Energy Calibration for a Sensitive Search for Neutrinoless Double-Beta Decay: Using the Cuoricino Experience to Prepare for CUORE

By

Larissa M. Ejzak

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Physics)

at the
UNIVERSITY OF WISCONSIN–MADISON
2013

Date of final oral examination: 05/07/13

The dissertation is approved by the following members of the Final Oral Committee:
Karsten Heeger, Professor, Physics
Baha Balantekin, Eugene P. Wigner Professor, Physics
Dan McCammon, Professor, Physics
Ellen Zweibel, William L. Kraushaar Professor, Astronomy & Physics
Yang Bai, Assistant Professor, Physics

ACKNOWLEDGMENTS

In addition to my advisor Karsten Heeger, I would like to thank all of the members of the UW CUORE group with whom I have worked: Adam Dally, Ian Guinn, Jackie Houston, Daniel Lenz, Reina Maruyama, Samuele Sangiorgio, Lauren Wielgus, Tom Wise, and Dan Zou. I am particularly grateful to Jess Clark for all her hard work on the calibration source carriers. I would also like to thank all of the other members of the CUORE Collaboration who have supported me in my work on the experiment, including Fabio Bellini, Adam Bryant, Luca Gironi, Elena Guardincerri, Laura Kogler, Maria Martinez, Maura Pavan, Claudia Tomei, and Marco Vignati.

DISCARD THIS PAGE

TABLE OF CONTENTS

Page

Appendix

Page

DISCARD THIS PAGE

LIST OF TABLES

Table Page

Appendix
Table Page

Appendix
Table Page

DISCARD THIS PAGE

LIST OF FIGURES

Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

ABSTRACT

Neutrinos, which were long believed to be massless particles, are now known to have a tiny finite mass. A thorough understanding of the properties of their masses may provide vital clues to the history of the development of the universe as we know it.

An intensive experimental search is underway for evidence of a Majorana mass component to the neutrino via many current and upcoming detectors seeking to observe the rare nuclear process neutrinoless double-beta decay ($0\nu\beta\beta$). These detectors must be able to achieve very low backgrounds and a precise understanding of their energy scales.

This paper presents the experimental approach of one of these $0\nu\beta\beta$ experiments, the Cryogenic Underground Observatory for Rare Events (CUORE), and the attendant challenges of achieving excellent energy calibration performance in the detector from the perspectives of both hardware design and analysis. Experience and data from Cuoricino, the predecessor of CUORE, have been extensively leveraged to prepare optimized operational procedures for CUORE. The expected sensitivity profile of CUORE as a function of time is also presented and compared with those of other leading $0\nu\beta\beta$ experiments.

Chapter 1

Introduction: The Role of Neutrinos

1.1 History of Neutrino Masses

The neutrino, by its very nature, is a somewhat mysterious particle. It interacts only weakly — it carries no electric charge or color charge — and is extremely light, so much so that for decades it was believed to be massless. With the revelation that neutrinos have mass comes the need to explain and characterize that mass as well as the possibility that neutrinos have had a more profound impact on the history of the Universe than had previously been suspected.

1.1.1 A Timeline of the Standard-Model Neutrino

In 1930, as a "desperate remedy" to salvage nuclear spin statistics and energy conservation in beta decay, Wolfgang Pauli proposed the existence of a light, electrically neutral, highly penetrating particle with spin $\frac{1}{2}$ that he called the 'neutron'. At the time, the nucleon now known as the neutron had not yet been discovered. Instead, it was believed that a nucleus of mass number A and charge Z was composed of A protons and $A - Z$ electrons and that one of these latter was simply emitted when the nucleus underwent beta decay [51]. However, this model was problematic in several respects. As protons and electrons both have spin $\frac{1}{2}$, a nucleus should contain $2A - Z$ spin-$\frac{1}{2}$ particles, and it therefore should obey Fermi statistics if Z is odd and Bose-Einstein statistics if Z is even. Instead, a nucleus obeys Fermi statistics if A is odd and Bose-Einstein statistics if A is even. Similarly, the emission of a single spin-$\frac{1}{2}$ electron in beta decay poses a problem of conservation of angular momentum. Meanwhile, the energy spectrum of the electrons emitted in beta decay is not a sharp peak at the energy difference between the initial and final nuclei, as would

be required by energy conservation if the electron were the only emitted particle, but rather a broad continuum with its upper endpoint at the total transition energy of the decay. Pauli's 'neutron' would provide the solution to all of these problems if one of them existed in the nucleus for each nuclear electron and if one was always emitted alongside the electron in beta decay.

By 1934, when Enrico Fermi formulated his theory of beta decay (an English translation of the original paper is provided in [133]), the discovery of the particle now known as the neutron had allowed for a different picture of the composition of the nucleus. Fermi proposed that the nucleus contained only protons and neutrons and that the beta decay process transformed a neutron into a proton with the emission of an electron and one of Pauli's particles, which he dubbed the 'neutrino'. Fermi's theory provided the basis for the modern quantitative understanding of the weak force, though he treated it as a contact interaction, while the weak force is now understood to be mediated by gauge bosons. In the same paper, Fermi presented a discussion of the shape of the endpoint of the beta decay spectrum in which he argued that the empirically observed curves imply a massless neutrino, or at least a neutrino mass of much smaller than the mass of the electron [133] (similar arguments now drive precision experiments seeking to directly measure the neutrino mass; see Section 1.1.3).

Although the neutrino was a neat solution to some of the mysteries of beta decay and a key element of Fermi's theory, the fact that no such particle had yet been observed meant that it must interact with matter only very rarely, making confirming the existence of the neutrino through direct observation a very difficult task. It was not until 1953 that Fred Reines and Clyde Cowan published results from an experiment that used a nuclear reactor at the Hanford Site in Washington as an intense source of neutrinos and attempted to detect those neutrinos via the inverse beta decay reaction, which was expected to occur due to crossing symmetry with beta decay:

$$\nu_- + p \rightarrow n + \beta^+, \tag{1.1}$$

where ν_- is the notation used by Reines and Cowan to indicate the neutrino that is produced accompanying electron (β^-) emission, in contrast to ν_+, the neutrino produced accompanying positron (β^+) emission [121]. By using the delayed coincidence between the annihilation of the positron and the capture of the neutron in the detector, Reines and Cowan believed that they were able to

identify neutrino-induced events and claim evidence for the direct detection of the neutrino; however, the detector was plagued by background from cosmic rays, so the experiment was repeated at the Savannah River Plant, where the detector could be placed underground for improved shielding, providing a clear confirmation of the results [57]. The first empirical observation of the neutrino thereby established two main experimental criteria, at least one of which has been necessarily fulfilled by every neutrino experiment in the ensuing decades: an intense neutrino source and very heavy shielding, often obtained in part by placing the detector underground, for stringent control of backgrounds.

In modern notation, the ν_- of Equation 1.1 is actually an antineutrino ($\overline{\nu}$) carrying lepton number $L = -1$ to balance the lepton number $L = 1$ carried by the electron emitted in β^- decay, while the so-called ν_+ is actually a neutrino carrying $L = 1$. Shortly after the first experimental observation of the antineutrino, a series of experiments by Chien-Shiung Wu and her collaborators demonstrated that if the spins of radioactive nuclei are polarized into a known orientation, there is a strong asymmetry in the preferred emission direction of the β, and the sign of the asymmetry depends on whether the emitted β is a β^+ (positron) or a β^- (electron) [20]. This observation established that the weak interaction maximally violates both parity (P) and charge-conjugation (C) invariance, and in doing so, it propelled into prominence a theory attributing this parity violation to the existence of only one possible helicity state (orientation of spin with respect to linear momentum) of the ν, as opposed to most particles, which can be either right-handed (spin vector parallel to linear-momentum vector) or left-handed (spin vector antiparallel to linear-momentum vector). In this theory, the $\overline{\nu}$ is strictly distinct from the ν, which has the opposite helicity, and helicity is in fact equivalent to chirality, a good quantum number, disallowing the possibility that ν and $\overline{\nu}$ could be equivalent; the mass of the neutrino and antineutrino also must be strictly equal to zero [109]. Several experiments in the late 1950's and early 1960's used conservation-of-momentum-and-angular-momentum arguments to establish the helicity assignments of ν and $\overline{\nu}$ by measuring the polarization of particles emitted alongside neutrinos or antineutrinos in various decay scenarios; for example, a measurement of the polarization of resonant-scattered γs from orbital electron capture showed that the emitted neutrinos were all left-handed [85], and observations

of pion decay,

$$\pi^- \rightarrow \mu^- + \overline{\nu}_\mu, \tag{1.2}$$

revealed that the emitted muons and therefore the emitted antineutrinos were right-handed [41].

As may be evident from the notation used in Equation 1.2, the measurement reported in [41] was partially motivated by an attempt to determine whether the neutrino associated with muon emission, ν_μ ($\overline{\nu}_\mu$), is the same as the neutrino associated with beta emission, ν_e ($\overline{\nu}_e$). While the helicity assignments of ν_e and ν_μ proved to be the same, the two neutrinos nevertheless soon proved to be different entities. In 1962, results were published from an experiment at Brookhaven using a high-energy neutrino beam, generated by the decay-in-flight of pions, incident on a beryllium target. The experiment observed 34 single-muon events, of which 5 were considered to be cosmic-ray background, leaving 29 attributed to muons produced by the interaction of the neutrino beam with the nuclei of the target; however, no electron showers of an energy consistent with electrons produced by the interaction of the neutrino beam with the target were observed, though they would have been expected to occur with a similar rate as the muon events if $\nu_e = \nu_\mu$ [59]. Thus neutrino interactions were seen to obey conservation of lepton flavor in addition to conservation of lepton number.

Thus the basic characteristics of the neutrino in the theory of the weak force were established and carried through the unification of the electrical and weak forces into electroweak theory in the 1960's (for which the 1979 Nobel Prize in Physics was awarded [3]) and the eventual combination of electroweak theory and the theory of the strong force, quantum chromodynamics, into the Standard Model of Particle Physics. In the Standard Model, there are three distinct left-handed massless neutrinos, each of which carries a flavor charge associated with one of the three generations of charged leptons: ν_e, ν_μ, and ν_τ. Similarly, there are also three generations of right-handed massless antineutrinos: $\overline{\nu}_e$, $\overline{\nu}_\mu$, and $\overline{\nu}_\tau$. The number of light neutrino species that interact via the weak force can be inferred to high precision from the width of the Z boson, one of the force carriers for the weak force; data from the Large Electron-Positron Collider (LEP) confirms that there are three such flavors of neutrinos [124]. Right-handed neutrinos and left-handed antineutrinos do not exist in the Standard Model, and neither does any mechanism by which the neutrino could gain a

mass (see Section 1.2.1 for more details on the neutrino mass formalism and why it is inconsistent with the Standard Model).

The maximal parity violation of the weak interaction is still preserved if right-handed neutrinos and left-handed neutrinos exist but do not interact with the weak force; however, as such particles were never experimentally observed, the Standard Model framework does not include them. At the close of the 20$^{\text{th}}$ century, however, several pieces of experimental evidence converged to indicate that neutrinos indeed have a very small, but non-zero, mass, meaning that the Standard Model description of neutrinos is no longer sufficient to explain experimental observations.

1.1.2 Neutrino Mixing and Masses

There are neutrinos from various different sources passing through the Earth all the time. Two of these sources that are well characterized by theory are the Sun, which produces electron neutrinos as it burns, and meson and muon decay in the upper atmosphere (e.g., $\pi^+ \rightarrow \nu_\mu + \mu^+ \rightarrow \nu_\mu + \bar{\nu}_\mu + \nu_e + e^+$), which produces muon neutrinos and electron neutrinos in a ratio of roughly $\nu_\mu : \nu_e \approx 2 : 1$ [129]. These solar and atmospheric neutrinos have provided strong evidence that neutrinos undergo flavor oscillations and therefore must have mass.

Ray Davis' solar neutrino experiment in the Homestake mine was built in the late 1960's to investigate the neutrino flux predicted by the solar model by collecting and counting ^{37}Ar nuclei produced by ν_es interacting with ^{37}Cl in the chlorine-based fluid in the detector tank via inverse beta decay. It provided the first indication that there were fewer electron neutrinos coming from the Sun than expected — about one-third of the predicted flux — and continuously collected data for several decades that all showed this result [55]. Similarly, in the late 1980's and early 1990's, the water-Cherenkov nucleon-decay experiments Kamiokande [96] and IMB [44] witnessed considerably lower atmospheric $\nu_\mu : \nu_e$ ratios than expected. It was not until 1998 that Super-Kamiokande reported a zenith-angle dependence of atmospheric ν_μ events that indicated that the farther the neutrinos had traveled, the more likely they were to have disappeared; then in 2001, the Sudbury Neutrino Observatory (SNO) published results showing that the rate of charged-current interactions in their detector (sensitive only to ν_e) agreed with the previously-observed solar neutrino deficit,

while the rate of neutral-current interactions (sensitive equally to ν_e, ν_μ, and ν_τ) concurred with the total (electron) neutrino flux predicted by the solar model, demonstrating that the 'missing' electron neutrinos had in fact changed flavor to become ν_μs and ν_τs instead [111].

Although for many years several possible solutions to the so-called 'solar neutrino problem' were considered, such as inadequacies in the solar model or neutrino decay, the results of Super-Kamiokande and SNO together constitute conclusive evidence that the observed deficit of solar ν_es is due to neutrino flavor oscillations. Neutrinos produced in one flavor state can oscillate into different flavor states as they travel, which is possible only because each flavor state is a superposition of different mass eigenstates. Super-Kamiokande established that the disappearance probability of a single neutrino flavor is a function of the distance traveled, as expected for oscillations; SNO showed that the total number of neutrinos from the sun is consistent with the predictions of the solar model, so the neutrinos are not decaying away into other particles on their journey to Earth. The neutrinos travel as the mass states, and therefore phase differences among the mass components develop as they travel, dependent upon the mass-squared differences between the states [111]. These phase differences are what allow the neutrinos, after traveling some distance, to interact as a different flavor state than that in which they began. If all the mass states had a mass of identically zero, there would be no mass-squared differences and therefore no oscillation.

The mixing among the flavor states and the mass states with masses m_j can be expressed as [37]

$$|\nu_f\rangle = \mathbf{U}|\nu_j\rangle, \tag{1.3}$$

where the mixing matrix $\mathbf{U}$ containing the oscillation parameters can be parametrized as follows:

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -\sin\theta_{13}e^{i\delta} & 0 & \cos\theta_{13} \end{pmatrix} \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.4}$$

where the θ_{jk} are mixing angles and δ is a CP-violating phase (two further possible complex Majorana phases are neglected here as they do not affect oscillations). Oscillation is generally considered in terms of survival probability (e.g., $P(\nu_e \rightarrow \nu_e)$), disappearance probability (e.g.,

$1 - P(\nu_e \rightarrow \nu_e)$), or appearance probability of a neutrino flavor that does not exist when the neutrinos are produced (e.g., $P(\nu_e \rightarrow \nu_\mu)$), and it depends on the distance traveled by the neutrinos, L; the energy of the neutrinos, E; and the mass-squared differences of the mass states, $\delta m^2_{jk} \equiv m^2_j - m^2_k$. For example, in natural units ($\hbar = c = 1$), the electron neutrino survival probability is [37]

$$P(\nu_e \rightarrow \nu_e) = 1 - \sin^2 2\theta_{13} \left[\cos^2 \theta_{12} \sin^2 \left(\frac{\delta m^2_{31}}{4} \frac{L}{E} \right) + \sin^2 \theta_{12} \sin^2 \left(\frac{\delta m^2_{32}}{4} \frac{L}{E} \right) \right] - \cos^4 \theta_{13} \sin^2 2\theta_{12} \sin^2 \left(\frac{\delta m^2_{21}}{4} \frac{L}{E} \right). \tag{1.5}$$

It should be noted that, as the dependence on the mass splittings δm^2_{jk} is always in terms of $\sin^2(\delta m^2_{jk})$, the oscillation probability is not sensitive to the sign of the mass splitting.

Depending on the distance L of a detector from the neutrino source it is studying and the energy E of the neutrinos it observes, an experiment may operate in a regime in which the three-flavor oscillation of Equation 1.5 effectively reduces to a much simpler two-flavor oscillation. Neutrino oscillation experiments have exploited this fact and have now successfully measured all three mixing angles and at least the magnitudes of both mass-squared differences. The combination of solar ν_e disappearance data with reactor $\overline{\nu_e}$ disappearance data from the KamLAND experiment establishes a value of the 'solar' mixing angle θ_{12} such that $\sin^2 \theta_{12} \approx 0.3$, and due to a shift in the mass splitting induced by the effective mass acquired by the solar neutrinos as they scatter off the dense matter of the solar core while traveling through the sun, in a phenomenon known as the Mikheyev-Smirnov-Wolfensten (MSW) effect, these combined data also determine not only the magnitude but the sign of the 'solar' mass splitting to be $\delta m^2_{12} \approx -7.5 \times 10^{-5}$ eV^2; the atmospheric data of Super-Kamiokande and accelerator-based data sensitive to the 'atmospheric' oscillation scale are not sufficiently strongly affected by matter effects to establish the sign of the relevant mass splitting, but these data have yielded values of $|\delta m^2_{32}| \approx 2-3 \times 10^{-3}$ eV^2 and a near-maximal $\sin^2 \theta_{23} \approx 0.4-0.6$; and the short-baseline reactor neutrino experiments Daya Bay, RENO, and Double Chooz have recently measured the final mixing angle to be $\sin^2 \theta_{13} \approx 0.1$ [37].

The measured mass splittings have established that two of the physical neutrino masses, m_1 and $m_2 > m_1$, are comparatively similar, while there is a larger gap between m_1, m_2 and m_3. It is

possible that both mass splittings are small compared to the overall mass scale of the neutrinos, a scenario called the 'degenerate mass hierarchy.' Otherwise, due to the remaining sign ambiguity of δm^2_{32}, it is possible that either $m_1, m_2 << m_3$ (called the 'normal hierarchy' by analogy with the pattern of masses in the quark sector [48]) or $m_1, m_2 >> m_3$ (called the 'inverted hierarchy'). Which hierarchy is the true one has implications for various physical processes, such as neutrinoless double-beta decay (see Section 2.2), and also for the best-fit values of some of the oscillation parameters.

Although neutrino oscillations are not sensitive to the overall neutrino mass scale, the measured mass splittings do at least establish a lower bound. Even if the lightest mass state has zero mass, at least one neutrino must still have a physical mass of at least $\sqrt{2-3\times10^{-3}\ \mathrm{eV}^2} \approx 0.05$ eV.

1.1.3 Direct Mass Measurements

Neutrino oscillations have established that neutrinos have mass, but because they can provide only a lower bound on the absolute scale of that mass, an experimental program is now under way in an effort to directly measure neutrino masses.

The kinematics of β decay, in which an electron (positron) and an $\overline{\nu_e}$ (ν_e) are emitted, provide an avenue for directly measuring the neutrino mass that has been extensively pursued. Some of the energy of the decay must go into the mass of the (anti)neutrino, thereby reducing the total energy available to be carried by the β and affecting the shape of the endpoint of its energy spectrum. If the CPT theorem holds, $m_{\nu_f} = m_{\overline{\nu_f}}$, so experiments probing β^- decays, β^+ decays, and electron capture should all be sensitive to the same quantity. Electron-capture-based direct neutrino mass measurements have set upper limits on the order of 10^2 eV, but so far experiments using β^- decay in tritium have been the most sensitive due to its fairly low endpoint energy (favorable for achieving good energy resolution), high decay rate per unit of source mass, and simple electronic structure yielding low systematic uncertainties in the measurement [66]; the most stringent limit set to date is $m_{\nu_e} < 2.05$ eV (95% C.L.) from the Troitsk experiment [32]. The Karlsruhe Tritium Neutrino (KATRIN) experiment, which will use a 9.8-m-diameter electrostatic spectrometer to obtain a high-resolution β-decay energy spectrum and expects to start data-taking in the latter half of 2015,

represents the next stage in tritium-based direct-mass-measurement experiments and is designed to improve the m_{ν_e} sensitivity by about an order of magnitude: it is expected to set a 90% C.L. upper limit of 200 meV if no neutrino mass signal is seen [66]. Accelerator-based experiments can use similar arguments to set limits on m_{ν_μ} and m_{ν_τ} from pion decay and tau decay, respectively, but have far inferior sensitivity: current limits are $m_{\nu_\mu} < 190$ keV at 90% C.L. and $m_{\nu_\tau} < 18.2$ MeV at 95% C.L. [48].

Of course, as discussed in the previous section, the neutrino flavor eigenstates are not the same as the neutrino mass eigenstates; therefore, any experiment that measures the mass of a flavor state is really measuring an effective mass resulting from the appropriate superposition of mass eigenstates (the physical masses). The effective 'average electron neutrino mass' measured by β-decay or electron-capture experiments can be written as follows [66]:

$$m^2_{\nu_e} = \sum_j |U^2_{ej}| m^2_j, \tag{1.6}$$

where the U_{ej} are elements of the neutrino mixing matrix given in Equation 1.4.

In principle, a direct determination of the neutrino mass can also be made from time-of-flight measurements. In practice, such a measurement requires a very long baseline and a very strong and well-localized (in both time and space) neutrino source to be able to distinguish the effect of the neutrino mass. Only astrophysical events like core-collapse supernovae can meet these requirements, and uncertainties in the theoretical supernova model and the rarity of such events make it difficult both to obtain high-statistics data sets and to extract high-precision neutrino mass evaluations from what data are collected; data collected from SN 1987A have allowed upper limits of only around 6 eV to be placed on the neutrino mass [66].

Meanwhile, cosmological data — from observations of the cosmic microwave background (CMB) and large-scale structures (LSS) — are sensitive to the *sum* of the physical neutrino masses. The relic density of neutrinos from the Big Bang indicates that neutrinos acted as hot dark matter in the early universe, smearing out, at small scales, the fluctuations that formed the bases of the large-scale structures now visible in the current universe; the effect of this smearing can be seen from large galaxy surveys in the present clumping scales of galaxies and other astrophysical systems,

and the magnitude of this effect is dependent upon the neutrino mass [66]. The neutrino mass constraints that can be obtained from these data correspond to the sum of the masses rather than to any particular neutrino state, and they are highly dependent upon the particular assumptions made for the cosmological model; with this caveat, these data currently imply the tightest constraints on neutrino masses, $\sum_j m_j \lesssim 0.2-0.3$ eV [63, 10].

1.2 Majorana Neutrinos

If neutrinos have mass, a theoretical possibility proposed by Ettore Majorana in the 1930's abruptly acquires keen physical interest. Neutrinos carry no electrical charge or color charge; they carry only lepton number ($+1$ for neutrinos, -1 for antineutrinos). Although no process has been observed thus far to violate total lepton number (note that, thanks to neutrino oscillations, lepton *flavor* number *is* known to be violated), baryon number B and lepton number L are only approximately conserved in the Standard Model; $B-L$ is the quantity that is strictly conserved [65] (see Section 1.2.3 for further discussion). It is possible, then, that the neutrino does not carry any good conserved-charge quantum number, and it may therefore be its own antiparticle.

For a massless neutrino, helicity is equivalent to chirality and a good quantum number — there is no frame in which the momentum of the particle can appear to flip signs, thereby flipping helicity — and because the weak force interacts only with left-handed particles and right-handed antiparticles, there is therefore no way for a (left-handed) neutrino to behave as a (right-handed) antineutrino, even if the two are otherwise indistinguishable apart from helicity [127]. For a massive neutrino, however, the question of whether the neutrino is a Majorana particle becomes pertinent again.

1.2.1 Dirac and Majorana Neutrino Mass Terms

From a field-theory perspective, the nature of neutrino mass is determined by the mass term(s) appearing in the Lagrangian. A mass term couples a right-handed field to a left-handed field; this kind of term can be constructed in two different ways.

For a massive particle ψ with right- and left-handed particle and antiparticle states, a Dirac mass term can be written, which has the following form:

$$-\mathcal{L}_{\psi_D} = m_D(\overline{\psi_L}\psi_R + \overline{\psi_R}\psi_L). \tag{1.7}$$

Most known massive particles have a mass term in this form; in particular, the Standard-Model Yukawa couplings that give the quarks and charged leptons their masses look just like Equation 1.7 with the Standard-Model Higgs field H_{SM} appearing in place of m_D [100]. These couplings represent Higgs interactions that reverse the handedness of a particle or antiparticle; for example, they can turn a left-handed particle into a right-handed particle or a right-handed antiparticle into a left-handed antiparticle. In the case of neutrinos, however, only left-handed neutrinos and right-handed antineutrinos are known to exist; if right-handed neutrinos and left-handed antineutrinos exist, they do not participate in weak interactions and have therefore never been observed. When neutrinos were believed to be massless, there was no need to consider whether it was possible to write a neutrino mass term. With the discovery that neutrinos have mass, however, it is necessary to postulate the existence of a (sterile) right-handed neutrino ν_R to write a Standard-Model neutrino mass term:

$$-\mathcal{L}_{\nu_D} = H_{SM}(\overline{\nu_L}\nu_R + \overline{\nu_R}\nu_L). \tag{1.8}$$

The Standard-Model Higgs is a complex scalar field with components with weak hypercharge $Y_W = \pm 1$:

$$Y_W = 2(Q - T_3), \tag{1.9}$$

where Q is electrical charge and T_3 is the third component of weak isospin, and Y_W and Q must both be conserved. Particles that form doublets, as charged leptons do with left-handed neutrinos, have $T_3 = \pm\frac{1}{2}$. Particles that are singlets, as a proposed right-handed neutrino would be, have $T_3 = 0$. A Dirac mass term for the neutrino, therefore, couples two fields that differ by one unit of Y_W, and this difference can be compensated by the Higgs field with a Standard-Model Higgs mass term like that in Equation 1.8.

Instead of calling upon two different fields of opposite chiralities to write a mass term, it is also possible to write a mass term using a single field and its own charge conjugate, which will

have the opposite chirality by definition. A mass term of this kind is a Majorana mass term; a massive particle ψ with both left-handed and right-handed associated particle fields can have both a left-handed and a right-handed Majorana mass term:

$$-\mathcal{L}_{\psi_M} = \frac{1}{2} m_{M_L} (\overline{\psi_L^c} \psi_L + \overline{\psi_L} \psi_L^c) + \frac{1}{2} m_{M_R} (\overline{\psi_R^c} \psi_R + \overline{\psi_R} \psi_R^c). \tag{1.10}$$

A mass term that takes this form represents an interaction that changes a particle into its charge conjugate, reversing not only the handedness but also the sign of every other quantum number carried by the particle; if the Lagrangian contains a term like this, then every mass eigenstate of the particle is some superposition of the left- and right-handed fields that is its own charge conjugate, meaning that the associated physical particle is its own antiparticle [100]. Effectively, a Majorana term mixes a right-handed antiparticle with a left-handed particle or a right-handed particle with a left-handed antiparticle. For any particle that carries a non-zero value of a conserved charge, such as a particle with $Q \neq 0$, such a term is plainly impossible, as it would violate the conservation of that charge. For neutrinos, which carry no electrical charge, Majorana mass terms may be valid if lepton number is not strictly conserved:

$$-\mathcal{L}_{\nu_M} = \frac{1}{2} m_{\nu M_L} (\overline{\nu_L^c} \nu_L + \overline{\nu_L} \nu_L^c) + \frac{1}{2} m_{\nu M_R} (\overline{\nu_R^c} \nu_R + \overline{\nu_R} \nu_R^c). \tag{1.11}$$

However, it is still necessary to conserve Y_W in this scenario. The mass mechanism represented by $m_{\nu M_L}$ must compensate for the net change of two units of Y_W in the left-handed Majorana neutrino mass term, while the mass mechanism represented by $m_{\nu M_R}$ must have $Y_W = 0$. The Standard-Model Higgs cannot provide the mass mechanism for either term: the left-handed term would require two instances of H_{SM} to compensate for the change in Y_W, making the term non-renormalizable, or a Higgs triplet with weak hypercharge 2, which does not exist in the Standard Model; the right-handed term does not require a contribution from a Higgs field at all [100]. Thus, if neutrinos are found to be Majorana in nature, the neutrino mass must at least partially arise from some mechanism that is not the Standard-Model Higgs.

1.2.2 Implications of Majorana Neutrinos

Neutrinos are massless in the Standard Model, so the discovery that neutrinos have non-zero mass is already an indication that some kind of physics beyond the Standard Model is involved in the neutrino sector. Exactly what form this new physics may take is unknown, but the possibility that neutrinos may be Majorana particles is considered very promising. Majorana neutrinos play a key role in several classes of beyond-the-Standard-Model theories that have been extensively studied.

In particular, if neutrinos have Majorana masses, it may help to explain another aspect of neutrino masses themselves: various observational and experimental limits constrain the neutrino mass to be less than a few eV, many orders of magnitude smaller than the other elementary fermions, the quarks and charged leptons, which have masses in the MeV and GeV ranges. It is possible that the neutrinos are so light merely as an arbitrary quirk of nature, but it would seem more natural if an explanation could be found that would relate the mass scale of the neutrinos to the mass scale of the quarks and charged leptons. As it happens, introducing a mechanism to do this into the Standard Model may also provide an explanation for the origin of the matter-antimatter asymmetry in the universe.

1.2.2.1 The Seesaw Mechanism for Explaining the Neutrino Mass Scale

The simplest way to allow for neutrino masses in the Standard Model is to add two or three singlet fermions, the right-handed neutrinos [61], in addition to the three left-handed neutrinos already present in the theory. With this addition alone, the Lagrangian can contain Dirac mass terms and right-handed Majorana mass terms for the neutrinos. For convenience, the neutrino mass terms of the Lagrangian can be written in the following form [111]:

$$-\mathcal{L}_{\nu_m} = \frac{1}{2}\overline{\nu^c} M_\nu \nu + h.c., \tag{1.12}$$

where 'h.c.' indicates the Hermitian conjugate of the first term (in Equations 1.7, 1.8, 1.10, and 1.11, the Hermitian conjugates are written out explicitly) and

$$\nu = \begin{pmatrix} \nu_L \\ (\nu_R)^c \end{pmatrix}, \qquad M_\nu = \begin{pmatrix} 0 & M^D \\ (M^D)^T & M_R^M \end{pmatrix}, \tag{1.13}$$

where ν_L and ν_R are arrays containing the left- and right-handed neutrinos, respectively; M^D is the Dirac neutrino mass matrix; and M_R^M is the right-handed Majorana neutrino mass matrix. The Lagrangian takes on a form that looks like a single Majorana mass term, and the eigenvalues of M_ν are the physical masses of the neutrinos in this system.

The natural expectation is for M^D to be on a similar scale as the Yukawa couplings to the Higgs of the quarks and charged leptons. However, the masses of the quarks and charged leptons fall within a range of about $10^6 - 10^{11}$ eV, while the upper limit on the neutrino mass scale is around 1 eV [48]. If the Dirac mass terms are the only ones contributing to the neutrino masses, they must be many orders of magnitude smaller than the analogous terms in the quark and charged lepton sectors.

On the other hand, if the right-handed Majorana mass terms contribute, there is no reason to expect M_R^M to also be at the weak scale; it is very reasonable to suppose that it could be much heavier. In this case, diagonalizing the mass matrix results in three light Majorana neutrinos, corresponding to the three light neutrinos that have already been observed, with masses roughly inversely proportional to $\frac{(M^D)^2}{M_R^M}$ [113]. This is called the Type I Seesaw Model, and it is considered to be a natural explanation for the small mass of the light neutrinos, especially considering that the scale of M_R^M that would be required to reproduce the masses of the light neutrinos for the 'natural' scale of M^D in this case is compellingly similar to the scale of grand unified theories (GUT scale) [113].

An extension of this model involves adding a left-handed Higgs triplet that would allow the left-handed Majorana neutrino mass terms to become non-zero. This is the Type II Seesaw Model, and it is especially relevant to models in which the electroweak framework is left-right symmetric at high energies [113].

If neutrinos could be shown to be Majorana in nature, it would be a strong plausibility argument in support of the Seesaw Model.

1.2.3 Leptogenesis

The visible mass of the universe is largely made up of baryons: in particular, protons and neutrons. To all observations, it appears that all of the massive objects in the universe consist of matter (protons and neutrons), and there is no evidence of similarly large concentrations of antimatter (antiprotons and antineutrons). The origin of this baryon asymmetry is not yet understood.

A mechanism for baryon number violation is present in the Standard Model [65]. The Standard Model, as a non-abelian gauge theory, has a vacuum structure that consists of an infinite set of ground state configurations separated by barriers. Within a given configuration, B and L are conserved separately, and at zero temperature (the current universe) the tunneling probability from one configuration to another is infinitesimally small. However, there are also finite-energy 'sphaleron' field configuration solutions that correspond to sitting on top of one of these barriers, and small fluctuations about these solutions lead to falling down into one of the adjacent ground state configurations. Transitions among ground-state configurations conserve only $B - L$; the separate baryon and lepton numbers are violated. The heights of the barriers are roughly of the order of the mass of the W boson, so in the high temperatures of the early universe when the baryon asymmetry was presumably established, these solutions were quite relevant, and the rate of such B-and-L-violating sphaleron processes was rapid.

In fact, the Standard Model technically satisfies all three conditions necessary to dynamically generate a baryon asymmetry, known as the Sakharov conditions:

1. Baryon-number violation: If there is no initial net difference in the number of baryons and antibaryons present in the universe, it must arise from somewhere. Sphaleron processes violate B in the Standard Model.

2. C and CP violation: Even if B is violated in individual processes, if processes involving baryons occur at the same rate as processes involving antibaryons, no net baryon asymmetry can be generated. In the Standard Model, weak interactions violate both C and CP.

3. Out-of-equilibrium dynamics: No asymmetries in non-conserved quantum numbers are present in a system that is in chemical equilibrium. The electroweak phase transition provides non-equilibrium conditions in the Standard Model.

However, the level of baryon asymmetry that would result only from what is present in the Standard Model is several orders of magnitude smaller than the observed asymmetry, meaning that some additional mechanism must be responsible [61].

The class of models that seeks to explain the dominance of matter over antimatter in the Universe by generating a net lepton number shortly after the Big Bang is called leptogenesis. Sphaleron transitions, because they conserve only $B - L$, are capable of converting such a net L into a net B such as is present in the universe today. Although there are leptogenesis scenarios in which Dirac neutrinos are capable of producing sufficiently high levels of baryon asymmetry, leptogenesis would benefit from the discovery of Majorana neutrinos, and in particular from the verification of some version of the Seesaw Model, because the fulfillment of the Sakharov conditions arises naturally from most seesaw scenarios [61].

Some of the simplest leptogenesis models depend on the Type I Seesaw Model. The heavy right-handed Majorana neutrinos used to achieve the light neutrino masses would be present in significant quantities in the high temperatures of the early Universe. Majorana neutrinos inherently violate L, and the existence of new neutrinos in general introduces new potential CP-violating phases in the neutrino sector. As the universe cooled and expanded, the heavy neutrinos would undergo L- and CP-violating decays; the expansion of the universe would provide the out-of-equilibrium dynamics in this scenario, so the relationship between the decay rate(s) and the expansion rate determines the specific conditions that would lead to sufficient buildup of net lepton number to generate the baryon asymmetry we see today [61].

Leptogenesis can also be driven by triplet scalar decays, such as are present in the Type II Seesaw Model. Although the introduction of a single triplet scalar is all that is strictly necessary to produce the three light neutrinos, however, the baryon asymmetry produced if a single triplet scalar is the only source of neutrino mass is far too small; some additional source of neutrino mass is necessary, such as the heavy right-handed neutrinos (Type I) or additional triplet scalars (Type II) [61].

Chapter 2

Double-Beta Decay

Neutrinoless double-beta decay ($0\nu\beta\beta$) (see recent reviews [70, 35, 38]) is a rare nuclear process hypothesized to occur if neutrinos are Majorana particles. The observation of $0\nu\beta\beta$ may also probe the absolute mass of the neutrino and the neutrino mass hierarchy.

At this time, the only feasible experimental approach to establishing the Majorana nature of the neutrino is to search for evidence of neutrinoless double beta decay. Other possible avenues of study have been suggested, such as two-particle interferometry [90] and atomic decays into $\nu\bar{\nu}\gamma$ [134], but thus far the practical requirements for these experiments are prohibitively difficult to achieve. Therefore, there is an ongoing program of $0\nu\beta\beta$ experiments throughout the physics community.

2.1 Two-Neutrino Double-Beta Decay

Nuclear beta decay is the familiar process in which a nucleon decays (a proton changes into a neutron or vice versa), releasing a positron (electron) and an electron neutrino (antineutrino). The formulation of a theory explaining the mechanism of beta decay was intimately involved in the history of the neutrino itself (see Section 1.1.1). In 1935, shortly after the publication of Fermi's theory of beta decay, Maria Goeppert-Mayer presented the first derivation of the decay rate of the process called two-neutrino double-beta decay ($2\nu\beta\beta$) [84], in which two nucleons undergo beta decay simultaneously; the Feynman diagram for the process is shown in Figure 2.1. Her calculation provided a quantitative understanding of the metastability of certain isobars, nuclei of different charge Z but the same mass number A, of which only the one with the lowest energy

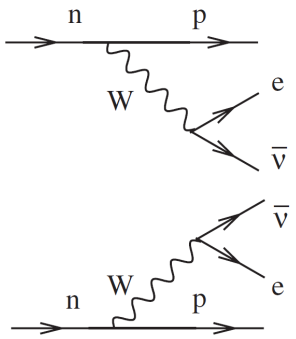


Figure 2.1: The Feynman diagram representing $2\nu\beta\beta$. On the quark level, the interaction that changes a neutron into a proton in fact changes a down quark into an up quark. Figure from [35].

would naïvely be expected to be stable, with all the others eventually decaying to it. The typical candidate for this kind of metastability is an even-even nucleus that, although not the isobar of its A with the lowest energy, still has lower energy than either of its odd-odd neighbors due to the effect of pairing forces [35]. For these nuclei, single-beta decay is energetically forbidden, as the resulting (A,$Z \pm 1$) nucleus would be less bound than the initial (A,Z) nucleus; however, if there exists a (A,$Z \pm 2$) nucleus that is more bound than the initial nucleus, it is possible to observe $2\nu\beta\beta$. One example is illustrated in Figure 2.2. Most nuclei that are considered to be candidates for the observation of $\beta\beta$ decay are specifically candidates for $\beta^-\beta^-$ decay analogous to the pictured case, as the phase space for this mode is more favorable than that for $\beta^+\beta^+$ decay [132].

Of course, an (A,Z) nucleus that is less bound than, e.g., both the (A,$Z+1$) and (A,$Z+2$) nuclei can still undergo $2\nu\beta\beta$ decay, but if the single-beta decay is allowed, the double-beta decay will be so highly suppressed in comparison that it will be difficult or impossible to observe. There is one $\beta\beta$ candidate nucleus, ^{48}Ca, for which the single-beta decay is energetically allowed, but it is instead suppressed due to angular momentum considerations; the ground state of ^{48}Ca has zero spin, while the ground state of its more-bound neighbor, ^{68}Sc, has spin 6, meaning that the single-beta decay would require two spin-$\frac{1}{2}$ particles to carry six units of angular momentum [68].

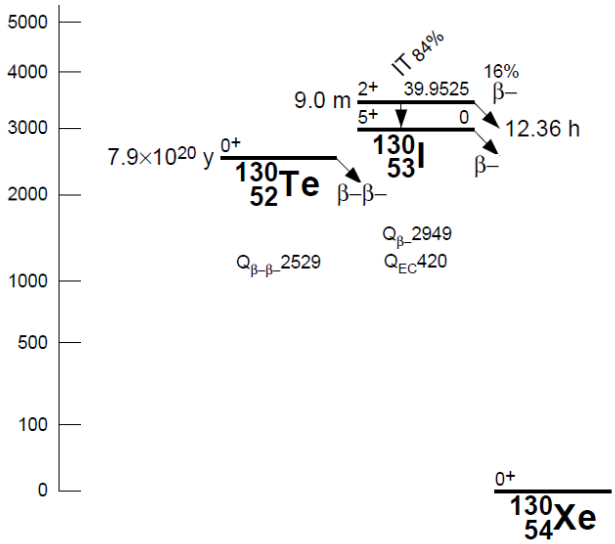


Figure 2.2: Energy level scheme of ^{130}Te $\beta\beta$ decay and the energetically forbidden β decay. The vertical (energy) axis is labeled in units of keV. Figure reduced to the relevant selection from the full A=130 summary figure of [68].

The half-lives for $2\nu\beta\beta$ are long — typically of order $10^{18} - 10^{21}$ years, and some are even longer — because it is a second-order weak process, but this decay has been observed in a number of isotopes [132]. The first empirical evidence for double-beta decay was observed in 1950 as an excess of ^{130}Xe in xenon gas extracted from a geological deposit of Bi_2Te_3, which was attributed to the $\beta^-\beta^-$ decay of ^{130}Te [97]. Such indirect geochemical measurements allow rare decays like $2\nu\beta\beta$ to be probed over very long time scales, but they are subject to various uncertainties due to factors such as how well the environmental conditions and age of the sample are known. The first laboratory measurement of $2\nu\beta\beta$ was achieved for ^{82}Se in 1987 with a time-projection chamber [69], and $2\nu\beta\beta$ has been directly measured in many other nuclei in the ensuing decades [35, 60].

The nuclear recoil is essentially negligible in $\beta\beta$ decay, meaning that the leptons (βs and neutrinos) combined carry the full transition energy of the decay, which is the energy difference between the initial and final nuclei, also called the Q-value of the decay [132]. Because neutrinos interact with matter only weakly, the energy spectrum of $\beta\beta$ decay that is typically measured is the sum energy of the two βs released in the decay. This energy spectrum is a continuous distribution between zero and the Q-value (see the continuum in Figure 2.4) for $2\nu\beta\beta$, as the neutrinos carry off some of the energy.

2.2 Neutrinoless Double-Beta Decay

If neutrinos are Majorana particles, there is the possibility that, some small fraction of the time, these isotopes could undergo $0\nu\beta\beta$ instead, in which the two (anti)neutrinos would disappear in a virtual particle exchange instead of being released as physical particles; the Feynman diagram for this process appears in Figure 2.3. The experimental signature for this process is a sharp peak in the beta energy spectrum at the Q-value of the decay, as the nucleus is so heavy that the recoil is negligible (see the peak in Figure 2.4). As it turns out, there are a number of questions about neutrino mass that $0\nu\beta\beta$ has the potential to address:

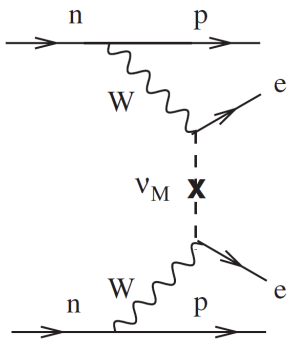


Figure 2.3: The Feynman diagram representing $0\nu\beta\beta$. On the quark level, the interaction that changes a neutron into a proton in fact changes a down quark into an up quark. The line labeled ν_M represents a virtual exchange of a light Majorana neutrino. Figure from [35].

