary, and can be described by a time independent covariance matrix.

We consider the situation where there is a signal on channel 0 with no signal on any other channel. This is a situation that is most often of physical interest. We can write the voltage on channel 0 as a sum of a signal response function $s_0(t)$ and a noise component $n_0(t)$, while the other channels contain only noise components:

$$
\begin{aligned}
y_0(t) &= a_0 s_0(t) + N_0(t) \\
y_1(t) &= N_1(t) \\
&\vdots \\
y_i(t) &= N_i(t)
\end{aligned}
\tag{B.1}
$$

The amplitude of the signal response, a_0, is the quantity in which we are interested in measuring.

Since we are sampling the waveforms in evenly spaced samples, we should consider the discrete version of the problem. Each waveform is composed of $M_S = T f_S$ samples, where

f_s is the sampling frequency. For simplicity of notation, we will use the indices i, j, k to indicate channels and p, q to indicate frequencies.

$$\begin{array}{rcl}
y_0(\omega_p) & = & a_0 s_0(\omega_p) + N_0(\omega_p) \\
y_1(\omega_p) & = & N_1(\omega_p) \\
& \vdots & \\
y_i(\omega_p) & = & N_i(\omega_p)
\end{array} \tag{B.2}$$

Dividing each channel by its own signal response, we get the 'normalized' waveforms:

$$\begin{array}{rcl}
c_0(\omega_p) & = & a_0 + n_0(\omega_p) \\
c_1(\omega_p) & = & n_1(\omega_p) \\
& \vdots & \\
c_i(\omega_p) & = & n_i(\omega_p)
\end{array} \tag{B.3}$$

Where we have defined $c_i(\omega_p) \equiv y_i(\omega_p)/s_i(\omega_p)$ and $n_i(\omega_p) \equiv N_i(\omega_p)/s_i(\omega_p)$.

In this form, we can rewrite channel 0 as

$$a_0 = c_0(\omega_p) - n_0(\omega_p) \tag{B.4}$$

We see that for each frequency ω_p, $c_0(\omega_p)$ provides an estimate of a_0 with noise $n_0(\omega_p)$. The goal is to build an estimator of a_0, (call it $\bar{a}_0$), that minimizes $\mathrm{Var}(\bar{a}_0)$, subject to the constraint that over many pulses $\langle \bar{a}_0 \rangle = a_0$. In [92], this is done on a channel by channel basis, however, here we include all channels in the estimator to account for correlations between channels.

We define an estimator for channel i, by a set of weights $w_{i,jp}^*$:

$$\bar{a}_i = \sum_{j,p} w_{i,jp}^* c_j(\omega_p) \tag{B.5}$$

To be explicit, w^*i, jp is a set of weights that can be different for each channel, i, that when summed over the waveforms on all channels, j, and over all frequencies, ω_p, provides an unbiased estimate of the amplitude for that channel, a_i, with minimum variance.

We write the constraint mathematically as

$$\begin{array}{rcl}
\langle \bar{a}_0 \rangle & = & a_0 \\
\sum_{j,p} w_{0,jp}^* \langle c_j(\omega_p) \rangle & = & a_0 \\
\sum_{j,p} w_{0,jp}^* a_j & = & a_0 \\
\sum_{j,p} w_{0,0p}^* a_0 & = & a_0
\end{array} \tag{B.6}$$

where the average is taken over many pulses of the same "true" amplitude. Between the second and third line, we took advantage of the fact that that the noise averages to zero:

$$\langle c_j(\omega_p) \rangle = a_j + \langle n_j(\omega_p) \rangle = a_j$$

And in the next step, we took advantage of our assumption that all the other channels had no signal ($a_j = 0$ for $j \neq 0$). We rewrite this constraint as a normalization on the weights.

$$\sum_p w^*_{0,0p} = 1$$

Keeping in mind that all variables are — in general — complex, the variance of $\bar{a}_i$ is given by

$$
\begin{aligned}
\text{Var}(\bar{a}_0) &= \langle \bar{a}^*_0 \bar{a}_0 \rangle - |\langle \bar{a}_0 \rangle|^2 \\
&= \left\langle \left(\sum_{j,p} w^*_{0,jp} c_j(\omega_p) \right)^* \left(\sum_{k,q} w^*_{0,kq} c_l(\omega_q) \right) \right\rangle - a^2_0 \\
&= \left\langle \left(\sum_{j,p} w^*_{0,jp} \left(a_j + n_j(\omega_p) \right) \right)^* \left(\sum_{k,q} w^*_{0,kq} \left(a_l + n_k(\omega_q) \right) \right) \right\rangle - a^2_0 \\
&= \left\langle \left(\sum_{j,p} w_{0,jp} n_j(\omega_p) \right)^* \left(\sum_{k,q} w_{0,kq} n_k(\omega_q) \right) \right\rangle \\
&= \sum_{j,p,k,q} w^*_{0,kq} w_{0,jp} \left\langle n_k(\omega_q) n^*_j(\omega_p) \right\rangle
\end{aligned}
$$

Where we have taken advantage of Eqn. B.6 and the fact that $\langle n_j(\omega_p) \rangle = 0$.

In matrix notation we could rewrite this for channel i as

$$\text{Var}(\bar{a}_i) = \boldsymbol{W}^\dagger_i \boldsymbol{V} \boldsymbol{W}_i \tag{B.7}$$

Where $\boldsymbol{W}_i$ is a column vector composed of the $N_C \cdot M_S$ weights for channel i, and $\boldsymbol{V}$ is the $(N_C \cdot M_S) \times (N_C \cdot M_S)$ covariance matrix between every channel and frequency,

$$
\begin{aligned}
V_{i,j}(\omega_p, \omega_q) &\equiv \langle n_i(\omega_p) n^*_j(\omega_q) \rangle \tag{B.8} \\
\boldsymbol{V} &\equiv \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle \tag{B.9}
\end{aligned}
$$

It is useful to point out that $\boldsymbol{V}$ is both hermitian and positive definite (therefore $\boldsymbol{z}^\dagger \boldsymbol{V} \boldsymbol{z}$, for every z[1]).