1. Whether neutrinos are Majorana particles — If $0\nu\beta\beta$ is observed, they are, but if it is not observed, information from other types of experiments will be needed and it may still not be possible to tell.

2. Improved constraints on the absolute neutrino mass scale — Oscillation experiments can give the mass-squared differences between neutrino mass states, but not the absolute mass.

3. Neutrino mass hierarchy — Current atmospheric oscillation data cannot give the sign of the associated mass difference, so which mass eigenstates are the heaviest/lightest are unknown; $0\nu\beta\beta$ may or may not be able to determine the mass hierarchy, depending on several factors (see discussion below and Figure 2.6).

The reason that $0\nu\beta\beta$ may be able to provide a handle on one or more of these questions is that, in the case that the exchange of a light Majorana neutrino is the dominant $0\nu\beta\beta$ mechanism, the $0\nu\beta\beta$ decay rate $\Gamma_{0\nu}$ is related to the neutrino mass. To be more specific[1],

$$\Gamma_{0\nu} = \ln(2)\ G^{0\nu}(Q,Z)\ |M^{0\nu}|^2\ \frac{|\langle m_{\beta\beta}\rangle|^2}{m_e^2}, \tag{2.1}$$

[1]The $\frac{1}{m_e^2}$ is a normalization convention common in the nuclear matrix element literature, e.g., [42].

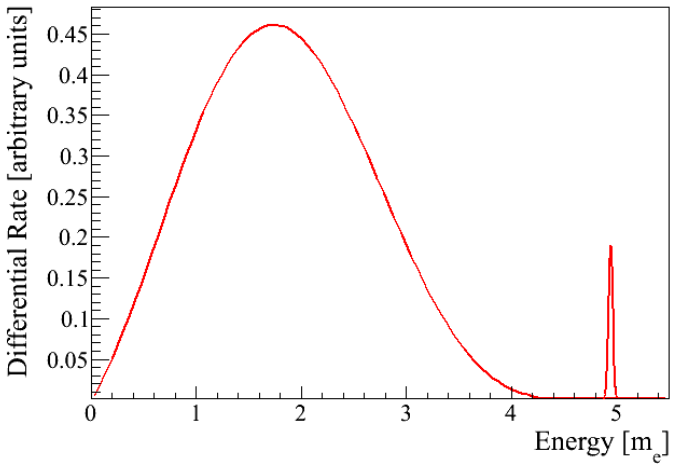


Figure 2.4: A cartoon of ^{130}Te $\beta\beta$ decay, neglecting backgrounds and assuming a smearing of the $0\nu\beta\beta$ peak due to a full-width-at-half-maximum energy resolution of 1% of the Q-value. Energy is given in units of electron mass. The Primakoff-Rosen approximation for the shape of the $2\nu\beta\beta$ continuum is used [117]. For visibility, the spectrum is pictured for a rate ratio $\Gamma_{0\nu}/\Gamma_{2\nu} = 1/100$; current limits in ^{130}Te correspond to $\Gamma_{0\nu}/\Gamma_{2\nu} \approx 1/4000$.

where the effective $0\nu\beta\beta$ neutrino mass is

$$\langle m_{\beta\beta} \rangle = \left| \sum_j |U_{ej}|^2 e^{i\phi_j} m_j \right|. \tag{2.2}$$

In the above equations, $G^{0\nu}(Q, Z)$ is a phase space integral, $M^{0\nu}$ represents nuclear matrix elements (NMEs), m_e is the electron mass, the U_{ej} are elements of the neutrino mixing matrix (see Equation 1.4), the ϕ_j are possible complex Majorana phases, and the m_j are the physical neutrino mass eigenvalues [4]. Experimental results are often expressed in terms of the $0\nu\beta\beta$ half-life $T^{0\nu}_{1/2} = \frac{\ln(2)}{\Gamma_{0\nu}}$, which can be determined from the number of events in the observed signal peak.

The effective Majorana mass of the electron neutrino can thus be inferred from the $0\nu\beta\beta$ half-life as follows [35]:

$$\langle m_{\beta\beta} \rangle = \frac{m_e}{\sqrt{F_N \cdot T^{0\nu}_{1/2}}}, \tag{2.3}$$

where $F_N \equiv G^{0\nu}(Q, Z)\ |M^{0\nu}|^2$ is a nuclear structure factor of merit that includes the NMEs and the phase space of the $0\nu\beta\beta$ transition.

It has long been said that the phase space factor $G^{0\nu}(Q, Z)$ can be 'accurately calculated' for a given nucleus, although in fact some approximations have been made in the calculations presented in the literature. A recent study has improved the phase space calculations by introducing a finite nuclear size into the electron wave functions used for the calculation and taking electron screening into account [108]. The corrections are greater for heavier nuclei; for example, the new calculation for the ^{130}Te $G^{0\nu}(Q, Z)$ differs by nearly 20% from the previous value. The finite nuclear size is implemented by considering an uniform spherical charge distribution of radius $R = r_0 A^{1/3}$ with $r_0 = 1.2$ fm. Based on comparisons to experimental measurements of nuclear radii, the authors estimate an uncertainty of 7% on their $0\nu\beta\beta$ phase space calculations arising from this choice.

By convention, a factor $\propto \frac{1}{R^2}$ also appears in $G^{0\nu}(Q, Z)$ that compensates for a factor $\propto R$ in $M^{0\nu}$, rendering both quantities dimensionless. Conventions for the choice of r_0 have historically differed in the literature; if one should desire to calculate F_N with a $G^{0\nu}(Q, Z)$ and an $M^{0\nu}$ obtained with different values of r_0, the majority of the mismatch can be corrected by simply scaling by this factor [58].

The authors of [108] follow a slightly different notation than that presented in Equation 2.1 above; they include an explicit factor of g_A^4, where g_A is the axial-vector coupling constant governing the β decay, which is often absorbed into $G^{0\nu}(Q, Z)$. The reasoning for this is that the strength of this coupling is known to be quenched to around 70% of its bare-nucleon value in single-β decay and appears to experience a similar quenching in $2\nu\beta\beta$ [42]; its true effective value in $0\nu\beta\beta$ is unknown. Authors typically calculate $0\nu\beta\beta$ NMEs for unquenched g_A, although different authors consider slightly different values (e.g., 1.25, 1.254, or 1.269) extracted from various experimental measurements of the neutron lifetime [48]. Some also evaluate $M^{0\nu}$ for some lower value, such as $g_A = 1.00$, to illustrate the potential effect of quenching. Explicitly writing the factor of g_A^4 simplifies the consideration of different values of g_A.

The reason that phase space factors are considered 'accurately calculable' despite approximations present in the calculations is that far larger theoretical uncertainty prevails with respect to the nuclear matrix elements. Unfortunately, nuclei are complex structures and NMEs are therefore difficult to calculate; a large range of values can be found in the literature, arising from variations in the details of the models and the approximations and assumptions made, though there are indications of an underlying mechanism correlating NMEs and phase space factors that may allow a reduction of this theoretical uncertainty in future [122]. A representative range of NME values can be found by considering recent calculations from five different methods: the quasiparticle random phase approximation (QRPA) (carried out by two different groups: the calculations of the Tübingen group are henceforth denoted by QRPA-T, and the calculations of the Jyväskylä group are henceforth denoted by QRPA-J), the interacting shell model (ISM), the microscopic interacting boson model (IBM), the projected-Hartree-Fock-Bogoliubov model (PHFB), and the generating coordinate method (GCM). For the QRPA-T calculations [73], ranges of values of the NMEs are reported depending on the choice of the input parameters in the model (including g_A; a range of $1.00 \leq g_A \leq 1.254$ is considered for every nuclide except ^{150}Nd, which is treated for a range of $0.94 \leq g_A \leq 1.25$ and is modeled differently because it is a strongly deformed nucleus) and on the model of short-range correlations. Following the expressed preference of the authors, we consider only the coupled-cluster method (CCM) short-range correlations here. The QRPA-J results [130]

use $g_A = 1.25$ and are presented independently for two different models of short-range correlations, the Miller-Spencer Jastrow (M-S Jastrow) correlator and the unitary correlation operator method (UCOM), with the exception of ^{48}Ca, for which only M-S Jastrow NMEs are reported. The ISM [112] values are also reported for $g_A = 1.25$ and for both M-S Jastrow and UCOM short-range correlations. IBM-2 [43] values are reported for $g_A = 1.269$, although a brief consideration of quenching as a function of mass number is presented; the NMEs are evaluated with both M-S Jastrow and CCM short-range correlations. The PHFB [119] authors calculate ranges of NMEs for both $g_A = 1.254$ and $g_A = 1.0$ considering M-S Jastrow and CCM short-range correlations and also considering CCM values only; as in the QRPA-T case, we consider only the CCM ranges here, following the preference of the authors. The GCM [123] calculations are performed for only $g_A = 1.25$ and UCOM short-range correlations. For each calculation (with the exception of GCM, for which only a single NME value is reported), we treat the maximum and minimum reported NME value as defining the range of values into which the true NME value is predicted to fall by that calculation. For some calculation methods (ISM, IBM-2), several single values are provided for different short-range calculations; for others (QRPA-J, QRPA-T, PHFB), calculations for different short-range calculations (and, in the case of QRPA-T, calculations methods) are presented as a central value and an error, in which case we take the maximum value plus error and the minimum value minus error. No statistical meaning is implied in the use of these ranges.

The factor of merit F_N encapsulates the information present in both $G^{0\nu}(Q, Z)$ and $M^{0\nu}$ and serves as an overall indicator of how favorable a given candidate nuclide is for $0\nu\beta\beta$. F_N values obtained from the $G^{0\nu}(Q, Z)$ and NME values discussed here are shown in Table 2.1 for selected $0\nu\beta\beta$ candidate nuclides that are the subject of active experimental efforts discussed elsewhere in this work (see, e.g., Sections 3.2 and 7.3.4). No scaling is necessary to match the phase space factors of [108], as all the NME values considered here are calculated for $r_0 = 1.2$ fm. Due to the differing treatments of the NMEs with regard to g_A and short-range correlations, some of the values are more directly comparable than others; Table 2.1 is subdivided so as to facilitate the direct comparisons that are possible. The full spread of NME values for each nuclide can be regarded as an expression of the overall present state of nuclear matrix element calculations.

Table 2.1: $0\nu\beta\beta$ nuclear factors of merit F_N, as defined in Equation (2.3), for selected candidate $0\nu\beta\beta$ nuclides, according to different evaluation methods and authors. QRPA: quasiparticle random phase approximation; ISM: interacting shell model; IBM: microscopic interacting boson model; PHFB: projected-Hartree-Fock-Bogoliubov model; GCM: generating coordinate method. The phase space values used in calculating F_N values are taken from [108]. The calculation methods are grouped by the g_A values and short-range correlations considered.

	$0\nu\beta\beta$ nuclear factor of merit F_N (10^{-13} y^{-1})					
	$g_A = 1.25$, M-S Jastrow + UCOM		$g_A = 1.25$, UCOM	$1.00 \leq g_A \leq 1.254$, CCM		$g_A = 1.269$, M-S Jastrow + CCM
Isotope	QRPA-J [130]	ISM [112]	GCM [123]	QRPA-T [73]	PHFB [119]	IBM-2 [43]
^{76}Ge	$0.628 - 1.89$	$0.305 - 0.456$	1.22	$1.15 - 3.06$	—	$1.80 - 2.33$
^{130}Te	$3.34 - 9.28$	$1.56 - 2.44$	9.14	$3.56 - 10.6$	$3.10 - 9.11$	$5.99 - 7.84$
^{136}Xe	$1.63 - 4.14$	$1.10 - 1.71$	6.28	$1.37 - 4.09$	—	$4.19 - 5.43$

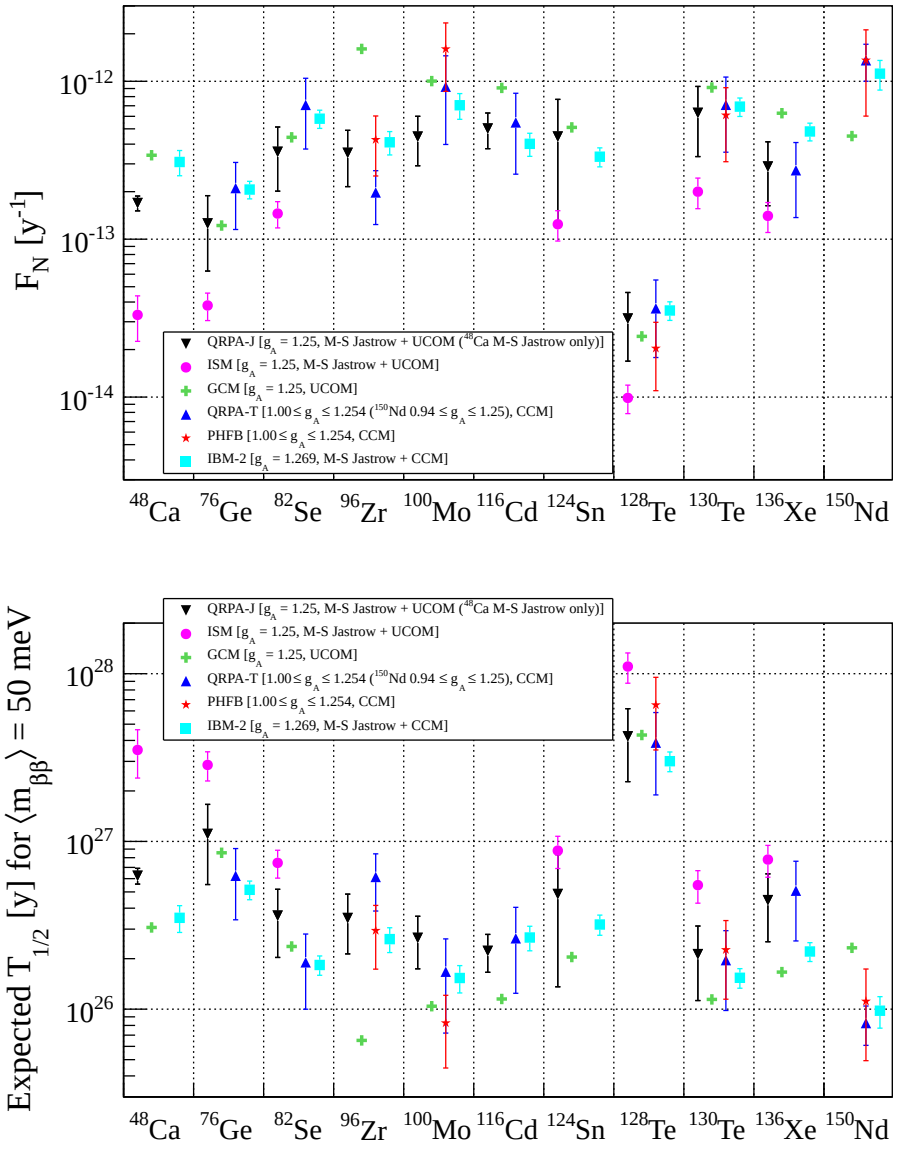


Figure 2.5: Graphical comparison of the $0\nu\beta\beta$ factors of merit (top) and expected $0\nu\beta\beta$ half-lives assuming $\langle m_{\beta\beta} \rangle = 50$ meV (bottom) for many $0\nu\beta\beta$ candidate nuclides. References are the same as for Table 2.1. Calculations are grouped by choice of g_A and short-range correlations. The central mark of each range indicates the mean of the maximum and minimum values of that range and is intended solely to guide the eye. The error bars show the extent of each range.

Figure 2.5 presents a graphical comparison of F_N for a broader selection of candidate $0\nu\beta\beta$ nuclides. It also shows a similar comparison of the corresponding expected half-lives for an effective double-beta mass value of $\langle m_{\beta\beta}\rangle = 50$ meV, on a similar scale to the physics reach expected to be achieved by experiments within the next few years (see Chapter 7). In principle, $0\nu\beta\beta$ should be 'easier' to observe in nuclides with shorter expected half-lives; however, not all detector technologies can currently achieve exactly comparable half-life sensitivities, nor are they all equally appropriate for the investigation of each nuclide.

As can be seen from the wide range of values present in Table 2.1 and Figure 2.5, there is significant theoretical uncertainty in determining the effective mass $\langle m_{\beta\beta}\rangle$ from the decay rate. This limits how precisely the quantity that is actually being measured, $\Gamma_{0\nu}$ or $T^{0\nu}_{1/2}$, can determine the true effective $0\nu\beta\beta$ mass $\langle m_{\beta\beta}\rangle$. Additional difficulty in using $\langle m_{\beta\beta}\rangle$ to set constraints on the physical masses comes from the fact that the ϕ_j are completely unknown [111]. These phases do not contribute to the neutrino oscillations that have been measured and are not directly measurable from $0\nu\beta\beta$, although a comparison between a measurement of $0\nu\beta\beta$ and a measurement of some process that depends on the neutrino masses in a different way, such as direct mass measurements with β decay, could begin to place some constraints on the ϕ_j; in principle, the cleanest measurement of the Majorana phases would come from observing the level of CP violation in a family of processes in which they would induce a CP-odd effect, such as $\nu \leftrightarrow \overline{\nu}$ oscillations, but this does not presently seem feasible to accomplish in laboratory experiments [62].

Nonetheless, it is possible to use the information about the neutrino mixing parameters that has been gained from oscillation experiments to determine the allowed phase space for $\langle m_{\beta\beta}\rangle$ with respect to the overall neutrino mass scale. In Figure 2.6, the allowed values of $\langle m_{\beta\beta}\rangle$ for the best-fit values and the 3σ ranges of the oscillation parameters are plotted as a function of the mass of the lightest neutrino mass state, $m_{lightest}$, and as a function of the sum of the neutrino masses. As can be seen from the figure, the allowed region splits into two distinct bands corresponding to the normal hierarchy and the inverted hierarchy. Most of the degenerate-hierarchy region, in which the two bands overlap, is disfavored by cosmology; this can be seen by comparing Figure 2.6b to cosmological limits, $\Sigma m_j \gtrsim 0.2-3$ eV.

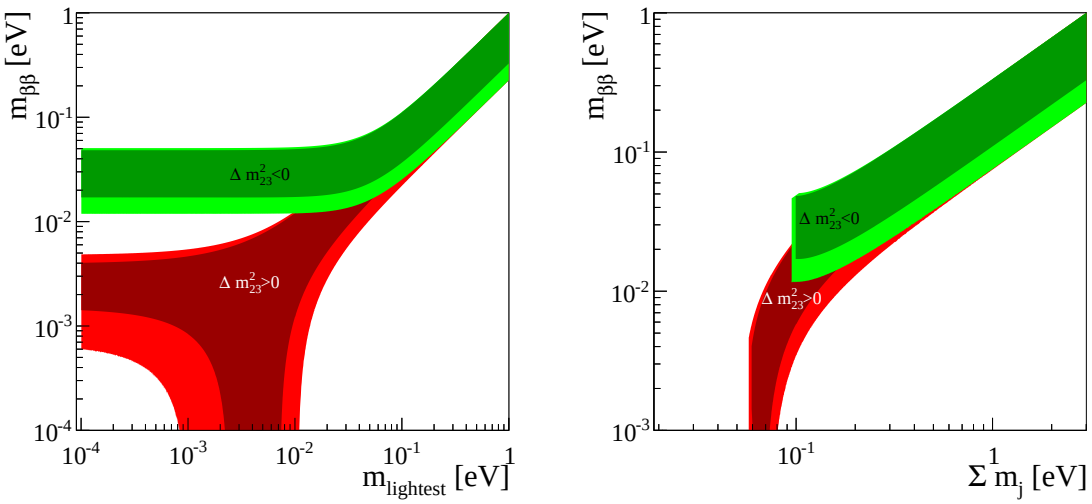


Figure 2.6: Values of the effective $0\nu\beta\beta$ neutrino mass allowed by recent neutrino oscillation data (inner bands represent best-fit data; outer bands represent data allowing 3σ errors) [81]. Both normal ($\Delta m^2_{23} > 0$) and inverted ($\Delta m^2_{23} < 0$) neutrino mass hierarchies are shown. (a) The coordinate plane represents the parameter space of $\langle m_{\beta\beta} \rangle$ and $m_{lightest}$, following the plotting convention of [128]. (b) The coordinate plane represents the parameter space of $\langle m_{\beta\beta} \rangle$ and Σm_j, following the plotting convention of [80]. The region corresponding to $\Sigma m_j \gtrsim 0.2$ eV is disfavored by cosmology.

The separation between the two bands in Figure 2.6 gives $0\nu\beta\beta$ experiments some discriminatory power regarding the mass hierarchy. For example, if $0\nu\beta\beta$ is observed for $\langle m_{\beta\beta}\rangle \gtrsim 10$ meV and direct neutrino mass experiments confirm the cosmological implication that the overall neutrino mass scale is too small to allow the degenerate hierarchy scenario, then neutrinos are Majorana particles and the hierarchy is inverted; or, if experiments successfully probe the entire inverted-hierarchy region without seeing $0\nu\beta\beta$ and the hierarchy is shown to be inverted by some other means, then neutrinos are *not* Majorana particles. More generally, if $0\nu\beta\beta$ is observed at all, the corresponding $\langle m_{\beta\beta}\rangle$ will determine a range within which the neutrino mass scale must fall. Direct neutrino mass experiments are complementary, as a direct measurement of the neutrino mass scale would in turn determine a range into which $0\nu\beta\beta$ must fall; the only conclusive proof that neutrinos are *not* Majorana particles would require a direct mass measurement to show that the only remaining allowed range of $\langle m_{\beta\beta}\rangle$ has been fully explored by $0\nu\beta\beta$ experiments.

Unfortunately, as Figure 2.6 illustrates, in the normal hierarchy case, it is possible for a cancellation to occur in the superposition described by Equation (2.2), causing the value of $\langle m_{\beta\beta}\rangle$ to vanish; this cancellation depends on the values of the phases ϕ_j, which are presently entirely unknown. It therefore may be impossible to observe $0\nu\beta\beta$ even if neutrinos are Majorana particles.

In the interest of completeness, it is important to note that the above discussion applies *only* in the simple case that $0\nu\beta\beta$ is dominantly mediated by the exchange of a virtual light Majorana neutrino (which will be assumed for the rest of this paper). If some other lepton-number-violating physics is responsible (e.g., exchange of a heavy Majorana neutrino or supersymmetric particles), it becomes much more difficult if not impossible to extract $\langle m_{\beta\beta}\rangle$ [5]. However, the Schechter-Valle Theorem states that no matter what mechanism is responsible, if $0\nu\beta\beta$ is observed then neutrinos are Majorana particles [125].

Chapter 3

Experimental $0\nu\beta\beta$ Situation

As $0\nu\beta\beta$, assuming it happens at all, is such a rare process — even more so than $2\nu\beta\beta$, which, as discussed previously, has quite a long half life to begin with — the attempt to observe it experimentally poses a challenge. One way to characterize the sensitivity of an experiment (see Chapter 7 for a more rigorous treatment of sensitivity) is to introduce a figure of merit F, defined as the ratio of the number of signal events to the Poisson fluctuation of the background:

$$F = \Gamma_{0\nu} \cdot \eta \cdot \varepsilon \cdot \sqrt{\frac{N \cdot t}{\Delta E \cdot \mathcal{B}}}, \tag{3.1}$$

where η is the stoichiometric coefficient of the $0\nu\beta\beta$ candidate (i.e., the number of nuclei of the candidate $0\nu\beta\beta$ element per molecule of the active mass), ε is the detector efficiency, N is the total number of nuclei (i.e., detector mass), t is the live time of the experiment, ΔE is the energy resolution of the detector, and $\mathcal{B}$ is the constant background rate per atom per energy interval [129]. Thus a successful $0\nu\beta\beta$ experiment will need to be a long-running, low-background experiment with good resolution and a large-mass detector. Backgrounds are of critical concern because other events in the same energy region could easily obscure the signal of interest, while good detector resolution is essential in order to distinguish the $0\nu\beta\beta$ peak from the tail of the irreducible $2\nu\beta\beta$ signal.

There are several techniques pursued by various groups to fulfill these requirements. The energy is the most crucial value that must be measured to identify an event as a $0\nu\beta\beta$ candidate, and various approaches can be followed to obtain an energy signal. Depending upon the characteristics of the detector, the ionization charge or scintillation light produced by the interaction of the βs in the detector can be collected, or the temperature rise due to the lattice vibrations induced by

the interaction can be measured; all of these signals are in some way proportional to the energy deposition of the event. These and other approaches form the bases of a number of past and future experiments. Some detectors can collect more than one of these signal types and use the differing responses to identify the type of particle that caused the event, providing a handle on background rejection.

Beyond this, detectors used for the search for $0\nu\beta\beta$ can be broadly categorized into two classes: those with the detector apparatus itself separate from the radioactive source material under investigation, and those with the radioactive source material embedded in the material from which the detector is constructed. The majority of $0\nu\beta\beta$ experiments follow the latter approach because it offers several advantages particularly well-suited to the application. It is often possible to achieve superior energy resolution in a configuration in which the source is embedded in the detector; β particles tend to travel only short distances in the detector material before interacting, meaning that detection efficiency is typically very high; and in general, a greater mass of source material can be investigated if it also constitutes part of the mass of the detector itself, especially considering that sources must be very thin in configurations in which the detector is a separate apparatus in order to allow the majority of the βs to escape the source and reach the detector. On the other hand, the requirement that the source material be embedded in the detector material restricts the particular source isotopes that are suitable for use with a given detector technology, and even for technologies that can accommodate several different $0\nu\beta\beta$ candidates, an entirely new detector must be built out of each one instead of using the same detector to investigate several different sources.

One notable class of $0\nu\beta\beta$ detectors consists of arrays of germanium diodes. These solid-state detectors collect the ionization charge produced by energy depositions in the detectors, and while the natural abundance of the $0\nu\beta\beta$ candidate ^{76}Ge is only 7.76% [46], germanium semiconductor crystals suitable for operation as diodes can be grown from enriched germanium to achieve a considerable source mass. Some past $0\nu\beta\beta$ experiments of this kind were Heidelberg-Moscow [104] and IGEX [6].

Some types of crystals can be operated as bolometers, which are calorimetric detectors that, when instrumented with temperature sensors and maintained at cryogenic temperatures, produce a

clean temperature pulse associated with the heat rise induced in the crystal by the energy deposited by a particle interaction. The total energy of the event can be determined from the amplitude of the observed pulse. Arrays of TeO_2 crystals have been operated as bolometers to investigate the $0\nu\beta\beta$ candidate ^{130}Te, as in the Cuoricino experiment [21].

Detectors based on the collection of scintillation light often consist of a volume of liquid scintillator instrumented with photomultiplier tubes. The DAMA collaboration has used liquid xenon scintillator to investigate the $0\nu\beta\beta$ candidate ^{136}Xe [49]. Other types of scintillator can also be loaded with some $0\nu\beta\beta$ candidate materials without unacceptably compromising the light yield of the scintillator; for example, the KamLAND-Zen collaboration has loaded scintillator consisting of a mixture of decane, pseudocumene, and PPO (the fluor) with enriched xenon gas, also to investigate ^{136}Xe [83] (see also Section 3.2.2), while the SNO+ collaboration originally intend to load linear-alkylbenzene-and-PPO scintillator with neodymium metal to search for $0\nu\beta\beta$ in ^{150}Nd [93] and now plans to use tellurium metal instead [1] (see also Section 3.2.3).

Scintillator-based measurements often have far worse energy resolution than can typically be achieved with ionization-based or bolometric measurements. In detectors that can obtain a scintillation signal *in addition to* an ionization signal or a thermal signal, however, the ratio of the strengths of the two signals observed for a given particle interaction event depends upon the mass and charge of the particle. This provides at least a partial discrimination capability between events caused by βs and background events such as α interactions. For the study of ^{136}Xe, the EXO and NEXT collaborations are pursuing xenon time projection chamber technology (liquid xenon in the case of EXO; high-pressure xenon gas in the case of NEXT), which allows the collection of both ionization and scintillation signals; see Sections 3.2.4 and 3.2.5. A number of crystals that can be operated as bolometers also scintillate; several such crystals that contain $0\nu\beta\beta$ candidates are under investigation for use in possible future experiments, including ZnSe (^{82}Se), $CdWO_4$ (^{116}Cd), and $ZnMoO_4$ (^{100}Mo) [56].

In contrast to the above-mentioned experiments, all of which use detector configurations in which the $0\nu\beta\beta$ source is embedded in the detector, the NEMO collaboration has pursued an extensive experimental program in which thin source foils are mounted inside an independent

detector apparatus. The NEMO 3 detector was used to measure $2\nu\beta\beta$ decay and search for $0\nu\beta\beta$ in seven different isotopes throughout the experiment's lifetime: ^{100}Mo, ^{82}Se, ^{130}Te, ^{116}Cd, ^{96}Zr, ^{48}Ca, and ^{150}Nd [60]. The detector consisted of a gaseous tracking volume comprising many individual drift cells, which collected ionization charge for the sole purpose of gathering position and tracking information, surrounded by a plastic scintillator calorimeter for the energy measurement [31]. The energy resolution of this design is not competitive with the semiconductor or bolometric detectors, but the tracking capability allows significant background rejection by analysis of the kinematics of the electrons, to the point that the radioactive background can be reduced to nearly zero, leaving the ability to resolve the $0\nu\beta\beta$ peak from the irreducible $2\nu\beta\beta$ continuum the sole limiting factor on the experimental sensitivity to $0\nu\beta\beta$. The flexibility in the choice of source isotopes that can be investigated by the detector is also a significant advantage.

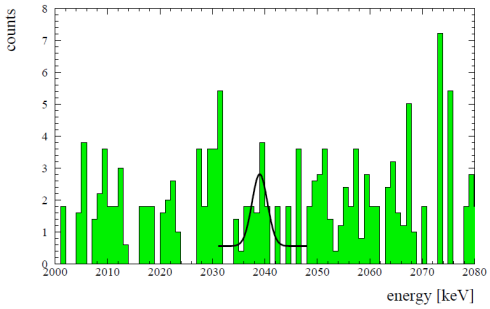
Until recently, experiments using ^{76}Ge (Heidelberg-Moscow, IGEX) and ^{130}Te (Cuoricino) held the best limits on $\langle m_\nu \rangle$ [129]. Results from EXO-200 (see Section 3.2.4) and KamLAND-Zen (see Section 3.2.2) have now been published that, when statistically combined, supercede these previous best limits in terms of $\langle m_\nu \rangle$ [72].

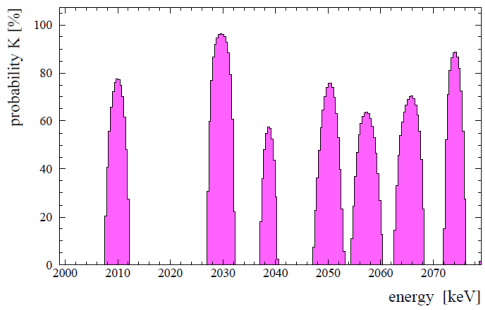
3.1 Klapdor-Kleingrothaus Claim of Observation

While various experiments have set limits on $0\nu\beta\beta$ in a number of different nuclei, one claim of observation of evidence of the decay in ^{76}Ge corresponding to the degenerate mass hierarchy region in Figure 2.6 has been reported by a subset of the Heidelberg-Moscow collaboration led by Hans Volker Klapdor-Kleingrothaus. Two updates to the initial results have been published, the first adding the analysis of additional data and the second utilizing a new analysis approach; spectra from each iteration are shown in Figure 3.1. This claim has stood for roughly a decade without being directly refuted, although a number of criticisms of the strength of the claim have been raised by the physics community; present $0\nu\beta\beta$ experiments are now beginning to achieve sensitivities that bring them into conflict with the claim.

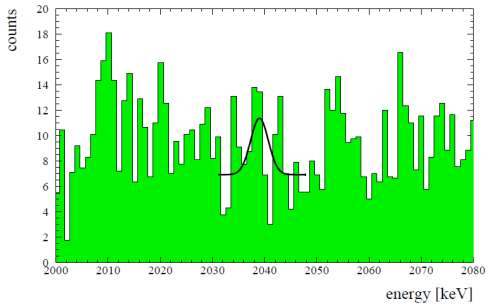
The claim was first published for the data collected from August 1990 to May 2000 by the Heidelberg-Moscow germanium diode $0\nu\beta\beta$ experiment, at which time evidence at $2-3\,\sigma$ significance was reported for $0\nu\beta\beta$ with a half-life of $1.5^{+16.8}_{-0.7} \times 10^{25}$ y (95% C.L. range) [102]. The analysis was predicated on a Bayesian peak search performed on the energy spectrum collected by the detectors; an scan over energy was performed in the $2000-2080$ keV region, determining the Bayesian probability distribution of the number of counts in a Gaussian peak with the known detector resolution at the considered energy superimposed on a constant background, and subsequently determining the probability of a line with non-zero contents at each energy for which the maximum of the probability distribution was non-zero. Using this procedure, evidence of several lines was obtained in the energy region, including a line near 2039 keV, the $\beta\beta$ Q-value in ^{76}Ge. A similar peak search was performed on the energy spectrum collected between November 1995 and May 2000 with only the three best-performing detectors and constructed only of events that passed a pulse-shape cut intended to discriminate highly localized single-site events in the detectors (characteristic of $\beta\beta$ events due to the relatively short scattering length of electrons and positrons) from multiple-site events (characteristic of the multiple Compton scatters that gammas are prone to undergo before being absorbed) [94] to verify that the 2039-keV peak, which should in fact be dominated by single-site events, was still present in this spectrum of single-site events (Figure 3.1a); however, the reported measurement was obtained from the full spectrum (Figure 3.1b). To extract the final result, the energy window was narrowed to include only the 2039 keV line so that the other lines that were found would not cause an overestimation of the background rate in the single-peak-plus-constant-background model.

A number of criticisms were levied at the claim shortly after its publication. Some dissatisfaction was expressed with regard to the peak-search method, due to the fact that it accounts for the situation in which no peak is present by embedding the possibility of a peak with zero amplitude into the peak-plus-background model instead of explicitly comparing an hypothesis in which a peak is present to a null hypothesis in which no peak is present [5]. The strongest criticism, however, regards the identification of the peaks that were found. The authors of [102] attributed several of these peaks to weak gamma lines from the decay of ^{214}Bi, which is typically present at some

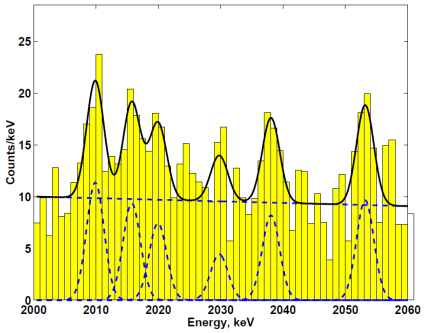




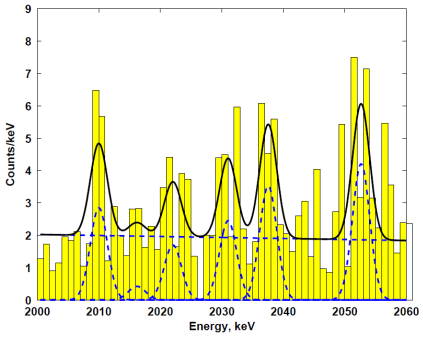
(a) The pulse-shape-selected spectrum published in 2001 and the results of the Bayesian peak search used to motivate the identification of the structure at 2039 keV as the $0\nu\beta\beta$ peak [102].



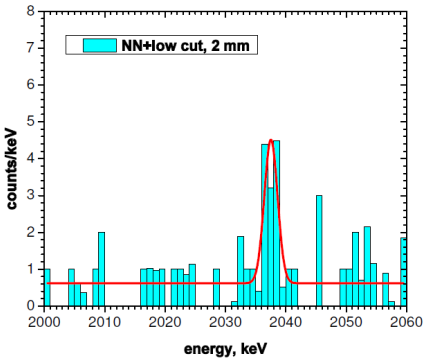
(b) The full spectrum published in 2001 from which the result was obtained [102].



(c) The full spectrum published in 2004 from which the result was obtained [103].



(d) The pulse-shape-selected spectrum published in 2004 used to motivate the identification of the $0\nu\beta\beta$ peak [103].



(e) The pulse-shape-selected and 'fiducial-volume'-selected spectrum published in 2006 from which the final result was obtained [105].

Figure 3.1: The development of the Klapdor-Kleingrothaus claim of positive results from the Heidelberg-Moscow search for $0\nu\beta\beta$ in ^{76}Ge.

level in any experimental setup due to contamination from naturally occurring decay chains; however, several other peaks (around 2030 keV and above $\sim$ 2060 keV) whose origins are unidentified appear with greater significance than the 2039-keV peak, and even those peaks ascribed to ^{214}Bi do not appear with relative strengths fully consistent with the expectations from simulations that account for both relative branching ratios and detector effects [5, 92]. Considering the unclear origin of the background lines, the decision to narrow the window to exclude them before extracting the final result is also problematic. An independent analysis of the published data performed a fit to the full $2000-2080$ keV window with a background model that included the four identified ^{214}Bi peaks with relative amplitudes set by their known branching ratios and found that the significance of the peak at 2039 keV was 1.5σ at most [75]. The same analysis also pointed out that, although the single-site event selection should be expected to induce a far greater suppression of the gamma lines than of the 2039-keV line if the latter is indeed due to $0\nu\beta\beta$, the observed suppression factor on the 2039-keV is in fact only slightly smaller than the observed suppression factor on the gamma lines, casting the origin of the 2039-keV line into further doubt.

The Heidelberg-Moscow detectors continued to operate through May 2003. After the full dataset was analyzed, the claim of observation of $0\nu\beta\beta$ was reiterated, now at a significance of 4.2σ for a half-life of $1.19^{+2.99}_{-0.90} \times 10^{25}$ y (3σ range) [103]. As in the previous publication, a peak-search scan in energy was performed in an energy window around the Q-value of the $0\nu\beta\beta$ decay on the full spectrum as well as several subsets of the data obtained with various data quality selections and pulse-shape analysis to confirm the presence and relative strengths of the located peaks in each spectrum, and the final reported result was obtained from the full spectrum; see Figures 3.1c and 3.1d.

Some of the earlier criticisms were at least partially addressed. Instead of considering a narrow energy window containing only the 2039-keV line, the fit used to extract the $0\nu\beta\beta$ result was performed over the $2000-2060$ keV energy window and included the ^{214}Bi peaks as well as a structure at $\sim$ 2030 keV superimposed on a linear background with a fixed slope estimated by simulations. A possible origin for the 2030 keV line was suggested to be electron conversion from the 2118-keV gamma line from ^{214}Bi, though the observed intensity was just barely consistent with

this hypothesis. Similarly, the intensities of two of the lines attributed to ^{214}Bi gammas remained consistent only within 2.5σ with expectations from simulations. The $2060-2080$ keV energy range, which contained unidentified peak structures in the previous publication, was not considered in this analysis and, indeed, was not shown in detail. A study of the peak suppression from the new pulse-shape analysis methods, which now included a neural-network approach trained on double-escape lines (expected to be primarily single-site events) and the strong ^{228}Th gamma line at 2614.5 keV (multiple-site events), showed much stronger suppression of gamma lines (reduced to around $15-30\%$ of the full-spectrum intensity) than of the 2039 keV line (reduced to around 62% of the full-spectrum intensity), though the suppression was still greater than expected for $0\nu\beta\beta$ events, of which approximately 5% are expected from simulations to manifest as multiple-site events.

Skepticism toward the $0\nu\beta\beta$ claim persisted in the physics community. Some of the authors of [75] repeated their independent analysis with the newly published data, now using a quasi-flat background shape in accordance with the approach in [103] and allowing the intensities of the ^{214}Bi lines to vary freely in deference to detector effects that they could not straightforwardly replicate; they found the significance of the 2039-keV line to be only 2.7σ, reducing to 2.2σ if the previously published data between 2060 and 2100 keV was scaled up to be consistent with the full detector exposure and included in the fit [129]. A contemporaneous publication from the Heidelberg-Moscow collaboration, presenting results for the data collected from November 1995 through August 2001 (after which the full collaboration was no longer involved in detector operations), suggested the presence of spurious peaks throughout the spectrum with intensities that varied arbitrarily between detectors and data samples; these peaks were found to be correlated with non-physical pulses of unknown origin appearing below the discriminator threshold in certain runs, and when this data was removed from the analysis, the evidence for a peak at 2039 keV was found to disappear [36]. However, the authors of the $0\nu\beta\beta$ claim maintained that no such effect that would suggest a non-physical origin for the 2039-keV peak was present in the data that they analyzed [103].

The most recent update of the claim of $0\nu\beta\beta$ observation in ^{76}Ge was the result of a reanalysis of the data considered in [103] conducted by two of the authors of that paper. They used a more highly selective adaptation of the neural-network approach to single-site event identification, presented in a preliminary manner in [103], in combination with a pulse shape library compiled from Monte Carlo simulations performed under the assumption that the entire energy of each event is deposited in a single point in the detector; moreover, they claimed that they were able to spatially localize the position of each event in the detector. The combination of the two event selection methods proved to nearly eliminate the background gamma peaks, while the observed structure near 2039 keV survived, though its position shifted downward somewhat to 2037.5 ± 0.5(stat) $\pm$ 1.2(syst) keV (Figure 3.1e). An additional 'fiducial volume' cut removing events identified as occurring within 2 mm of the surface of the detector yielded a peak significance of 6.8σ and a $0\nu\beta\beta$ half-life measurement of $2.23^{+0.44}_{-0.31} \times 10^{25}$ y [105]. It has been suggested that the behavior under the various pulse-shape selections of the ~ 2039-keV structure observed in the Heidelberg-Moscow detectors can be understood in terms of two small sum peaks at 2035.5 keV and 2039.5 keV that could not be resolved separately from the $0\nu\beta\beta$ line in the full spectrum but were rejected by pulse-shape selection [101], lending some support to the hypothesis of a $0\nu\beta\beta$ origin for this structure; nevertheless, the author of [101] observes that the pulse-shape analysis is presented in [103] in insufficient detail to truly allow an evaluation of the validity of the claim.

Two ^{136}Xe experiments, EXO-200 (see Section 3.2.4) and KamLAND-Zen (see Section 3.2.2), have already published results that bring them into conflict with the claim. When comparing $0\nu\beta\beta$ results in different nuclei, it is necessary to consider the spread of the nuclear matrix elements and the correlations between the NME calculations for the two nuclei. When this is done for the half-life range claimed in [105], it is found that the combined results of EXO-200 and KamLAND-Zen are in tension with the majority of the range into which the $0\nu\beta\beta$ half-life would be expected to fall in ^{136}Xe to be consistent with the ^{76}Ge results [72]. A Bayesian reanalysis of the pulse-shape-selected spectrum analyzed in [105] using an explicit comparison of an hypothesis in which a peak is present to a null hypothesis in which no peak is present finds a weaker significance for the signal but also a broader preferred half-life range; a Bayesian evaluation of the compatibility

of this result with the ^{136}Xe results similarly accounting for NME correlations finds a probability of compatibility of only $1.3-2.5\%$ for equal prior probabilities, and there is no evidence for $0\nu\beta\beta$ if the ^{136}Xe and ^{76}Ge results are nevertheless combined [47]. Section 7.3.2 contains some further discussion of experimental sensitivity to the Klapdor-Kleingrothaus claim.

While the implications of the claim of observation of $0\nu\beta\beta$ would be exciting and far-reaching if unambiguously confirmed, a complementary, independent measurement is necessary before the discovery can be regarded as fully credible. It is particularly important to test the claim using a different $\beta\beta$ isotope, with different systematic concerns and nuclear matrix element calculations. Thus the goal of all next-generation $0\nu\beta\beta$ experiments is twofold: to test the Klapdor-Kleingrothaus claim and to extend sensitivity into the inverted-hierarchy region.

3.2 Major Current and Upcoming Experimental Efforts

CUORE, the successor to Cuoricino, will investigate ^{130}Te as its predecessor did. In addition, several other isotopes will be the subjects of upcoming $\beta\beta$ experiments now in various stages of planning and construction. These experiments span a range of detector technologies appropriate to the various isotopes under investigation; each carries different advantages and disadvantages.

Brief descriptions of a number of the major current and upcoming $0\nu\beta\beta$ experiments follow. A self-consistent comparison of the sensitivities of most of these experiments is presented in Section 7.3.4.

3.2.1 GERDA/Majorana

GERDA and Majorana are two experimental programs developing the capabilities of high-purity germanium semiconductor detectors, which read out ionization signals to obtain energy spectra, to study the $0\nu\beta\beta$ decay of ^{76}Ge. Because the natural abundance of ^{76}Ge is low, the germanium must be enriched before it is made into diodes to obtain sufficient active mass. Germanium detectors must be operated at cryogenic temperatures to keep electrical noise low; however, when this is done, they can achieve very good resolutions of a few keV at the Q-value of ^{76}Ge. The GERDA and Majorana collaborations are following two different paths toward developing

background suppression techniques, and once both detectors have operated and demonstrated their capabilities at total operating masses of $30-40$ kg, the two collaborations intend to combine their expertise to construct a single tonne-scale experiment.

Phase I of the GERDA experiment has been operating at the Laboratori Nazionali del Gran Sasso since November 2011. Several bare, refurbished detectors from previous germanium experiments are immersed in liquid argon, which is in turn surrounded by a volume of ultra-pure water instrumented with photomultiplier tubes. This setup serves the combined purposes of cooling the detectors, providing a Cherenkov muon veto, and shielding the detectors from environmental neutrons. The experiment has now collected sufficient statistics to allow the measurement of the ^{76}Ge $2\nu\beta\beta$ half-life [11]. Phase II will more than double the active mass with the addition of new segmented germanium detectors; the segmented design allows for greater background reduction via the identification of multi-site background events from multiple Compton scattering [8]. Phase II data-taking is anticipated for early 2013.

Majorana will utilize point-contact germanium detectors, which have an intrinsically high probability of detecting individual signals in the case of a multiple-site interaction even without segmentation [40]. The detectors will be hosted in a vacuum cryostat with a shield composed of lead, copper, and scintillator. The experiment will follow a phased schedule similar to that of GERDA: a small prototype cryostat, set to contain two strings of natural-germanium detectors, is near completion; the first full cryostat, containing the first enriched detectors, is expected to be deployed towards the end of 2013; and the second cryostat, containing only enriched detectors, will follow approximately one year later [76].

3.2.2 KamLAND-Zen

The KamLAND-Zen experiment uses the KamLAND liquid-scintillator detector, which was originally built as a reactor-neutrino oscillation experiment. A 3-m-diameter mini-balloon filled with xenon-loaded liquid scintillator has been suspended inside the main liquid scintillator balloon of the original detector; the xenon is enriched in the $0\nu\beta\beta$ candidate isotope, ^{136}Xe. As it exploits

an existing large low-background detector, KamLAND-Zen was able to achieve a quick and profitable start; it accumulated enough data to measure $2\nu\beta\beta$ decay in ^{136}Xe and set improved limits on its $0\nu\beta\beta$ decay within only a few months after its September 2011 start [83, 82]. Although this technology is suitable for investigating a large mass of isotope with relative ease, it is limited by its comparatively poor energy resolution: ~ 250 keV at the ^{136}Xe Q-value.

It appears that there is a significant ^{110m}Ag contamination in the xenon-loaded liquid scintillator, which may have its origins in the March 2011 Fukushima reactor accident. The liquid scintillator was extracted for purification in June 2012 to attempt to remove this contamination. Data-taking is expected to resume some time during summer 2013 [99]. The installation of a larger, cleaner inner balloon containing approximately twice as much xenon as the current inner balloon is foreseen as a future upgrade to the experiment.

3.2.3 SNO+

Like KamLAND-Zen, SNO+ [93] is a scintillator-based adaptation of a large existing neutrino detector, in this case the Sudbury Neutrino Observatory. This detector was originally built as a water Cherenkov detector, but the SNO+ upgrade will completely replace the heavy water in the 6-m-radius inner acrylic vessel with liquid scintillator. The scintillator will be loaded with natural tellurium to allow the investigation of the ^{130}Te $0\nu\beta\beta$ decay [1].

Following an initial phase of data-taking with the detector filled with light water, the water will be gradually replaced with purified scintillator over a period of a few months [93]. The addition of the tellurium and should follow shortly thereafter; the start of $0\nu\beta\beta$ data taking can be anticipated for near the end of 2013. A further upgrade to the experiment with an order-of-magnitude increase in the level of tellurium loading is under consideration [1].

3.2.4 EXO-200

The EXO-200 experiment uses a time projection chamber filled with liquid xenon enriched to 80.6% in $0\nu\beta\beta$ candidate ^{136}Xe. This technology can achieve an energy resolution of around 100 keV, closer to that of large scintillation detectors than to that of bolometers or germanium

diodes, and it collects both ionization and scintillation signals, providing event identification capabilities that allow the rejection of multiple-site gamma interactions. The xenon itself is used for shielding, and the analysis is performed with a fiducial volume cut.

EXO-200 began collecting data in May 2011 and, within a bit over one month, had measured the $2\nu\beta\beta$ spectrum [9]. Following this, the xenon flow in the chamber was tuned to improve the electron lifetime and thereby reduce the loss of charge affecting the ionization signal over the drift length. An analysis of data collected under the new operating conditions obtained a limit on the $0\nu\beta\beta$ decay of ^{136}Xe that rejects the Klapdor-Kleingrothaus discovery claim at 90% C.L. for most NME calculations and at 68% C.L. for all NME calculations [34]. Data collection is expected to continue for several years; eventually, a ton-scale experiment utilizing the same experimental approach is foreseen.

3.2.5 NEXT-100

Like the EXO program, the NEXT program will use xenon time projection chambers that collect both scintillation and ionization signals to investigate the $0\nu\beta\beta$ decay of ^{136}Xe. However, NEXT will use high-pressure xenon gas instead of liquid xenon, which may allow an improvement in resolution of up to an order of magnitude over EXO. NEXT will also be capable of full event topology tracking and will be surrounded by lead and copper shields instead of using part of the detector volume for self-shielding, in the hope of achieving a fully active detector volume.

A prototype called NEXT-DEMO [19] is operating and has already demonstrated the claimed tracking and resolution capabilities of the technology. NEXT-100, an experiment on a similar physical scale to EXO-200, will soon enter the construction phase and is expected to begin to take data in mid-2014 [87], paving the way for an eventual ton-scale experiment.

3.2.6 SuperNEMO

SuperNEMO is the only experiment presented here for which the source is not embedded in the material of the detector itself. Instead, the detector consists of a gas tracking chamber surrounded by a calorimeter; source films thin enough to allow the βs to escape from the source

and pass through the detector are mounted inside the tracking volume. A great advantage of this setup over experiments in which the detector itself is built of the source material is that different source isotopes can be investigated over its lifetime, including isotopes that are not suitable to be manufactured into other kinds of detectors, by simply replacing the films, as indeed was done in NEMO 3, the predecessor of SuperNEMO. Another advantage is that the event topology reconstruction allows excellent background rejection, such that the experiment is essentially operating in the zero-background limit. However, the energy resolution of the calorimeter is relatively poor, comparable to that of the large scintillator detectors, so the $2\nu\beta\beta$ decay still constitutes a significant background to $0\nu\beta\beta$. The thin-film source requirement also somewhat restricts the scalability to large masses, and the physical efficiency is relatively low compared to detectors in which the source is embedded of the bulk of the detector.

A demonstrator module for SuperNEMO is now under construction. It will contain $5-7$ kg of source and constitute a test of the new planar geometry of the $\sim$20-module SuperNEMO detector (as opposed to the cylindrical construction of NEMO3), which is anticipated to allow improvement on the resolution and efficiency achieved in NEMO3 and to facilitate modular scaling to larger masses [39]. Data-taking with the demonstrator module is anticipated for 2014 [115], at which time construction on the full SuperNEMO will proceed in parallel. The baseline $0\nu\beta\beta$ candidate isotope choice is ^{82}Se, although the option to use other isotopes still remains [39].

Chapter 4

The CUORE Family: Cuoricino, CUORE-0, and CUORE

4.1 Detector Technology

In 1984, two independent publications appeared that signalled the beginning of two extensive and successful experimental programs based on thermal detectors that are operated at cryogenic temperatures and use a temperature sensor to measure the phonons induced by an energy deposition in an absorber. One of these urged the use of these detectors, called bolometers, in X-ray spectroscopy [114]; the other proposed that such detectors could allow $0\nu\beta\beta$ experiments in which the source was embedded in the detector material for a wider range of candidate nuclides, as up to that point only experiments with ^{76}Ge (in lithium-doped germanium diode detectors) and ^{136}Xe (in xenon time-projection chambers) had been able to take advantage of this kind of detector configuration [77].

The first results of a search for $0\nu\beta\beta$ in ^{130}Te using a TeO_2 crystal operated in a cryostat located in an underground laboratory to provide the shielding from cosmic rays necessary to achieve the requisite low backgrounds were published in 1992 [12]. Since that time, a series of experiments has continued to expand the capabilities of the technology with larger arrays of crystals [13, 14, 25, 23], leading up to the upcoming ton-scale Cryogenic Underground Observatory for Rare Events (CUORE). These constitute the CUORE family of $0\nu\beta\beta$ experiments.

4.1.1 Tellurium-130

There are a number of nuclides that are suitable experimental candidates for the observation of $0\nu\beta\beta$ decay. Several are the subject of one or more experimental efforts at present, including ^{76}Ge, ^{136}Xe, and ^{130}Te (see Chapter 3).

CUORE-family experiments use ^{130}Te as their source, which has several advantages. Tellurium contains a relatively high natural isotopic abundance — 34.167% [74] — of the $0\nu\beta\beta$ candidate, which means that enrichment is not necessary to achieve a reasonably large active mass. Also, the Q-value of the decay is $\sim$ 2528 keV [120, 126, 118], which falls between the peak and the Compton edge of the 2614.5-keV gamma line of ^{208}Tl, the highest-energy gamma from the natural uranium and thorium decay chains, which constitute the majority of the detector background; this leaves a relatively clean window in which to look for the signal. Another advantage of this fairly high Q-value is that the phase space factor $G^{0\nu}(Q, Z)$ introduced in Equation (2.2) is proportional to Q^5 [132], so a high Q-value leads to a favorable nuclear structure factor of merit F_N for $0\nu\beta\beta$ in ^{130}Te (see Table 2.1).

Finally, tellurium can be grown as a TeO_2 crystal that is suitable to be operated as a bolometer, serving as source and detector at the same time, as discussed in the following section.

4.1.2 Bolometers and Thermistors

The design of CUORE-family detectors is based on the bolometric principle (see Figure 4.1). The apparatus consists of an array of dielectric and diamagnetic TeO_2 crystals, so that the detector itself contains the source (^{130}Te) that it is studying. There are three basic elements to each detector: the crystal, which serves as an absorber for the energy deposition of each measured event; a temperature sensor glued to the face of the crystal, which measures the temperature changes of the crystal due to the lattice vibrations induced by the energy deposition of an event; and a heat bath to which the absorber and temperature sensor are connected via a thermal link, allowing the temperature of the crystal to decrease back to its base temperature following each event in preparation for another measurement. Such detectors combine excellent energy resolution with low intrinsic background, and they have been operated in stable conditions underground for several years [12, 13, 14].

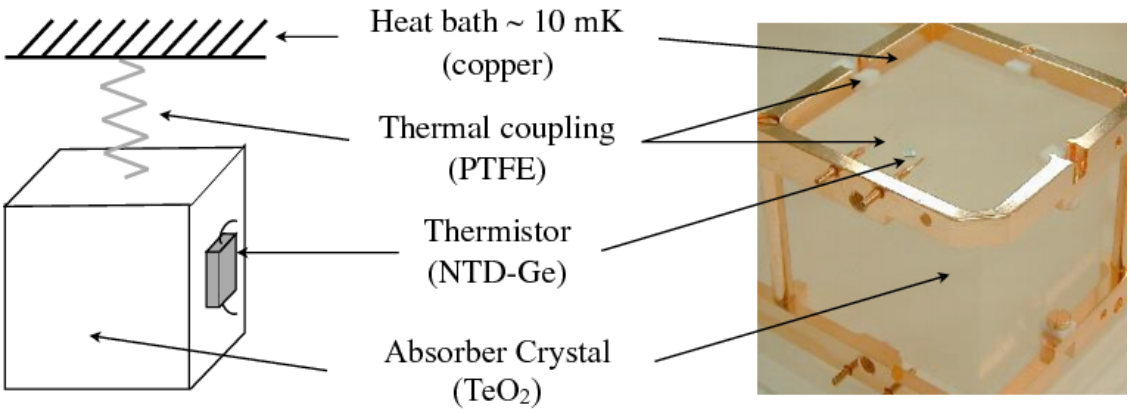


Figure 4.1: A side-by-side comparison of a schematic bolometer with a close-up photograph of a Cuoricino crystal mounted in its copper support structure. The chip glued to the crystal is the thermistor; the white block labeled in the figure as the 'thermal coupling' is a PTFE standoff; some contribution to the thermal coupling also comes from the thermistor wires. Figure from M. Vignati.

Let C be the heat capacity of one crystal, which is thermally coupled to a heat bath of temperature T_b by a thermal conductance of G. Under the assumption that the crystal is a perfect calorimeter and $\Delta T(t) \ll T_b$ for all times t, whenever an event occurs in a crystal, the energy $\Delta\mathcal{E}$ it deposits is converted to phonons and manifests as a small temperature rise

$$\Delta T(t) = \frac{\Delta\mathcal{E}}{C} e^{(-\frac{t}{\tau})} \tag{4.1}$$

with recovery time constant

$$\tau = \frac{C}{G}. \tag{4.2}$$

According to the Debye law, at low temperature $C \propto T^3$; for this reason, the array is housed in a cryostat that keeps the crystals at $8-10$ mK, reducing their heat capacity and thereby improving both the magnitude of the temperature response and the recovery time constant [23]. Each crystal is equipped with a thermometer, or, to be more specific, a thermistor — a resistor whose resistance changes rapidly with temperature:

$$R(T) = R_0 e^{\sqrt{T_0/T}}. \tag{4.3}$$

Here, R_0 and T_0 are characteristic parameters of the thermistor that depend upon its dimensions and the material from which it is constructed [131]. A bias current I is applied across the thermistor, and the change in the resultant voltage V across it as the thermistor's resistance changes with temperature is read out. The thermistors therefore produce voltage data, which must be calibrated to signals of known energies to obtain a voltage-to-energy conversion.

The particular detector configuration used in CUORE-family experiments is illustrated in Figure 4.1. A TeO_2 crystal serves as the energy absorber, and a thermistor chip made out of neutron-transmutation-doped (NTD) germanium semiconductor is glued to the face of the crystal to serve as the temperature sensor. The chip is glued to the crystal with a 3×3 matrix of nine individual glue spots to create a good thermal link that is flexible enough to compensate for the differences in thermal contraction among the different materials without cracking during thermal cycling. The crystal is mounted in a copper support structure with polytetrafluoroethylene (PTFE) standoffs between the copper and the crystal. The copper structure serves as the heat bath at the base temperature of the detector, and the PTFE standoffs provide a thermal link between the crystal and

the heat bath. The wires of the thermistor are also thermalized to the copper structure, providing an additional thermal link to the heat bath. Overall, the thermal conductance between the detector and the heat bath is weak enough that the temperature of the crystal will visibly rise due to an energy deposition of a few keV but strong enough that the heat of each event will be carried away from the crystal within a few seconds. As a result, each particle event is observed as a pulse with a sharp rise time and a more gradual decay time. The few-second response time of bolometers is slow in comparison to other types of particle detectors, but it is appropriate for the low event rates expected in studies of very rare processes such as $0\nu\beta\beta$.

Individual TeO_2 crystals can be produced with masses up to ~ 1 kg, allowing for the construction of close-packed large-mass arrays of individually instrumented detectors. Bolometric detectors enable precision measurement of the energy spectrum of events inside the crystals, allowing the search for an excess of events above background in a narrow window around the transition energy of the isotope of interest.

4.2 Cuoricino

Cuoricino [25, 26, 30] was the first large-scale implementation of TeO_2 bolometers as $0\nu\beta\beta$ detectors. It achieved the greatest sensitivity of any bolometric $0\nu\beta\beta$ experiment to date and served as a prototype for the CUORE experiment. Cuoricino took data from 2003 to 2008 in the underground facilities of the Laboratori Nazionali del Gran Sasso (LNGS), Italy. The data acquisition was separated by an extended maintenance period into two major runs, Run I and Run II.

4.2.1 The Experimental Apparatus

The Cuoricino detector consisted of 62 TeO_2 bolometers with a total mass of 40.7 kg (approximately 11 kg of ^{130}Te). Of these, 18 crystals had dimensions of $3\times3\times6$ cm^3 (330 g); 14 were made of natural TeO_2 (the 'small' crystals), while two were enriched to 75.0% in ^{130}Te and two were enriched to 82.3% in ^{128}Te (the 'enriched' crystals) to facilitate the study of $2\nu\beta\beta$ [107]. These crystals were first operated in the Mi-DBD experiment [24], and they were lapped and remounted

in Cuoricino to provide continuity with the previous experiment. The remaining 44 crystals (the 'big' crystals) were newly produced for Cuoricino with a larger size of $5 \times 5 \times 5$ cm^3 (790 g) and were grown from natural tellurium.

The physical detection efficiencies ε of the two crystal shapes, which are primarily affected by the geometrical effect of β particles escaping the detector and by radiative processes, were estimated from Monte Carlo simulations [21]. They are shown in Table 4.1.

The full-width-at-half-maximum (FWHM) resolution of each detector during each dataset (see Section 5.1.1) was found by fitting the 2614.5-keV gamma line of ^{208}Tl, the nearest strong peak to the $0\nu\beta\beta$ transition energy, in the calibration data for that detector and dataset with a Gaussian peak shape. The average FWHM for each crystal type is presented in Table 4.1.

4.2.2 Results

With a total ^{130}Te exposure of 19.75 kg y, Cuoricino set a limit of $T^{0\nu}_{1/2} > 2.8 \times 10^{24}$ y (90% C.L.) [21] on the $0\nu\beta\beta$ half-life of ^{130}Te. This is the present best limit achieved for this nuclide.

The data was treated as a collection of subsets, each corresponding to a single channel (individual detector) and dataset. A simultaneous fit to a 60-keV-wide region centered at the Q-value on all subsets was performed, in which each subset was fit with a Gaussian peak shape whose FWHM was fixed equal to the FWHM resolution obtained from calibration data as described in Section 4.2.1. The flat background rate was left free in the fit but was constrained to be the same for all crystals of the same type (big, small, or enriched[1]). The background rates obtained from the $0\nu\beta\beta$ best fit, along with the total exposure corresponding to each crystal type, are tabulated in Table 4.1

[1]Here, 'enriched' crystals refers only to crystals enriched in ^{130}Te, as those crystals enriched in ^{128}Te contain a negligible amount of ^{130}Te and therefore were not used for the ^{130}Te $0\nu\beta\beta$ fit.

Table 4.1: Physical detection efficiencies, FWHM resolutions, exposures, and fitted flat background rates for Cuoricino by crystal type. All values are reported in [21]. Note: Unlike the other values reported in this table for the 'enriched' crystals, the average FWHM resolution incorporates both the ^{130}Te-enriched and the ^{128}Te-enriched crystals; the remaining values refer to the ^{130}Te-enriched crystals only.

Crystal Type	ε [%]	Average FWHM Resolution [keV]	Exposure [kg(^{130}Te)·y]	Flat Background Rate [cts/(keV kg y)]
Big	87.4 ± 1.1	6.3 ± 2.5	0.94 (Run I)	0.20 ± 0.02 (Run I)
			15.80 (Run II)	0.153 ± 0.006 (Run II)
Small	84.2 ± 1.4	9.9 ± 4.2	0.09 (Run I)	0.20 ± 0.02 (Run I)
			2.02 (Run II)	0.17 ± 0.02 (Run II)
Enriched	84.2 ± 1.4	13.9 ± 5.3	0.15 (Run I)	0.8 ± 0.4 (Run I)
			0.75 (Run II)	0.35 ± 0.05 (Run II)

4.2.3 Experience in Preparation for CUORE

The Cuoricino experience has provided a wealth of information on the design, construction, operation, analysis, and performance of a large-mass TeO_2 bolometer array. CUORE will constitute more than an order-of-magnitude increase in active detector mass over Cuoricino; at the same time, the knowledge gained from Cuoricino has been applied to preparations for CUORE in order to optimize the capabilities of the experiment.

4.2.3.1 Crystal Dimensions

Before Cuoricino was constructed using mainly $5 \times 5 \times 5$ cm^3 crystals, the $3 \times 3 \times 6$ cm^3 crystals of Mi-DBD were the largest TeO_2 crystals that had ever been operated as bolometers [15]. The detection efficiency of the larger, symmetrical crystals is somewhat higher than for the smaller crystal dimensions, and using larger crystals also allows the investigation of a larger detector mass for the same level of electronics complexity; however, there was concern that detector performance would deteriorate for a more massive crystal.

As can be seen from Table 4.1, the Cuoricino experience demonstrated that in fact the opposite is true: the energy resolutions achieved in the big crystals are better on average and have a smaller spread than those achieved in the small crystals under the same operating conditions, meaning that the larger crystals are superior in every respect. The CUORE array will be entirely composed of crystals with the same nominal dimensions as the big Cuoricino crystals.

4.2.3.2 Backgrounds

One of the most critical of the experimental parameters that determine an experiment's sensitivity (see Chapter 7) is the background rate. The collaboration's experience with Cuoricino led to an extensive R&D campaign aimed at reducing the background that will be seen in CUORE. The only real handle on particle identification that CUORE-like bolometers provide is the energy measurement obtained by measuring the temperature pulses caused by the interactions of particles in the crystals. While multi-detector coincidences can be used to veto events caused by penetrating particles such as muons that have evaded the natural shielding provided by placing the array

underground, events due to gammas or alphas emitted by radioactive contaminants in the detector apparatus are indistinguishable from $0\nu\beta\beta$ events on an event-by-event basis if they fall within the energy range at which the $0\nu\beta\beta$ events are expected to be observed. Thus, efficient shielding and strict material selection, production, cleaning, and handling protocols are crucial to reduce, as much as possible, the number of background events that ever reach the detectors.

The background rates observed in Cuoricino listed in Table 4.1 can be considered as a starting point for understanding what we expect to see in CUORE, with one caveat. The values in the table should be corrected for instrumental efficiency, as this efficiency cannot be assumed to be the same in CUORE as it was in Cuoricino. For Cuoricino, this efficiency includes contributions from analysis effects such as pulse-shape cuts, the anti-coincidence cut, noise, and pile-up with reference pulses from the heater mounted on each crystal to aid in stabilizing detector response, and it has been evaluated to be $(94.7 \pm 0.3)\%$ [21] for real particle events[2].

An analysis of the background sources responsible for the flat background in the ROI has been performed on a partial set of Cuoricino statistics [23, 30], following the technique and the model developed for the Mi-DBD experiment [52]. The result of this analysis was the identification of three main contributions: $30 \pm 10\%$ of the measured flat background in the ROI is due to multi-Compton events due to the 2614.5-keV gamma ray from the decay chain of ^{232}Th from the contamination of the cryostat shields; $10 \pm 5\%$ is due to surface contamination of the TeO_2 crystals with ^{238}U and ^{232}Th (primarily degraded alphas from these chains); and $50 \pm 20\%$ is ascribed to similar surface contamination of inert materials surrounding the crystals, most likely copper. Other sources that could contribute to the detector background are muons [22] and neutrons, but simulations indicate that these have only a minor effect.

On the basis of this result, the R&D for CUORE has pursued two major complementary avenues: one, the reduction of surface contamination, and two, the creation of an experimental setup

[2]The particle and background efficiencies for the pulse-shape and anticoincidence cuts were evaluated by performing an extended-likelihood simultaneous fit to the 2614.5-keV ^{208}Tl peak plus linear background in the spectrum of events accepted by the cut and the spectrum of events rejected by the cut, using the efficiencies as parameters in the fit. The 'background' efficiency of the pulse-shape cut is considerably lower than the particle efficiency, but this is attributed to the intentional removal of spurious, non-physical pulses. As we are interested in the real particle background rate, the particle efficiency should be used for the efficiency correction.

in which potential background contributions are minimized by the selection of extremely radiopure construction materials and the use of highly efficient shields. The required surface contamination levels are extremely low, on the order of $1-10$ nBq/cm^2, nearly undetectable with any standard technique used in surface analysis. In most cases, only bolometric detectors are sufficiently sensitive to measure contaminations at this level; at this time, our understanding of these contaminations comes only from the statistics-limited data sets collected by small test detectors constructed from CUORE materials (see [16] for the contract requirements on and measurements of the contamination levels of the crystals).

Nevertheless, a rough conservative estimation of the background reductions achieved for the CUORE crystals and copper is possible. A preliminary evaluation of the surface contaminations of the final CUORE crystals [27] indicated a lower limit on the reduction with respect to the contamination seen in Cuoricino of a factor of 2; the measurement was statistics-limited, so the true reduction factor may be greater. As for the copper, extensive efforts have been dedicated to the study of different treatment procedures able to reduce the copper surface contamination, culminating in the simultaneous operation in the Cuoricino cryostat of three arrays of detectors, each prepared with one of the final candidate background-suppression techniques, known as the Three Towers Test (TTT) [18]. On the strength of the results of this test, a technique that proved to be capable of reducing the copper surface contamination also by at least a factor of 2 as compared with that observed in Cuoricino has been selected by the collaboration as the baseline for the CUORE copper treatment.

4.2.3.3 Calibration

Each detector's exact response depends on a number of factors. The parameters R_0 and T_0 of each thermistor can be expected to be slightly different, and they are further modified by the stresses exerted on the thermistor chips by the glue attaching them to the crystals [131]. The readout of the voltage across the thermistor is dependent on the details of the biasing circuit; in particular, it can shift based on the point on the $V-I$ load curve at which it is operating, known as the working point, which in turn depends on the base temperature of the bolometer. To

mitigate this effect, heaters are mounted on the detector towers that constantly provide some level of heating to maintain the detectors at a base temperature a few degrees milliKelvin higher than the lowest base temperature of which the cryostat is capable; the output of these heaters can be increased or decreased to compensate for few-mK shifts in the base temperature of the cryostat. The shape of the temperature pulse is affected by the values of the various thermal couplings in the system, including those between the crystal and the copper, the crystal and the thermistor, and the thermistor and the copper; each of these is also expected to differ slightly from one detector to the next, especially the glue spot pattern that couples the thermistor to the crystal. The pulse shape is also mildly energy-dependent; the rise and decay times of the pulse are both correlated with the event energy [131].

To achieve the optimum statistical power in the search for $0\nu\beta\beta$, the spectra acquired by each detector in the array must be summed together and a fit to the sum spectrum performed; however, to do this for spectra obtained with detectors whose responses can significantly differ, each detector must be well calibrated individually to convert its pulse-amplitude spectrum into an energy spectrum. In addition, as the search for $0\nu\beta\beta$ is performed by fitting the spectrum with a peak centered at the known Q-value for the decay, a high accuracy in the energy determination is necessary to ensure that the peak is correctly placed. A high-precision recent measurement has determined the ^{130}Te Q-value with an error of ± 0.013 keV [120], more than an order of magnitude smaller than the calibration uncertainty achieved in Cuoricino or that anticipated for CUORE (see Section 6.3.1); the expected calibration uncertainty is comparable to the uncertainties on two other recent measurements of the 130Te Q-value, ± 0.32 keV [126] and ± 0.23 keV [118]. The calibration uncertainty is therefore potentially a non-negligible contributor to the systematic uncertainty of the experimental results and should be carefully taken into account, although as long as the calibration uncertainty is small in comparison with the ≈ 5-keV energy resolution of the detectors, the effect will be mitigated to some degree.

Performing reliable energy calibrations of the detectors using a radioactive source with a selection of gamma lines at known energies spanning the gamma region of the spectrum (roughly $0-3000$ keV) is therefore a crucial task for the proper understanding and analysis of the data

collected by CUORE-family detectors. Chapter 5 discusses in detail many of the challenges and design efforts involved in developing the calibration system for CUORE, while Chapter 6 explains the calibration analysis tools that were used for Cuoricino and presents a study of the calibration performance in Cuoricino and an extrapolation of similar expectations for CUORE as well as some refinements and adaptations to be implemented for CUORE.

4.2.3.4 Enrichment

The enriched detectors that operated in Cuoricino were originally prepared for Mi-DBD by growing crystals from isotopically enriched material. It was found that the level of enrichment achieved in the powder is decreased somewhat by the process of crystallization, which requires seeds of natural telluride; the final crystals were measured by mass spectrometer to be enriched to 82.3% in ^{128}Te or 75.0% in ^{130}Te, while the powder used to grow them was enriched to 94.7% or 92.8%, respectively [24].

For CUORE, all crystals are grown from natural tellurium because the natural abundance of ^{130}Te is sufficiently high to provide considerable source mass even without the added expense and complication of isotopic enrichment. As can be seen from Table 4.1, bolometers constructed from enriched crystals appear to suffer some deterioration in performance compared with natural crystals, and they also seem to experience intrinsically higher background rates, possibly due to the greater contamination of the crystals themselves implied by the greater ^{238}U and ^{232}Th surface-level contamination of the enriched crystals indicated by Monte Carlo studies [24]. Nevertheless, the Cuoricino experience has established that isotopic enrichment is a viable option for a future, more sensitive CUORE-scale experiment that can operate a greater $0\nu\beta\beta$ candidate mass without increasing the physical mass of the detector.

4.3 CUORE-0 and CUORE

The Cryogenic Underground Observatory for Rare Events (CUORE) [29, 23], the follow-up experiment to Cuoricino, is designed to search for $0\nu\beta\beta$ in ^{130}Te. It is currently under construction and will exploit the experience and results gained from its predecessor.

The CUORE collaboration is now operating a single CUORE-like tower in the former Cuoricino cryostat. This configuration, named CUORE-0, served as a validation of the CUORE assembly procedures and will allow some early evaluation of the readiness of the background reduction measures.

4.3.1 The Experimental Apparatus

CUORE will consist of an array of 988 TeO_2 cubic detectors, similar to the $5 \times 5 \times 5$ cm^3 Cuoricino crystals described above, although the average mass of a crystal will be 750 g instead of 790 g due to more strictly controlled production procedures. The total mass of the detectors will be 741 kg. The detectors will be arranged in 19 individual towers and operated at ≈ 10 mK in the Gran Sasso underground laboratory. The expected FWHM energy resolution of the CUORE detectors is ≈ 5 keV at the Q-value. This resolution represents an improvement over that seen in Cuoricino and has already been achieved in tests performed in the CUORE R&D facility at LNGS.

With its 988 detectors and a mass of ≈ 206 kg of ^{130}Te, CUORE will be larger than Cuoricino by more than an order of magnitude. Background rates are also expected to be reduced by approximately an order of magnitude with respect to Cuoricino.

The design and construction of the cryostat that will be used to maintain the CUORE detectors at the necessary cryogenic temperatures is a rather unique undertaking. It is based on the comparatively recently developed technology of pulse tube (PT) cooling, which uses mechanical cooling to reach a temperature of 4 K for the precooling of the helium in the ^{3}He/^{4}He dilution unit; in contrast, a more traditional refrigerator uses a bath of liquid helium, which boils off over time and must be periodically refilled, causing vibrational disruption to the operating conditions of the cryostat. This 'dry cooling' should allow an improved stability of the base temperature of the detectors [23]. The CUORE cryostat is the first cryostat of its kind large enough to house and cool the large detector mass represented by the CUORE array.

The ^{3}He/^{4}He dilution unit achieves the base temperature of the cryostat by exploiting the fact that a mixture of liquid ^{3}He and ^{4}He will separate into two phases at very low temperatures: a 'concentrated ^{3}He' phase containing $\approx 100\%$ ^{3}He and a 'dilute ^{3}He' phase containing $\approx 6.6\%$

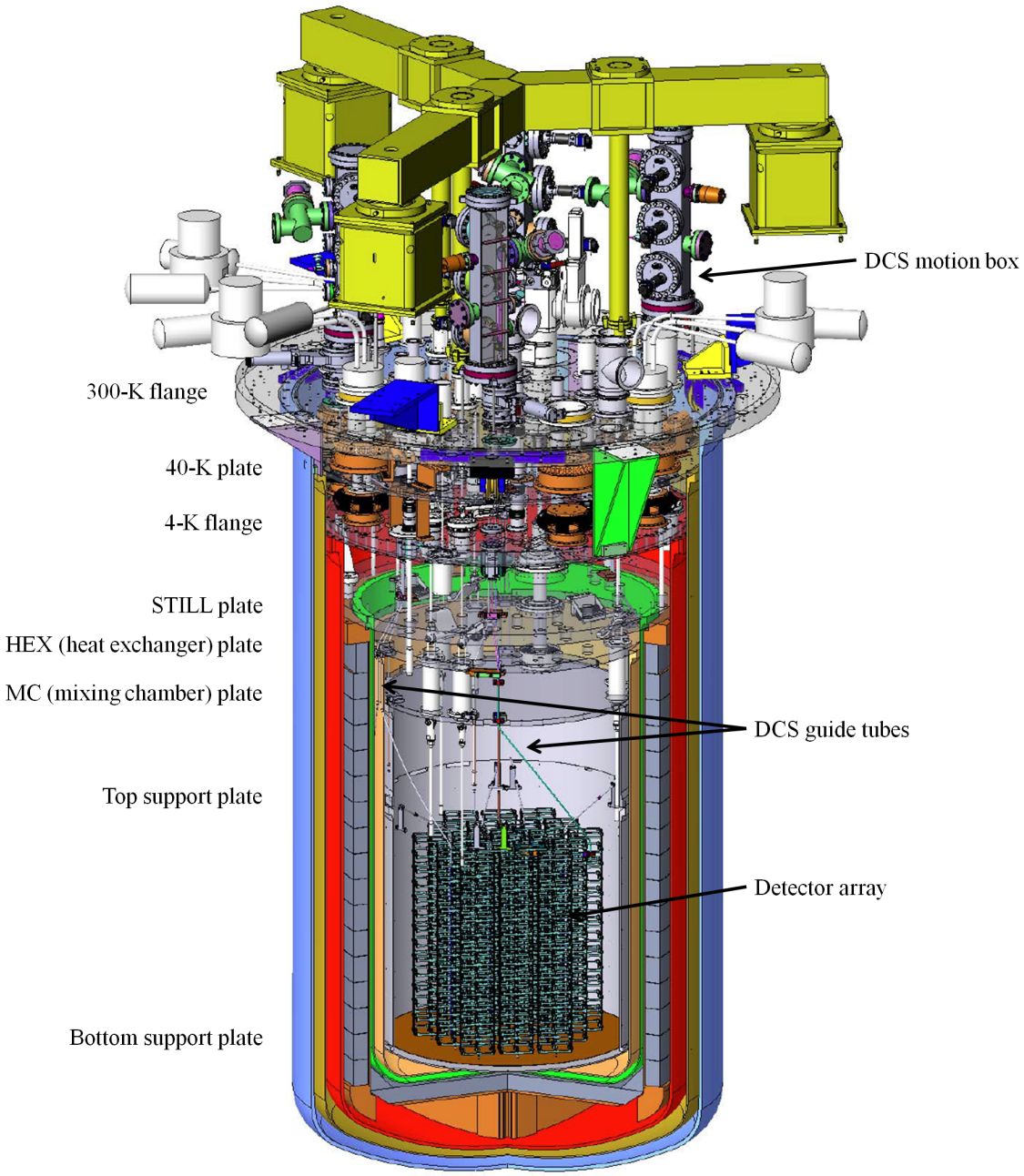


Figure 4.2: A 3D scale model of the CUORE detector and cryostat. Some elements of the detector calibration system (DCS) are labeled. The grey blocks are the inner lateral lead shielding. Not pictured is the inner top lead shield, which will be a thick disc in the volume just above the crystal array.

^{3}He [116]. The concentration of ^{3}He in the dilute phase remains finite even at absolute zero, and the specific heat of the ^{3}He in the dilute phase is larger than the specific heat of the ^{3}He in the concentrated phase, meaning that the latent heat of mixing leads to cooling whenever ^{3}He passes from the concentrated to the dilute phase. In a ^{3}He/^{4}He dilution unit, the lower-density concentrated-^{3}He phase floats on top of the higher-density dilute phase of the ^{3}He/^{4}He mixture in the mixing chamber. The lower portion of the mixing chamber is connected to a second chamber called the still, which is kept at a temperature of 0.7 K so that the vapor pressure of ^{3}He is much higher than that of ^{4}He; pumping on the still thus predominantly removes ^{3}He from the mixture, and the resulting osmotic pressure pulls ^{3}He from the dilute phase in the mixing chamber and drives more ^{3}He from the concentrated phase into the dilute phase, causing cooling. The ^{3}He pumped from the still is then condensed and returned to the mixing chamber, creating a closed system that can provide continuous cooling down to temperatures of less than 10 mK.

CUORE-0 consists of CUORE crystals mounted in CUORE-style frames as a single tower. The construction of the CUORE-0 tower served as a trial of the new handling and assembly procedures developed for CUORE prior to the launch of the full CUORE assembly line. The Cuoricino cryostat has been refurbished to ensure that it will operate reliably until CUORE data-taking begins, allowing a high-statistics test of CUORE-like detector performance before CUORE construction is complete and serving as a sensitive $0\nu\beta\beta$ experiment in its own right.

4.3.2 Schedule

CUORE-0 is now operating in the underground laboratory at LNGS. It will continue to collect data until the start of CUORE.

The CUORE detector is currently under construction. The experimental hut that will house the cryostat at LNGS is complete, and the commissioning of the cryostat and its subsystems is well underway. The gluing of chips to the faces of the crystals that have already been delivered to and stored at LNGS has begun, and tower assembly will proceed in parallel with the cryostat commissioning. The final installation of the full detector array in the cryostat and first data-taking

are scheduled for 2014, and the experiment is expected to accumulate data or about 5 years of total live time.

4.3.3 The Future of TeO_2 Bolometers

The CUORE experiment will thoroughly explore the parameter space for $0\nu\beta\beta$ in ^{130}Te that is implied by the Klapdor-Kleingrothaus claim of observation in ^{76}Ge and, depending upon the true value of the nuclear matrix element, may begin to achieve sensitivity to the inverted-mass-hierarchy region of $\langle m_{\beta\beta} \rangle$ allowed for the case in which the neutrinoless double-beta decay rate is dominated by the exchange of a light Majorana neutrino; see Section 7.3 for a calculation of the sensitivities of CUORE-0 and CUORE as a function of time. The physics potential of CUORE is very similar to that of a number of other $0\nu\beta\beta$ experiments using various isotopes that have begun taking data recently or will be coming online within the next few years.

However, if no evidence of $0\nu\beta\beta$ is seen by the current generation of detectors, the effort to push experimental sensitivities to longer and longer half-lives will remain necessary in the foreseeable future; indeed, even if $0\nu\beta\beta$ *is* discovered in the coming decade, it will still likely be desirable to build a new, more sensitive generation of experiments to measure the decay rate with high precision in multiple nuclides.

To this end, the CUORE collaboration is pursuing several avenues for achieving greater physics reach with detectors on a similar physical scale to CUORE. One option for SuperCUORE is to construct the detector array with crystals grown from enriched, rather than natural, tellurium (see Section 4.2.3.4). A parallel effort is focused on detector adaptations that will allow the incorporation of active background rejection techniques. One possible method involves coupling individually instrumented bolometric absorber panels to the surfaces of each crystal, allowing events originating from the surfaces of the crystal to be identified by comparing the signals from the surface panels to those from the main crystal; such surface events can be attributed to surface contamination and therefore rejected [79, 78]. Another method is based on supplemental light collection; although TeO_2 crystals do not scintillate, Cherenkov light is emitted by electrons traveling through the crystals, while the energies of the typical α particles that constitute a significant background

for CUORE-family experiments are below threshold for Cherenkov emission, providing at least some mild discrimination of α events from β/γ events depending on the efficiency of Cherenkov light collection that can be achieved [56].

Thus the CUORE experiment sits at the forefront of the present search for $0\nu\beta\beta$ decay, while the detector technology on which it is based is poised to pursue that search toward new frontiers of experimental sensitivity.