Now, our challenge is to solve for the set of weights, $\boldsymbol{W}_i$, that minimizes $\text{Var}(\bar{a})$. We use the method of Lagrange multipliers. The equation to optimize is given in matrix form by

$$\boldsymbol{W}^\dagger_i \boldsymbol{V} \boldsymbol{W}_i - \boldsymbol{\lambda}^T_i (\boldsymbol{U}_i \boldsymbol{W}_i - \boldsymbol{\delta}_i) \tag{B.10}$$

---

[1] In general, any covariance matrix must be Hermitian and positive *semi*-definite. The more strict condition here that it be positive definite stems from our assumption that all channels are linearly independent. That is, two channels can have correlated noise, but they cannot be a strict multiple of each other.

Where $\boldsymbol{\lambda}_i$ is a column vector of N_C real Lagrange multipliers, $\boldsymbol{\delta}_i$ is a column vector of the Kroniker δ_{ij}. $\boldsymbol{U}_i$ is an $N_C \times (N_C \cdot M_S)$ matrix of ones that is defined as follows

$$\boldsymbol{U}_0\boldsymbol{W}_0 = \begin{pmatrix} \sum_p w_{0,0p} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and our constraint equation is written explicitly as

$$\begin{pmatrix} \sum_p w_{0,0p} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Notice that the subscript i on $\boldsymbol{U}_i$ denotes that a separate column vector should exist for each channel i.

Differentiating with respect to W_i, and λ_i we get the system of equations:

$$\begin{aligned} \boldsymbol{W}_i^{\dagger}\boldsymbol{V} - \boldsymbol{\lambda}_i^T\boldsymbol{U}_i &= 0 \\ \boldsymbol{U}_i\boldsymbol{W}_i - \boldsymbol{\delta}_i &= 0 \end{aligned}$$

And solving for $\boldsymbol{W}_i$, we get:

$$\boldsymbol{W}_i = \boldsymbol{V}^{-1}\boldsymbol{U}_i^T\left(\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\right)^{-1}\boldsymbol{\delta}_i \tag{B.11}$$

This equation is hard to parse at first sight, but is in fact a glorified solution to a least squares problem. We can calculate the expected variance of the final result as

$$\begin{aligned} \mathrm{Var}(\bar{a}_i) &= \boldsymbol{W}_i^{\dagger}\boldsymbol{V}\boldsymbol{W}_i \\ &= \boldsymbol{\delta}_i^T\left(\left(\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\right)^{-1}\right)^{\dagger}\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\left(\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\right)^{-1}\boldsymbol{\delta}_i \\ &= \boldsymbol{\delta}_i^T\left(\left(\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\right)^{-1}\right)^{\dagger}\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\left(\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\right)^{-1}\boldsymbol{\delta}_i \\ &= \boldsymbol{\delta}_i^T\left(\left(\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\right)^{\dagger}\right)^{-1}\boldsymbol{\delta}_i \\ &= \boldsymbol{\delta}_i^T\left(\boldsymbol{U}_i\boldsymbol{V}^{-1}\boldsymbol{U}_i^T\right)^{-1}\boldsymbol{\delta}_i \end{aligned}$$

Inverting the Covariance Matrix $\boldsymbol{V}$

The general solution requires finding the inverse of the covariance matrix $\boldsymbol{V}$. However, in CUORE-0 this matrix consisted of $(626 \cdot 51)^2$ elements, and in CUORE will consist of
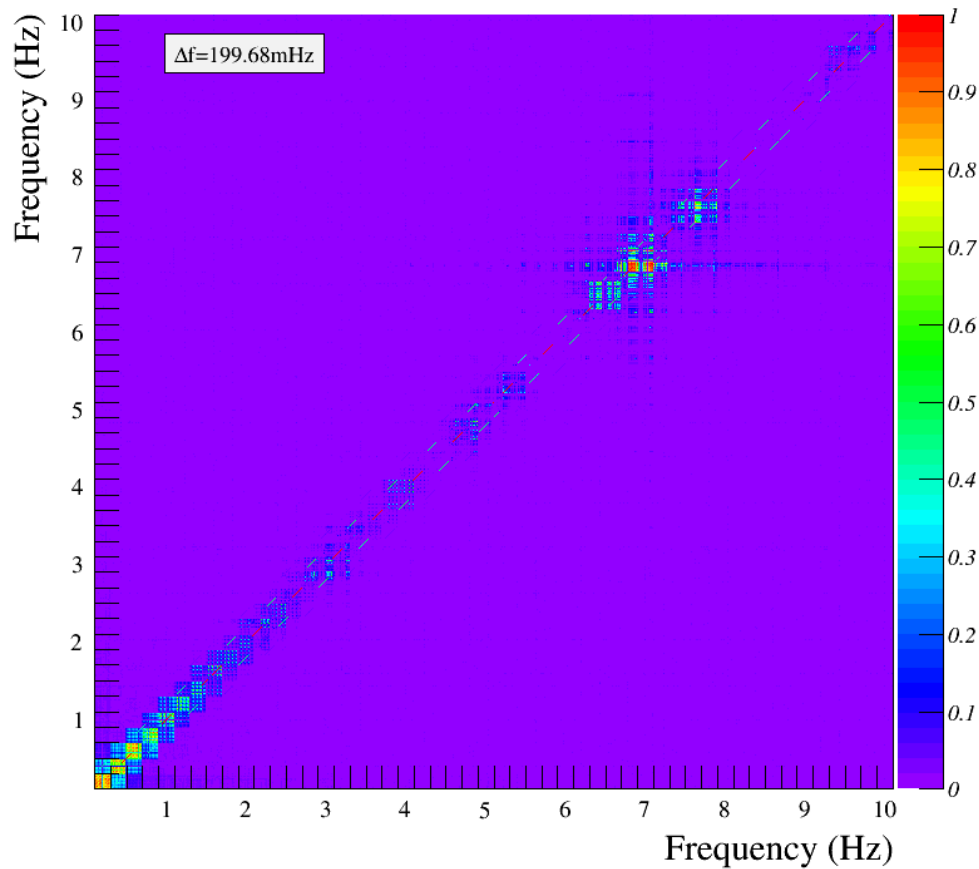
Figure B.1: The correlation matrix $\rho_{ij} = |V_{ij}|/\sqrt{V_{ii}V_{jj}}$ over the range 0-10.2 Hz for dataset 2073. Each frequency bin is $\sim$.2 Hz wide, and each square frequency box in the figure contains a 51×51 sub-matrix of the channel-channel correlations at that frequency pair. The majority of the correlation exists at low frequency, with a noise peak at $\sim$ 7 Hz. Most of the cross frequency correlations are zero, making the covariance matrix extremely sparse.