Chapter 5

Calibration Design in CUORE-Family Detectors

The search for neutrinoless double-beta decay depends upon the ability to observe an excess of events that are consistent with two simultaneously emitted β particles at a known energy. If characteristics can be identified on an event-by-event basis that allow discrimination among the interactions of different types of particles (e.g., αs, βs, and γs), events that can be unambiguously attributed to sources other than $0\nu\beta\beta$ can be rejected before performing the search for excess events at the $0\nu\beta\beta$ Q-value.

CUORE-family detectors are close-packed multi-bolometer arrays. To some extent, a multi-detector array can self-veto against penetrating background particles like muons (which would cause simultaneous events in multiple adjacent crystals), and spurious events due to detector noise can be filtered out through pulse-shape analysis; however, in the end, the only real data, and therefore the only real handle on event identification, that bolometers provide is energy information. This is quite suitable for a $0\nu\beta\beta$ experiment, as an energy signal is what one is looking for, but it means that reliable, precise energy calibration is absolutely essential to the experiment's ability to provide meaningful data.

Because the thermistors read out the temperature changes in the bolometer crystals as voltage data, energy calibration must be performed to determine the relationship between the energy deposited in the crystal and the voltage signal subsequently obtained. This must be done for each detector (the crystal plus the readout chain, including the thermistor and the electronics) individually, before the spectra can be summed together and analyzed for evidence of $0\nu\beta\beta$.

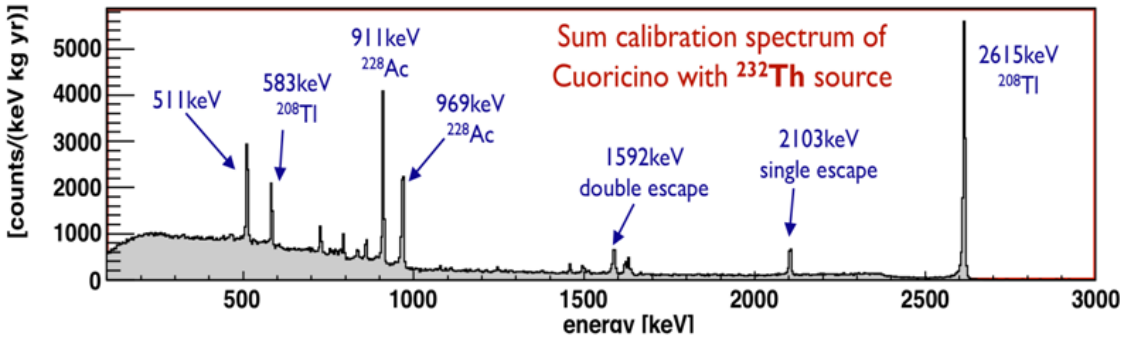


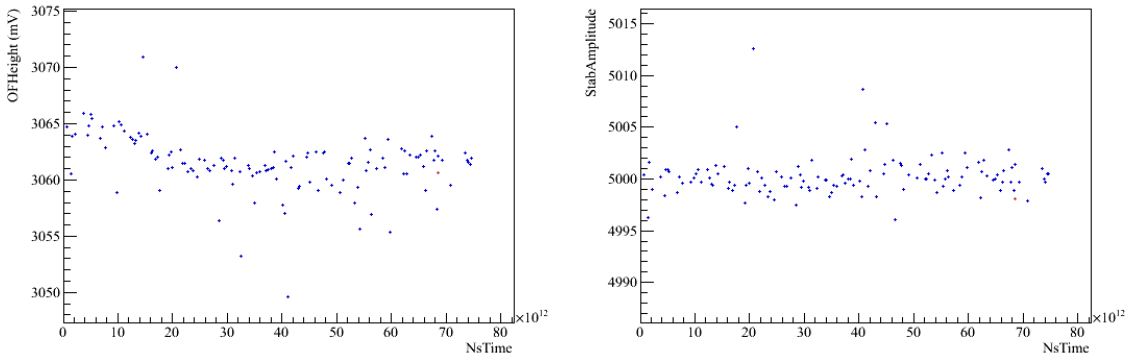
Figure 5.1: The summed spectrum of all Cuoricino bolometers, from calibration data collected with ^{232}Th. Each bolometer has been calibrated individually, then summed together after the calibration was applied. Note the strong 2614.5-keV gamma line from the decay of ^{208}Tl, which provides a solid handle on the calibration near the $0\nu\beta\beta$ Q-value, ≈ 2528 keV. Figure from S. Sangiorgio.

5.1 General Calibration Approach

5.1.1 Acquisition of Calibration Data

In order to obtain the energy calibration of the detectors, a gamma source with a known spectrum is used to illuminate the crystals. Although the most critical energy region that must be calibrated is the region of interest around the $0\nu\beta\beta$ Q-value, the whole gamma spectrum up to approximately 3 MeV should be calibrated as well as possible for reliable identification of backgrounds. For this purpose, Cuoricino used ^{232}Th as its calibration source because it is sufficiently long-lived ($T_{1/2} = (1.405 \pm 0.006) \times 10^{10}$ y [33]) that the calibration sources should never need to be replaced, and its decay chain produces a number of gamma lines up to 2615 keV that are strong enough to be used for calibration (see Section 5.1.2.1); the 2614.5-keV peak is particularly strong and can be used to ensure solid calibration in the region of interest (see Figure 5.1).

The response of the bolometers is not entirely stable with time; as the temperature of the experimental space of the cryostat drifts, the base temperature of the detectors also drifts and so does the response. Changes over short time scales can be at least partially corrected with the constant-energy stabilization heater pulses that are injected into each crystal at regular intervals

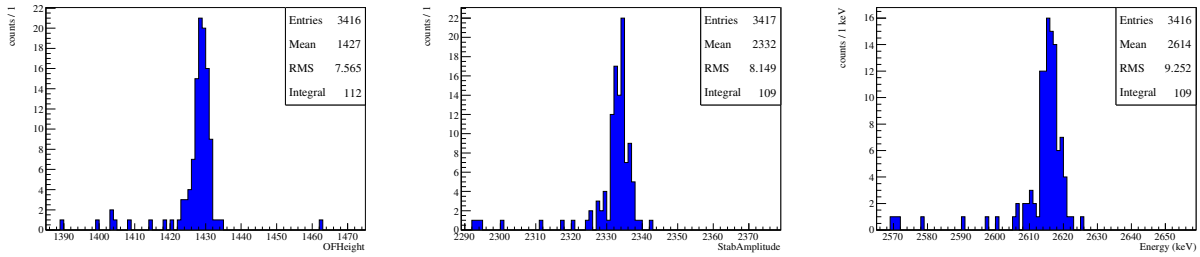


(a) (b)

Figure 5.2: Amplitude vs. time (in ns) of the stabilization heater pulses in a measurement taken with a Cuoricino detector. (a) Before stabilization. (b) After stabilization.

(approximately once every 5 minutes) by a heater chip glued to its face; Figure 5.2 illustrates the measured amplitudes of the heater pulses over time before and after stabilization in a measurement on a Cuoricino detector. However, there is still a response drift over long time scales that must be corrected by periodic calibration.

In Cuoricino, the default procedure was to collect new calibration data approximately once per month, although this did not always occur. In CUORE, it is possible that the pulse-tube cryostat will be able to maintain a more stable base temperature than Cuoricino's traditional refrigerator, as it will no longer be necessary to disrupt the cryostat by refilling a liquid helium bath every two days. Some members of the collaboration are also investigating the possibility of acquiring and tracking thermistor parameters to allow some further stabilization of the detector response via analysis of the thermistor behavior [131]; the data processing and analysis chain that follows this approach is henceforth referred to as the Thermal Response (TR) analysis. In future, the frequency of calibration will be optimized based on the operation of CUORE; the baseline schedule is still to calibrate once per month, but there is the possibility that experience with the final detector will allow a move to bimonthly calibration.



(a) Pulse amplitudes (b) Stabilized pulse amplitudes (c) Calibrated event energies

Figure 5.3: The 2614.5-keV calibration line as it appears in spectra constructed with the quantities sequentially obtained during the data processing of a calibration measurement taken with a Cuoricino detector.

Each period of data collected between two adjacent calibration periods is called a dataset. Each dataset is calibrated by combining the data collected during the calibration period immediately proceeding it with the data collected during the calibration period immediately following it; the sum spectrum of the two calibration periods is used to extract a calibration function for each channel (in Cuoricino and CUORE, each channel corresponds to a unique detector), which is then applied to all data for that channel in that dataset. This approach allows the calibration for each dataset to 'average' any response drift that occurred across that entire dataset.

5.1.2 Automatic Calibration with Diana Analysis Software

What follows in this section is a brief description of the operation of the calibration module in the Diana analysis software package, which is a framework developed by the collaboration to sequentially perform necessary preprocessing tasks (such as pulse fitting, stabilization, calibration, and coincidence analysis) on the event data collected by the detectors to produce the final spectra that can be used for physics analysis. The calibration module is applied after the voltage pulse for each event has been fitted with the average pulse shape for the detector to extract the baseline voltage level of the detector readout prior to the pulse and the amplitude of the pulse itself, and after

that amplitude has been stabilized by comparing its associated baseline with the amplitude-vs.-baseline trend observed in the heater pulses collected during the measurement. The stabilization of the heater pulses is illustrated in Figure 5.2. Figure 5.3 shows the 2614.5-keV calibration peak in a calibration measurement taken with a Cuoricino detector in spectra produced with the three quantities obtained at these three stages of processing: the raw amplitudes extracted by fitting the pulse associated with each event, the stabilized amplitudes of the events, and the calibrated energies of the events. The behavior of the calibration module is documented in greater detail in the Ph.D. thesis of the primary author of the module, L. Kogler [107].

5.1.2.1 Calibration Peak Finding

The peaks produced by the ^{232}Th decay chain that are used as calibration peaks are summarized in Table 5.1. Each is labeled as either a 'primary' or 'secondary' peak based on its treatment in Cuoricino[1]. Primary peaks are the first peaks that the calibration module tries to identify; they are expected to be larger and clearer than the secondary peaks, and they are intended to be identifiable on every channel. Secondary peaks are smaller; the calibration module attempts to find them by interpolating from the identified positions of the primary peaks, and it is not expected that it will always be possible to find the secondary peaks on every channel.

There are two ways in which the calibration module may attempt the first-pass peak-finding of primary peaks: it may use a seed file (usually the calibration functions obtained for the channels in the preceding dataset) to obtain guesses for the locations of the peaks, or it may use an AutoSeed algorithm that is based on the peak-finding capability of the TSpectrum class in ROOT [2]. In either case, each of the peak location guesses obtained in this way is then used to define a window in which the module will attempt to find and fit the most prominent peak as described in Section 5.1.2.2. For the final Cuoricino result, the seed-file method was used to calibrate all datasets.

[1]Because the CUORE sources are inside the shields, the 583 line appears to be considerably stronger relative to the rest of the spectrum than it was in Cuoricino. It is conceivable that the 583 line may be treated as a primary peak in CUORE (see Section 6.4), but this possibility has not yet been fully evaluated.

Table 5.1: Gamma lines used for calibration. Only primary peaks are used to evaluate calibration times from the simulations presented in Section 5.5.

Label	Energy	Source	Cuoricino Classification
2615	2614.533 keV	^{208}Tl	Primary
2104	2103.533 keV	^{208}Tl single escape	Secondary
1593	1592.533 keV	^{208}Tl double escape	Secondary
	1588.19 keV	^{228}Ac	
969	968.971 keV	^{228}Ac	Secondary
	964.766 keV	^{228}Ac	
911	911.204 keV	^{228}Ac	Primary
583	583.191 keV	^{208}Tl	Secondary
511	511.0 keV	annihilation gamma	Primary
	510.77 keV	^{208}Tl	

Once a structure has been identified with a particular calibration peak and fitted, that peak is considered to be 'found' if the fitted peak shape passes a significance-above-background test. Otherwise, the calibration module considers itself unable to find the peak, and no point corresponding to that peak is present in subsequent functional fits.

Once the first-pass peak-finding and fitting is complete, the calibration module fits an interpolation function to points in energy-vs.-stabilized-amplitude space corresponding to the primary peaks that were located in the first pass. This fitted interpolation function is then used in the second-pass peak-finding as a seed function to obtain guesses for the locations of the secondary peaks. If any primary peaks were not found in the first pass, new guesses for the locations of those peaks are also obtained from the fitted interpolation function at this time and used for a second attempt to find those peaks. Peak fitting then proceeds from the location guesses as in the first pass.

If too few peaks are located by the calibration module, the user is prompted to set peak windows manually with a graphical user interface (GUI) tool.

5.1.2.2 Peak Fitting

Once a starting guess is obtained for a peak's location, a 'wide' window of width $\pm W$ (where W is typically approximately 7 times the heater resolution, the resolution of the peak corresponding to the flagged stabilization heater pulses on the channel, or 2/3 the distance to the next peak, whichever is smaller) centered on the location guess is defined. The data in this window are histogrammed into 80 bins, and the maximum bin is taken as the new peak location guess.

A new 'narrow' window of width $\pm w$ (where w is typically 5 times the heater resolution for the channel, or 2/3 the distance to the next peak, whichever is smaller) centered on the new location guess is defined. An unbinned maximum likelihood fit is performed on this data range, with the initial value of the peak location set to the center of the range. The line shape used for the fit is either a single or double Gaussian peak[2] plus a linear background. For the double-peak structures,

[2]Only the peaks labeled 1593 and 969 in Table 5.1 are fit with a double Gaussian shape. The two gammas contributing to the 511 peak are so close in energy that they cannot be resolved.

only one peak mean is left as a free parameter in the fit; the mean of the other peak is fixed in relation to that parameter by the ratio of the nominal energies of the two peaks.

The fit functions are built from components that are parametrized so as to be normalized to a total area of 1. The sum of the background and peak components is then parameterized in terms of the background fraction f_{bkg}; the background term is multiplied by f_{bkg} and the peak term by $(1 - f_{bkg})$. For the double-peak shape, it is further parameterized by the relative amplitudes of the peaks; the peak term is constructed as a sum of two normalized Gaussian shapes, one multiplied by $amp2$ and one multiplied by $(1 - amp2)$.

5.1.2.3 Calibration Function Fitting

Once all peak-fitting passes are complete, the calibration module fits the calibration function to points in energy-vs.-stabilized-amplitude space corresponding to all calibration peaks that passed the significance test to be considered successfully found. The χ^2 of the fit is then checked; if it is too high, and if one peak can be thrown out without ending up with fewer peaks than the number of free parameters in the calibration function and at least one peak has a deviation greater than 3σ, the module throws out the point with the greatest deviation and tries the fit again. The deviation (in units of σ) is defined as

$$\frac{abs([\text{calibration function solved for peak mean at peak energy}] - [\text{peak mean}])}{[\text{peak mean error}]}.$$

If the χ^2 of the calibration fit is still too high, the calibration module generates a warning.

There are currently a few non-optimal behaviors in the way the calibration module handles throwing out the highest-deviation point from a poor calibration function fit that should be addressed in preparation for CUORE. These are discussed in Appendix A.1.

5.2 Calibration Hardware

5.2.1 Cuoricino and CUORE-0

In the Cuoricino cryogenic apparatus, which can host a narrow vertical detector array such as the single tower of Cuoricino, two radioactive sources are inserted inside the external shielding on

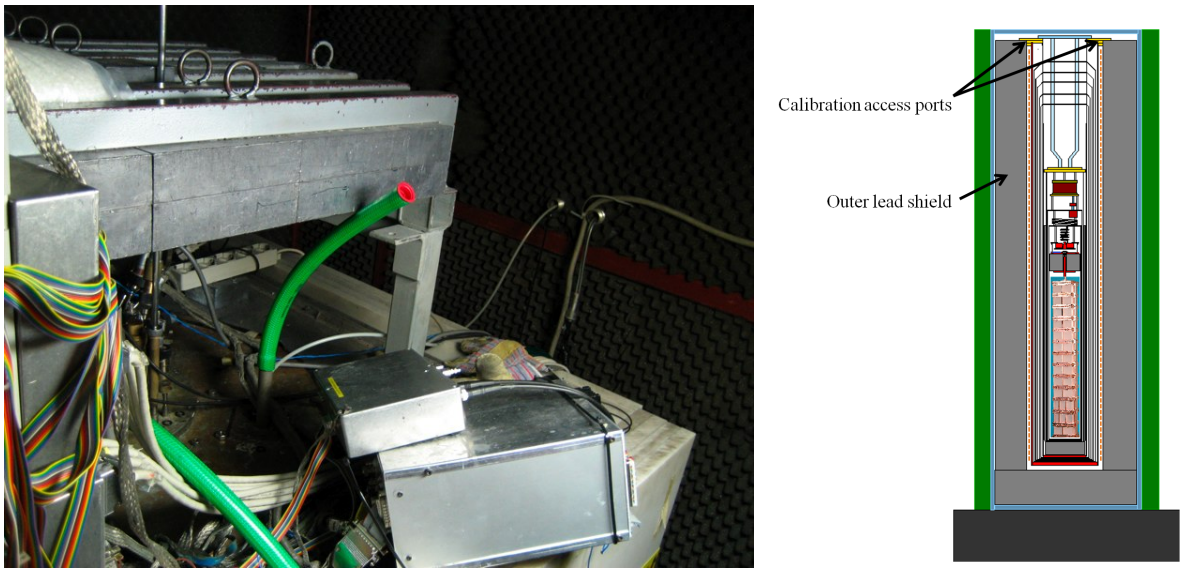


Figure 5.4: (a) A photograph of the top of the Cuoricino cryostat taken from the access ladder inside the Faraday cage. The green hose attached to the vertical pipe in the center of the frame is the nearer calibration access port for insertion of a calibration source; part of the hose leading to the farther port is visible in the bottom left corner. Photograph from R. Maruyama. (b) A schematic of the Cuoricino tower inside the cryostat and outer shielding. Orange dashed lines indicate the placement of the calibration sources.

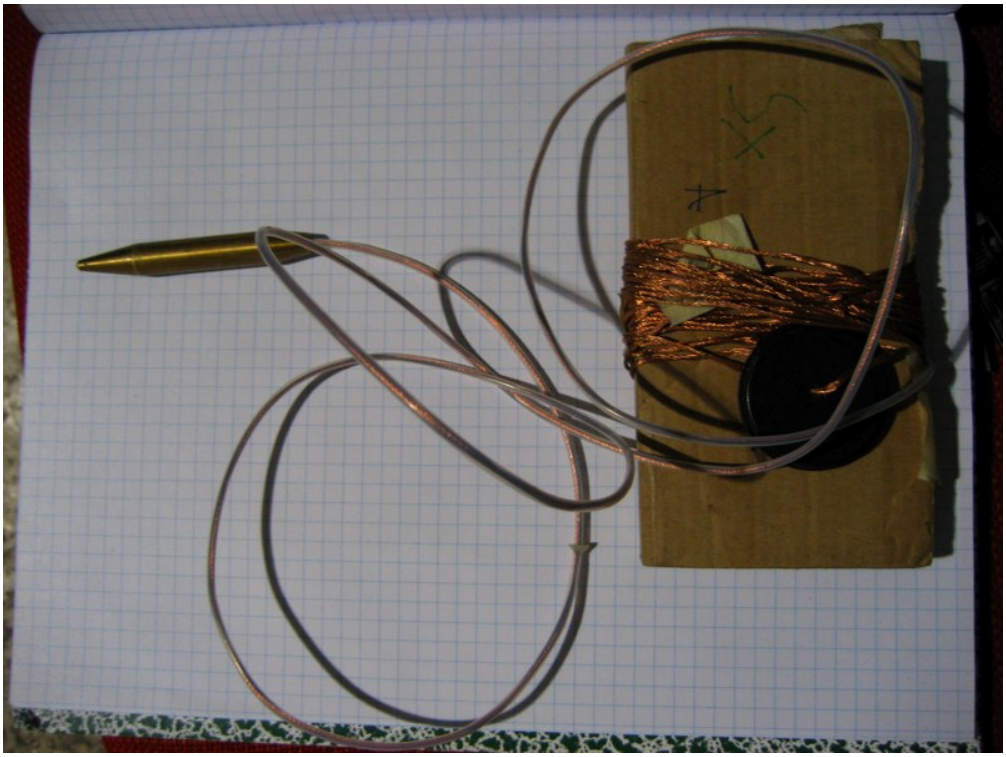


Figure 5.5: A photograph of one of the Cuoricino calibration sources, partially unwound from its storage card. The active region of the source is enclosed in a PTFE sleeve. The black plastic cap serves as an insertion stop to correctly position the source in relation to the detector area of the cryostat. Photograph from R. Maruyama.

either side of the cryostat. The sources do not travel inside the cold vacuum region of the cryostat itself; two pipes on the top plate provide access to the space between the external lead shield and the outer vessel of the cryostat, and a flexible hose attached to each pipe allows the operator to guide the source into position. Figure 5.4 shows one of these calibration access ports and the placement of the calibration sources in the Cuoricino setup.

The radioactive material for each source is provided by six strands of Goodfellow W99/Th1 W 145370 thoriated tungsten wire; each strand has a thoria content of 1%, a diameter of 0.2 mm, and a length of approximately 110 cm, ≈ 25 cm longer than the total height of the Cuoricino detector tower. The mechanical strength of the source is provided by a flexible stranded copper cable, about 1 mm in diameter; the tungsten wire is loosely wrapped around the lower portion of the cable and held in place with a PTFE sleeve. The cable terminates in a brass weight to help guide the source into position and keep it hanging straight during calibration. A plastic cap at the upper end of the cable sits against the mouth of the insertion hose when the source is inserted and serves as a stop to ensure the correct vertical positioning of the source. Figure 5.5 shows the source for the farther port partially unwrapped from its storage card; the two sources are not interchangeable due to the differing placements of the plastic stops.

The radiation from the sources is somewhat attenuated by the internal shielding inside the cryostat before reaching the detector region. Because the detector suspension leaves the array free to rotate, the relative positioning of the calibration sources with respect to the crystals is not precisely known; however, the symmetry is such that the change in illumination with the rotation of the detector array is small.

The CUORE-0 detector tower is operated in the same cryostat as was used for Cuoricino, so it is calibrated in the same manner.

5.2.2 CUORE

One of the most dramatic changes that has to be made in scaling up from the single tower of Cuoricino to the 19 towers of CUORE is that the calibration system must grow in complexity to enable calibration of each individual detector. In CUORE, the outer towers will shield the inner

towers, so some calibration sources must be routed all the way into the detector area and between the towers to achieve even illumination, which is important to allow the successful calibration of the entire detector in a reasonable amount of time without causing an excessively high event rate in some crystals — because bolometers are inherently slow (each pulse lasts several seconds), a high rate causes pileup and raises the baseline temperature of the detectors, leading to increased dead time and degradation of the energy resolution.

The thermistors that are used to read out the crystals are very sensitive and their response will change with small variations in the working point of the detector, making it necessary to perform a calibration every month or two with minimal disruption of the detector and cryostat. 'Minimal disruption' means that

- the detector calibration system (DCS) must not compromise the low-background environment of the detector as a result of
  - activity in the materials used to construct various elements of the calibration system,
  - gammas from the calibration sources reaching the detector while the sources are stored in normal data-taking position, or
  - active source material either escaping into the detector or otherwise becoming unable to be extracted from the detector (e.g., through breakage); and
- the thermal load placed on the cryostat as a result of the calibration sources moving between room temperature (300 K) and the detector region ($8-10$ mK) must not cause the detector to warm up and change its working point, or the calibration will be invalid. There are two main types of thermal load contributed by the DCS,
  - static loads from stationary elements of the system and
  - transient loads while the source carriers are moving (e.g., friction).

The CUORE calibration system poses a design challenge for a number of reasons. It is only one of many systems that must share space inside the cryostat; the calibration sources must follow

complicated routes around these other systems to reach the detector region so they can provide the calibration signal, and again when they are removed to allow normal data taking. Motion in cryogenic and vacuum conditions is difficult on its own, because of the mechanical and thermal effects of friction and vibration; additional complication arises from the fact that the calibration sources must travel through regions of differing temperatures, from 300 K to $8-10$ mK, without failing under thermal cycling or thermally overloading the cryostat. Finally, the low-background requirements for the experiment place additional constraints on materials and techniques that can be used in the system, meaning that alternate solutions must sometimes be found in place of the traditional methods of addressing the above issues.

The design of the DCS tries to address all of these concerns. As a general overview, the DCS consists of 12 flexible source carriers, routed through the levels of the cryostat by means of guide tubes, and stored and deployed by four motion boxes containing three spools each that sit on top of the 300-K flange of the cryostat. Many elements of the system, including the motion boxes and parts of the guide tube routes, are visible in Figure 4.2.

5.2.2.1 Source Carrier

The source carrier is conceived as a collection of small, individual active sources, chained together to form a single flexible unit that is capable of sliding down through the guide tubes into calibration position under its own weight when fed off a spool. The source carrier is built by attaching individual source capsules to a continuous string; this reduces the risk of source carrier failure that would result in a piece of active material breaking off and remaining in the detector area after extraction of the carrier. Each capsule houses one or more radioactive source inserts.

The radioactive source used to calibrate the array will be ^{232}Th, as in Cuoricino. The long half-life of ^{232}Th is especially beneficial to the CUORE DCS concept, as it means that the sources should never need to be replaced under normal circumstances, and so by virtue of the motion box system the sources should generally be able to be inserted for calibration and extracted for normal data taking without ever opening up the cryostat to the outside. Thoriated tungsten wire inserted into the source capsules provides the necessary source material. Thoriated tungsten is produced by

Table 5.2: Relevant parameters for the calculation of the calibration source activity provided by a thoriated tungsten wire insert.

Half-life of ^{232}Th	Molar mass of ThO_2	Density of thoriated tungsten
$T_{1/2}(^{232}\text{Th})$	$M(\text{ThO}_2)$	ρ
1.405×10^{10} y	264.04 g/mol	19.3 g/cm^3

introducing a small amount of thoria (thorium dioxide, ThO_2) into the tungsten. The activity can be tuned by selecting the thoria content and dimensions of the wire inserts. The activity per length as referred to the parent nuclide, ^{232}Th, of a piece of thoriated tungsten wire can be calculated as follows:

$$\text{activity/length} = \frac{ln(2)}{T_{1/2}(^{232}\text{Th})} \frac{N_A}{M(\text{ThO}_2)} (\text{thoria content}) \pi (\frac{d}{2})^2 \rho,$$

where $T_{1/2}(^{232}\text{Th})$ is the half-life of ^{232}Th (1.405×10^{10} y), N_A is the Avogadro constant, $M(\text{ThO}_2)$ is the molar mass of ThO_2 (264.04 g/mol), d is the diameter of the wire, and ρ is the density of the wire (19.3 g/cm^3 for thoriated tungsten). The stoichiometric correction and isotopic-abundance correction that would be present in the general case are omitted here because there is only one Th atom per molecule of ThO_2 and the isotopic abundance of ^{232}Th is $\approx 100\%$.

From here on, all activities will be referred to the parent nuclide as in the expression above. If the chain is in secular equilibrium, the total number of events in the ^{232}Th chain is approximately 10 times the activity referred to the parent. Thoriated tungsten wire was also used as the calibration source for Cuoricino, and the Cuoricino calibration sources were measured to be in approximate secular equilibrium.

5.2.2.2 Guide Tubes

The guide tubes route the source carriers through the cryostat and also provide a thermal connection to various stages of the cryostat. In order to follow the complicated routes necessary to reach the calibration positions of the sources, there are a number of bends in the tubes. These bends are of critical consideration, both to the reliable motion of the source carriers through the tubes and to the amount of heat dissipated into the cryostat by the friction of the source carriers moving through the bends.

Additionally, because the guide tubes essentially form a penetration through all the thermal stages of the cryostat, the thermal gradient that will develop along them represents another thermal load on the cryostat; this necessitates careful choices of thermalization locations and physical gaps that separate the tubes into three distinct sections[3]: one runs between 300 K and 4 K, one runs from 4 K down to the detector region, and one is thermalized to 10 mK along its entire length in the detector region. At each of these gaps, the ends of the tubes are flared to help guide the source carrier, absorb any possible misalignment between the two tube sections that may arise during cooldown of the cryostat, and prevent the source carrier from becoming caught against hard edges of the tubes.

Between the 300-K and 4-K flanges, the guide tubes follow an S-shaped path. These S-tubes consist of commercial stainless steel bellows to safely accommodate the anticipated relative motion between the flanges, which may be as much as 10 mm. The bellows are lined with flared PTFE tubes to prevent the ridged interior of the bellows from interfering with the source carrier motion. Below the 4-K flange, the guide tubes are made of copper to respect radioactivity requirements as they approach the detector region, and the guide tubes in the detector region itself are made of the same high-purity copper as the support structure of the crystal towers. Six of the guide tube routes pass between the detector towers and entirely enclose the (internal) source carriers in the detector region, while the other six wrap around the outside of the HEX vessel of the cryostat to reach the

[3]In fact, only the internal tube routes are separated into these three sections. The guide tubes for the external calibration sources extend only to the top of the detector region; the strings then hang free from the bottoms of these tubes when they are in the fully deployed calibration position.

final calibration position and then allow the (external) source carriers to extend out the bottom of the tube and hang freely.

5.2.2.3 Motion Boxes and Drive Spools

Each motion box houses three individual drive spool assemblies, which can be dismounted and replaced from outside in the unlikely case that a new source needs to be installed. There is also a gate valve between the motion box and the cryostat that can be closed if necessary when the source carriers are fully extracted. Each spool is individually controlled by a computer control system; it controls the speed and direction of the motor, reads out the tension on the string and issues a stop signal if the tension is too high, and reads out the number of turns the spool has undergone as a measure of the location of the source carrier inside the guide tubes. Each motion box is also instrumented with a proximity sensor capable of counting each capsule as it passes.

5.2.2.4 Thermalization

During normal physics data-taking, the calibration sources are stored in the motion boxes at 300 K. When they are deployed for calibration, however, they travel into the cryogenic environment of the cryostat to their final calibration positions in the detector area, at a temperature of $8-10$ mK. The internal sources will be located inside guide tubes thermally anchored to the same temperature as the detector towers. The source strings must be thermalized to around 4 K during deployment so that they do not radiate excessive heat at colder stages, placing too much heat load on the cryostat and warming up the detectors. To achieve this, a clamping system has been designed to squeeze on each source as it travels downward into the cryostat, creating good thermal contact between the thermalized clamp and the capsules and thereby cooling the capsules. One clamping unit for each set of three guide tubes traveling from a single motion box is mounted below the 4-K flange and operated by a counterweight system located above the 300-K flange. The control software for the thermalizer is integrated with the drive spool control. Figure 5.6 shows one thermalizer clamp assembly. The clamps will also be closed on the bare strings when they are in the fully deployed calibration position to provide a thermal anchor point at 4 K.

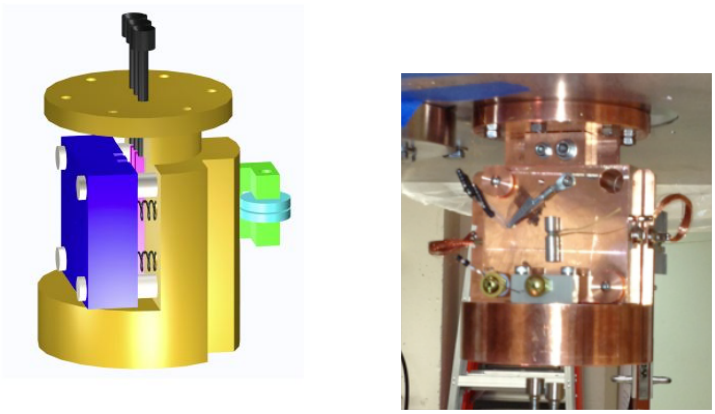


Figure 5.6: Left: A model of a single thermalizer clamp assembly. Right: A photograph of a production clamp mounted in a test stand.

5.3 Design and Testing of CUORE Calibration Sources

The design of the CUORE source carrier is the result of several iterations and extensive prototyping and testing. It was driven by two main motivating requirements: it must be capable of navigating the complicated routes inside the cryostat from the 300 K flange down into the detector region, and it must be fully extracted from the cryostat during normal data-taking to ensure that no radiation from the sources compromises the background rates observed by the detectors.

The requirement for a source carrier anchored at only one end and capable of moving downward through the cryostat under its own weight was established very early on. The alternative would be a semi-rigid source carrier that can be both pushed and pulled from above through guide structures, which would cause excessively high friction and place an unsupportably high heat load on the cryostat, or a source carrier that travels in a continuous loop through the cryostat in such a way that it can be pulled on from either end to bring it into position, which would prevent the radioactive source carrier from ever being fully retracted from the cryostat volume and would also be impossible to replace without warming up the cryostat. The original concept for the source carrier was based upon the flexibility and mass distribution of a commercial ball chain. A ball chain is composed many identical solid pieces chained together in such a way as to allow considerable

bending movement between adjacent pieces; in this way, a ball chain is capable of traveling under its own weight through the complicated bends of the guide tube routes that we use to maneuver calibration source material into position near the detectors.

This section presents the development of the CUORE calibration source carrier design as well as the materials selection and manufacturing approach for the final production source strings. For many of the parts discussed in this section, once the quality assurance guidelines and procedures for the final string production were established, the majority of the final measurement, testing, and assembly tasks were performed by J. Clark.

5.3.1 Early Source Carrier Prototyping, Spooling Tests, and the Final Design

The earliest source carrier prototyping tests were carried out with commercial ball chain. As a result of these tests, it was soon realized that while the ball chain could be easily deployed through the guide tubes, the friction caused by an attempt to extract the ball chain was extremely high due to the large amount of metal-to-metal surface contact made between the chain and the inner surfaces of the guide tubes, especially on the bends. An attempt was made to coat a length of commercial ball chain using a liquid suspension of PTFE to reduce the friction of the surface, but the result was unsatisfactory due to the complicated structure of the ball chain and the large relative motion among many small, adjacent parts. Loading commercial ball chain with radioactive material also posed a challenge, so the decision was made to move to a custom-built approach.

Prototyping of a chain of spherical PTFE beads loaded with radioactive metallic cores and chained together using beading techniques was briefly pursued, but this experience soon raised concerns about the labor-intensive manufacturing task posed by custom building a chain of several hundred beads and the risk that if any of the ‘links’ in such a chain failed, a piece of radioactive material could become trapped in the cryostat and require the interruption of the running of the experiment so that the cryostat could be warmed up and opened to remove the resulting contamination.

These concerns led to what turned out to be the final concept for the CUORE source carrier, illustrated diagrammatically in Figure 5.7 and photographically in Figure 5.8. The source carrier

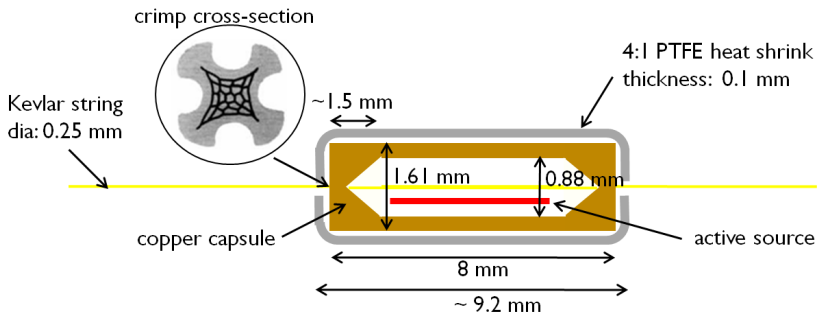


Figure 5.7: The anatomy of the source carrier concept with the relevant dimensions. The total length of 9.2 mm is based on measurements of prototypes; there is some variation due to the behavior of the ends of the heat shrink.

Figure 5.8: A photograph of a completed prototype source capsule. The darkening of the silicon-wax coating of the string is due to the heat used to shrink the PTFE sleeve.

is built on a single, continuous piece of Kevlar string. Kevlar has a high tensile strength, with a typical breaking strength of around 300 N, and a low thermal conductivity of 0.04 W/(m K) [67], which will minimize the heat load on the cryostat due to the fact that the source carrier will have an attachment point at higher temperature when the active region is positioned in the $\approx$ 10 mK detector region. Small copper tubes are crimped onto the Kevlar string, creating hollow interiors that can host small active source wires. As the tubes constitute rigid sections along the otherwise flexible Kevlar, a length of 8 mm was chosen to allow as much room as possible for wire insertion while avoiding compromising the movement of the string around the bends in the guide tubes. These crimp tubes are then covered by PTFE heat shrink sleeves, which cover the sharp edges left by the crimping process and also provide a PTFE surface to reduce the friction of the source carrier against the guide tubes through which it will travel.

The 29-mm spacing (measured from the center of one capsule to the center of the next) between capsules was designed to leave plenty of bare Kevlar to maintain the flexibility of the string, to keep the total capsule mass that will need to be thermalized small, and to ensure that the activity distribution seen by each crystal along the length of the string is as similar as possible. Figure 5.9 illustrates the relationship between the crystal geometry and the capsule spacing.

The prototype source carriers that have been produced in the lab show great mechanical reliability. A load test was performed on early prototype crimps by attaching weight to the bottom of a length of Kevlar to which copper tubes had been crimped and lifting by gripping only a single tube with pliers; the crimps held 2.85 kg with no failure, slippage, or evident damage to the Kevlar. Thermal cycling by way of repeated immersion in liquid nitrogen has caused no failure or weakening of the PTFE, the Kevlar, or the crimps. Several prototype strings have been subjected to thousands of cycles of motion tests through prototype guide tube routings, and all significant wear and damage to the Kevlar that has been observed has been inflicted in the process of identifying and resolving failure modes of the apparatus and software controls that resulted in highly abnormal malfunction, such as the string becoming tangled and stuck around the guide tubes; the source carrier can easily survive several CUORE lifetimes of normal operation.

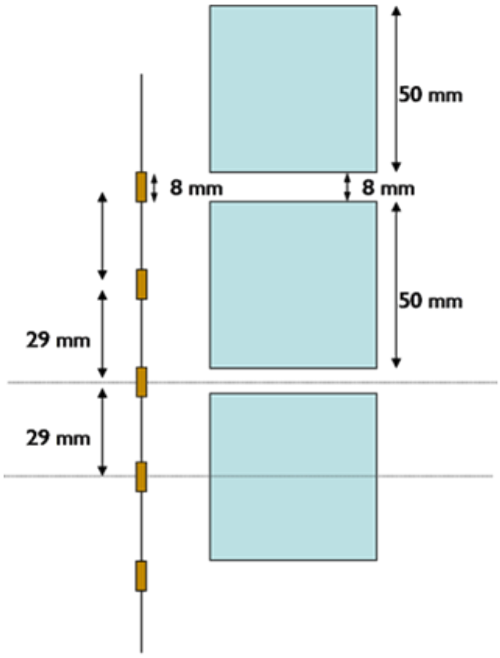


Figure 5.9: A schematic illustration of the nominal vertical positioning of the source capsules with respect to a crystal tower.

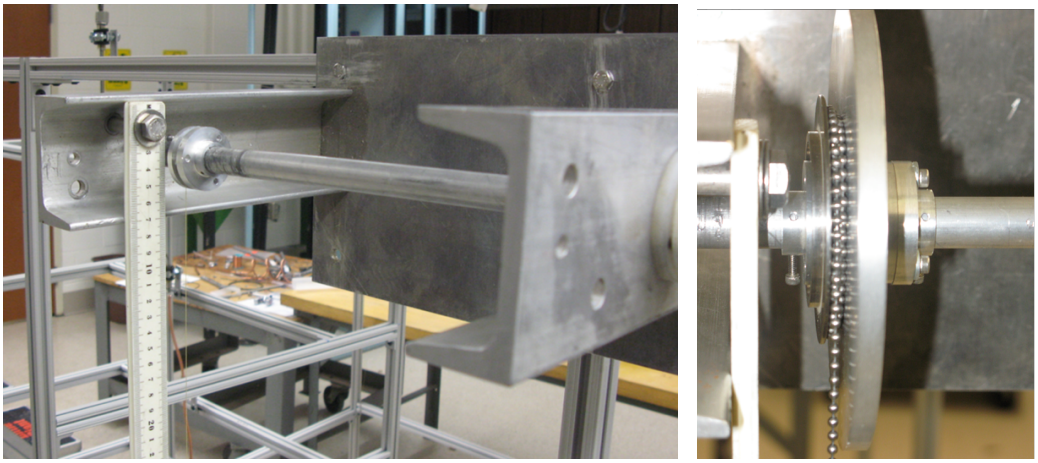


Figure 5.10: Photographs of setup for early spool tests. The discs were positioned separately on the shaft to create different effective spool widths. The larger discs in the right-hand image allowed for a greater spool depth.

To perform motion tests with source carrier prototypes to refine the string design, it is necessary to have a prototype spool from which to deploy them. The design of the spool is itself dependent on the geometry of the source carrier, and it is an important factor in the spooling/unspooling uncertainty (specifically, how well the position of the source carrier inside the guide tubes is known after a given number of turns of the spool). Because we want to be able to control the position of the source carrier with very high accuracy and precision, to within a few millimeters, spooling tests were carried out from very early in the design of the source carrier, starting when string prototyping still focused on commercial ball chain.

For the first tests of the spooling uncertainty, a horizontal shaft was set up so that prototype source carriers could be spooled and unspooled with the loose end hanging freely under gravity (see Figure 5.10). This setup was used to emulate a simple spool, a cylinder with two flat end pieces. At first, the shaft itself was used as the spool diameter, and two discs were fixed in position along the shaft to define the spool width; eventually a set of discs of varying diameter and thickness was produced that could be slid onto the shaft in between the larger end discs to replicate different

spool parameters. Measurements of the unspooled length of the carrier were taken at each entire turn of the spool.

In early tests with both ball chain and Kevlar string, measurements were taken both as the carrier was spooled and as it was unspooled; these data indicated that, within measurement uncertainty, the way in which the carrier is wound completely determines the way in which it will subsequently unwind. After this, measurements were taken only during the unspooling. Early tests similarly indicated that the initial winding achieved when attaching the carrier to the spool, while maintaining tension in the carrier by hand, was likely to be quite different from a winding achieved with the carrier hanging freely, so in later tests no measurements were taken until the carrier had been spooled and unspooled in place once or twice. The ball-chain data shown in Figure 5.11 clearly illustrates the difference in behavior between the initial unwinding following a manual winding of the source carrier and subsequent windings and unwindings with the source carrier hanging freely; by extension, this demonstrates the importance of the specific winding conditions to the spooling behavior. Additionally, the differences among repeated windings and unwindings are greater when more total chain is spooled up, as the cumulative uncertainty in the increase of the effective diameter of the spool due to the wound-up portion of the chain increases as the total amount of wound-up chain increases.

Once the crimp-tube-based source carrier design was reached, several spooling tests were performed with a prototype source carrier, with and without a short piece of ball chain tied to the end to provide extra weight. Due to the discretization of the mass distribution along the string, the additional end weight proved necessary to keep the string hanging straight with only one or two capsules hanging free of the spool, and a similar end weight was expected to be necessary to ensure reliable motion of the source carrier through the guide tube system. The preferred spool configuration was determined to be a spool width of approximately 1.5 times the diameter of the crimp tubes, allowing the source carrier as little freedom as possible in how it will lie on the spool without creating a danger of the capsules becoming stuck. It can be seen from Figure 5.12 that not only does the spooling uncertainty increase for a larger spool width, but the greatest uncertainty is due to the spooling and unspooling of the active region, as the increase in the effective diameter

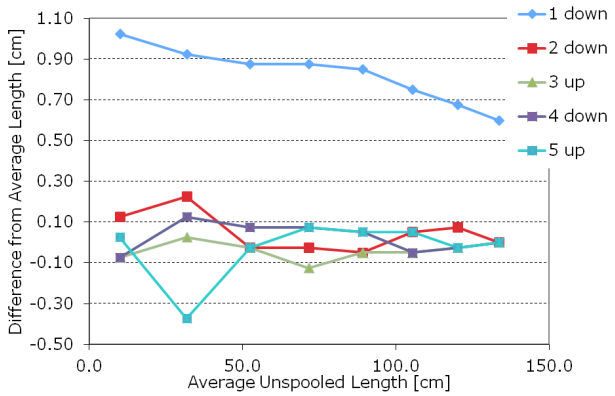


Figure 5.11: Early spooling data collected with ball chain. Values on the x-axis are the mean unspooled length measurements at each full turn of the spool; values on the y-axis are the differences between each actual measurement and that mean. The initial unwinding (denoted '1 down'), after manually winding the spool, differs considerably from subsequent winding/unwinding; the spread increases as more chain is spooled because there is more possible variation in how the spooled chain can change the effective diameter of the spool when more total chain is spooled.

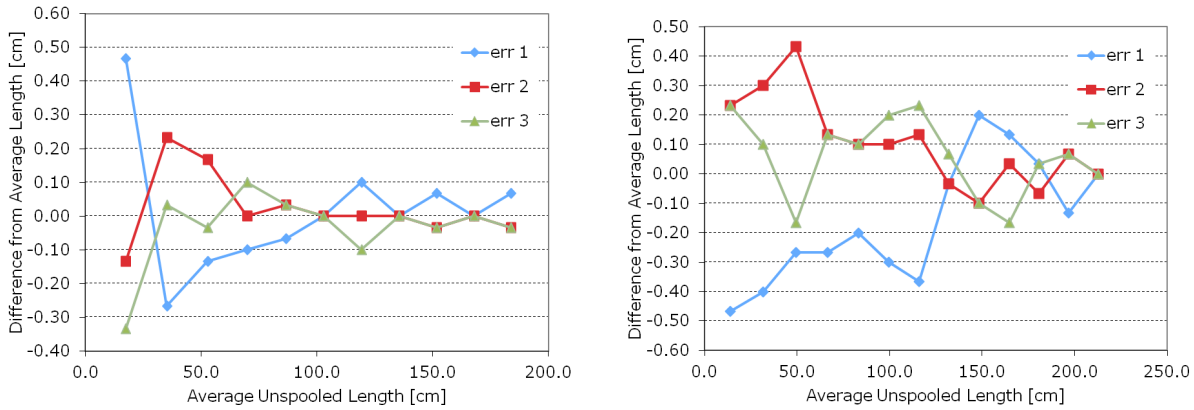


Figure 5.12: Spooling uncertainty in tests with spool diameter 5 cm, crimp tube diameter 1.83 mm (increases to approximately 2 mm with PTFE heat shrink), and capsule spacing 3 cm. All source capsules are off the spool for unspooled lengths of $\gtrsim$ 100 cm, for which the spread decreases considerably. Left: Spool width 3.1 mm. Right: Spool width 3.9 mm.

of the source carrier caused by the capsules means that the carrier length stored in a turn of the spool is more sensitive to small changes in the way the carrier distributes itself on the spool during winding; outside the active region, the source carrier is only a Kevlar string of small, constant diameter, so the uncertainty is much reduced.

When the full prototype drive spool went into production, a secondary guide pulley feeding the spool had been added to the design. The guide pulley was intended to serve several purposes. It provided positioning control of the string as it fed onto the spool, in another effort to maximize the repeatability of the string winding via the spool design; similarly, it provided positioning control as the string fed into the top of the guide tube system to help ensure that the string would not miss the tube. It was also mounted on a load cell to serve as a tension gauge for the string. To confirm that the pulley would have at least no detrimental effect on the spooling uncertainty in the tests with the free-hanging string, a grooved metal rod was adopted as a stand-in for the guide pulley in a spooling test using only a Kevlar string, with similar results to an equivalent test without the grooved rod (see Figure 5.13). Once the prototype spool was built, it was found that the pulley did not reliably rotate with the motion of the string, so the design was adapted into a stationary guide elbow instead, very similar to this preliminary rod test.

Once the prototype drive spool was complete, motion tests with a prototype string moving through a prototype guide tube route confirmed that the spooling uncertainty remained within a few millimeters in this more representative setup. The motor driving the spool is equipped with an encoder to count the number of rotations of the motor, which is directly proportional to the number of rotations of the spool; a camera was set up at a gap in the tube system and triggered to take a picture at the same rotation count on every cycle, at which the end weight was visible in the gap; photo analysis software was then used to determine the position of the weight in each picture. Figure 5.14 shows the camera setup and some of the photographs taken by it.

In the final system, the encoder remains the primary method by which the positioning of the string is controlled, but an array of other instrumentation provides cross-checks and helps to mitigate the uncertainty in spooling behavior established by these early tests that the spool design was unable to eliminate. The first of these is the string tension gauge provided by the load cell, which

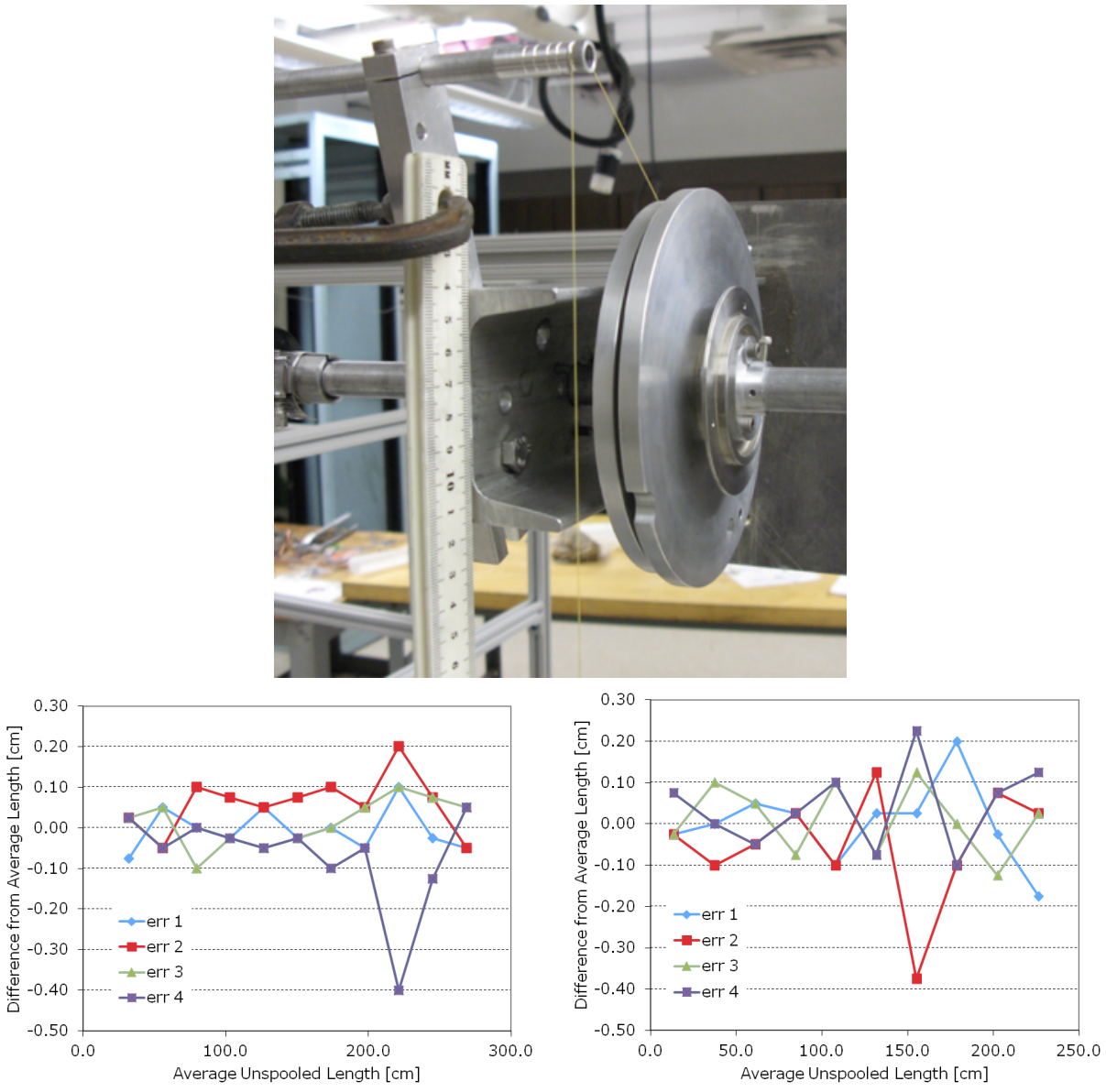
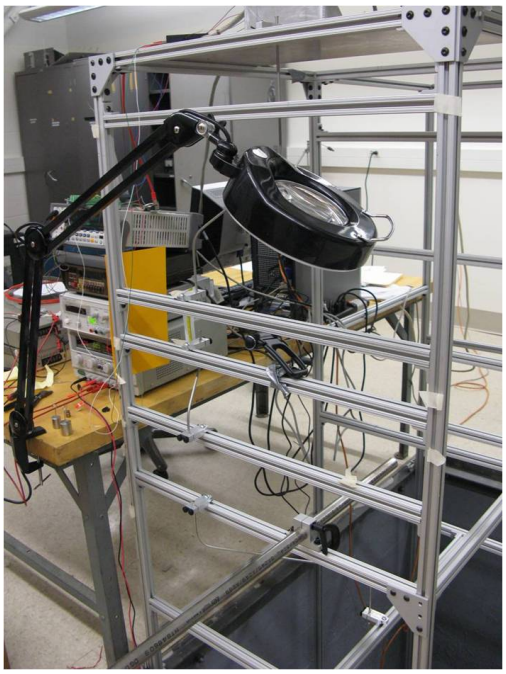
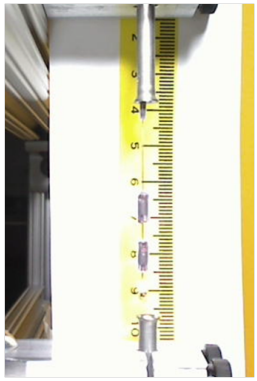


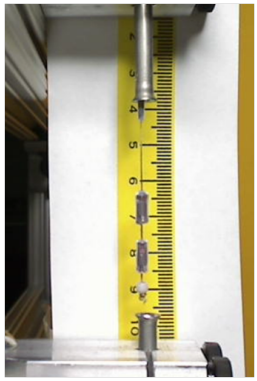
Figure 5.13: Top: Photograph of experimental setup with a grooved rod feeding Kevlar string onto a spool with a 7.5-cm diameter and a 3.4-mm width. Bottom left: Spooling performance without feeder. Bottom right: Spooling performance with feeder.



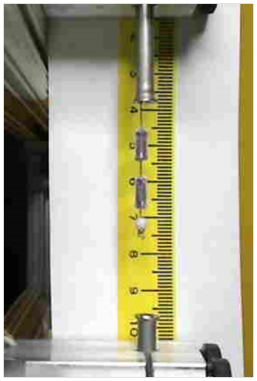
(a) Photograph of prototype guide tube system with webcam.



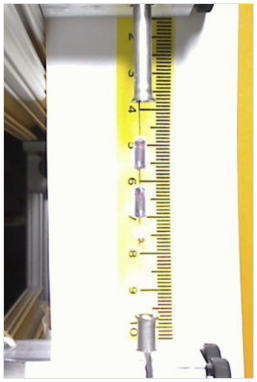
(b) First down



(c) Last down



(d) First up



(e) Last up

Figure 5.14: Prototype guide tube setup for early motion tests and the photographs taken in the setup by the webcam on the first and last cycles of one test run to investigate the repeatability of the spooling. Due to a delay between the signal sent to the camera and the actual acquisition of the picture, only pictures taken when the string was traveling in the same direction are comparable.

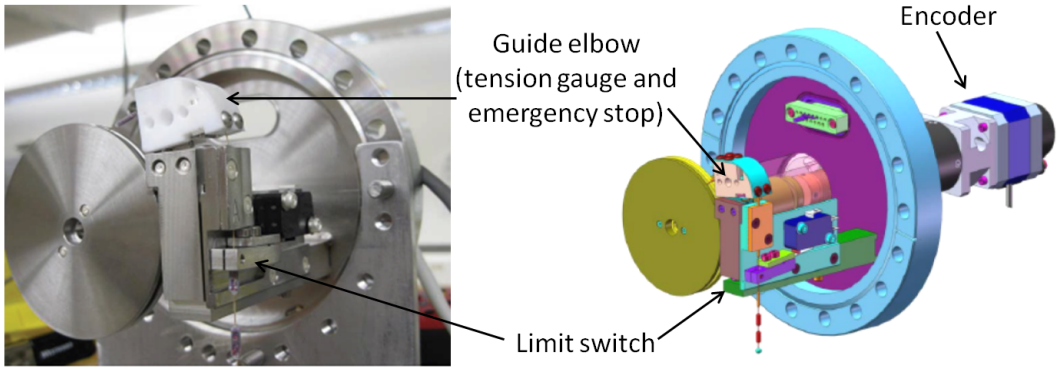


Figure 5.15: The final design of the drive spool used to store and deploy the calibration source carrier.

was found to be sufficiently sensitive to observe a unique tension profile for each guide tube route, with identifiable characteristic features that can be associated with the position of the string in the tube system, such as when the bottom capsule first arrives at a bend. The primary purpose of this monitoring is that if the load cell reading goes outside its expected range at any given point along the profile, an emergency stop signal is sent to the drive spool, providing protection against the string becoming inextricably caught or tangled inside the cryostat. An additional benefit, however, is that the load cell profiles collected during the commissioning of the calibration system should allow a new source carrier to be loaded, wound, and associated with a known positioning without warming or opening the cryostat if a source carrier should need to be replaced during the lifetime of CUORE. Secondly, a proximity sensor near the base of each motion box counts each capsule as it passes through the sensor. This sensor was originally added as a way to be certain that all strings had completely cleared the gate valve at the base of the motion box after retraction at the end of a calibration run before closing the gate valve for normal data-taking, but it is now also used to re-zero the encoder on each drive spool as the top source capsule passes through the proximity sensor, when the active region of the string — known to be the greatest source of spooling uncertainty — is guaranteed to be completely off the spool. Finally, each drive spool assembly is equipped with

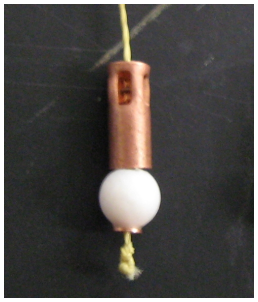


Figure 5.16: Photograph of a prototype end weight. The PTFE ball with the press-fit copper core is the same as in the final source carrier design. The core is crimped onto the string, and the support lip (visible at the bottom of the ball) positions the ball on the core. A knot is placed below the ball as a failsafe.

a limit switch through which the source capsules pass, but the larger weight capsules that make up the end weight on each string cannot. This sends a hard stop signal to the drive spool to cease the spooling up of the source instead of relying on the encoder reading to bring the string back to its home position. The final drive spool design is illustrated in Figure 5.15.

As can be seen from Figure 5.14, at the time the photographs were taken, the end weight design consisted of two weight capsules and a small PTFE ball to guide the end of the string into the flared ends of the guide tubes. The weight capsules are constructed from commercial Nicopress crimp sleeves with a length of 6.35 mm, an outer diameter of 2.81 mm, and an inner diameter similar to that of the source capsules; the end weight was designed this way to allow active source material to be loaded into the weight capsules if necessary to extend the active length of the source carrier. The PTFE ball has a copper core because the crimps were found to be a very mechanically reliable way to attach material to the Kevlar; a copper core with a support lip at the bottom edge is crimped onto the Kevlar, then the PTFE ball is press-fit onto the copper so that it rests against the lip. A knot is tied below the ball as an additional failsafe against anything falling off the string in the unlikely case that one or more crimps should fail. Figure 5.16 shows a close photograph of a similar prototype end bead in which the support lip of the core can be clearly seen.

Because of the limit switch, the entire length of the source carrier that is strung with weight capsules remains off the spool at all times, and its weight is responsible for ensuring the proper unwinding of the source carrier at the beginning of deployment. Extensive motion testing of prototype carriers with the prototype drive spool and guide tube system established that eight weight capsules at a spacing of 13 mm[4] are necessary to ensure reliable unspooling and motion through the guide tubes. The PTFE ball is placed at a separation of 7 mm between the bottom of the copper shelf on the ball's core and the bottom of the bottommost weight capsule to ensure that the proximity sensor will never count the copper core of the ball as a separate object from the weight capsule.

The combination of the large necessary number of weight capsules and the small available vertical space between the bottom of the crystal towers and the inner surface of the innermost cryostat vessel requires the loading of all weight capsules with radioactive source material to ensure proper illumination of the lowest detectors. The bottoms of the guide tubes sit 5 mm from the inner surface of the vessel, and end caps on the tubes will occupy ≈ 2 mm of the guide tube length. These restrictions and the desired symmetry with respect to the detector towers illustrated in Figure 5.9 defined the final dimensions of the calibration source carrier, each of which is constructed with 25 source capsules at 29-mm spacing, 8 weight capsules at 13-mm spacing, and the PTFE guide ball. The nominal calibration positions of one internal source carrier and one external source carrier are shown in Figure 5.17.

5.3.2 Source Carrier Materials and Parts

The choice of materials for the CUORE source carrier was driven by several considerations: mechanical reliability, motion and friction concerns, radioactivity requirements, and thermal behavior.

The source carrier is assembled from many small parts, and the quality of a completed capsule depends on a number of operations that require good alignment precision either between the tool

[4]The spacing between the bottom of the topmost weight capsule and the bottom of the bottommost source capsule is also 13 mm.

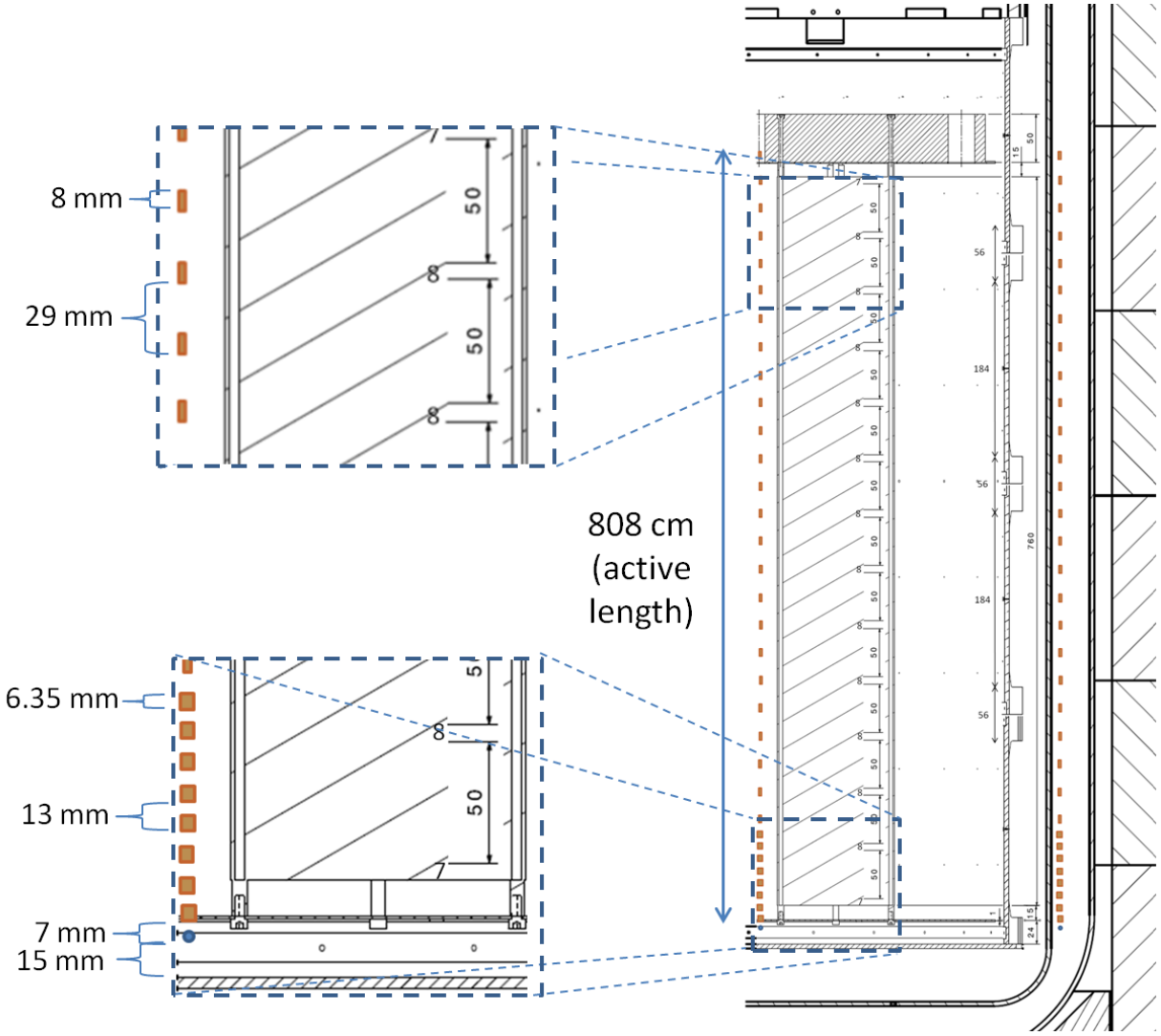


Figure 5.17: Diagram assembled from technical drawings of the cryostat systems showing the dimensions of the source carrier and the nominal placement of one internal and one external source carrier with respect to a detector tower. The orange rectangles represent source capsules (thin rectangles) and weight capsules (thick rectangles); the blue circles represent PTFE guide balls. Scaling is approximate.

Table 5.3: Parts acquired for the construction of the final calibration sources for CUORE.

Part	Supplier	Catalog Number	Specifications
String	W. F. Lake	R722-70	PTFE-coated Kevlar; 0.010" (0.254 mm) dia.
Guide balls	—	—	Solid PTFE; 3 mm dia.
Custom cores	UW Physics Instrument Shop	—	Copper; 1.52 mm OD, 3.26 mm length
Tubing	McMaster-Carr	7190K51	Copper alloy 122; $1/16$" OD, 0.0345" ID
Stop sleeves	Nicopress	871-32-B	Copper; $1/8$" OD, $1/4$" approx. length; for use with $1/32$" cable
Heat shrink tubing (small)	Zeus & Small Parts	—	PTFE; 4:1 shrink ratio; $5/64$" ID
Heat shrink tubing (small)	Small Parts	SM4T-0125-24-05	PTFE; 4:1 shrink ratio; $5/64$" ID
Heat shrink tubing (large)	Zeus	—	PTFE; 4:1 shrink ratio; $1/8$" ID
Thoriated tungsten wire	Goodfellow	W145370	0.2 mm dia.; 1% ThO_2 content
Thoriated tungsten wire	Goodfellow	W145400	0.25 mm dia.; 1% ThO_2 content
Thoriated tungsten wire	Goodfellow	W145440	0.38 mm dia.; 1% ThO_2 content
Thoriated tungsten wire	Toshiba	—	0.35 mm dia.; 2% ThO_2 content

and the piece (e.g., the placement of the crimps on the copper crimp tubes) or between two pieces (e.g., the PTFE heat shrink sleeve's coverage of the crimped tube). The manufacturing procedure is described in Section 5.3.3, and Figure 5.30 illustrates problems that can be caused by misalignment during manufacturing. The dimensions of the various components of the source carrier must be highly consistent to ensure the reliability of both the manufacturing procedure and the motion of the source carrier through the guide tubes; uniformity in capsule dimensions is also desirable for the operation of the thermalizers, which are planned to squeeze on multiple capsules at once, requiring that the diameters of the capsules be as similar as possible and that the relative positioning of the capsules and the thermalization clamp be well known.

This section discusses the materials selection for the CUORE source strings and the quality control procedures applied during the preparation of components prior to the assembly of the final CUORE calibration sources. Table 5.3 summarizes the materials acquired for the production of the sources; dimensional tolerances and other quality control concerns are discussed in the sections below that describe the individual source carrier components. The full specifications of the sources can be found in the technical design document for the source strings [54].

5.3.2.1 Kevlar String

As mentioned in Section 5.3.1, Kevlar has high tensile strength and low thermal conductivity; it also has a very small, negative coefficient of thermal expansion (around $-(5-4)\times10^{-6}$ cm/cm/°C) and does not become brittle or show signs of damage when exposed to liquid nitrogen temperatures (around 77 K) [67]. All of these properties make it very well suited to our application.

The majority of the source carrier prototyping work was carried out with 0.014"-diameter Kevlar string obtained from McMaster-Carr (product #8800K41). This string was coated with silicon wax, which exhibited blackening and some slight stiffening under direct exposure to the heat gun used to shrink the PTFE sleeves around the copper capsules (see Figure 5.30). Kevlar itself does not melt, and its decomposition temperature is between 800 and $900\,^{\circ}$F [67], somewhat above the $750\,^{\circ}$F temperature used to shrink the PTFE sleeves. Extended exposure to temperatures above $300\,^{\circ}$F is known to cause strength loss in Kevlar [67], but lab tests demonstrated that about

3 minutes of constant exposure to the heat gun was necessary to cause the Kevlar to break under a load of nearly 1 kg; this effect is considered insignificant for our application, as the mass of the completed source carrier is only a few grams and an exposure of only $10-30$ seconds is typically required to shrink each PTFE sleeve. A sample of uncoated Kevlar was obtained in the hope that the apparent damage to the coating could be avoided, but the uncoated string was very 'fuzzy,' with many fine loose fiber ends sticking out all along the length of the string that made string production difficult and increased the friction of the string moving against the guide tube surface. A sample of PTFE-coated Kevlar was also obtained, but it was thicker and stiffer than the silicon-wax-coated Kevlar and was not sufficiently flexible to allow the source carrier to travel reliably through the guide tubes.

When colleagues in the CUORE collaboration proposed parylene coating the copper support structure for the crystal towers as a method of preventing alphas from surface contamination of the copper from reaching the detectors, we returned to the evaluation of alternate finishes for the string. Parylene coating can be deposited in very thin layers on the order of a few microns for textile conservation applications [91] and is also used for friction-reduction applications. We investigated the possibility of applying a layer of parylene to the silicon-wax-coated Kevlar to test whether this treatment could reduce the coefficient of friction of the source carrier. One sample was prepared at the University of California at Berkeley on the same apparatus used to test the coating of the copper frame pieces; a second was prepared by J. Williams' biomedical engineering lab at UW. Figure 5.18 is a photograph of the sample coated at Berkeley, still mounted on the support frame, which was designed to expose the greatest possible surface area of the string while accommodating a significant length of string in the available volume of the deposition chamber.

The W. F. Lake Corporation, a supplier of coated textile products, manufactures PTFE-coated Kevlar (product #R722-70) in a similar diameter to the silicon-wax-coated Kevlar, 0.010" (0.254 mm). A sample obtained from the company proved to have a similar flexibility to the silicon-wax-coated Kevlar, as well.

To provide context for the final choice of Kevlar material for the production of the CUORE source carriers, we report the results of tests performed by J. Clark once all string samples were

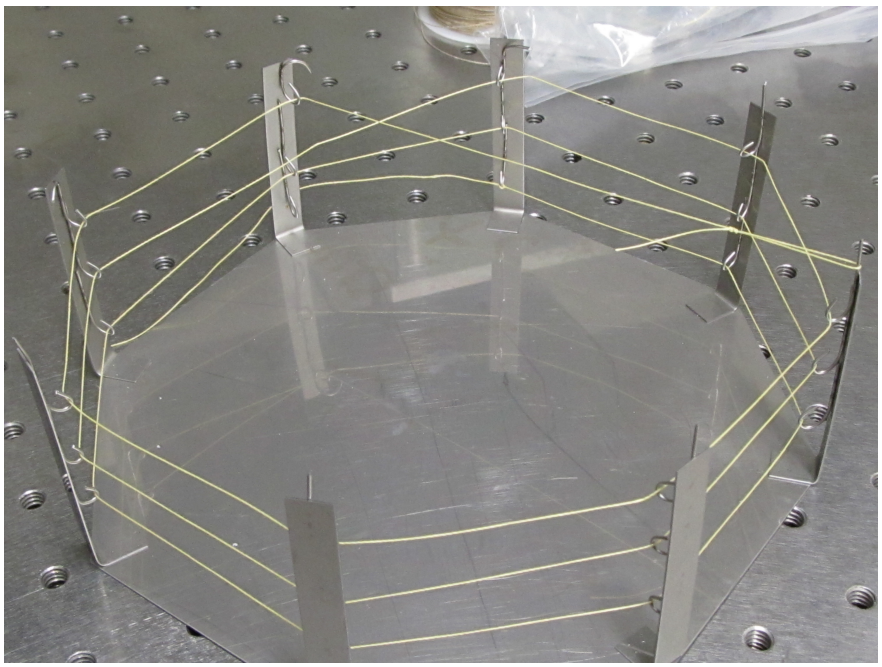
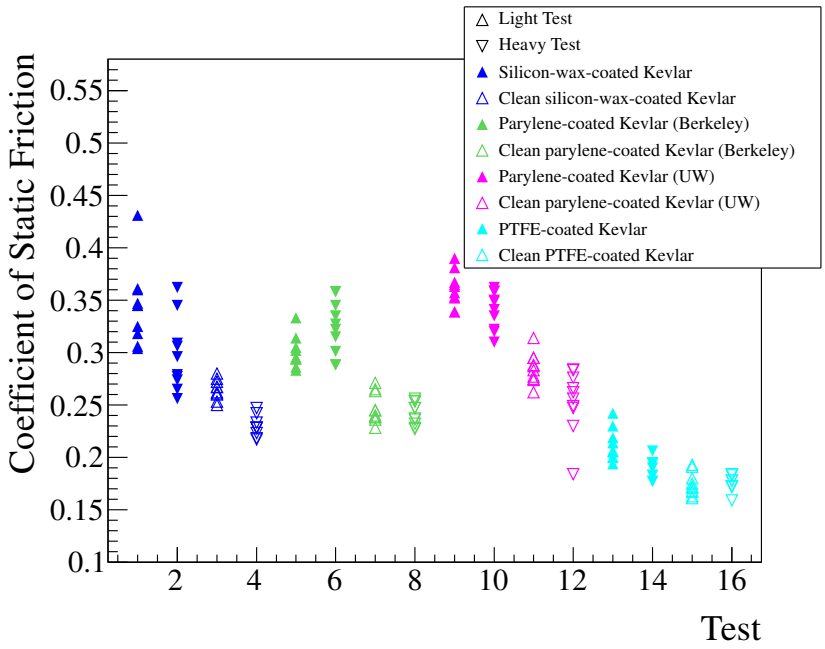


Figure 5.18: Photograph of the frame designed to support the Kevlar string for parylene coating. The string was coated at the University of California at Berkeley. Kinks in the string are visible at the points where the string was contacting the support hooks during coating, illustrating the slight stiffness in the string caused by the coating.

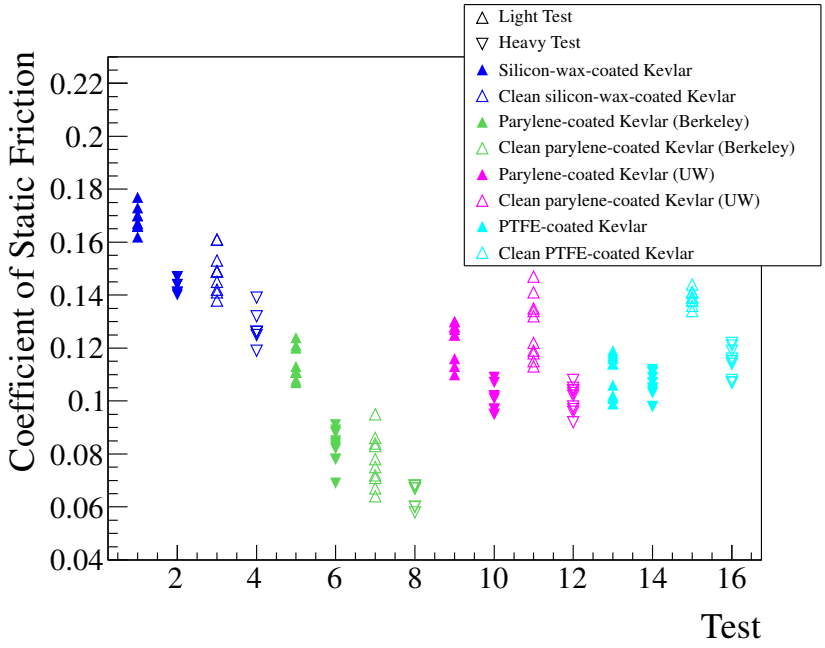
obtained. The coefficients of static friction of all four coated Kevlar strings were measured on rods of copper and PTFE, the two materials that constitute the inner surfaces of the guide tubes (PTFE between the 300-K and 4-K flanges; copper everywhere else). Tests were performed under mass loads in two different ranges: 'light' tests were performed with masses of approximately $12-24$ g, roughly corresponding to the mass of a source string, and 'heavy' tests were performed with masses of approximately $65-130$ g to simulate periods of higher string tension during source extraction. Each string was also tested both before and after cleaning with the ultrasonic procedure defined for the preparation of all source carrier materials for assembly (see Section 5.3.3 and Appendix B) to verify that the cleaning did not compromise the performance of the coating. The friction measurement results are presented in Figure 5.19.

In most cases, there is a significant difference between the 'light' and 'heavy' results on the PTFE rods for the same string, which we believe can be attributed to deformation of the soft plastic under the increased weight. Cleaning appears to appreciably decrease the friction of the Berkeley parylene string on PTFE, while it has little effect on and may even slightly increase the friction of the other strings. However, the friction of all four strings on PTFE is much lower than the friction of the strings on copper, which is expected; due to this and the fact that the PTFE guide tubes are mounted only in the upper levels of the cryostat, where the cooling power is greatest and the friction is therefore of least concern, our choice of coating is driven by the friction on copper. Cleaning unambiguously reduces the friction of each string on copper, and there is little difference between the 'light' and 'heavy' test results. The lowest friction on copper is seen with the PTFE-coated Kevlar, and in fact, on copper, the parylene-coated strings do not show any significant improvement with respect to the silicon-wax-coated string.

Following the friction study, some of the mechanical reliability studies performed earlier on the silicon-wax-coated string were repeated on the PTFE-coated string to verify that the PTFE-coated string was the best choice for final source carrier production. No visible damage to or flaking of the PTFE coating was observed after thermal cycling in liquid nitrogen or abrasion, and the silicon-wax-coated string unraveled more easily under forced fraying than the PTFE-coated string. Prolonged exposure to the heat gun caused far less discoloration to the PTFE-coated string



(a) Static coefficients of friction of coated Kevlar strings measured on copper.



(b) Static coefficients of friction of coated Kevlar strings measured on PTFE.

Figure 5.19: Measurements of the static coefficients of friction of Kevlar string with different coatings on rods of copper and PTFE. Tests were performed with mass loads of $12-24$ g ('light') and $65-130$ g ('heavy'), and both before and after ultrasonic cleaning. Figures from J. Clark.

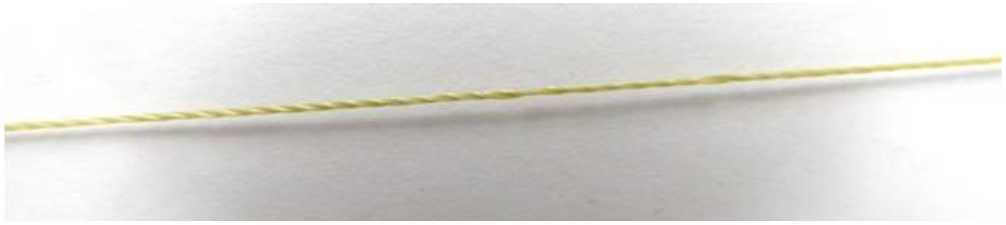


Figure 5.20: Photograph of PTFE-coated Kevlar after prolonged ($\approx$ 3 minutes) exposure to heat gun. Discoloration is less pronounced than that observed on silicon-wax-coated Kevlar after similar exposure (cf. Figure 5.30). Photograph from J. Clark.

(see Figure 5.20) than the silicon-wax-coated string showed under similar exposure, supporting the hypothesis that the discoloration is an indication of damage to the coating rather than to the Kevlar itself, and the string manufacturing procedure (see Section 5.3.3) proved equally successful with the PTFE-coated Kevlar as with the silicon-wax-coated Kevlar. A prototype string built with PTFE-coated Kevlar has now been extensively motion tested in the UW lab. As a result of this testing, the PTFE-coated Kevlar was chosen for the production of the final CUORE sources. Each source string was constructed on a piece of PTFE-coated Kevlar of a total length of 4.5 m.

5.3.2.2 PTFE Guide Balls and Copper Cores

The design of the bottom guide ball proceeded directly from the original ball chain source concept; the idea is to provide a gradually curved surface that is the first part of the source carrier to enter any guide tube section during downward motion, minimizing the likelihood that the string will become stuck and fail to feed into the tube. A low-friction outer surface of the ball is desirable for the same reason, leading to the choice of the PTFE bead. The copper core allows the guide ball to be attached with a crimp just like the capsules, as discussed in Section 5.3.1. All crimp tube material on the string is copper; copper was chosen because it is relatively radioclean, it is suitable for and widely used in cryogenic environments, and it is soft enough that it can be easily crimped by hand.

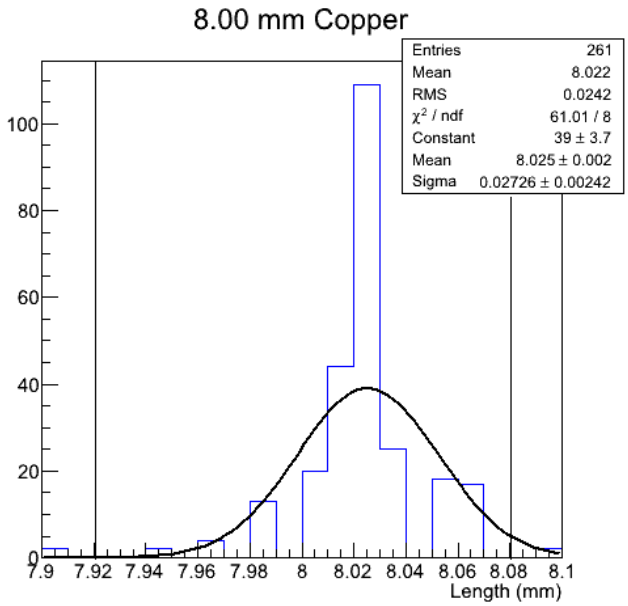
Solid PTFE balls were acquired and drilled out into beads by hand at UW. Every ball used in production of the final CUORE sources was measured to have an outer diameter of 3.01 mm with a tolerance of ± 0.03 mm, or $\pm 1\%$, and verified that the holes drilled through the balls were centered.

Because the PTFE is soft, the press-fit between the ball and the core can accept a slightly larger than 1% tolerance on the copper cores. The relevant concern is merely that the core should not be long enough to extend past the top of the ball. The final dimensions and tolerances measured for the cores are a length of 3.26 ± 0.05 mm and an outer diameter of 1.52 ± 0.03 mm. The inner diameters were simply verified to be large enough to be easily threaded onto the PTFE-coated Kevlar.

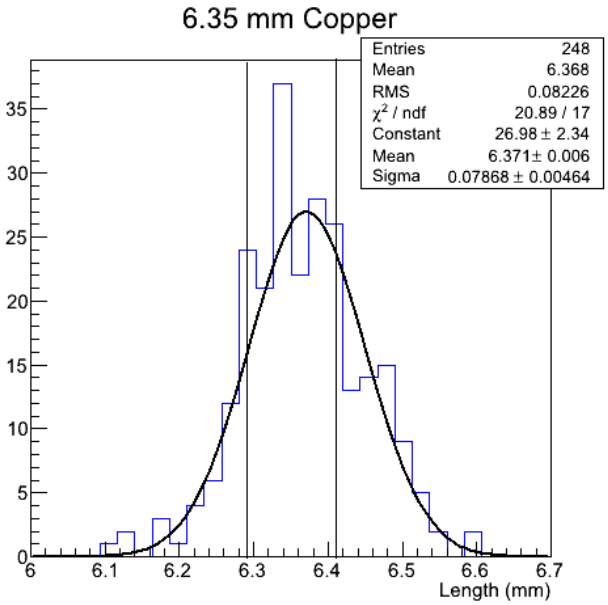
5.3.2.3 Copper Crimp Tubes

The baseline tolerances for the crimp tube dimensions were set at 1% of the nominal part dimensions to ensure good consistency among the capsules. For the source capsules, which are cut to length from 1-ft-long pieces of copper tubing (McMaster-Carr product #7190K51), the nominal dimensions and tolerances were set to a length of 8.00 ± 0.08 mm and an outer diameter of 1.59 ± 0.02 mm; for the weight capsules, which are the commercially available Nicopress stop sleeves intended for use with $1/32$" cable (Nicopress stock #871-32-B), the nominal dimensions and tolerances were set based on the specifications reported by the manufacturer to a length of 6.35 ± 0.06 mm and an outer diameter of 3.18 ± 0.03 mm. Before placing the orders for the final production materials, however, a sample of the remaining stock that had been acquired for prototyping was measured to determine the real dimensions, estimate rejection rates in case any overpurchasing might be necessary, and establish whether any tolerances could be realistically reduced. The sample consisted of 261 source-capsule-sized crimp tubes and 248 weight-capsule-sized crimp tubes.