$\sim (988 \cdot 1000)^2$ elements. Computationally, this problem is essentially impossible to solve and if not done properly in danger of becoming numerically unstable and yielding spurious results.

As can be seen in Fig. B.1, the correlation matrix is actually quite sparse. There are obviously very high correlations between channels, but the correlation between frequencies is small. In the following sections, we will consider special cases, that reduce the size of the problem by considering various limiting simplifications of the covariance matrix, V.

B.2   Special Cases

In Sec. B.1 we demonstrated the most general set of weights to solve this problem. In this section we will restrict to specific subcases and rederive both the filter described in [92, Appendix C] and the Optimal Filter used in Cuoricino.

Zero Cross-Channel Correlation

Here we study the specific case in which the channels are not correlated, but within any single channel the frequencies are cross correlated.

The majority of cross-frequency covariance will average to zero. This can be seen from a simple time averaging argument:

$$
\begin{aligned}
\langle n_i^*(\omega_p) n_i(\omega_q) \rangle \quad &\propto \quad \langle e^{-i\omega_p t} e^{i\omega_q t} \rangle \\
&\propto \quad \langle e^{i(\omega_p - \omega_q)t} \rangle \\
&\propto \quad \delta(\omega_p - \omega_q)
\end{aligned}
$$

However, any noise sources that produce harmonics can produce correlations across multiple frequencies. Another possibility, though likely less significant, is that non-linearities in the gain circuit lead to intermodulation of two uncorrelated noise sources.

In the limit of no cross channel correlations, we can block diagonalize V into N_C submatrices: one $(M_S \times M_S)$ submatrix for each channel. Since our constraint equation can also be separated by channel, the entire problem can be split into N_C *separate* problems that can be solved individually. The minimization condition for each channel now becomes

$$
\boldsymbol{w}_i^\dagger \boldsymbol{V}_i \boldsymbol{w}_i - \lambda_i (\boldsymbol{u}^T \boldsymbol{w}_i - 1) \tag{B.12}
$$

Were $\boldsymbol{w}_i$ is a column vector of M_S weights for channel i, $\boldsymbol{V}_i$ is the submatrix of $\boldsymbol{V}$ corresponding only to channel i, and $\boldsymbol{u}$ is now a $M_S \times 1$ column vector of ones.

Minimizing this equation, yields the result in [92, Appendix C]:

$$
\boldsymbol{w}_i = \frac{\boldsymbol{V}_i^{-1} \boldsymbol{u}}{\boldsymbol{u}^T \boldsymbol{V}_i^{-1} \boldsymbol{u}} \tag{B.13}
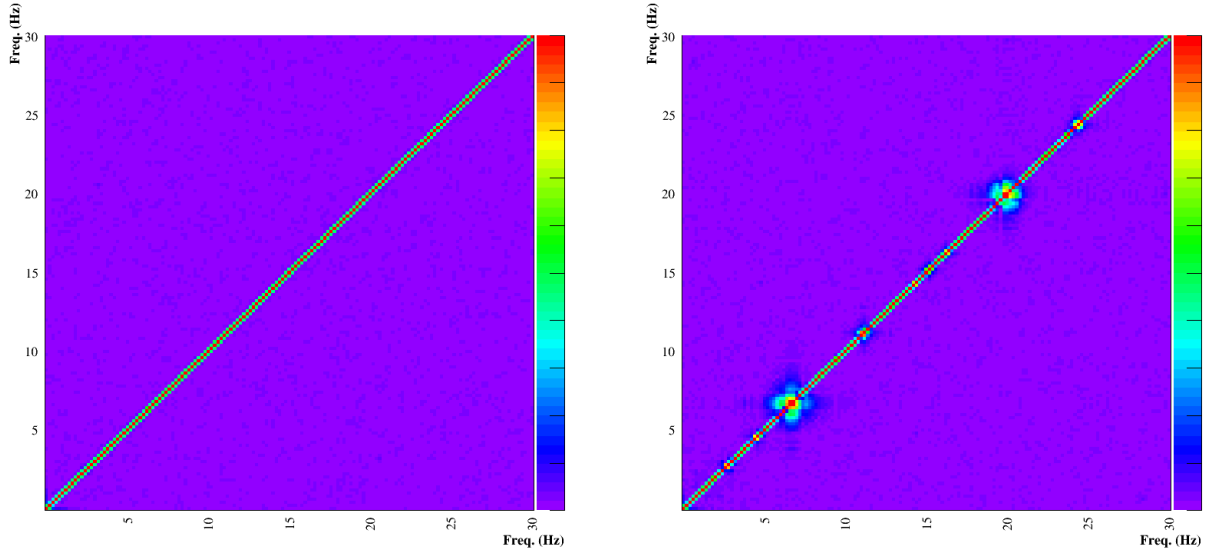$$

Figure B.2: Frequency-frequency covariance matrices from ds2073 for channel 18 (*left*) and 43 (*right*). The majority of the frequency-frequency covariance matrices showed similar characteristics to the one on the left, with little or no cross frequency noise, however, a few channels demonstrated a few harmonic peaks like the matrix on the right.

with the variance of $\bar{a}_i$ given by

$$\mathrm{Var}(\bar{a}_i) = \frac{1}{\boldsymbol{u}^T \boldsymbol{V}_i^{-1} \boldsymbol{u}} \tag{B.14}$$

Zero Cross-Frequency Correlation, Zero Cross-Channel Correlation

We can further simplify the zero cross-channel correlation by adding the condition that there is also no cross-frequency correlations. This is the limit where every channel and every frequency are independent of every other. In this limit our covariance matrix is really just a diagonal vector of variance, expressed as

$$V_{ij}(\omega_p, \omega_q) = \langle n_i(\omega_p) n_j^*(\omega_q) \rangle \propto \delta_{ij}\delta_{pq}$$

This is the limit that leads to the Optimum Filter that was used in Cuoricino and CUORE-0.