The results of the length measurements are summarized in Figure 5.21. Based on these results, the original length tolerances of 8.00 ± 0.08 mm and 6.35 ± 0.06 mm for the source-capsule copper and weight-capsule copper, respectively, were confirmed. In the case of the source-capsule copper, this tolerance led to a rejection rate of only 3%. The central value of the distribution was



(a) Small copper tubes used for source capsules.



(b) Large copper tubes used for weight capsules.

Figure 5.21: Measured lengths of the sample of copper crimp tubes used to verify tolerances prior to the final materials purchase. The vertical lines indicate the boundaries of the accepted tolerance range for each tube size. Figures from J. Clark.

Table 5.4: Final dimensions and tolerances defined for the copper crimp tubes used to assemble the final CUORE source carriers. All values are in units of millimeters.

Capsule type	Length	Length tolerance	Outer diameter	Outer diameter tolerance
Source	8.00	$\pm$0.08	1.61	$\pm$0.01
Weight	6.35	$\pm$0.06	2.81	$\pm$0.01

8.02 mm instead of 8.00 mm, but we chose not to adjust the central value of the tolerance range accordingly; we expect that any future batch should have a similar standard deviation but might have a slightly different central value because the cutting setup would have to be reestablished. In the case of the weight-capsule copper, the 6.35 ± 0.06 mm tolerance led to a high rejection rate of 34%, and the measured central value was 6.37 mm. This slight shift from the nominal central value was considered insignificant in comparison to the observed standard deviation, so we chose not to adjust the central value of the tolerance range in this case, either. We chose to maintain the ± 0.06 mm tolerance to ensure consistency in the final manufactured weight capsules and overpurchased the Nicopress sleeves for final CUORE source production to compensate for the anticipated rejection rate.

The great majority of the measurements of the outer diameters of both sizes of tubes were within the ± 0.01 mm precision of the digital caliper used for the measurements of the central value of the measurements, which was 1.61 mm for the source-capsule copper and 2.81 mm for the weight-capsule copper. Although the measured diameters differed from the nominal diameters given by the manufacturers, they were extremely self-consistent; in fact, none of the small crimp tubes and only two of the large crimp tubes were more than 0.01 mm different from the central diameter. Based on these results, we defined final outer diameter tolerance ranges of 1.61 ± 0.01 mm and 2.81 ± 0.01 mm for the source-capsule and weight-capsule crimp tubes, respectively, based on the measured dimensions and the precision of the measuring device. This very high diameter precision will be beneficial to the operation of the thermalization mechanisms.

The final dimensional tolerances are summarized in Table 5.4. Every copper crimp tube used in the production of the final CUORE calibration sources was measured and was rejected if it fell outside the defined tolerance range in either length or outer diameter. Crimp tubes that were out-of-tolerance due to excessive length were saved to be cut down in case of a shortage of materials, but this contingency never arose.

5.3.2.4 PTFE Heat Shrink Sleeves

PTFE heat shrink tubing is available in several shrink ratios, where the ratio refers to the diameter of the tubing before and after shrinking. During early prototyping, tubing with the smallest available diameter that would fit over the copper and shrink ratios of both 2:1 and 4:1 was tested; it was found that the 4:1 tubing provides good coverage of the copper crimp tubes with no cracking, splitting, or other visible stress, while the 2:1 tubing did not shrink tightly enough to provide well-shaped capsule ends. This result was carried over to the weight capsules. Before shrinking, the heat shrink tubing used for final production is slightly less than 0.4 mm larger in diameter than the copper for both the source capsules and the weight capsules.

Heat shrink tubing is available in several lengths; batches in lengths of 1, 2, and 4 ft were acquired for CUORE calibration source production and cut into sleeves for the capsules. Cutting precision at the same level as the length tolerances of the copper crimp tubes is not necessary, as heat shrink sleeve length differences of a few tenths of a millimeter can be absorbed by the ends of the capsule — as long as the hard edges of the crimps are fully covered, the ends of the sleeves may, but do not need to, shrink fully down to contact the Kevlar. PTFE heat shrink is soft and flexible, and careful hand-cutting with a meter stick and a precision knife is sufficient for the preparation of the sleeves. However, several prototype strings were constructed over time out of different batches of heat shrink, and this experience showed that the tubing shrinks lengthwise as well as radially and that the lengthwise shrink ratio differs slightly from one batch to another. The optimum pre-shrunk sleeve length can differ by as much as $1-2$ mm depending on the tubing batch from which the sleeve was cut, and while the small PTFE ‘step’ formed on one or both ends of a capsule by an excessively long sleeve is preferable to the incomplete copper coverage

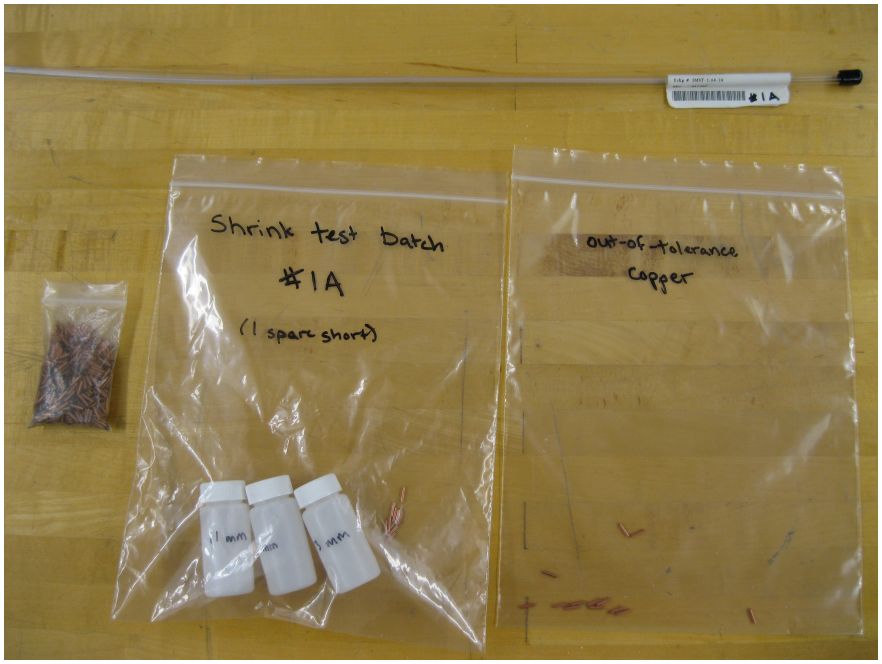


Figure 5.22: Photograph showing the preparation of the shrink test of the first batch of PTFE heat shrink tubing. Each PTFE batch was assigned a batch label (in this case, #1A), and the test parts and the remaining production tubing were both marked with this label; test sleeves were stored by length in individually labeled bottles and bagged with confirmed in-tolerance copper crimp tubes for construction of test capsules. The "(1 spare short)" notation indicates that one piece was miscut in the process of preparing the shrink test parts, leaving the remaining production tubing short by the length of approximately one sleeve.

resulting from an excessively short sleeve, it can still cause enough string motion issues that it is not acceptable to choose a single long sleeve length to cut from all batches.

Instead, a part of each heat shrink batch acquired for CUORE calibration source production was set aside for shrink testing to determine the optimum sleeve length for each batch individually. The measurement of the copper tube sample presented in Section 5.3.2.3 was used to set aside sufficient in-tolerance crimp tubes to construct test capsules for these shrink tests, and there was sufficient silicon-wax-coated Kevlar remaining from prototyping to serve as test string. For each batch of PTFE tubing, nine test capsules were manufactured: for the small tubing used for the source capsules, three sleeves each at lengths of 11, 12, and 13 mm were cut, while three sleeves each at lengths of 10, 11, and 12 mm were tested from each batch of the larger tubing used for the weight capsules. The test lengths were chosen based on the typical optimum sleeve lengths identified during prototyping. Figure 5.22 shows the preparation of parts for the shrink test of the first batch of small tubing. After the initial nine test capsules were manufactured, the length that resulted in the best-quality capsule coverage (full coverage of copper crimp edges and no PTFE 'step') was designated as the correct sleeve length for that batch; if an intermediate length appeared to be preferred, two additional test sleeves at that length could be cut and tested for confirmation. Ultimately, one of the original test lengths was identified as the optimum sleeve length for all but two PTFE batches: 11.5 mm was chosen for one batch of small tubing, and 12.5 mm was chosen for one batch of large tubing [54].

As in the case of the copper crimp tubes, the length of each heat shrink tube was measured with a digital caliper. Because the PTFE is both slick and easily crushed, the manipulation of the caliper for the measurement is difficult, and two different people often measure the same sleeve slightly differently. To quantify this measurement effect, all test sleeves cut for the first nine batches of the small tubing were measured independently by two different people; the results are shown in Figure 5.23. The relative measurement uncertainty was determined to be small by comparison to the cutting precision, so a single person cut and measured the sleeves for the remaining shrink tests. The distributions of sleeve lengths for all shrink tests are shown in Figure 5.24. These results show that the cut lengths skew slightly longer than the nominal lengths, which is intentionally done to

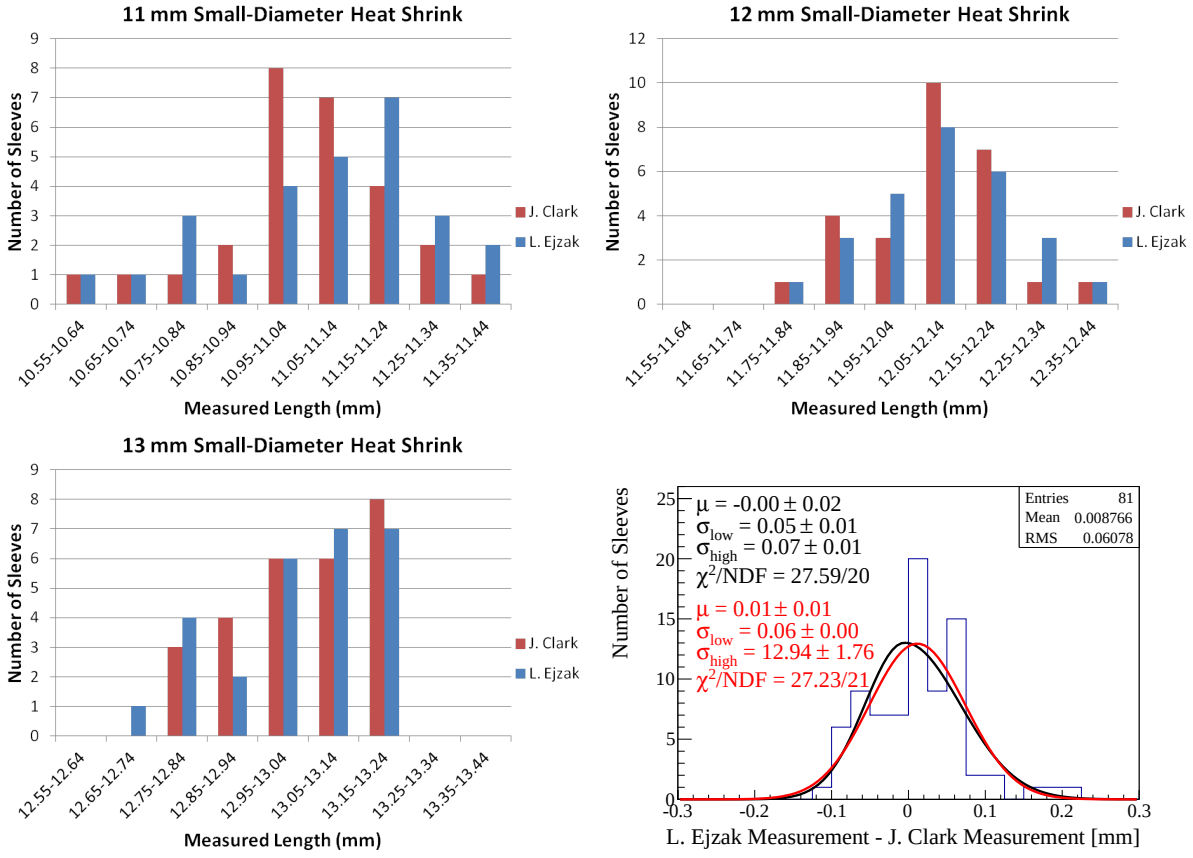


Figure 5.23: Lengths of the PTFE heat shrink sleeves prepared for the shrink tests of the first nine batches of small-diameter PTFE heat shrink tubing, as measured by two different people. The bottom right plot shows the distribution of the difference between the two length measurements for all sleeves, fit with both a Gaussian (red) and an asymmetric Gaussian (black); the means of both fits are consistent with zero, indicating no overall bias, and the spread is somewhat smaller than the spread in absolute lengths.

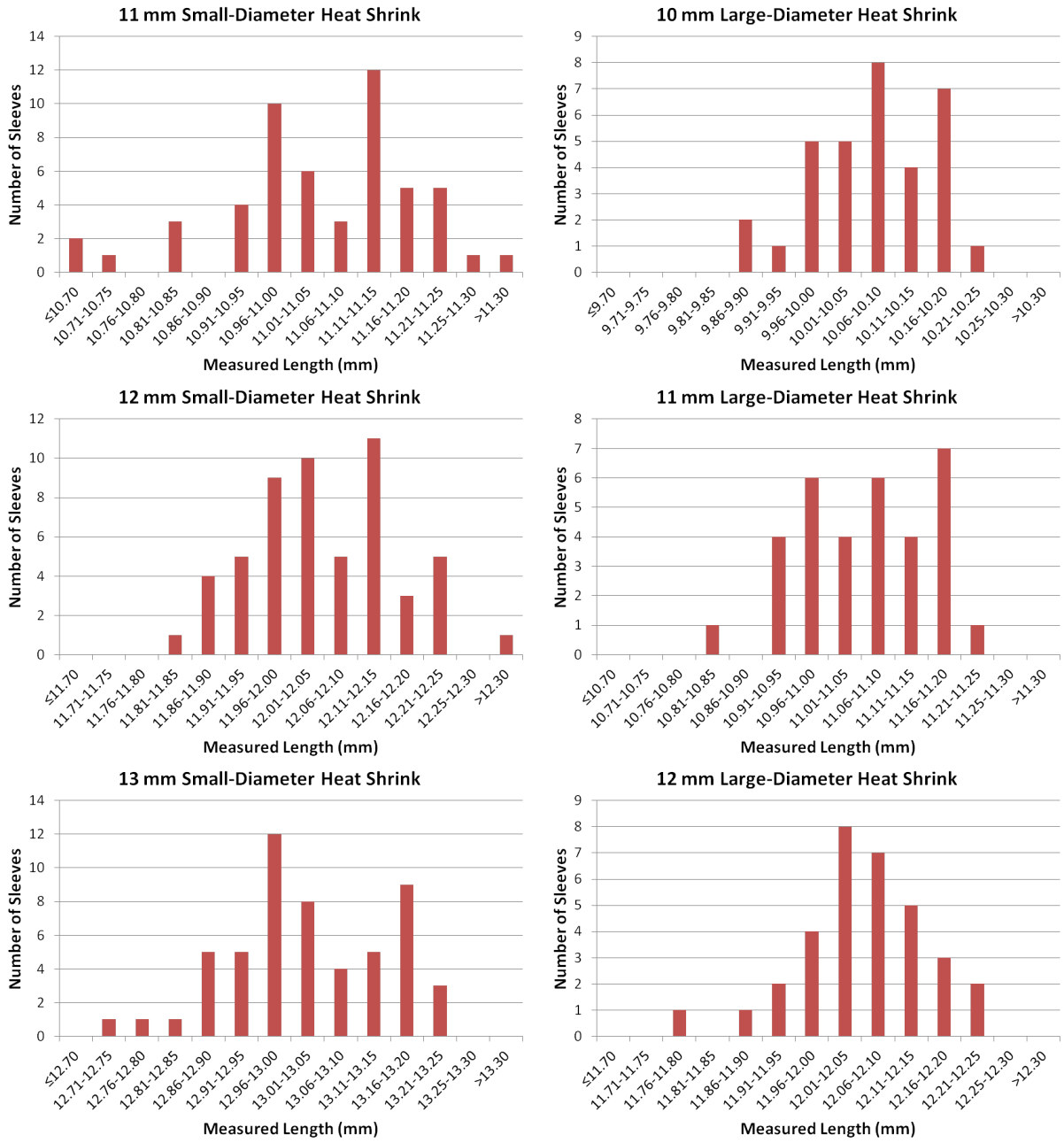


Figure 5.24: Lengths of the PTFE heat shrink sleeves prepared for all shrink tests. Based on these results, a cutting tolerance of ± 0.3 mm was determined to be realistic for the final production sleeves. The cut lengths skew slightly longer than the nominal lengths, which is intentionally done to avoid the risk of leaving exposed copper after shrinking.

avoid the risk of leaving exposed copper after shrinking. They also confirm hand-cutting precision of a few tenths of a millimeter for the PTFE heat shrink sleeves, sufficient for production. Due to the measurement uncertainty illustrated in Figure 5.23, however, each sleeve cut for final CUORE calibration source production was independently measured by two different people to confirm a length within ± 0.3 mm of the intended length.

5.3.2.5 Radioactive Source Wire Inserts

As discussed in Section 5.2.1, the calibration sources used throughout the lifetime of Cuoricino use thoriated tungsten wire from Goodfellow to provide their radioactive material. We have chosen to continue to use thoriated tungsten wire in CUORE because the collaboration has gained considerable expertise in performing calibration analysis with this source, and the 1.4×10^{10} y half-life of ^{232}Th [68] means that the sources should never need to be replaced due to natural loss of activity. However, cutting thoriated tungsten wire to lengths of a few millimeters, as required by the CUORE calibration source design, poses a manufacturing challenge. The wire is produced by mixing thorium dioxide powder into tungsten powder, then pressing, heating, and extruding the material. The resulting wire is very brittle; when cut by a tool that puts the wire under compression, such as a typical pair of wire cutters, the cut ends of the wire crack, split, and fray. These rough edges catch on the Kevlar during insertion into the capsules, making it difficult to correctly position the inserts. Tungsten wire is also very stiff and springy, making handling and manipulating it difficult.

For early testing, some tungsten wire pieces were cut by hand in the UW lab using a cutting jig, shown in Figure 5.25. It allowed several strands of wire to be stretched under tension, then tightly sandwiched between two aluminum blocks, with a sheet of paper pressed between the wires and each clamping surface to reduce the possible rolling and shifting of the wires. The width of the clamping pieces was equal to the desired final length of the wires. The wires were rough-cut with a rotary grinding tool, buffed flush with the sides of the clamp with very fine-grit emery paper, then released from the clamp and ultrasonically cleaned to remove any residual powder. This procedure allowed for reasonable cutting precision and produced clean, good-quality wire ends, but it was

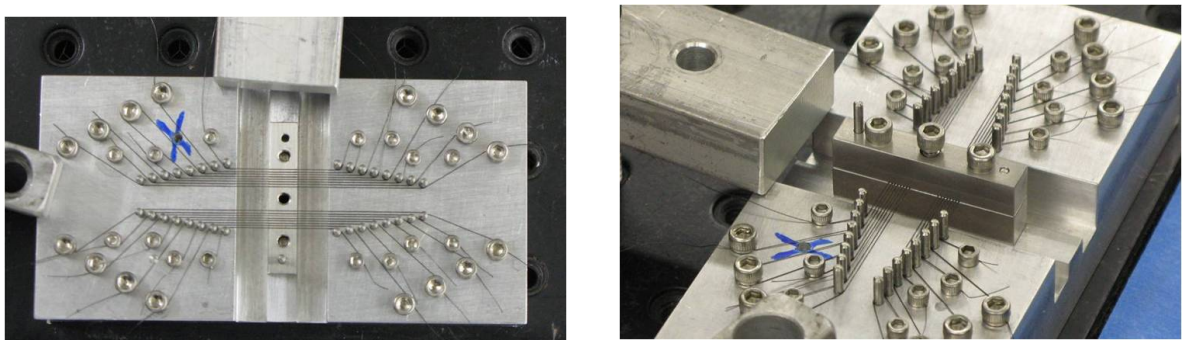


Figure 5.25: Two views of the jig used for hand-cutting thoriated tungsten wire inserts. Left: Top view after stringing wire over bottom clamp block. Right: Angle view after clamping wire between both clamp blocks. After this, a rotary grinding tool was used to rough-cut the wires, then the clamp block assembly was removed from the main jig to buff the wire ends flush with the sides of the blocks.

extremely labor intensive and did not easily allow for flexibility in cutting lengths. The waste ratio was also very high, as can be seen from Figure 5.25.

For final CUORE source production, the tungsten wire was cut by wire-electro-discharge machining (wire EDM) by Wire Works Engineering in Madison, WI. The cutting path in wire EDM is programmable and software-controlled, providing the desired flexibility in cutting lengths, and a prototype job submitted to the shop demonstrated a cutting precision within a few hundredths of a millimeter, corresponding to an activity uncertainty of only around 1% for wire lengths of a few millimeters, and resulted in wire ends with clean, hard corners and no fraying, shown in Figure 5.26. The wires were submitted for cutting loaded into jigs designed following a similar principle to the hand-cutting jig, firmly holding multiple wires under compression inside a larger, more easily manipulated piece. These jigs were single-use, however, as they were cut to size along with the wires according to a prescribed cutting path, and they allowed for the cutting of a far greater number of pieces, producing much less waste as a result. Figure 5.27 shows photographs of the jig submitted for the prototype job after the EDM cutting was performed.

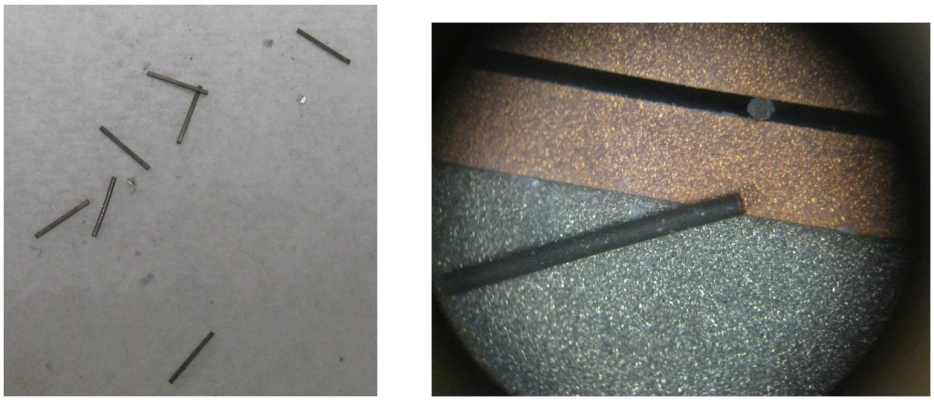


Figure 5.26: Prototype wire inserts cut with wire EDM by Wire Works Engineering. Left: Several inserts after removal from the cutting jig as seen with the naked eye. Right: Microscope view. One wire can be seen still clamped between the copper surfaces of the prototype cutting jig; the jig and the wire were both cut simultaneously. The cut end of the free wire has clean, hard corners and no apparent fraying.

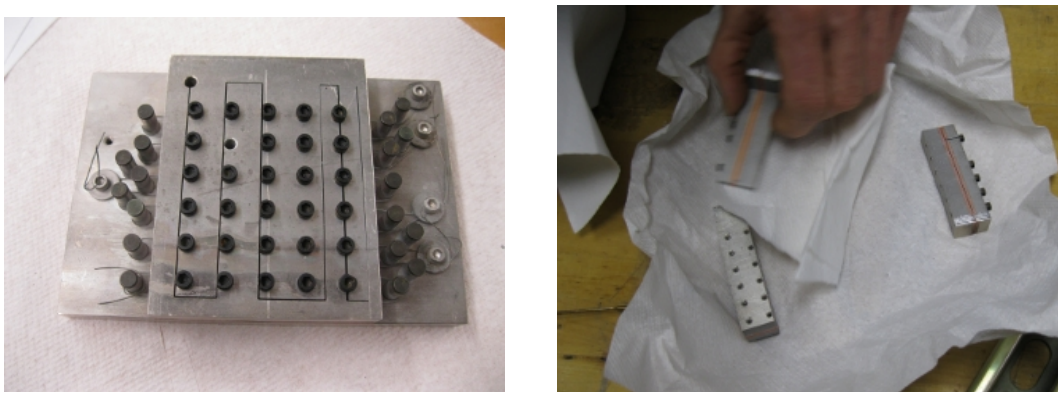


Figure 5.27: Prototype cutting jig submitted to Wire Works Engineering after the cutting job was performed. Left: Jig as returned by the shop. A mistake in the programming of the EDM apparatus resulted in the cut path passing through the centers of one row of screws. Right: Jig after being cut into sections at UW in preparation for disassembly and recovery of cut wire pieces.

The thoria content, size, and number of wire inserts in each capsule determines the activity distribution carried by the source carrier. Section 5.5.3 shows the results of simulations with the intended source activity distribution, which provides reasonably even illumination of all detectors and event rates consistent with the limits established in the high-rate calibration tests presented in Section 5.4. The activity per length as referred to the parent nuclide, ^{232}Th, of a piece of thoriated tungsten wire can be calculated as follows:

$$\text{activity/length} = \frac{ln(2)}{T_{1/2}(^{232}\text{Th})} \frac{N_A}{M(\text{ThO}_2)} (\text{thoria content}) \pi (\tfrac{d}{2})^2 \rho,$$

where $T_{1/2}(^{232}\text{Th})$ is the half-life of ^{232}Th (1.41×10^{10} years), N_A is the Avogadro constant, $M(\text{ThO}_2)$ is the molar mass of ThO_2 (264.04 g/mol), d is the diameter of the wire, and ρ is the density of the wire (19.3 g/cm^3 for thoriated tungsten). The stoichiometric correction and isotopic-abundance correction that would be present in the general case are omitted here because there is only one Th atom per molecule of ThO_2 and the isotopic abundance of ^{232}Th is $\approx 100\%$.

Some tradeoff is possible among thoria content, length, diameter, and number of pieces to achieve the desired activities per capsule. However, the options are constrained by the available wire diameters and thoria content, as well as by ease of handling and the space available inside each capsule. For the reliable manual manipulation of the wire inserts with tweezers, it is desirable that the wire diameters be no smaller than 0.2 mm and the lengths no smaller than 2 mm. The internal diameter of the capsules is around 0.88 mm, but this space must host the Kevlar string as well as any active wire inserts, so to avoid pushing the Kevlar off-center in the capsule as much as possible, a wire diameter of no larger than 0.4 mm is preferred. The maximum insert length that can be accommodated inside a capsule is somewhat dependent on the diameter of the wire, due to the deformation of the inside of the ends of the capsule from the crimps, and on the number of wires, due to the possibility for a single wire to sit in the capsule at an angle; however, to avoid the risk that a wire may become caught in the top crimp of a capsule, we prefer that the wire length not exceed half the total length of the copper crimp tube, opting for multiple shorter insert wires instead when possible.

Figure 5.28: Wire insertion test for source capsules. After inserting the wires and crimping the capsule closed, the capsule was cut open and the wires removed and inspected under magnification. In this case, three of the four inserts remained in the left-hand side of the capsule (the bottom end during insertion) and the fourth remained in the right-hand (top) side when the cut halves were separated. A successful insertion test is one in which the wires can be pulled free of the capsule easily and show no deformation on the end, indicating that they were not caught in the top crimp as it was placed.

Tungsten wires with 1% thoria content are readily available from Goodfellow in a variety of diameters; the two that best suit our physical constraints as outlined above are 0.25 mm and 0.38 mm. Wires with greater thoria content than 1% are difficult to find, however. 1.6%, 1.7%, and 2% thoriated tungsten is most widely available for welding purposes in rods of 6 mm in diameter or more, and it does not appear to be possible to acquire thoriated tungsten with thoria content greater than 2%. The smallest 2% thoriated tungsten wire that we were ultimately able to acquire was produced by Toshiba in a diameter of 0.35 mm.

Although the weight capsules are shorter than the source capsules, their much closer spacing means that the weight capsules also require less than half the activity per capsule required by the source capsules to achieve the same overall activity per length of source carrier. The maximum linear activity limit is therefore set by the amount of material that can be accommodated by a source capsule. When Goodfellow was our only source of thoriated tungsten wire, wire insertion tests with the silicon-wax-coated Kevlar showed that the practical maximum amount of material that could be inserted in a source capsule corresponded to 4 pieces of approximately 3.6-mm-long, 0.254-mm-diameter wire (see Figure 5.28), indicating that it was impossible to build strings with the activity distribution desired for the external sources with the available wire. Once the 2%

wire was obtained from Toshiba and the baseline string was changed to the PTFE-coated Kevlar, however, a similar test demonstrated that a source capsule could now accommodate 3 pieces of 3.9-mm-long, 0.35-mm-diameter wire. The wire specifications required to achieve the desired activity distributions for CUORE calibration sources are summarized in Table 5.7; these are the specifications used to construct the final CUORE sources.

5.3.3 Source Carrier Manufacturing

Once all components of the source carrier have been prepared, the full source carrier must be assembled. For each capsule (of both types, source and weight), this requires the following steps:

1. Thread copper crimp tube onto Kevlar string; place in desired position.
2. Crimp one edge ("bottom" edge) of tube to string.
    - Crimp footprint: A four-pin indenter provides a reliable crimp with little deformation of the overall shape of the capsule. See Figure 5.7 for a schematic illustration of the crimp cross-section and Figure 5.29 for photos of the tool used.
    - Crimp depth: The crimp tool has an adjustable crimp depth setting. To minimize risk of potential damage to the Kevlar, the shallowest crimp depth setting at which the crimped piece will not slide along the Kevlar is chosen.
3. Insert active source wire into crimp tube.
4. Crimp other edge ("top" edge) of tube to string, securing source.
5. Thread PTFE heat shrink sleeve onto Kevlar string; slide over crimped copper tube and center.
6. Heat PTFE sleeve to shrink onto copper.
    - Heating method: A variable-temperature, variable-fan-speed heat gun is used to shrink the PTFE sleeves. Due to the moving air from the heat gun, some care is necessary in maintaining the positioning of the sleeve.

Figure 5.29: Photos of the crimp tool used in the UW lab for source carrier prototyping and production. Left: Open; ready to insert piece for crimping. Right: Fully closed; the four indenter pins at their deepest extent in the crimp depth setting used to manufacture the source carriers.

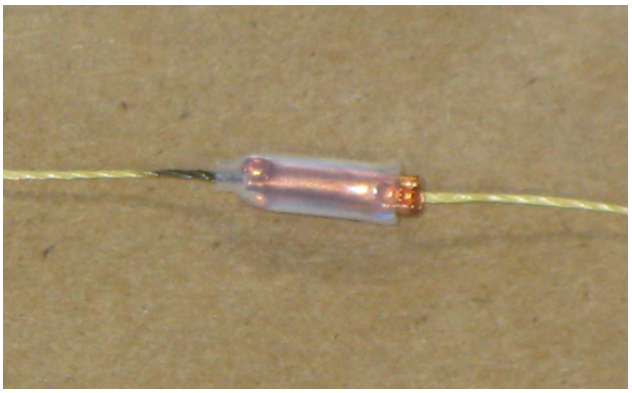


Figure 5.30: A poor-quality source capsule; compare with Figure 5.8. This capsule was assembled on silicon-wax-coated Kevlar instead of the PTFE-coated Kevlar used for final source carrier production; blackening of the wax coating from the heat gun can be seen on the left-hand side of the capsule. There is also a 'step' formed in the PTFE on the left-hand side of the capsule due to excess PTFE extending past the copper tube. On the right hand, the PTFE fails to cover the sharp edges of the crimp on the right-hand side and also forms another hard edge. Additionally, the crimp on the right-hand side is positioned too far from the edge of the tube, restricting the space available to host the active source material inside the capsule.

- Temperature: The nominal shrink temperature for the PTFE sleeves is $660\,^{\circ}$F; however, lab tests showed that the sleeves do not shrink tightly enough around the crimp tubes at this temperature. $750\,^{\circ}$F was chosen as a good compromise temperature between achieving a quick, tight shrink and avoiding as much potential damage to the Kevlar string as possible.

All of these tasks can be performed free-hand for the source capsules, and indeed this is how early prototyping work was done. However, quality control is very important for the capsules, as a misalignment of the crimp tool or the PTFE sleeve can create hard edges that interfere with source carrier motion, and the shorter length of the weight capsules makes them very difficult to position in the crimp tool by hand. Figure 5.30 demonstrates many of the problems that must be avoided when manufacturing a capsule.

To avoid many of these problems, a manufacturing jig was designed to provide the operator much better repeatability of the positioning of the various components of the source carrier fabrication process. The design of the manufacturing jig for the source carrier (see Figure 5.31) is based on a double-rail system. The Kevlar string is mounted vertically in such a way that the entire active length of the source carrier is accessible at once, and the elements of the jig move along the length of the string to attach and assemble the capsules. All the necessary tools are mounted on traveling support arms that slide up and down one rail or the other, except the heat gun, which is manipulated separately. A single main reference bar, with a hole through the center to allow the Kevlar string to pass through it, is attached to both rails and is used as the reference point for positioning the tool arms for each operation by means of placing a spacer between the reference bar and the tool support arm; the main reference bar remains in one position until the completion of one capsule, then moves to the next capsule position, and so on. This ensures that the positioning of every element of a single capsule will be referenced to the same zero point. In addition, the main reference bar itself is positioned by a scale on each rail (see Figure 5.32); each end of the reference bar is set to the same position on its corresponding scale to ensure that the bar remains level, and a predefined set of capsule positions is defined based on the desired spacing between capsules. This ensures that the positioning of each capsule is also referenced to a common zero, defined by the

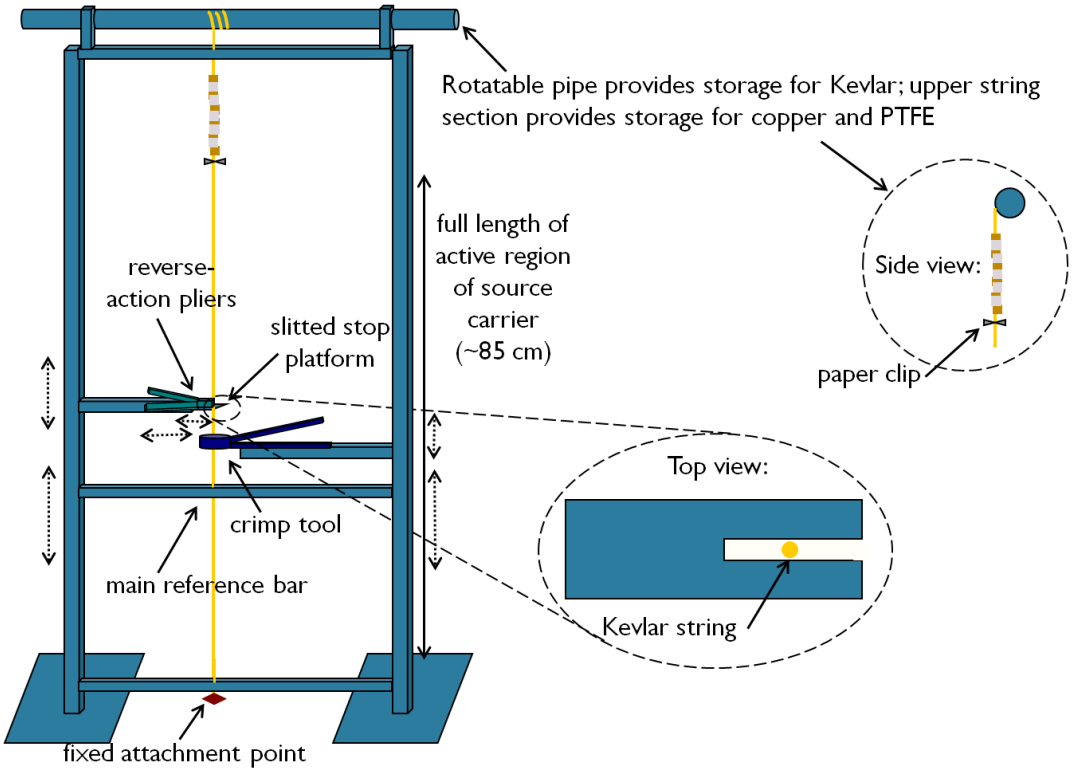
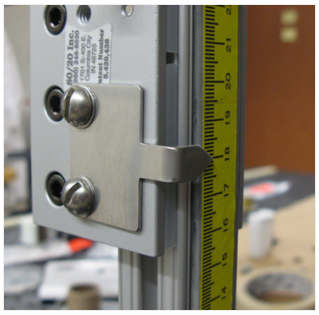
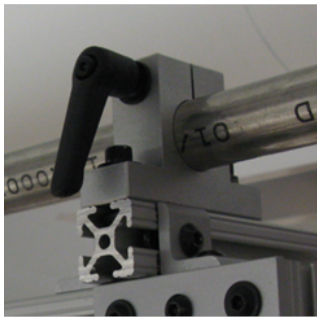


Figure 5.31: The schematic design for the source carrier jig. The string is affixed to the top and bottom attachment points and held vertically under tension. Each tool is supported on its own support arm and can be moved out of the way when necessary. The slitted stop platform positions the copper crimp tube in the pliers; the pliers in turn position the crimp tube in the crimp tool to place the first (bottom) crimp. Spacers between the crimp tool support arm and the main reference bar set the crimper positions for both the top and bottom crimps; a spacer between the pliers support arm and the main reference bar sets the position for the slitted stop platform to support the PTFE sleeve during shrinking with the manually manipulated heat gun.



(a) Each side rail is marked with a scale in units of millimeters, and each end of the main reference bar is equipped with a pointer for alignment at predefined capsule positions.

(b) The non-active Kevlar is stored on the top pipe during source carrier assembly.



(c) Quick-release clamps hold the pipe in position and allow it to be easily shifted or rotated to adjust the centering or tension of the string.

Figure 5.32: Photographs of elements of the source carrier assembly jig.

scales on the jig, instead of indexing each capsule off the placement of the previously assembled one, which would compound the placement error of each successively assembled capsule.

The jig is constructed out of 10 Series 80/20 aluminum framing because there are many pieces of commercial standard hardware readily available that are well suited to the requirements of the design, including attachment hardware that provides a stable $90\,^{\circ}$ junction between two pieces of framing, quick-release stanchion clamps, and linear bearing assemblies for sliding carriages. Holes are drilled through the jig's three horizontal crossbars to form a path for the string, and quick-release clamps support a pipe at the top of the jig that provides storage space for the non-active length of string (see Figure 5.32). After attaching the string at the bottom fixed attachment point and feeding it through the hole on the main reference bar and the crimp tool, all copper and PTFE parts are loaded onto the string; the string is then fed through the hole in the top crossbar, affixed to the pipe with masking tape, and wound up on the pipe until the pipe is clamped in place with the string is under tension. All copper and PTFE parts are stored at the top of the string, with a paper clip supporting them and keeping them out of the working area; capsules are assembled

starting with the bottom capsule, and each part is slid down the string from the storage area above the paper clip as needed.

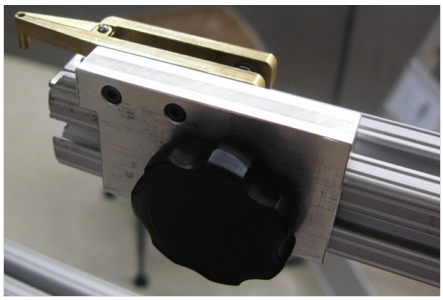
There are three main types of tools mounted on two support arms, one on each rail on its own sliding carriage:

- Crimp tool: It is a closed ring, so the Kevlar string must pass through the center of it at all times. It is rigidly fixed to the right-hand tool support arm, or crimp tool support arm. Care must be taken when sliding it up or down the string that it does not catch on completed or partially completed capsules.

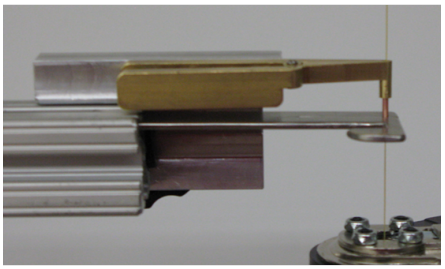
- Reverse-action pliers: Pliers originally custom-designed to fit the copper tubes for the source capsules for positioning in the crimp tool; they are also used to position the larger-diameter weight capsules, although the operator must compensate for the fact that the grip of the pliers on the larger copper is less secure and somewhat prone to slippage. They are mounted on a small sliding carriage on the left-hand tool support arm; this carriage allows the pliers to slide out to grip a copper crimp tube and position it in the crimp tool for the placement of the bottom crimp, then slide back to clear the crimp tool.

- Slitted stop platform: A flat stop platform with a slit wide enough to allow the Kevlar string to pass through but narrow enough to stop the crimp tubes and the heat shrink sleeves. It is mounted on the left-hand tool support arm, also called the pliers support arm. It slots into the end of the 80/20 and can slide out to position a copper or PTFE sleeve and slide back to clear the crimp tool.

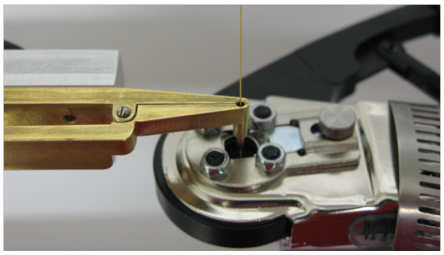
When the stop platform and the pliers are both extended and the set screw on the pliers carriage is fully tightened, the relative vertical alignment of the stop platform and the pliers positions the copper tube in the pliers such that when the platform is retracted and the pliers are lowered until they rest on the surface of the crimp tool, the copper tube is correctly positioned in the crimp tool for the placement of the bottom crimp. Therefore, it is not necessary to use a spacer to position the pliers support arm in reference to the main reference bar for the placement of the bottom crimp; only the crimp tool support arm must be positioned with a spacer for the bottom crimp. The crimp



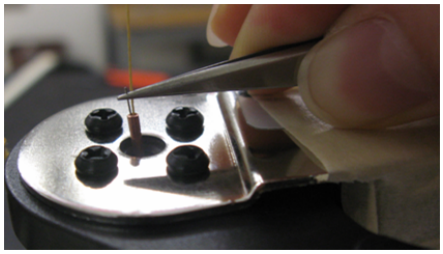
(a) Back view of pliers carriage showing knob for set screw.



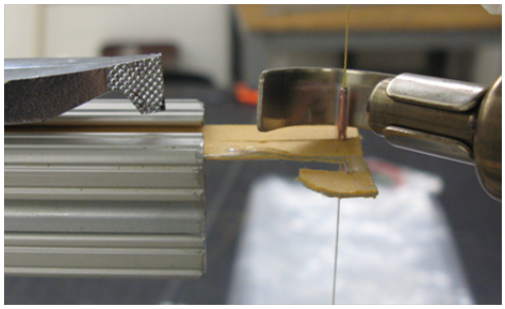
(b) Positioning copper tube in pliers with slitted stop platform.



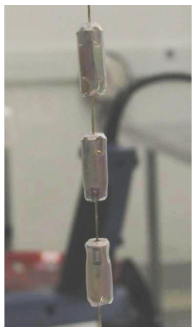
(c) Lowering copper tube into crimp tool for bottom crimp.



(d) Inserting source wire into capsule.



(e) Positioning PTFE sleeve for shrinking after placing top crimp and moving crimp tool.



(f) Jig-assembled weight capsules.

Figure 5.33: The assembly of a capsule. (d) and (e) were taken on the first prototype jig, which used commercial pliers, a cardboard stop platform, and an 8-pin crimp tool to test the proof-of-concept for the jig operation. (a)-(c) were taken after the aluminum stop platform and custom pliers were produced and the 4-pin crimp tool was mounted. (f) was taken after the dimensions of the jig were finalized and the jig was moved into the clean room. Note that the crimps of the bottom capsule in (f) are misaligned due to slippage of the copper in the pliers, while for the top two capsules, the jig operator successfully prevented this slippage.

tool support arm is positioned with a different spacer for the top crimp, and the pliers support arm is positioned with a spacer for the shrinking of the PTFE sleeve. Figure 5.33 shows a series of photographs taken at various stages of the jig development illustrating the assembly of a capsule. The heat gun is stored in a separate stand within easy reach of the jig operator and left running throughout source carrier assembly to maintain temperature. A pair of tweezers is also needed for manipulation of the source wires, the insertion of which is entirely manual.

As the source carrier will be moving into and out of the cryostat on a regular basis, it must respect the cleanliness requirements of the cryostat. No surface contaminants from the source carrier should be left behind on the inner surface of the guide tubes to constitute a permanent contamination in the cryostat. Due to the hollow interiors of the capsules, cleaning an assembled source carrier is undesirable because it would be difficult to ensure that all cleaning solutions were fully rinsed away and the source carrier was fully dried. Instead, the final source manufacturing jig was fully disassembled, cleaned, and reassembled bolted to a custom table inside the clean room, and all source carrier components are cleaned individually before being assembled in the clean room and stored under nitrogen. The heat gun manufacturer provided a procedure for cleaning the heat gun for use in a clean room environment; it is reproduced in Appendix B. Figure 5.34 is a photograph of the clean source carrier manufacturing setup.

The cleaning of all source carrier components is performed in three stages in beakers suspended in an ultrasonic bath:

1. **Degreasing** in a solution of 20% acetone and 80% ethyl alcohol
2. **Cleaning** in a solution of 5% citric acid, 1% H_2O_2, and 94% deionized water, the same solution used to clean the copper cryostat vessels
3. **Rinsing** in 100% ethyl alcohol

After both the degreasing stage and the cleaning stage, the pieces are drained and rinsed with deionized water. After the final rinse, the beaker containing the pieces is carried into the clean room before being drained so that the pieces are exposed only to clean room air from that time forward. All pieces are thoroughly dried under nitrogen flux in a desiccator cabinet in the clean

Figure 5.34: A photograph of the final source carrier jig in the UW clean room. A spacer can be seen positioning the pliers arm, and the heat gun in its stand is visible behind the jig frame.

(a) Copper pieces after being drained following cleaning in acid solution.

(b) Beakers of pieces in the clean room, just before being drained following final rinse. The bowl and sieve used for draining are visible, as is the previous batch of copper pieces drying in petri dishes inside the desiccator cabinet in the background.

(c) Discoloration of copper pieces visible after drying.

Figure 5.35: Photographs from the cleaning of parts for the PTFE shrink tests. These parts were used to optimize the cleaning procedures for the final production parts for the CUORE source carriers. The discoloration visible in (c) was present to a greater or lesser extent in a number of test batches and was determined to be caused by poor-quality deionized water.

room before final source assembly. These cleaning procedures were tested on all copper and PTFE pieces that were prepared for the PTFE shrink tests that established the correct pre-shrunk sleeve lengths for each batch of PTFE (see Section 5.3.2.4 for shrink test details and Figure 5.35 for photographs from the cleaning of the shrink test pieces); the final rinse was originally intended to be performed with deionized water, but experience with the shrink test pieces showed that the quicker evaporation of the alcohol was necessary to avoid discoloration of the copper. These tests also established that using poor-quality deionized water could result in discoloration of the copper, and as a result, a requirement that the measured conductivity of the water must be less than 1.0 μS was set.

Careful quality control of the materials and parts ensures that it is possible to construct highly uniform calibration source carriers, and the manufacturing jig provides a great deal of control and consistency to the production of the strings; however, the carrier manufacture remains a skilled task and requires some experience on the part of the operator to be able to successfully build a high-quality source carrier that mechanically performs reliably in the spool-and-guide-tube system. Multiple people have used the jig to produce numerous prototype and dummy source carriers for motion tests as well as all twelve final source carriers plus two spares since the jig was first built; all of their experience has been condensed into a highly detailed procedure document intended to provide both a reference and a basis for any necessary future training. This document and the detailed cleaning procedure are presented in Appendix B.

5.4 High-Rate Calibration in the Three Towers Test

Before the activity distributions that were loaded into the CUORE calibration source carriers were finalized, a dedicated study was performed to confirm the maximum acceptable event rate on the detectors during calibration data taking. Highly active calibration sources may cause excessively high event rates on the detectors; this increases dead time by causing pileup events in which two ore more events occur so close together that the pulse shapes are spoiled and the energies of the events cannot be reliably extracted, and it may even warm up the detectors sufficiently to shift the working points of the thermistors, effectively changing the calibration in the process

of attempting to measure it. However, insufficiently active calibration sources will not allow us to collect sufficient statistics to identify the peaks in the calibration spectrum without sacrificing too much of the detectors' live time. The multi-tower geometry of the CUORE array makes it effectively impossible to achieve perfectly even illumination of each crystal; therefore, we are interested in how high an event rate we can tolerate during calibration on the hottest crystals without sacrificing calibration performance in order to collect calibration statistics as quickly as possible on the coldest crystals. The nominal maximum event rate during calibration defined for Cuoricino was 100 mHz, but the geometry constraints of CUORE make it desirable to be able to relax this requirement.

The Three Towers Test (TTT) was primarily intended as a high-statistics comparison of the reduction in radioactive surface contamination achieved by three copper treatment techniques under evaluation for use in CUORE: wrapping in polyethylene film, simple cleaning in ultra-clean acids, and a multi-step process culminating in plasma etching. Three mini-towers, one prepared with each procedure, were mounted and operated simultaneously in the Cuoricino cryostat from September 2009 through mid-January 2010 to allow comparison of the observed background rates in the three towers under identical operating conditions [18]. Figure 5.36 shows the orientation of the towers as they were mounted in the cryostat. The Three Towers Test also provided the opportunity to perform calibration tests. During its run, we collected several days of calibration data with sources inserted in the cryostat that provided several levels of activity, all equal to or greater than the standard source activity used in Cuoricino, and tested the performance of the calibration module on these data.

Initially, the intent was also to test the new design for the CUORE source carriers and empirically confirm that the discretization of the activity into small capsules separated by intervals of ≈ 30 mm has a negligible effect on the distribution of activity induced on the crystals. Each Cuoricino source was constructed with six continuous strands of 1%-thoriated tungsten wire for a nominal activity of ≈ 50 Bq per source. Loading a CUORE-style source string (described in detail in Section 5.3.3) with as much activity as a Cuoricino source would require concentrating

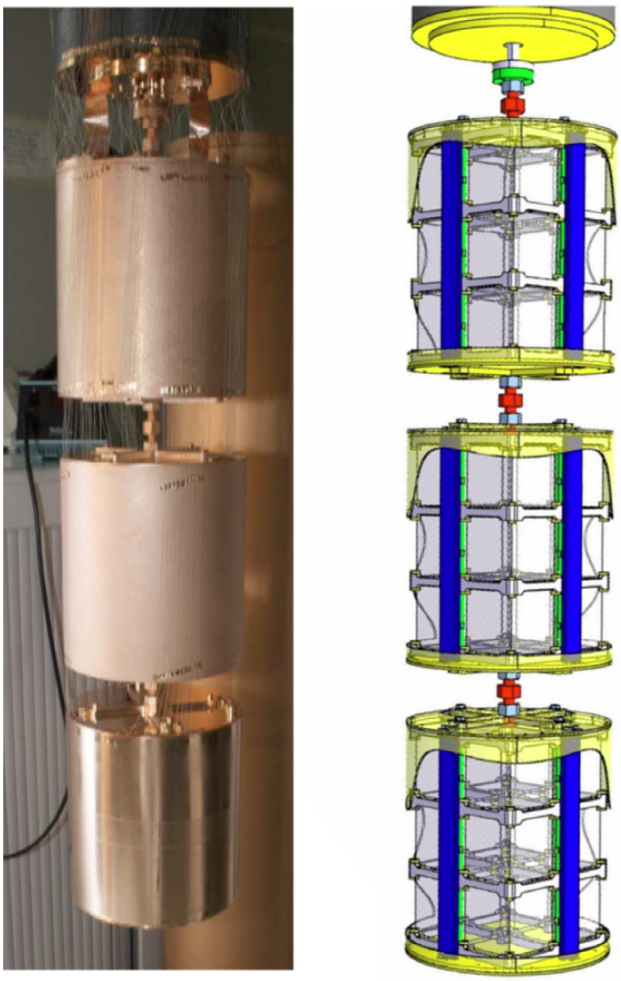


Figure 5.36: Left: A photograph of the detector towers operated in the Three Towers Test, mounted and shielded in preparation for insertion into the Cuoricino cryostat. Right: A schematic of the TTT detector towers depicted without the copper shielding. Figure from [18].

that activity within approximately 1/5$^{\text{th}}$ of the original active length to account for the discretization of the activity in a way that can be accommodated by the space available inside the CUORE source capsules, but the thoriated tungsten wire that serves as the radioactive inserts for the source capsules is not readily available with a thoria content of greater than 1%[5]. With this restriction, it is not physically possible to load a CUORE-style source string with even the same activity as a Cuoricino source, much less any greater activity. We were instead obliged to construct test sources with a similar physical design to the usual Cuoricino calibration sources.

5.4.1 Source Preparation and Data Taking

To determine the range of rates we wanted to test, we considered the theoretical improvement in calibration time that we could achieve assuming that increased pileup is the only significant effect of increased event rate and assuming a calibration time of 48 h (the Cuoricino baseline) at an event rate of 50 mHz (the 'typical' event rate on a Cuoricino detector during calibration). This simple situation can be estimated by applying an exponential correction to the number of counts C collected during calibration time t_c:

$$C_{\text{good}} = C\, e^{-R\, w_{\text{acq}}} = R\, t_c\, e^{-R\, w_{\text{acq}}},$$

where R is the event rate on the channel and w_{acq} is the width of the pulse acquisition window, which is 5 s for the CUORE data acquisition system. Substituting the appropriate values into this expression, we obtain an equation for t_c as a function of the event rate expressed as a rate factor F_R referred to the normal Cuoricino calibration event rate, 50 mHz (e.g., $F_R = 2$ corresponds to an event rate of 100 mHz):

$$t_c = \frac{(48\text{ h})\, e^{-0.25}}{F_R}\, e^{0.25\, F_R}. \tag{5.1}$$

Equation 5.1 is shown graphically in Figure 5.37, where it can be seen that t_c is minimized at a calibration event rate of 4 times the typical Cuoricino calibration rate; most of the gain has already been achieved by a rate factor of 2.5 – 3, however, and there is little point in concerning ourselves

[5] We were ultimately able to acquire 2% thoriated wire for the production of the final CUORE external sources, but we were not yet aware of the availability of this wire at the time that we were constructing the sources for this test, and the lead time was far greater than we could have tolerated within the time frame of the Three Towers Test in any case.

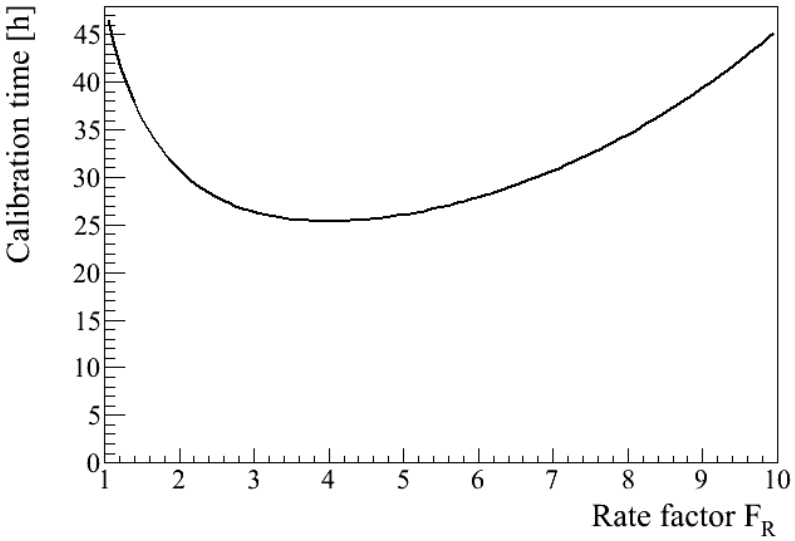


Figure 5.37: Estimated expected calibration time for detectors subject to calibration event rates F_R times the typical Cuoricino calibration event rate (50 mHz), assuming that increased pileup is the only significant effect of increased event rate and assuming that the calibration time at an event rate of 50 mHz is 48 h (the Cuoricino baseline).

Figure 5.38: Close view of part of the active region of a ‘double-strength’ source produced for the high-rate calibration test in the Three Towers setup.

with rates above 5 times the typical Cuoricino rate. Based on this consideration, we built two pairs of sources, each of which was approximately twice the activity of a standard Cuoricino calibration source. In combination with the standard Cuoricino sources, these sources allowed us to take data with source configurations that provided 2, 3, 4, and 5 times the typical Cuoricino calibration event rates.

The ‘double-strength’ sources were built with a design similar to that of the original Cuoricino sources. Four 4.5-m-long pieces of 1.7-mm-diameter window cable, consisting of stranded copper wrapped around a cotton core, were obtained to serve as the basic structure for the sources. Twelve 110-cm-long pieces of 0.2-mm-diameter thoriated tungsten wire with 1% thoria content were wrapped around the bottom ≈ 110 cm of each cable and held in place with PTFE heat shrink tubing of the same type used to cover the source capsules of the CUORE-style source carriers. This thoriated tungsten wire had the same specifications as the wire used to construct the standard Cuoricino sources, allowing us to simply use twice as many pieces to achieve a source with twice as much activity. For the first source constructed, the bottom ends of all twelve wire pieces were taped to the cable with masking tape before sliding the cable into the PTFE tubing. When the tubing was shrunk, the top end of one wire punctured the tubing, requiring that the sharp protrusion be covered with masking tape; for subsequent sources, both ends of the wires were taped to the cable before the addition of the PTFE tubing. Figure 5.38 shows a photograph of part of the active region of one of the sources after shrinking the PTFE tubing.

The bottoms of the sources needed to be weighted to allow the sources to be successfully guided into the calibration source positions in the experimental apparatus; however, it was also necessary that the weights take up as little room as possible to allow multiple sources to be inserted simultaneously for the higher-activity configurations. For this reason, we machined two narrow tapered-cylindrical brass weights. An 8-32 stainless steel screw protruded from the top of each weight, and a 2-mm-diameter hole was drilled crosswise through the shaft of the screw to accommodate a split ring. Sufficient bare cable was left below the active region of each source to form an eye-loop, on which a split ring was also strung; the eye-loop was held in place with a copper oval crimp sleeve. In this way, the split rings on two sources could be simultaneously attached to the split ring on a single weight for the simultaneous deployment of the sources. Photographs of a prototype eye-loop and a weight are shown in Figure 5.39.

On the final sources, the crimp sleeves anchoring the eye-loops were covered by masking tape to soften the hard edges of the crimp sleeves and prevent the bare ends of the cables from unraveling. A piece of masking tape was also placed as a marker near the top of each cable indicating the insertion extent for correct placement in the cryostat, with placements measured from the original Cuoricino sources. These marked the positions for plastic discs that served as insertion stops. The calibration source positions are not symmetric, so once the sources were marked in this way, they were specific to either the left or right position. Figure 5.40 shows a photograph of a completed source. The parts acquired for the construction of the sources are summarized in Table 5.5.

Although the Three Towers Test was conducted in the Cuoricino cryostat, its operating conditions differed from those of Cuoricino such that it was not immediately evident that the Cuoricino baseline ≈ 48 h calibration time was equally applicable to the TTT. The greatest difference was that a number of the chips attached to the faces of the crystals were lost in the cooldown of the Three Towers, including a number of thermistors and most of the heater chips attached to the faces of the crystals that were to be used for stabilization of the detectors. As a result, only 39 measurement channels were live, and a ^{40}K source was inserted into each calibration position during data taking for the entire first TTT dataset to provide a line at a known energy on which to stabilize;

(a) (b)

Figure 5.39: (a) Prototype eye-loop like those placed at the bottom of each ‘double-strength’ TTT calibration source. On each final source, a small split ring was strung on the eye-loop for attachment to the brass weight. (b) Brass weight for ‘double-strength’ TTT calibration sources.

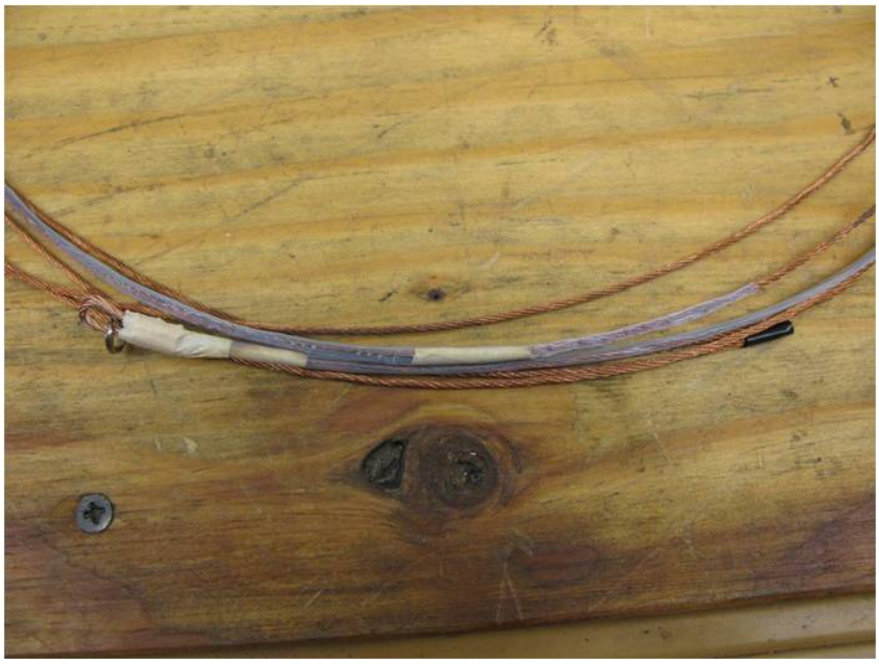


Figure 5.40: A completed ‘double-strength’ TTT calibration source. The taped wire ends are visible under the PTFE tubing, and the split ring can be seen on the eye-loop.

Table 5.5: Parts acquired for the construction of the TTT ‘double-strength’ calibration sources.

Part	Supplier	Catalog Number	Specifications
Cable	UW Physical Plant	—	Stranded copper w/ cotton core; 0.068” (1.7 mm) dia.
Thoriated tungsten wire	Goodfellow	W145370	0.2 mm dia.; 1% ThO_2 content
Heat shrink tubing	Small Parts	SM4T-0125-48	PTFE; 4:1 shrink ratio; 0.125” (3.2 mm) ID, 0.010” (0.254 mm) wall thickness
Oval crimp sleeves	McMaster-Carr	3897T22	Copper; 0.375” (9.5 mm) length, for $1/16^{\text{th}}$” rope dia.
Drilled-shaft screws	—	—	Stainless steel; size 8-32; $5/64$” (2 mm) hole drilled through shaft near head
Split rings	McMaster-Carr	90990A140	18-8 stainless steel; 0.334” OD, 0.242” ID, 0.068” thickness

the dataset was then processed two ways, the standard processing stabilized on the ^{40}K line and the Thermal Response (TR) processing described in [131]. In principle, the TR processing is inherently capable of correcting for the usual gradual baseline drifts observed in the detectors during operation and does not require the stabilization of the data on events of a known energy; after the first dataset, the ^{40}K sources were removed and subsequent datasets were processed with the TR method only.

In order to confirm the data collection time necessary to obtain a good calibration in the Three Towers setup, we investigated the calibration data collected at the end of TTT dataset 1001 (measurements 100151 and 100153), which totalled approximately 87.5 h and around 610,000 events. The Diana software has an option that allows the processing of only the first N events of the given data, where N can be specified by the user. We used this feature to emulate calibration data collected for a variety of different total times in increments of roughly 4 h, approximated by processing in increments of 28,000 events. For this dataset, both the standard processing and the TR processing were stabilized with the ^{40}K source, and the 1461-keV ^{40}K peak was also used as a calibration peak. We also investigated measurement 100184, from the calibration data collected at the end of TTT dataset 1002, which was taken with no ^{40}K source in the cryostat and was processed with only the TR processing with no stabilization. This measurement totalled approximately 47 h and around 290,000 events, so in this case we approximated increments of roughly 4 h by processing in increments of 25,000 events, where the lower event rate can be attributed to the removal of the ^{40}K source. For each 'calibration time', we applied the calibration module to the data and used the module's automatic peak-finding success as a simple indicator of calibration quality. We investigated the calibration performance for both types of processing on the Dataset 1001 calibration data and for the Dataset 1002 calibration data; the results are shown in Figure 5.41.

Based on the results presented in Figure 5.41, it was determined that the calibration quality has mostly flattened out after approximately 36 h of calibration data have been collected. Therefore, due to scheduling constraints, we chose to replace the '48 h' specification in Equation 5.1 with 36 h to determine the approximate length of time we would need to collect data for each calibration test

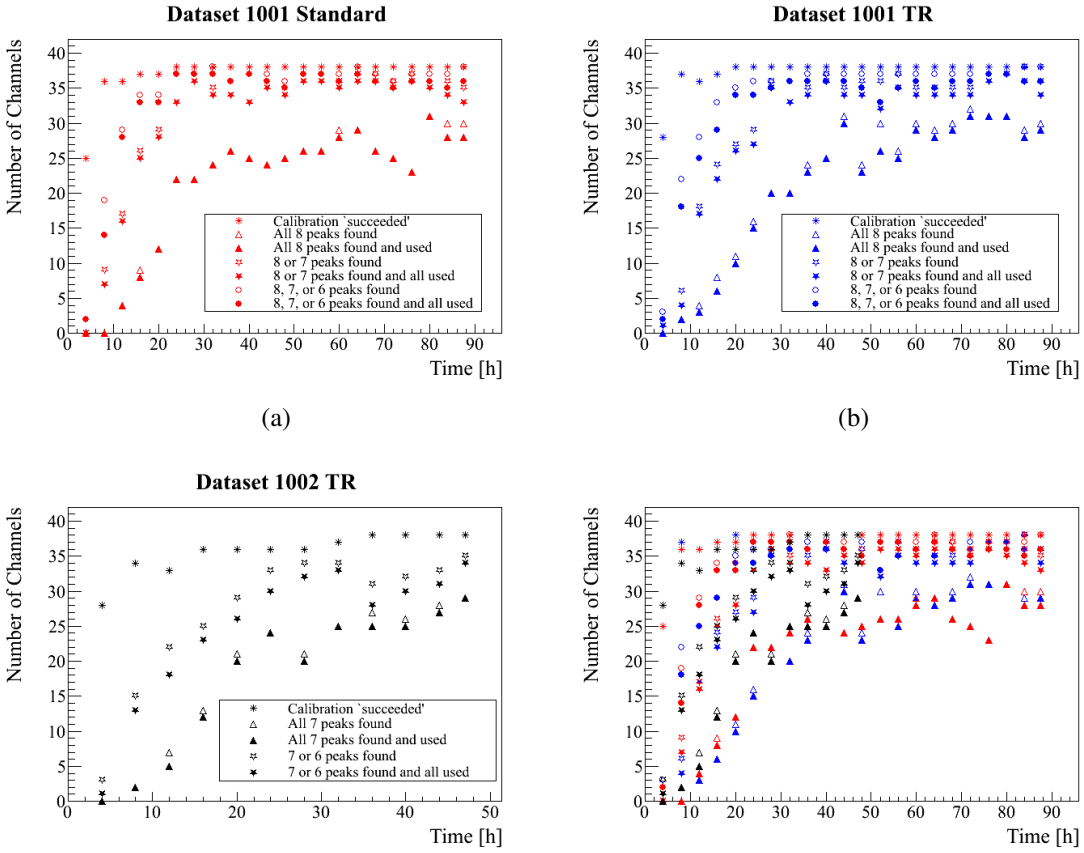


Figure 5.41: Automatic peak-finding success as a function of approximate calibration time as estimated from several TTT calibration measurements. In (a) and (b), the 1461-keV line of ^{40}K is used as a calibration peak in addition to the usual calibration peaks, which are summarized in Table 5.1. All points from (a)-(c) are overlaid on the same axes in (d).

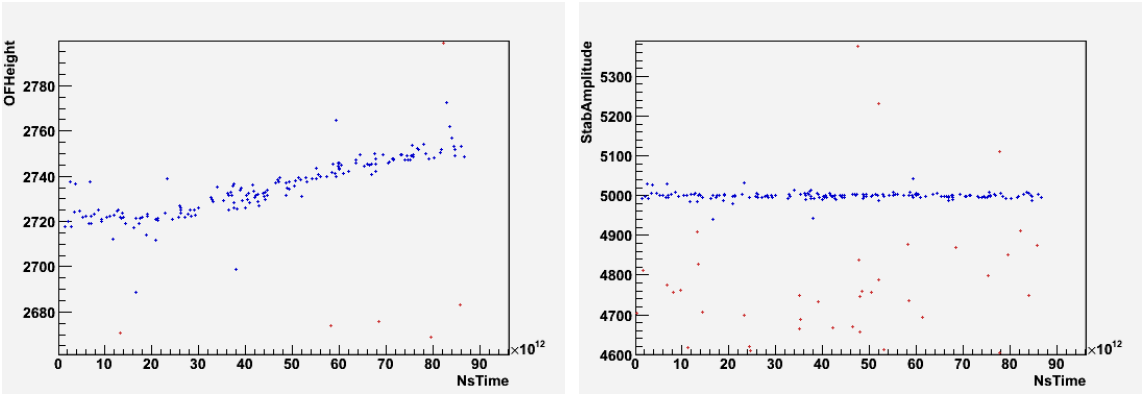
measurement to achieve statistics in each measurement roughly equivalent to a ‘good’ standard calibration.

The high-rate calibration tests belong to their own dedicated TTT dataset, Dataset 1005. In addition, two normal calibrations were interspersed with the high-rate test calibrations in this dataset. The ^{40}K sources were inserted alongside the calibration sources in each test measurement, although not in the normal calibrations, with the intention of allowing for greater flexibility of analysis. However, when the ‘5x’ measurement (i.e., with five times the calibration source activity of a standard Cuoricino/TTT calibration) was conducted, the event rates in the detectors were found to be much lower than expected. This is attributed to the sources becoming stuck upon insertion into the calibration positions and not being fully deployed due to the presence of three source weights (one for the standard Cuoricino source, one for the two ‘double-strength’ sources, and one for the ^{40}K source). A second 5x measurement was therefore conducted with no ^{40}K sources inserted. The first and second 5x measurements will henceforth be labeled 5x(1) and 5x(2), respectively.

A second 2x measurement (2x(2)) was also conducted due to severe baseline drift observed in the first 2x measurement (2x(1)) due to the cryostat base temperature passing outside the range in which it can be compensated by the heaters mounted on the towers (see also Section 4.2.3.3).

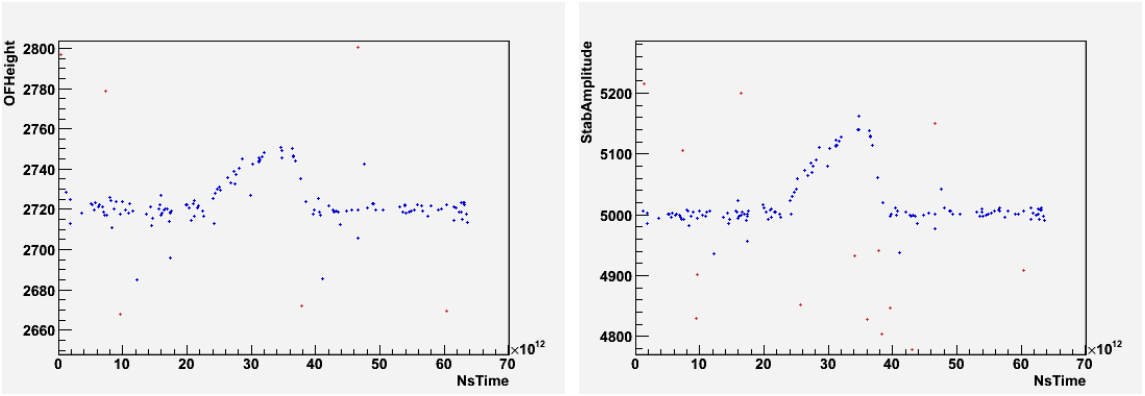
It was found that the 1461-keV line from the ^{40}K sources was not really strong enough in comparison to the ^{232}Th-chain spectrum to stabilize on in the high-rate measurements. It was also not valuable to use the 1461-keV peak as a calibration peak, as the only usable 5x calibration and the normal calibrations did not include the ^{40}K source, and we wished to process all measurements identically for the purposes of comparison. It did prove largely successful to stabilize the 2x(1) measurement on the 2614.5-keV peak; see Figure 5.42 for an example. Unfortunately, there was also excessive baseline drift in the 5x(2) measurement that could not be satisfactorily corrected by stabilizing on the 2614.5-keV peak, and a bad interval midway through the measurement had to be defined and rejected from analysis; this is illustrated in Figure 5.43.

All Dataset 1005 measurements are summarized in Table 5.6.



(a) (b)

Figure 5.42: Amplitude vs. time (in ns) on TTT Channel 1 in the 2x(1) run. The points in blue were identified as belonging to the 2614.5-keV line for the purposes of stabilization. (a) Before stabilization. (b) After stabilization.



(a) (b)

Figure 5.43: Amplitude vs. time (in ns) on TTT Channel 1 in the 5x(2) run. The points in blue were identified as belonging to the 2614.5-keV line for the purposes of stabilization. The 'peak' visible in the blue points indicates a period of severe baseline drift that could not be corrected through stabilization; this interval was rejected as bad for analysis. (a) Before attempted stabilization. (b) After attempted stabilization.

Table 5.6: Summary of measurements in TTT Dataset 1005, the dedicated dataset for high-rate calibration tests.

Run #	Name	Real time [h]	1x equiv. time [h]	Rates [mHz]	^{40}K?	Notes
100277	3x	≈ 20	≈ 36	$\approx 120-180$	Yes	—
100279	5x(1)	≈ 8	≈ 15	$\approx 100-200$	Yes	Lower rates than expected; sources believed stuck; not used for analysis
100281	4x	≈ 17	≈ 32	$\approx 170-270$	Yes	—
100283	2x(1)	≈ 24	≈ 38	$\approx 100-140$	Yes	Severe baseline drift; can be stabilized
100290	1x(1)	≈ 46	≈ 46	$\approx 50-100$	No	Normal calibration
100292	5x(2)	≈ 18 (≈ 13)	≈ 33 (≈ 24)	$\approx 230-350$	No	Severe baseline drift; cannot be stabilized; rejecting bad interval reduces statistics (marked in parentheses)
100293	2x(2)	≈ 23	≈ 36	$\approx 100-140$	Yes	—
100294	1x(2)	≈ 44	≈ 44	$\approx 50-100$	No	Normal calibration

5.4.2 Results

Before investigating the relative calibration performance of the high-rate calibration data, we would like to confirm that the pileup behavior with increasing event rate conforms to what we expect. Recall that, based on the estimated pileup correction, the fraction f of 'good' signal events on a channel is expected to be

$$f = e^{-R\, w_{\text{acq}}}, \tag{5.2}$$

where the width of the acquisition window w_{acq} is 5 s. For each channel in each measurement, f can be found by dividing the number of signal events that pass both the NumberOfPulses==1 cut (selecting only events in which only one energy deposition pulse appears in the acquisition window) and the Filter_Retrigger cut (rejecting any energy deposition pulse that may appear in more than one event acquisition due to multiple triggers) by the total number of signal events. Both NumberOfPulses and Filter_Retrigger are values filled by the Diana data processing that characterize each event. The results of calculating f this way are plotted as a function of rate in Figure 5.44, along with a set of points representing the expected f at each rate as calculated from Equation 5.2. It can be seen that the empirical results agree fairly well with Equation 5.2, although at higher rates, the scatter becomes wider and the 'expected' fraction behaves more as an upper bound.

Because of the difficulty with the ^{40}K sources, we chose to process all data in Dataset 1005 with the TR processing only, which does not require stabilization. Although 2x(1) can be used only when stabilized on the 2614.5-keV line, all the other measurements would also need to be stabilized on the 2614.5-keV line to be comparable, and as that line is heavily relied upon for calibration quality checks, we prefer to avoid the possibly confounding influence of stabilizing on the same line. From here on, all data presented in this section have been processed with the TR processing without any preliminary stabilization.

The 1x(1) and 1x(2) runs were used to obtain the official calibration for Dataset 1004; the calibration module was run on the sum data of the two measurements. We treat the calibration obtained this way as the 'official' calibration for the data collected in Dataset 1005 as well, meaning that

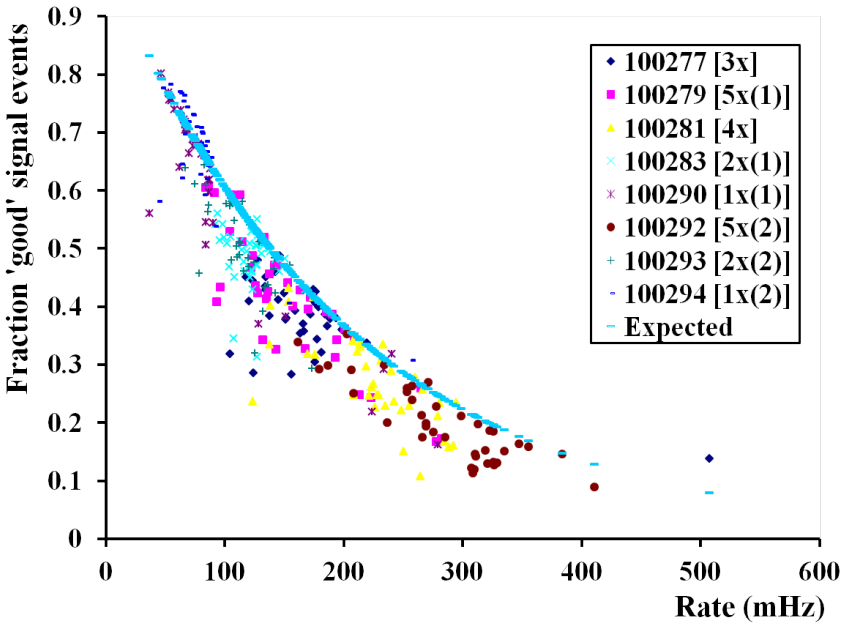


Figure 5.44: Fraction of ‘good’ signal events (one pulse in window, not retriggered) on each channel in TTT Dataset 1005 compared with the fraction expected from the event rate on the channel.

the quality of any calibration obtained from the high-rate test data can be judged by its consistency with this 'official' calibration.

We wish to confirm that the spectral shape is not spoiled by these high event rates (e.g., by widening the resolutions of the peaks by causing changes in baseline levels, either from heating the detectors or by distorting event pulse shapes in ways not recognized by Diana). Figure 5.45 shows the sum spectra of all channels from each of the six measurements from Dataset 1005 that can be used without stabilization; the data have all been calibrated with the 'official' calibration obtained from measurements 100290 and 100294. Examples of uncalibrated single-channel spectra are shown in Figure 5.46 for the high-rate measurements.

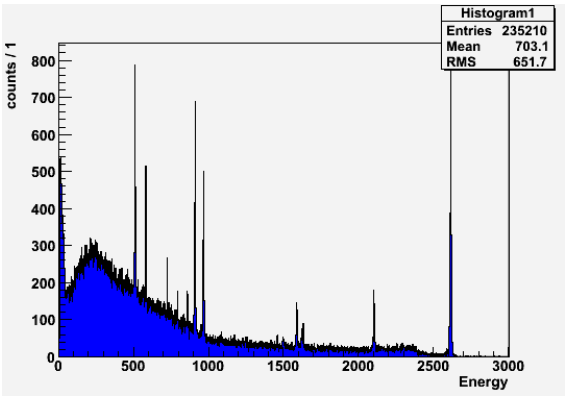
With the proper spectral shapes confirmed, at least upon first inspection, we now wish to evaluate the quality of the calibrations that we can obtain from these spectra. As done previously when estimating optimum calibration time from the TTT Dataset 1001 calibration data, some simple information can be extracted from the behavior of the calibration function[6]:

- Whether the calibration succeeds according to the calibration module (i.e., whether a calibration function is produced)
- How many peaks are found
- How many of the located peaks are used for the calibration function fit (i.e., not rejected for high deviation)

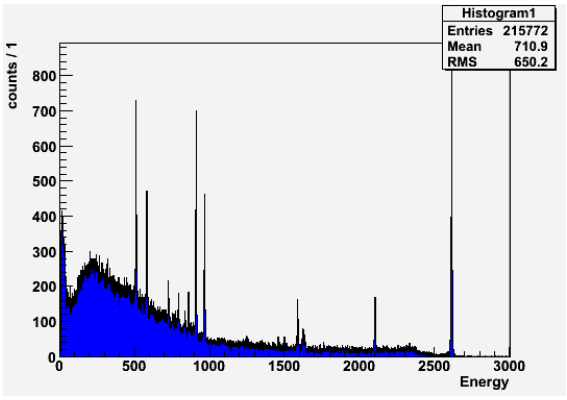
We can also define several quality checks adapted from the quality checks that were applied to each Cuoricino calibration in order to determine whether the associated dataset was good for analysis:

Residual check The calibration obtained from each test measurement is applied to the data of that measurement. Each channel passes this check if the residual of the calibrated energy of the 2614.5-keV line is < 1 keV. For Cuoricino data, any channel and dataset for which the residual at 2614.5-keV was larger than 1 keV was considered to have excessively high calibration uncertainty and was not used for analysis [45].

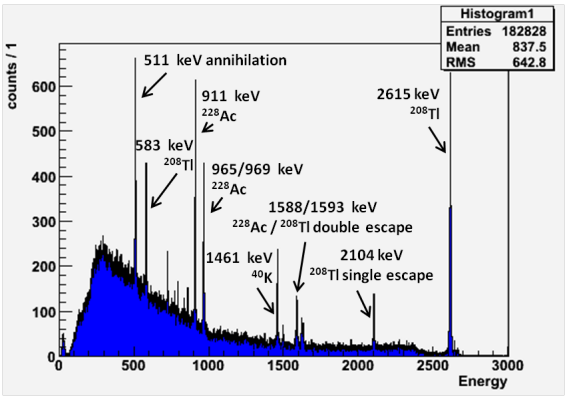
[6] Once again, all calibrations are evaluated only on the basis of the automatic peak-finding of the calibration module. No manual calibration repair is attempted.