We start with the result of Sec. B.2. In this limit the covariance matrix for the ith channel, $\boldsymbol{V}_i$, is diagonal, with the pth frequency element given by $|n_i(\omega_p)|^2 = |N_i(\omega_p)/s_i(\omega_p)|^2$. Plugging this into Eqn. B.14, we obtain the familiar equation for the Optimum Filter for channel i:

$$w_{ip} = \frac{|s_i(\omega_p)|^2 \,/\, \langle |N_i(\omega_p)|^2 \rangle}{\sum_p |s_i(\omega_p)|^2 \,/\, \langle |N_i(\omega_p)|^2 \rangle} \tag{B.15}$$

Non-Zero Cross-Channel Correlation

Here we study the alternate case where the cross-frequency correlations are negligible but the cross-channel correlations are significant. Unfortunately, in this limit we can not factor the constraint equation and so we are forced to revert back to Eqn. B.11.

However, since the cross frequency correlations are negligible, we block diagonalize the covariance matrix, V. Now, instead of inverting one $(N_C \cdot M_S) \times (N_C \cdot M_S)$ matrix, we need only invert M_S independent $N_C \times N_C$ matrices. Since this is the situation we find in CUORE-0, we go into a bit more details.

Since the covariance matrix can be block diagonalized into M_S, $N_C \times N_C$ covariance matrices, we rewrite our equations in terms of these smaller matrices, $V(\omega_p)$. They are defined symbolically from the block diagonalization,

$$
V = \begin{pmatrix}
V(\omega_0) & 0 & 0 & \dots & 0 \\
0 & V(\omega_1) & 0 & \dots & 0 \\
0 & 0 & V(\omega_2) & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \dots & V(\omega_{M_S})
\end{pmatrix}
$$

In practice, in CUORE-0, we never explicitly calculate either $c_k(\omega)$ or V. Instead, we work in terms of the *non*-normalized variables $y_k(\omega)$ and C. Where the un-normalized covariance matrix, C, is given by

$$
C_{ij}(\omega_p, \omega_q) \equiv \langle N_i(\omega_p) N_j^*(\omega_q) \rangle \tag{B.16}
$$

as opposed to V which was given by

$$
V_{ij}(\omega_p, \omega_q) = \left\langle \frac{N_i(\omega_p)}{s_i(\omega_p)} \frac{N_j^*(\omega_q)}{s_j^*(\omega_q)} \right\rangle
$$

We can now rewrite our solutions, in terms of these variables by explicitly pulling out the expected signal response terms

$$
V_{ij}(\omega_p, \omega_q) = s_i^{-1}(\omega_p) C_{ij}(\omega_p, \omega_q) (s_j^{-1}(\omega_q))^*
$$

In the case where V has negligible cross-frequency correlations, we can write

$$
\begin{aligned}
V_{ij}(\omega_p) &= s_i^{-1}(\omega_p) C_{ij}(\omega_p)(s_j^{-1}(\omega_p))^* \\
V_{ij}^{-1}(\omega_p) &= s_i(\omega_p) C_{ij}^{-1}(\omega_p) s_j^*(\omega_p)
\end{aligned}
$$

Where $C_{ij}(\omega_p) \equiv C_{ij}(\omega_p, \omega_p)$.

Our general solution from Eqn. B.11, can now be written as

$$
\bar{a}_i = A \sum_{j,p} s_i^*(\omega_p) C_{ij}^{-1}(\omega_p) y_j(\omega_p) \tag{B.17}
$$

where A is a normalization constant, given by

$$A = \left(\sum_p s_i^*(\omega_p) C_{ii}^{-1}(\omega_p) s_i(\omega_p) \right)^{-1} \tag{B.18}$$

The variance of $\bar{a}_i$ is given by

$$\mathrm{Var}(\bar{a}_i) = \left(\sum_p s_i^*(\omega_p) C_{ii}^{-1} s_i(\omega_p) \right)^{-1} = A \tag{B.19}$$

And this, of course, is our "Decorrelating Optimum Filter" used in CUORE-0.

Decorrelation Filter

Here we derive the "decorrelation" filter from [103]. This filter is similar to the Decorrelating Optimum Filter, in that it subtracts noise from neighboring channels, however, it does not make use of an expected signal response, $s_i(t)$. Instead, it seeks to preserve the spectral shape of the channel being filtered by setting all the weights for a channel on itself to unity (i.e. $w_{i,ip} = 1$ for all i).

Mathematically, this can be written by specifying that $s_i(\omega_p) = 1$ for all i and ω_p. Our constraint matrix U_i is modified to a $M_S \times (N_C \cdot M_S)$ matrix such that

$$\boldsymbol{U}_0 \boldsymbol{W}_0 = \begin{pmatrix} w_{0,00} \\ w_{0,01} \\ \vdots \\ w_{0,0M_S} \end{pmatrix} \tag{B.20}$$

so that our constraint takes the form

$$\begin{pmatrix} w_{0,00} \\ w_{0,01} \\ \vdots \\ w_{0,0M_S} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{B.21}$$

In this case, Eqn. B.11 can be written as

$$w_{i,jp} = \frac{C_{ij}^{-1}(\omega_p)}{C_{ii}^{-1}(\omega_p)} \tag{B.22}$$

Notice that we are using the unscaled covariance matrix $C_{ij}(\omega_p)$, rather than $V_{ij}(\omega_p)$. In fact, in the limit that $s_i(\omega_p) = 1$, they are equal.

Our filtered signals take the form

$$y_i^f(\omega_p) = y_i(\omega_p) + \sum_{j \neq i} \frac{C_{ij}^{-1}(\omega_p)}{C_{ii}^{-1}(\omega_p)} y_j(\omega_p) \tag{B.23}$$

which is the "Decorrelating Filter" found in [103].

B.3 Waveform Filtering in CUORE-0

One of our implicit assumptions, seen in Eqn. B.1, was that the measured signals were actually composed of noise terms and a single "signal" pulse that started at $t = 0$. In practice, it's actually difficult to determine the start time of the signal, since the position within the acquired window depends on when the signal had risen enough to pass the threshold and trigger the acquisition. This gives the start time of the signal a dependence on both the amplitude and noise.

If we include this uncertainty in the start time of the signal, our equation for channel 0 might look more like this

$$y_0(t) = a_0 s_0(t - \Delta t) + N_0(t)$$

and in frequency space, this adds a frequency dependent phase to the amplitude components:

$$y_0(\omega_p) = a_0 e^{-i\omega_p \Delta t} s_0(\omega_p) + N_0(\omega_p)$$
$$y_0(\omega_p) = a_0(\omega_p) s_0(\omega_p) + N_0(\omega_p)$$

The problem is, of course, since we assumed a_0 to be real our summation of estimates of estimates of a_0, will now include complex phasors, which will add noise and worsen our resolution.

There are several approaches to solving this issue. The first is that we could include this uncertainty in our minimization and solve for Δt. However, since the Δt appears in the exponential, this would make the minimization non-linear forcing a numerical approach and a non-analytic solution. A second approach would be to shift the template pulse to align with each signal pulse we wish to filter. Computationally, this is feasible, but would require deciding on an alignment parameter, such as a χ^2 statistic to determine when the template is "most" aligned with the signal pulse. Computationally, this adds a large amount of complexity, due to repeatedly evaluating this alignment parameter.

In CUORE-0, we approach this problem by solving for the M_S amplitude estimates, $\bar{a}_i(\omega_p)$, without summing over frequency. In CUORE-0, Eqn. B.17 is actually calculated as

$$\bar{a}_i(\omega_p) = A \sum_j s_i^*(\omega_p) C_{ij}^{-1}(\omega_p) y_j(\omega_p) \tag{B.24}$$

The $\bar{a}_i(\omega_p)$ are then inverse Fourier transformed back into time domain to yield a filtered amplitude waveform, $\bar{a}_i(t)$.

The filtered waveform, $\bar{a}_i(t)$, has a noise power spectrum given by

$$\text{NPS}(\omega_p) = A^2 \left(s_i^*(\omega_p) C_{ii}^{-1}(\omega_p) s_i(\omega_p) \right) \tag{B.25}$$

At this point, the problem becomes one of locating the peak of the filtered waveform and evaluating its height. This is the equivalent problem to aligning the template pulse with the signal pulse or evaluating Δt, but computationally much simpler. In CUORE-0, we do this interpolating a parabola through the maximum point in the waveform and its

immediate to neighbors. The amplitude and time of the pulse is given by measuring the maximum of this parabola. The resulting evaluated amplitude, $\bar{a}'_i$, still has a variance given by Eqn. B.19. Though, the amplitude evaluation process may add a relatively small amount of added noise due to the uncertainty in Δt. One possible channel to improve this in the future is trigonometric interpolation rather than parabolic fitting.
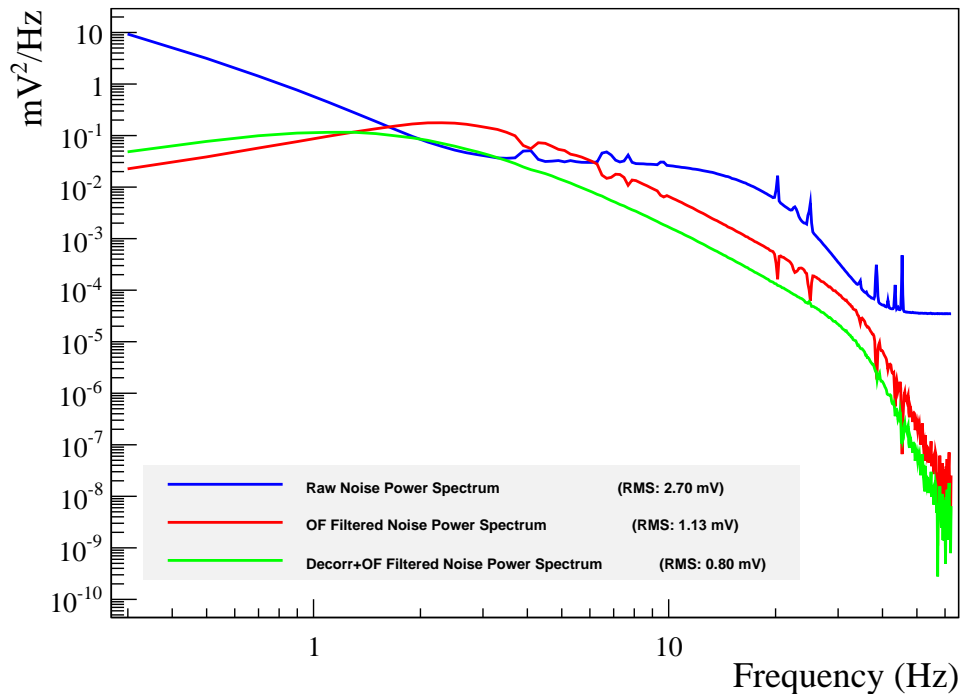

Figure B.3: Noise power spectra for Channel 13 in Dataset 2073. *Blue:* Raw noise power spectrum calculated from averaging the noise events. *Red:* The predicted noise power spectrum after having been passed through the new Optimum Filter given in § B.2. *Green:* The predicted noise power spectrum after having been passed through the Decorrelating Optimum Filter given in § B.2.

Appendix C

Bayesian Nuisance Parameters

.

In this Appendix, we discuss a more general approach to dealing with systematic uncertainties and derive the approach that we used in Chapter 6 as a limiting case of the general approach. Broadly speaking, our approach to including systematic uncertainties is to modify our profile likelihood function, $\mathcal{P}_{\text{stat}}(\hat{\Gamma}|\Gamma)$, to a profile likelihood which includes our systematic uncertainties, $\mathcal{P}_{\text{stat+syst}}(\hat{\Gamma}|\Gamma)$.
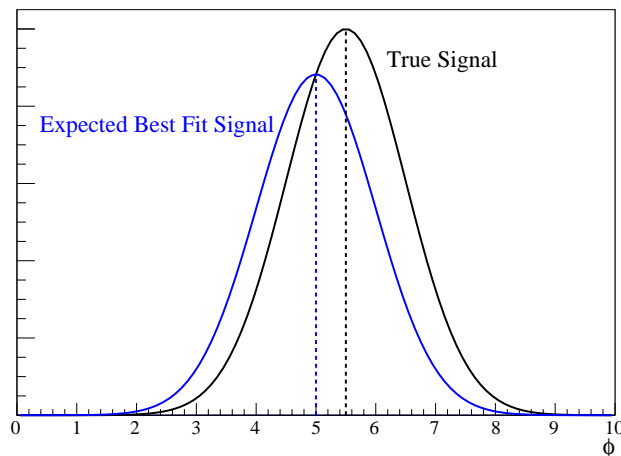
Figure C.1: Toy example of a relative bias caused by a nuisance parameter. Here, because we are fitting the peak at a wrong location, we expect a bias in our fit rate. The bias will also be proportional to the true decay rate, $\langle \hat{\Gamma} - \Gamma \rangle \propto \Gamma$.

We consider some nuisance parameter, ϕ, with a probability distribution, $\pi(\phi)$, which has some effect on our measured decay rate, $\hat{\Gamma}$, *given* a true value of the decay rate, Γ. We define the systematic uncertainty on our measurement to be the expected bias from our lack of knowledge of the true value of ϕ. For a fixed value of ϕ, we can write the expected bias

on our measurement $\hat{\Gamma}$ as

$$\delta(\phi, \Gamma) \equiv \langle \hat{\Gamma} - \Gamma \rangle(\phi, \Gamma) \tag{C.1}$$

We then create our profile likelihood which accounts for this systematic, $\mathcal{P}_{\text{stat+syst}}(\hat{\Gamma}|\Gamma)$, by apply the bias correction $\delta(\phi, \Gamma)$ and integrated over the probability distribution of ϕ, $\pi(\phi)$:.

$$\mathcal{P}_{\text{stat+syst}}(\hat{\Gamma}|\Gamma) = \int \mathcal{P}_{\text{stat}}(\hat{\Gamma} - \delta(\phi, \Gamma)|\Gamma)\pi(\phi)d\phi \tag{C.2}$$

This approach requires two things: the expected bias, $\delta(\phi, \Gamma)$, and the probability distribution of our nuisance parameter, $\pi(\phi)$. We outlined an approach to calculated the first term in Chapter 6. There we calculated $\delta(\phi, \Gamma)$, for a single value of ϕ over a range of Γ. The value of ϕ was the 1σ deviation, but generically, $\delta(\phi, \Gamma)$ can be scanned over both ϕ and Γ — this requires many toy Monte Carlos. The second term, $\pi(\phi)$, is a prior on the distribution of ϕ. For nuisance parameters like the efficiency, we can make a good argument that $\pi(\phi)$ should be gaussian distributed, but for other nuisance parameters (e.g. the choice of line shape) it is difficult to assign prior probabilities.

Combining Statistical and Systematic Errors in the Gaussian Limit

To connect this to the approach we used in Chapter 6, we look at the case were all probability distributions are gaussian, and the bias is linear in ϕ.

Assume that our bias takes the form:

$$\delta(\phi, \Gamma) = c_1\phi + c_2\phi\Gamma \tag{C.3}$$

Further, assume that ϕ is gaussian distributed, $\pi(\phi) \sim \text{Gauss}(0, \sigma_\phi)$. In this case, our combined profile likelihood can be calculated as

$$\mathcal{P}_{\text{stat+syst}}(\hat{\Gamma}|\Gamma) \propto \int \mathcal{P}_{\text{stat}}(\hat{\Gamma} - c_1\phi - c_2\phi\Gamma|\Gamma)e^{-\frac{1}{2}\phi^2/\sigma_\phi^2}d\phi \tag{C.4}$$

This can be numerically evaluated to solve for the total profile likelihood. However, in the case where our statistical profile likelihood is also gaussian distributed (i.e. $\hat{\Gamma} - \Gamma$ is distributed as $\text{Gauss}(0, \sigma_{\text{stat}})$), this can be written

$$P_{\text{stat+syst}}(\hat{\Gamma}|\Gamma) = \frac{1}{2\pi\sigma_{\text{stat}}\sigma_\phi} \int e^{-\frac{1}{2}(\hat{\Gamma} - \Gamma - (c_1 - c_2\Gamma)\phi)^2/\sigma_{\text{stat}}^2} e^{-\frac{1}{2}\phi^2/\sigma_\phi^2}d\phi$$

We solve this by making the integration substitution $x = (c_1 + c_2\Gamma)\phi$ and identifying the sigma for x as

$$\sigma_{\text{syst}} \equiv c_1\sigma_\phi + c_2\sigma_\phi\Gamma = \sigma_{\text{abs}} + \sigma_{\text{rel}}\Gamma \tag{C.5}$$

And solve the convolution to yield our new distribution

$$P_{\text{stat+syst}}(\hat{\Gamma}|\Gamma) = \frac{1}{\sqrt{2\pi(\sigma_{\text{stat}}^2 + \sigma_{\text{stat}}^2)}}e^{-\frac{1}{2}(\hat{\Gamma} - \Gamma)^2/(\sigma_{\text{stat}}^2 + \sigma_{\text{stat}}^2)} \tag{C.6}$$

Which is, of course, the expected result of the convolution of two gaussians. One thing to notice, is that this is not truly a gaussian, because σ_{stat} depends on Γ. However, in our case we can ignore this fact, since our statistical uncertainty is significantly larger than our systematic uncertainty. This function only deviates from gaussian far out in out in the wings.

Since the result of our calculation was just to add two gaussian uncertainties in quadrature, we can derive a faster way to do this. We can define the two parameters

$$\chi^2_{\text{stat}} \equiv -2\log\left[P_{\text{stat}}(\hat{\Gamma}|\Gamma)\right] \tag{C.7}$$

$$\chi^2_{\text{syst}} \equiv \frac{(\hat{\Gamma} - \Gamma)^2}{\sigma^2_{\text{syst}}} \tag{C.8}$$

Of course in the gaussian limit, the first parameter can be evaluated to

$$\chi^2_{\text{stat}} = \frac{(\hat{\Gamma} - \Gamma)^2}{\sigma^2_{\text{stat}}}$$

And we can simply achieve our convolution of statistical and systematic uncertainties using

$$\frac{1}{\chi^2_{\text{stat+syst}}} = \frac{1}{\chi^2_{\text{stat}}} + \frac{1}{\chi^2_{\text{syst}}} \tag{C.9}$$

Which is exactly the equation that we used in Chapter 6.

And of course, this process can be repeated over and over for each additional systematic uncertainty. Each one adding in quadrature

$$\sigma^2_{\text{total}} = \sigma^2_{\text{stat}} + (\sigma_{\text{abs},1} + \sigma_{\text{rel},1}\Gamma)^2 + (\sigma_{\text{abs},2} + \sigma_{\text{rel},2}\Gamma)^2 + \dots \tag{C.10}$$

Note: It's not uncommon to see the combination

$$\sigma^2_{\text{syst}} = \sigma^2_{\text{abs}} + \sigma^2_{\text{rel}}\Gamma^2$$

but this is only valid in the case where σ_{abs} and $\sigma_{\text{rel}}\Gamma$ are uncorrelated uncertainties.

Appendix D

The Coldest Cubic Meter in the Universe

The full CUORE detector consists of 988 bolometric modules for a total mass of 741 kg. When operational, this mass, plus an additional 2 metric tonnes of supporting material, will need to be cooled to 10 mK, producing the largest region in the Universe at that temperature. In this section, I briefly discuss the temperature of the Universe and some of the naturally occurring cold places in it. I introduce the CUORE detector and cryostat and describe the working volume and its temperature. I also compare the CUORE detector to some of the other large low temperature experiments that are operating or being built.

D.1 Low Temperature Regions in Nature

In comparison to the CUORE detector, the Universe is actually quite warm. Its temperature is dominated by the Cosmic Microwave Background (CMB) photons that pervade all empty space. This thermal bath of photons exists everywhere throughout the Universe and has a well defined temperature which has been measured with extreme accuracy to be $T_{\mathrm{CMB}} = 2.72548 \pm 0.00057$ K [118].

Many regions of space are heated above T_{CMB} by structure formation and the radiation this gives off, but there is currently only one known naturally occurring region *below* T_{CMB} and that is the Boomerang Nebula [119]. This proto-planetary nebula (PPN) consists of a central star surrounded by an envelope of molecular gas. The Boomerang Nebula is unique among known PPN in that it has produced an extremely massive and rapidly expanding envelope of gas. The high opacity of this envelope absorbs CMB photons in the outer layers, shielding the inner regions and allowing them to cool via adiabatic expansion. By combining radio measurements with radiative modeling, the authors of [120] place the kinematic temperature at about 1 K, but possibly as cold as 0.3 K. This makes the Boomerang Nebula the coldest known object in the Universe *outside the laboratory*. But here on Earth we routinely achieve temperatures colder than 300 mK.

Figure D.1: Image of the Boomerang Nebula taken on the Hubble Space Telescope. Photo credit: NASA/ESA.

D.2   The CUORE Cryostat

The CUORE detector is hosted in one of the largest cryostats ever constructed and is cooled by a ^{3}He/^{4}He dilution refrigerator that was designed and built by Leiden Cryogenics and is one of the most powerful in the world. A detailed description of the CUORE cryostat can be found here [121], and a paper describing its commissioning is in preparation. The cryostat is built as a series of nested vessels that step the temperature down from 300 K to $\sim$ 40 K, $\sim$ 3.5 K, $\sim$ 800 mK, $\sim$ 50 mK and finally the detector temperature of 10 mK. Each stage is connected to a cooling unit and has a radiation shield that thermally isolates the enclosed volume.

The largest stage of the cryostat below 1 K is the Still. The temperature of this stage will be adjusted to optimize the temperature of the coldest stage, but it is typically maintained between about $\sim$800-900 mK. It is composed of a radiation shielding copper can 112 cm in diameter by 185 cm in height, mounted to the bottom of the copper Still plate that has a diameter of 133 cm and a thickness of 4.3 cm. The total enclosed volume at or below $\sim$ 1 K is $\sim$1890 L.

Inside the still shield is the next colder stage of the cryostat, the Heat Exchanger. The temperature of this stage will also be adjusted to maintain the base temperature but it is typically maintained at $\sim$50 mK. This stage consists of a radiation shielding copper can of diameter of 103 cm and height of 165 cm, it is mounted to the bottom of a copper plate 107 cm in diameter and 2.8 cm thick. The total enclosed volume at or below $\sim$50 mK is

Table D.1: The stages of the CUORE cryostat below 1K and the volumes and masses colder than that temperature. The values are cumulative and should be read as the 'total volume/mass colder than.' (Numbers are approximate.) [122–124]

Stage	Temperature (mK)	Volume (L)	Mass (kg)
4K Stage	3.5 K	3340	16000
Still	850	1890	14100
Heat Exchange (HEx)	50	1440	5900
Base (no load)	10	990	400
Base	10	636	2000

1435 L.

Inside the Heat Exchanger (HEx) vessel is the coldest stage of the cryostat, the Mixing Chamber (MC) plate, the lead shielding, and the CUORE detector itself. The MC plate is suspended from the Heat Exchanger plate and hosts the final stage of the dilution refrigerator. The plate itself is 98 cm in diameter and 1.8 cm thick and supports a radiation shielding copper vessel below, which is 94 cm in diameter and 130 cm in height. The MC stage is cooled to 10 mK and encloses a volume of 990 L.

However, inside this volume things get slightly complicated. Below the 10mK plate, there will be ∼2.6 metric tonnes of lead and copper shielding, which, for reasons of cooling power, is thermalized to 50 mK. This shielding takes up a volume of 235 L inside the 10 mK shielding. So to be conservative, we will take the operating volume at 10mK to be only that of the shielded detector itself, 636 L. These sizes and volumes are summarized in Tab. D.1.

D.3 Discussion

The current record for the coldest cubic meter in the Universe was set in the first of the CUORE cryostat commissioning runs without the lead shielding mounted [paper in preparation]. When the CUORE detector is fully commissioned and running (2015), the detector will be held stably at the operating temperature of ∼10 mK for the duration of CUORE data taking – which is expected to be ∼ 5 years. During this time, both the 636 L held at 10 mK and the 1435 L held entirely below 50 mK will be the coldest volumes of those respective sizes in the known Universe. This gives CUORE cryostat the distinction of being the *Coldest Cubic Meter* in the known Universe.

In Tab. D.2, I list a few, but certainly not all, of the larger currently running experiments below 100 mK. I list their approximate cold volumes and operating temperatures. Not surprisingly, the largest coldest experiments are often rare-event searches.

Table D.2: A (non-exhaustive) list of currently running large volume experiments with operating temperatures below 100mK. The mass represents the *target* or detector mass, excluding any supporting material or structure. CUORE (HEx) is not a separate experiment, but just the 50 mK stage of the CUORE detector. All numbers are approximate, and only meant to give a sense of scale. [a]Auriga/Nautilus operated at $T \sim 100$ mK from 1997-99, but are currently taking data at 4.4 K.

Experiment	Mass (kg)	Size (L)	Temperature	Physics Goal	Location	Ref.
CUORE	741	636	10mK	$0\nu\beta\beta$	Gran Sasso, Italy	[121]
CUORE-0	39	27	12mK	$0\nu\beta\beta$	Gran Sasso, Italy	
CRESST-II	10	24	15mK	Dark Matter	Gran Sasso, Italy	[125]
Edelweiss	32	50	20mK	Dark Matter	Modane, France	[126]
SuperCDMS	10	21	40mK	Dark Matter	Soudan, SD	[127]
CUORE (HEx)	-	1435	50mK	$0\nu\beta\beta$	Gran Sasso, Italy	[121]
Auriga/Nautilus[a]	2200	848L	100mK	Gravity Wave	Italy	

D.4   Caveats, Qualifications, Ifs and Buts..

There are several caveats and assumptions that should be noted. Many of these are technicalities, but they deserve mentioning:

- The CUORE detector, like everything on Earth, is bathed in a constant flux of neutrinos both from the sun and earth's core. The solar neutrinos were last thermalized in the core of the sun to temperatures of order 10^7 K and neutrinos from the earth's core were last thermalized to temperatures of order 10^4 K. However, neutrinos interact so infrequently — we expect of order 100 solar neutrinos per year to interact in the CUORE detector — that they never reach a thermal equilibrium with anything on Earth (thankfully). However, all of these neutrinos are still technically present inside the volume of the CUORE detector, so I explicitly ignore them here.

- Another very interesting source of neutrinos that are also present inside the volume of the CUORE detector is the Cosmic Neutrino Background (CνB). These are the relic neutrinos from the Big Bang, and like the photons in the CMB they are expected to pervade the entire Universe. Unlike the photons in the CMB, these neutrinos interact so infrequently that they have not actually been detected yet. These neutrinos are expected to be warmer than the CUORE detector but like the solar and geo-neutrinos, never come into thermal equilibrium, so I explicitly ignore these as well.

- CUORE will have a significant amount of lead shielding sitting inside the 10 mK shield and thermalized to 50 mK. Lead becomes superconducting around 7 K and as a result, as the temperature drops below $\sim T_c/10 = 700$ mK, the thermal conductivity becomes very poor and the lead begins to self insulate. It will eventually reach 50 mK, but it is difficult to say on what timescale. Neglecting the lead in the 10 mK volume, the remaining volume at 50 mK is about 1245 L.

- The definition of the cubic meter can be made fuzzy. Here I have considered only a simple contiguous convex volume of space.

- Proving the non-existence of a cubic meter in the Universe colder than CUORE is, of course, an impossible task. I have implicitly restricted the discussion to known or discovered phenomena. But, I admit the possibility of another planet somewhere in an infinite Universe, which is entirely identical to Earth in every way, except that their CUORE collaboration has decided to operate their CUORE detector at 9 mK.

D.5   More on the CMB

The CMB sets the standard for what we consider the temperature of the Universe, and in this appendix we consider the possibility of a purely *statistical* fluctuation of the CMB temperature down to 10 mK somewhere in the Universe. If we broaden our working definition of temperature, we can calculate the probability that the *mean* kinetic energy of all the CMB photons in a single cubic meter of space fluctuates down to the point that their mean energy is consistent with ~ 10 mK. (I should warn, that the following argument will be *very* approximate.)

Starting from the Planck distribution

$$n(E)dE = \frac{1}{\pi^2 \hbar^3 c^3} \frac{E^2}{e^{E/k_B T} - 1} dE$$

we can integrate this to find the photon number density at 2.7 K, $n \sim 4 \times 10^8$ m^{-3}, the mean photon energy $\langle E \rangle \approx 0.7$ meV, and the RMS $\sigma_E \sim 0.4$ meV. Due to the extremely large number of photons in a given cubic meter, the magnitude of fluctuations of *mean* kinetic energy is extremely suppressed, $\sigma_{\langle E \rangle} \sim \frac{\sigma_E}{\sqrt{\langle N \rangle}} \sim 20$ neV. A fluctuation down to a mean kinetic energy consistent with $T \sim 10$ mK (or $\langle E \rangle \sim 2.3$ μeV), amounts to a $\sim 31,000\sigma$ downward fluctuation.

If one were to assume a Gaussian probability distribution, this calculation would imply that the fraction of cubic meters in the Universe that had fluctuated down this far would be about 1 in $\sim 10^{10^8}$. For reference, there are $\sim 10^{80}$ cubic meters in the observable Universe. However, we are extrapolating this distribution down so far that considering it Gaussian is certainly no longer warranted.

So instead we ask the question what *is* the coldest cubic meter in the CMB — according to this interpretation of mean energy density. Or in other words, what temperature corresponds to the mean energy density which we expect fewer than 1 in 10^{80} cubic meters of space to be colder than. Assuming a Gaussian distribution, this would take us down only 19σ — a much more modest distance to extrapolate than $31,000\sigma$. This corresponds to a fractional difference in temperature of $\frac{\Delta T}{T} \sim 6.1 \times 10^{-4}$. Or in other words, for statistical fluctuations alone, we expect that at any given time the coldest cubic meter anywhere in the CMB is only ~ 1.6 mK colder than the average temperature of the CMB.