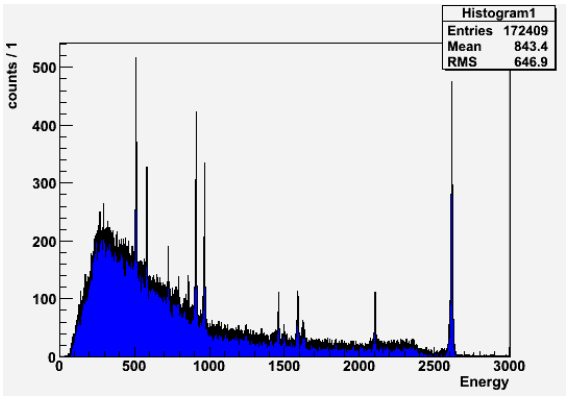
(a) 1x(1), run 100290



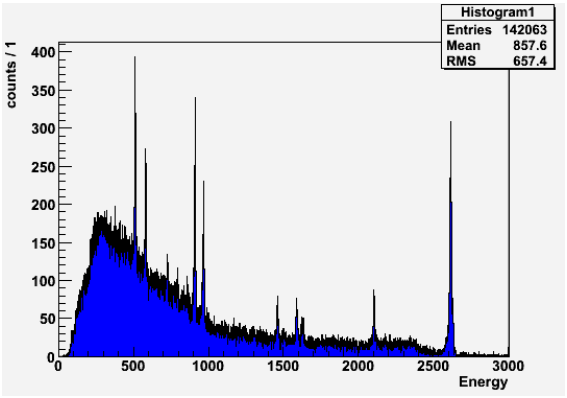
(b) 1x(2), run 100294



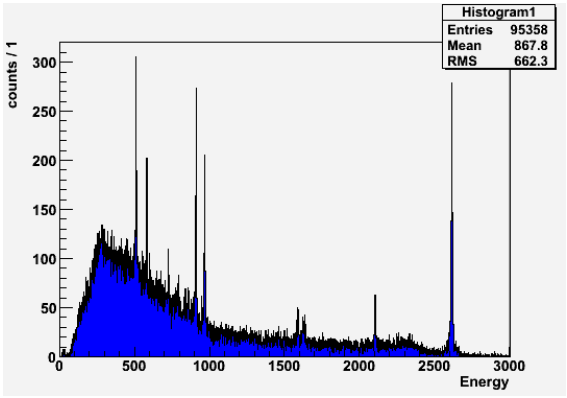
(c) 2x(2), run 100293



(d) 3x, run 100277

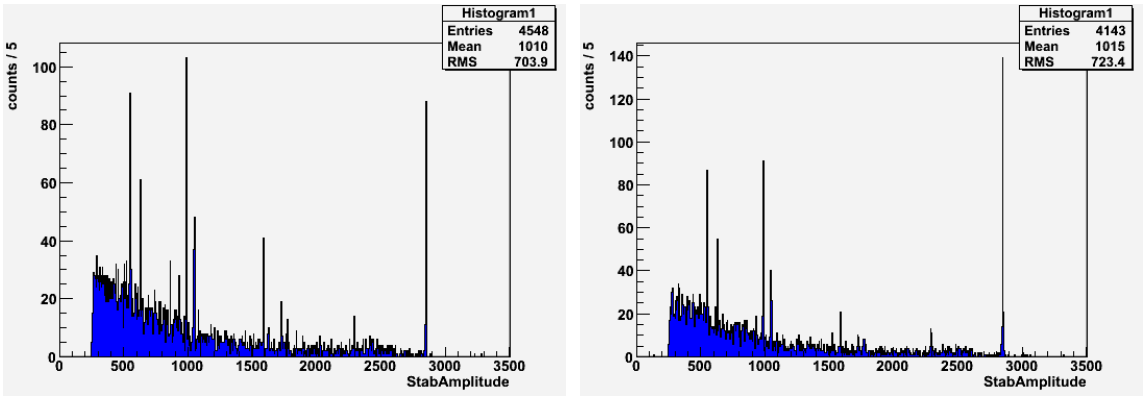


(e) 4x, run 100281



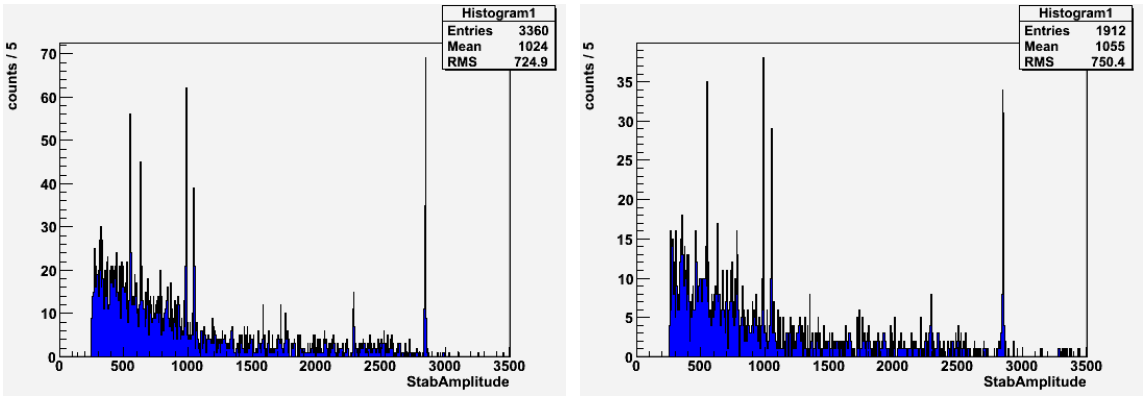
(f) 5x(2), run 100292

Figure 5.45: Sum spectra of runs in TTT Dataset 1005 calibrated with 'official' calibration obtained from runs 100290 and 100294. Notice the presence of the 1461-keV ^{40}K line in the 2x(2), 3x, and 4x runs. The fall-off at low energies is due to the application of an analysis threshold.



(a) 2x(2), run 100293

(b) 3x, run 100277



(c) 4x, run 100281

(d) 5x(2), run 100292

Figure 5.46: Uncalibrated Channel 7 spectra from high-rate calibration measurements in TTT Dataset 1005.

Resolution check The calibration obtained from each test measurement is applied to the data of that measurement. Each channel passes this check if the full-width-at-half-maximum (FWHM) resolution of the 2614.5-keV peak is < 11 keV. A study was performed on Cuoricino data that showed that a cut of FWHM < 11 keV for the large crystals and FWHM < 20 keV for the small crystals should optimize the sensitivity of the experiment [45]. Although the final Cuoricino analysis did not adopt this data quality cut, it is used here as an additional indication of calibration quality.

Consistency check The official calibration obtained from measurements 100290 and 100294 is applied to the data of each test measurement. Consistency between the official calibration and the test measurement is determined from the officially calibrated position in keV of the 2614.5-keV peak in the test measurement, x_{test}, the officially calibrated position in keV of the 2614.5-keV in the official calibration spectrum, x_{official}, and the mean FWHM energy resolution of the channel as measured on the 2614.5-keV calibration peak throughout all TTT datasets, $\overline{\text{FWHM}}$, as follows:

$$\frac{|x_{\text{test}} - x_{\text{official}}|}{\overline{\text{FWHM}}} \leq 1.$$

This is analogous to the consistency check performed between the initial and final calibrations of Cuoricino datasets to confirm that the detector response had not drifted excessively over the course of the dataset [45]. It checks the relative positions of the $2614.$-keV peak in the pulse-amplitude spectra of each measurement before calibration rather than comparing the calibration results; the purpose is to confirm that the high event rates in the test measurements have not caused baseline buildup in the detectors or induced a drift in the detector responses.

The results of applying all these checks to the test measurements are shown in Figure 5.47. The test measurements used are 1x(1), 1x(2), 2x(2), 3x, 4x, and 5x(2). 1x(1) and 1x(2) are treated individually in the same way as the other test measurements in order to obtain a basis for comparison of calibration performance on normal calibration measurements with approximately comparable

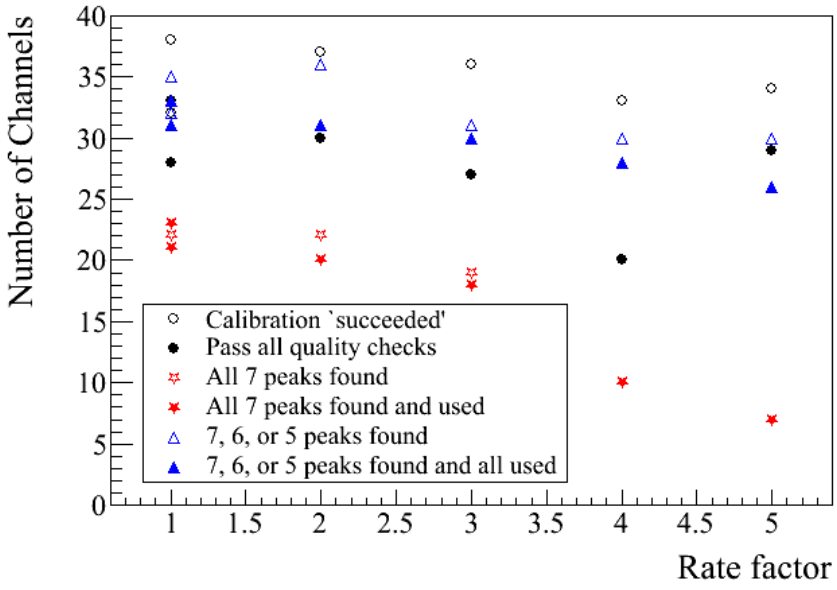


Figure 5.47: Number of channels in each test measurement that pass each quality check in the TTT high-rate calibration study. 1x(1) and 1x(2) are the two normal calibrations that were combined to provide the official calibration for the dataset; here, they are calibrated individually like the other test measurements to provide a basis for comparison to normal calibrations with similar statistics to those collected in the high-rate runs.

statistics to the high-rate measurements. In Figure 5.47, each point corresponds to a measurement; the value plotted is the number of channels in that measurement that pass the indicated check.

It can be seen from Figure 5.47 that while calibration performance declines only mildly up through the 3x run, there is a sharp drop-off at the 4x run, especially in the number of channels on which all 7 calibration peaks are found by the calibration module. Defining channel groupings by measurement in this way, however, ignores the spread in rates evident in each measurement. Also, although the residuals and resolutions of the 2614.5-keV peak may remain within the defined acceptable limits, there is still the possibility that increasingly higher event rates may cause a small but monotonic shift or a greater spread in one or both, which could still adversely affect the calibration accuracy or uncertainty. A change in calibration resolutions would also directly affect the $0\nu\beta\beta$ analysis, as the resolutions measured on calibration data are taken to be the fixed resolutions for the $0\nu\beta\beta$ fit. We must therefore consider the behavior of the residual and resolution of the 2614.5-keV peak as a function of event rate.

The calibrated positions of the 2614.5-keV peak in the test measurements for all channels that pass the residual check, each calibrated on itself, are plotted versus event rate in Figure 5.48. The mean and spread of the residuals appear stable for rates less than about 150 mHz or even 200 mHz. For rates greater than 200 mHz, however, the spread increases, and there is some indication that the mean calibrated position of the peak may actually increase for rates greater than 300 mHz.

Although the resolution of a given channel is expected to remain approximately the same over time, the resolution is not necessarily expected to be the same between one channel and another. For this reason, a scatter plot like Figure 5.48 will not be so informative for considering the change in resolution with rate. Instead, we construct a quantity analogous to that used for the consistency check by taking the difference in resolutions ΔFWHM between the resolution in the test measurement and the resolution in the official calibration and dividing by $\overline{\text{FWHM}}$; we then divide the channels of the various measurements that passed all quality checks in Figure 5.47 into several rate ranges and histogram the results. These histograms are shown in Figure 5.49. It is clear from this figure that while there is no significant difference between the resolutions in the test measurements and the resolutions from the official calibration for rates up to 150 mHz, the channels in the

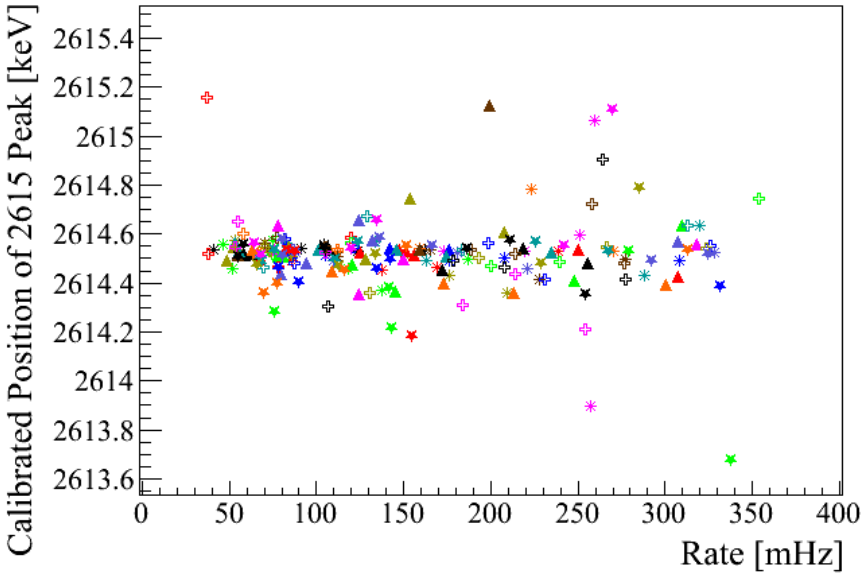


Figure 5.48: Calibrated positions of the 2614.5-keV peak vs. event rate in TTT high-rate test measurements. The calibration applied to each measurement is obtained from that measurement itself. Each individual marker type (color and shape) corresponds to a different channel.

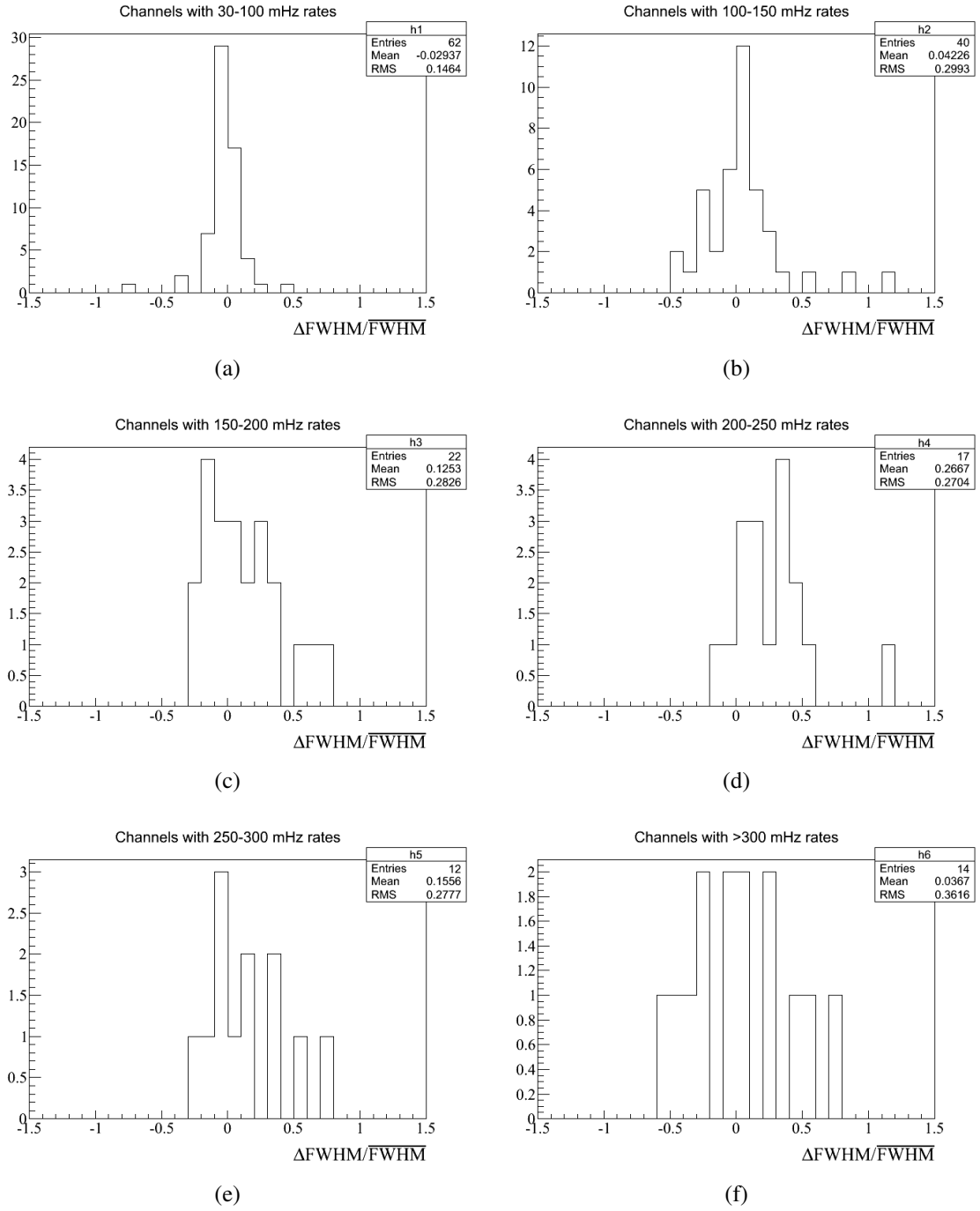


Figure 5.49: Comparison of resolutions in TTT high-rate test measurements to official channel resolutions for several ranges of channel event rate.

$150-200$ mHz and $200-250$ mHz ranges manifest a considerable worsening of resolution. This worsening is no longer evident in the $250-300$ mHz and > 300 mHz rate ranges, but the disappearance of the effect could be attributed to the loss of statistics, both in the number of channels that fall into these ranges and in the calibration statistics in the spectra actually used for calibration due to the significant bad interval removed from the 5x(2) measurement.

Considering the results of the calibration quality checks and the apparent behavior of the residuals and resolutions of the 2614.5-keV calibration peak with rate, we conclude that calibration performance remains acceptable for event rates of up to around three times the 'normal' Cuoricino calibration rate, or 150 mHz. This will allow the maximum rate requirement to be somewhat relaxed with respect to the originally defined maximum rate of 100 mHz when defining the optimum calibration source activity distribution for CUORE.

5.5 CUORE Calibration Simulations

Although approximate expectations for calibration performance in CUORE can be formed based on experience from detectors housed in the Cuoricino cryostat, such as Cuoricino and the Three Towers, the differences in geometry between the Cuoricino cryostat and the CUORE cryostat will result in somewhat different calibration spectrum shapes in CUORE due to the closer placement of the sources to the detectors. To determine the actual activity distribution to be loaded into the calibration source carriers and confirm that this activity distribution will produce the desired results, two rounds of simulations have been performed.

Preliminary GEANT4 simulations performed by S. Sangiorgio demonstrated that the effect of 'discretizing' the activity is negligible, so all subsequent simulations have been performed with continuous source strings to reduce computational complexity. These preliminary simulations also investigated the optimal ratio of the activity of the external strings to the activity of the internal strings under the assumption of a constant activity distribution along the strings.

5.5.1 Calibration Requirements and Constraints

The tension between two opposing effects drives the optimization:

1. The more quickly counts accumulate in each calibration peak, the faster a reliable calibration can be achieved and the less live time is lost to calibration. We want the calibration data collection to require no longer than **two days**. (At least **100 counts above background** in a peak is necessary to ensure that the peak can be found.)

2. The higher the event rate above threshold, the more detection efficiency is lost to pileup (events occurring so close together that the thermistor pulses cannot be effectively reconstructed and the event energies cannot be determined) and the greater the risk that the detector will begin to heat up, causing 'baseline buildup' and spoiling the very temperature response that we are trying to calibrate. The conservative hit rate limit is **100 mHz** on each detector, although the high-rate test in TTT demonstrated that this limit can be relaxed to around 150 mHz if necessary (see Sec. 5.4).

As a result of these considerations, the optimal activity distribution is the one that, as much as possible, reduces the need to either push the hottest detectors to higher rates or increase the calibration time to allow the coldest detectors to collect sufficient calibration data to be reliably calibrated — in other words, the one that achieves the greatest possible evenness of illumination across all 988 crystals. The preliminary simulations found that the optimum external:internal activity ratio was approximately 5:1. By imposing an upper limit of 100 mHz on the above-threshold rate on the hottest detector, preliminary simulations found that the optimal ^{232}Th activity[7] of each internal string was 3.9 Bq (making the optimal activity of each external string 19 Bq).

At the same time, it was seen that the top and bottom layers saw significantly fewer events than the middle layers due to 'edge' effects from the finite source length. Further simulations indicated that this effect could be largely corrected by applying z-compensation to the activity distribution along each source string: the optimal correction was to distribute an additional 1% of the total (12-string) source activity across the top 10 cm of all six internal strings (equivalent to distributing an additional 1% of the total activity of one external-internal string pair across the top 10 cm of a single internal string) and an additional 15% of the total (12-string) source activity across the top

[7]This is the activity of the parent of the decay chain, ^{232}Th, only. The total activity of the chain is an order of magnitude higher, assuming that the chain is in approximate secular equilibrium.

10 cm of all six external strings (equivalent to distributing an additional 15% of the total activity of one external-internal string pair across the top 10 cm of a single external string), plus the same for the bottoms. The final optimal total activity of the sources was therefore determined to be as follows:

One internal string $3.9\text{ Bq} + 2 * 0.01 * (3.9\text{ Bq} + 5 * (3.9\text{ Bq})) = 4.4\text{ Bq}$

One external string $5 * (3.9\text{ Bq}) + 2 * 0.15 * (3.9\text{ Bq} + 5 * (3.9\text{ Bq})) = 27\text{ Bq}$

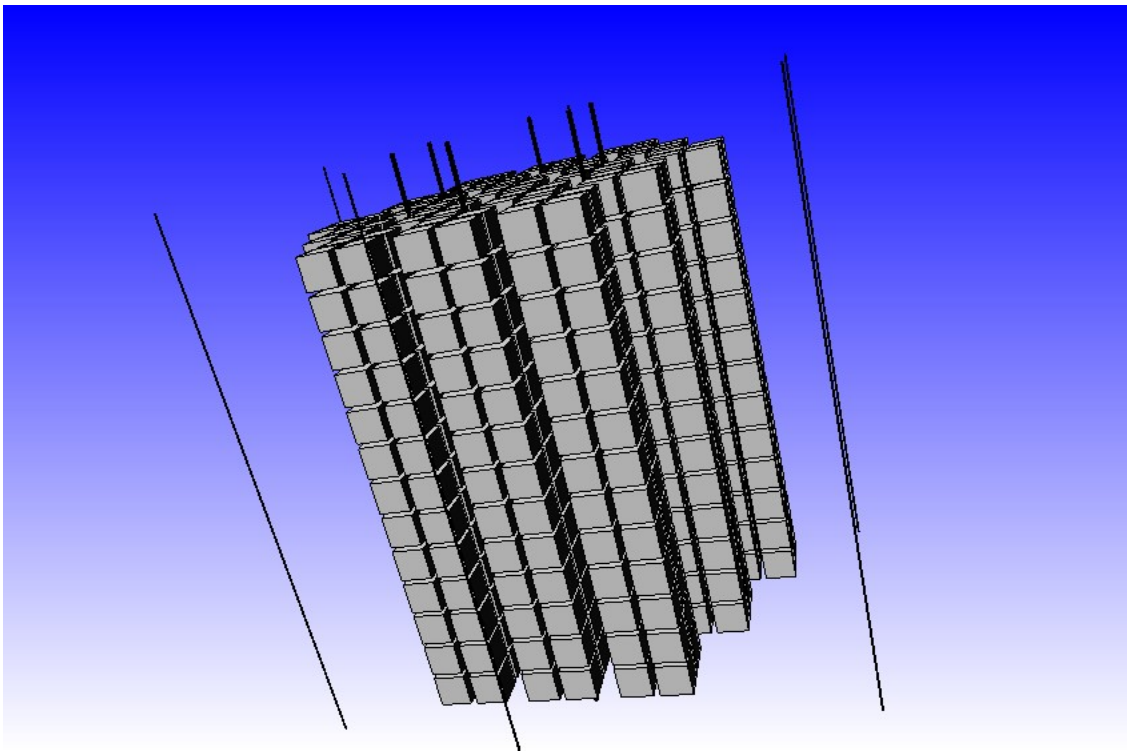
The physical implementation of this activity distribution requires cutting thoriated tungsten wire to appropriate lengths and inserting them into the capsules on the source carriers. The practical dimensional constraints on the wires are discussed in Section 5.3.2.5.

The internal string activity described above can be easily achieved within the physical confines of the source string design. The top 4 source capsules roughly correspond to the top 10 cm of the string, and the 8 weight capsules roughly correspond to the bottom 10 cm of the string; the additional activity will therefore be inserted into these capsules. Thoriated tungsten wire with 1% thoria content is readily available in a variety of diameters. The desired activity per capsule for the three different capsule types and the wire specifications required to achieve those activities are summarized in Table 5.7. Diameters and lengths are tuned for ease of handling and insertion.

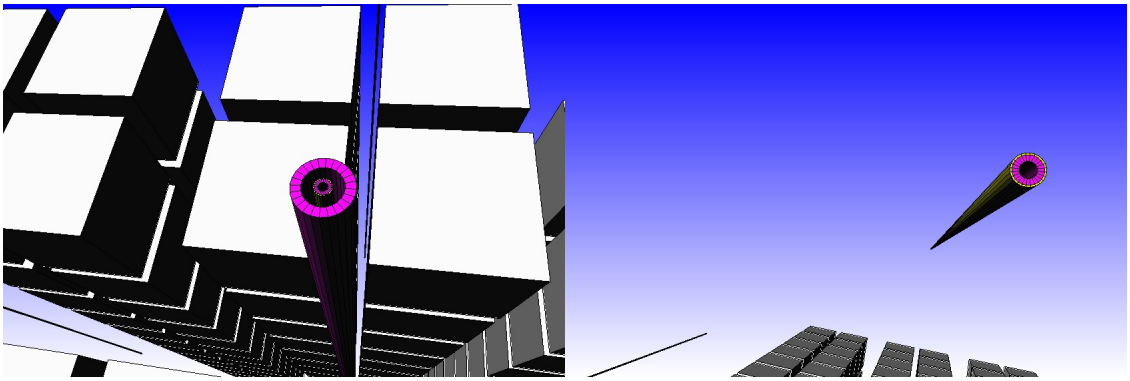
Producing source strings with 5 or more times this much activity is more of a challenge. Ultimately, however, we were able to acquire 2% thoriated tungsten wire. This allows us to achieve the full optimal activity distribution on our source strings with the wire specifications given in Table 5.7, so it is not necessary to evaluate non-optimal activity scenarios.

5.5.2 Analysis Approach

Now that the geometry of the cryostat is essentially final, a high-statistics simulation of the calibration sources has been performed in the official GEANT4-based CUORE Monte Carlo to confirm that the optimal configuration as previously determined still meets the desired requirements for successful calibration of CUORE. Figure 5.50 shows some visualizations of the simulation geometry. For simplicity, the calibration sources are considered to be continuous. Their



(a) Overview of source and tower geometry



(b) Internal string

(c) External string

Figure 5.50: Views of Monte Carlo geometry (tungsten wire is not visible in (b) and (c) because the copper and PTFE volumes extend somewhat past the ends of the active source)

Table 5.7: Per capsule specifications of active wire for production.

Location		Activity/cap.	% Th	Diam.	Length	#/cap.
Internal String	Top 4 Capsules	0.193 Bq	1%	0.38 mm	2.47 mm	1
	Middle 21 Capsules	0.134 Bq	1%	0.25 mm	3.96 mm	1
	Bottom 8 Capsules	0.096 Bq	1%	0.25 mm	2.85 mm	1
External String	Top 4 Capsules	1.55 Bq	2%	0.35 mm	3.90 mm	3
	Middle 21 Capsules	0.672 Bq	2%	0.35 mm	2.54 mm	2
	Bottom 8 Capsules	0.775 Bq	2%	0.35 mm	2.92 mm	2

geometry mirrors the geometry of the source capsules: a 0.35-mm-thick layer of copper with an outer diameter of 1.6 mm is surrounded by a 0.1-mm-thick layer of PTFE. The internal sources are additionally enclosed by copper guide tubes with an outer diameter of 6 mm and a wall thickness of 1 mm. The total active length of each source is 810 mm, extending 20 mm below the bottom and 30 mm above the top of the tower. The true diameter of the tungsten wire will differ for different capsule types; in the simulation, the diameter of the tungsten is approximated as 0.2 mm. Each source is separated into three sections: top, middle, and bottom. The top and bottom sections are each 100 mm long, while the middle section is 610 mm long. Six separate simulations were conducted:

- Internal Tops
- Internal Middles
- Internal Bottoms
- External Tops
- External Middles
- External Bottoms

For each Internal simulation, 10,000,000 total decays (1,000,000 parent decays) were simulated. For each External simulation, 50,000,000 total decays (5,000,000 parent decays) were simulated. For analysis of different activity distributions, the activity distribution is parametrized in terms of

a the base external:internal activity ratio,

b the fraction of the total base internal + external activity that is added to the tops of the internal strings,

c the fraction of the total base internal + external activity that is added to the bottoms of the internal strings,

d the fraction of the total base internal + external activity that is added to the tops of the external strings, and

e the fraction of the total base internal + external activity that is added to the bottoms of the external strings.

In all following discussions, b and c are set equal to each other and d and e are set equal to each other; however, it is, in principle, possible to treat them separately. The events from the Internal Middles simulation are assigned a weight of 1, and the rest are weighted accordingly to reflect the desired activity distribution.

The analysis of the calibration simulations is based on the calculation of a few key quantities.

Simulation time, t Because the Internal Middles simulation is always assigned a weight of 1, it is easiest to set the overall activity (as opposed to the relative activity distribution, which is set by parameters a-e listed above) by setting the base activity of a single internal string (before applying z-compensation), A_i, and then deriving the activities of all other string sections by applying parameters a-e. The base activity of a single external string therefore also defines the amount of calibration-data collection time that the simulation represents:

$$t = \frac{\text{number of simulated } ^{232}\text{Th decays}}{6 \times A_i \times (610 \text{ mm})/(810 \text{ mm})}$$

This quantity gives an idea of how the statistics of the simulation compare to the statistics expected in a full calibration run. For $A_i = 3.9$ Bq and 1,000,000 simulated ^{232}Th decays, which is the case for the analysis presented in this document,

$$t = 5.7 \times 10^4 \text{ s} = 0.66 \text{ days}$$

Total hit rate above threshold for detector D, r_D The typical energy threshold for CUORE-like detectors is 300 keV. This quantity is simply found by taking the correctly weighted sum spectrum of the six string-section simulations for the detector in question, extracting the number of events with energy greater than 300 keV, and dividing that number by t:

$$r_D = \frac{1}{t} \sum_{i=1}^{6} w_i N_{iD}(E > 300 \text{ keV}),$$

where the sum is over the six simulations, w_i is the weight applied to simulation i (determined by the desired activity distribution), and $N_{iD}(E > 300 \text{ keV})$ is the number of events with energy greater than 300 keV in detector D in simulation i. This quantity must be checked in the hottest detector to confirm that it does not exceed approximately 100 mHz (this requirement is somewhat fluid, as high-rate calibration tests in the Three Tower Test showed that the calibration module is still able to calibrate detectors with rates up to approximately 150 mHz), and it is also used to apply a pile-up correction to the simulated spectra.

Number of counts in peak P on detector D, N_{DP} The simulation produces all events at their nominal energies; it does not apply any smearing to reproduce the real detector resolution. The number of counts in peak P is estimated by finding the number of counts in a 1-keV-wide bin centered on the Q-value of the peak, Q_P, and subtracting the number of counts in

the left-by bin:

$$N_{DP} = \sum_{i=1}^{6} w_i [N_{iD}((Q_P - 0.5\text{ keV}) < E < (Q_P + 0.5\text{ keV})) \\ -N_{iD}((Q_P - 1.5\text{ keV}) < E < (Q_P - 0.5\text{ keV}))]$$

For two of the three ^{232}Th calibration peaks that actually comprise two adjacent peaks (see Table 5.1), the peaks are separated by more than 2 keV, so N_{DP} can be found by simply summing the two bins that are centered on the two Q-values and subtracting the two corresponding left-by bins. For the 511 peak, the peak-fitting procedure followed in Cuoricino neglected the ^{208}Tl gamma entirely in favor of the annihilation gamma; simulations indicated that the ^{208}Tl gamma contributed only about 6% due to the sources being located outside the shields. This is no longer true in CUORE. However, the two adjacent peaks are so close together that for the purposes of determining the number of counts in the peak from this simulation, results obtained by treating it as a single peak centered at 511.0 keV should differ negligibly from the results of a more careful treatment.

Probability of pileup on detector D, p_D Instead of tracking real timing information for the simulated events, an estimated pileup correction is applied based on the total hit rate above threshold on the detector.

$$p_D = 1 - e^{-r_D * (5\text{ s})},$$

where 5 s is the width of the pulse acquisition window.

Of the N_{DP} counts in peak P on detector D, approximately $p_D N_{DP}$ of them will occur in the same acquisition window and therefore be rejected from the analysis. At least 100 counts in the peak must survive ($(1 - p_D)N_{DP} = 100$) to guarantee that the calibration module can successfully identify the peak.

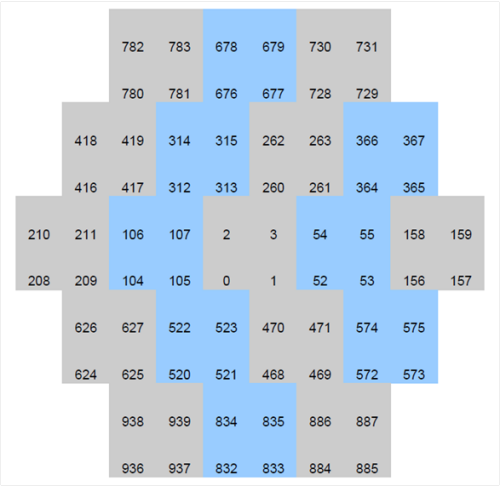


Figure 5.51: Map of detector numbers on the top layer of the array, Layer 0

Calibration time on detector D, T_D In the end, this is the main quantity of interest. The calibration time for peak P is found by calculating the time required to acquire 100 counts in the peak after correcting for pileup:

$$T_D P = \frac{100}{(1-p_D)N_{DP}/t}$$

T_D is then equal to the longest $T_D P$ among the peaks that are required to be found for a successful calibration.

As discussed in Section 5.1.2.1, only the primary peaks are expected to be found in every calibration. For this reason, the analysis of the calibration simulations focuses on the peaks that were treated as primary peaks in Cuoricino (see Table 5.1).

Each detector is identified by a unique detector number between 0 and 987. For transparency of detector positioning, this identifier can be translated into a tower number, a layer number, and a crystal number. Figure 5.51 shows the map of detector numbers corresponding to the crystals of the top layer, Layer 0 (layer number increases from the top of the array to the bottom). From this map, it can be seen that

$$\text{detector number} = 13 * 4 * \text{Tower} + 4 * \text{Layer} + \text{Crystal}.$$

5.5.3 Results

The simulation finds that the hottest crystal is Tower 1 Crystal 1 on Layer 11, which is immediately facing one of the internal sources. The hit rate above threshold on this detector is approximately 108 mHz, which is acceptable for calibration performance.

Figure 5.52 shows the total hit rate distribution across the whole array. The crystals can be broadly categorized into a 'hotter' group and a 'colder' group, which essentially correspond to the crystals that are adjacent to internal sources and those that are not, respectively. Sub-categories of these two groups of crystals can also be seen; there are several rather sharp peaks, each representing a large number of crystals that see a very similar level of illumination.

Figure 5.54 shows the total-hit-rate-above-threshold maps for the top, middle, and bottom layers of the array. The total hit rate is normalized to the rate on the hottest detector (108 mHz), which is not shown in these maps; it is on Layer 11, the second from the bottom.

Layers 2 through 10 look very similar to Layer 6, while the Layer 1 and 11 maps appear midway between the maps of the middle layers and the top or bottom layer, respectively. Each layer of the array can be treated as four symmetric quadrants; Figure 5.53 shows the average rate vs. layer for each of the 18 crystal positions in a quadrant, evaluated from the Poisson-error-weighted average of the number of counts above threshold in the spectra of the four crystals that occupy that crystal position on that layer, one from each quadrant. It can be seen that the rates on the hotter crystals actually slightly increase on the top and bottom layers due to the z-compensation. Although a small drop-off in rate at the top and bottom layers is evident in the colder crystals, the rate vs. layer in the colder crystals is still very flat for the middle 9 layers.

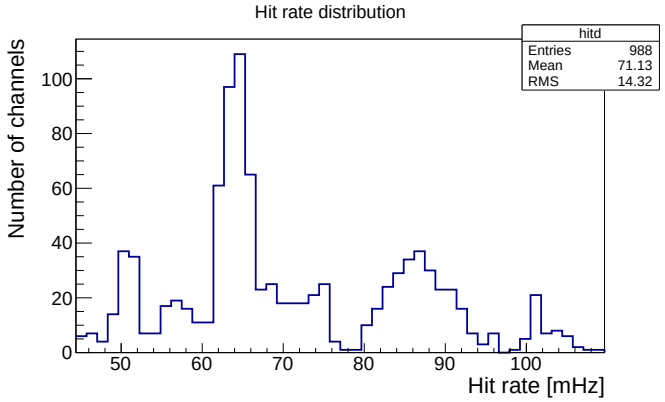


Figure 5.52: Distribution of total hit rate above threshold on all detectors in the CUORE array.

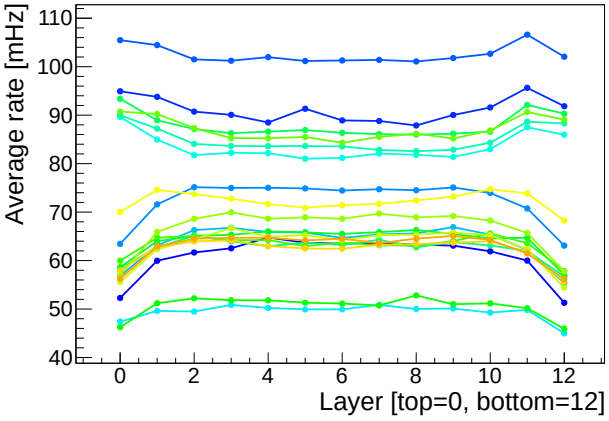


Figure 5.53: Average total hit rate above threshold vs. layer for the 18 unique crystal positions corresponding to one quadrant of the detector, calculated from the weighted average of all four quadrants.

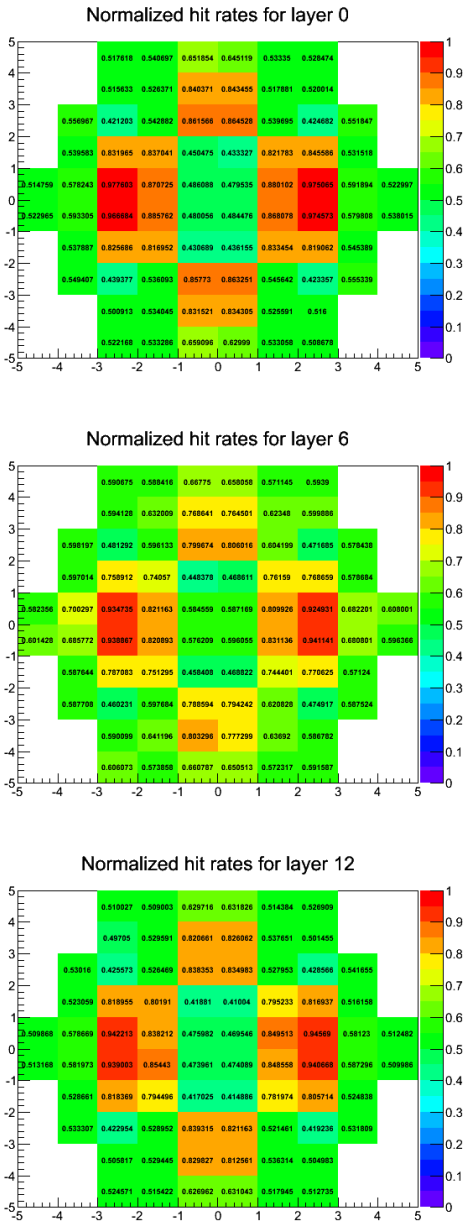


Figure 5.54: Total hit rate maps for the top (Layer 0), middle (Layer 6), and bottom (Layer 12) layers normalized to the hottest detector.

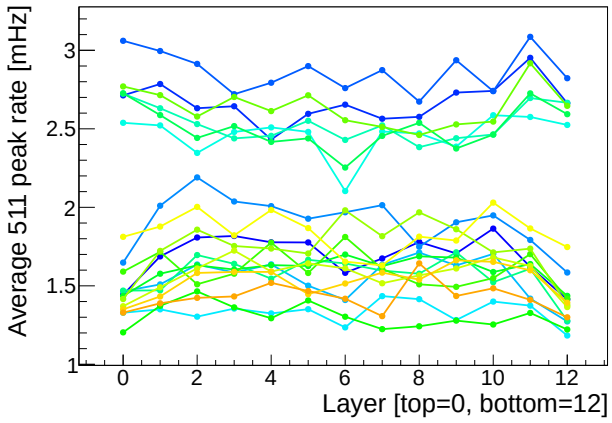


Figure 5.55: Average 511-keV-peak hit rate above threshold vs. layer for the 18 unique crystal positions corresponding to one quadrant of the detector, calculated from the weighted average of all four quadrants.

Figures 5.56 and 5.55 are the equivalents of Figures 5.54 and 5.53 for the hit rates in the 511-keV peak (before pileup correction). The normalization in Figure 5.56 is to the detector with the highest rate in the 511 keV peak, which is Tower 1 Layer 9 Crystal 3. For this low-energy peak, it is especially clear that the outermost detectors have significantly lower hit rates than the inner detectors; this is especially true on the top and bottom layers, but it is still evident even on the middle layers. Due to the overall lower statistics, there is greater variation among detectors in similar positions and also greater variation around the general rate vs. layer trend, but overall behavior similar to that seen in the total-rate maps is exhibited.

Figures 5.57, 5.58, 5.59, and 5.60 show the hit rate (uncorrected for pileup) and calibration time (including pileup correction) distributions for the 2614.5-keV, 969-keV, 911-keV, and 511-keV calibration lines, respectively. The distributions for each calibration line are clearly separated into two peaks, corresponding to the 'hotter' and 'colder' crystal groups. In general, this separation is more pronounced for lower-energy lines; this is attributable to the fact that the lower-energy gammas are less penetrating, so those issuing from the internal sources will have a greater tendency to stop in the crystals closest to those sources, while those coming from the external sources are already attenuated by the time they reach the detector.

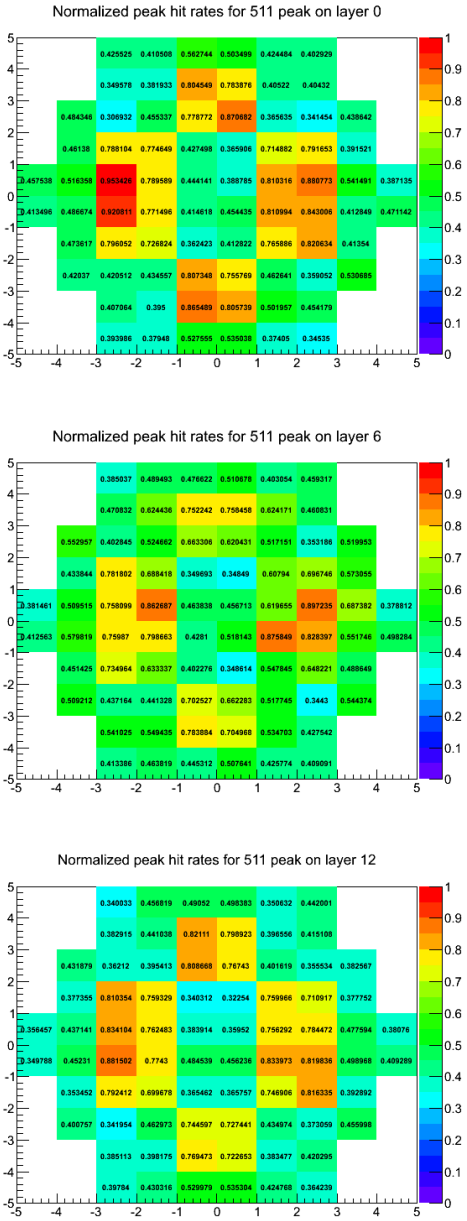


Figure 5.56: 511-keV-peak hit rate maps for the top (Layer 0), middle (Layer 6), and bottom (Layer 12) layers normalized to the detector with the highest rate in the 511-keV peak. The pileup correction has not been applied.

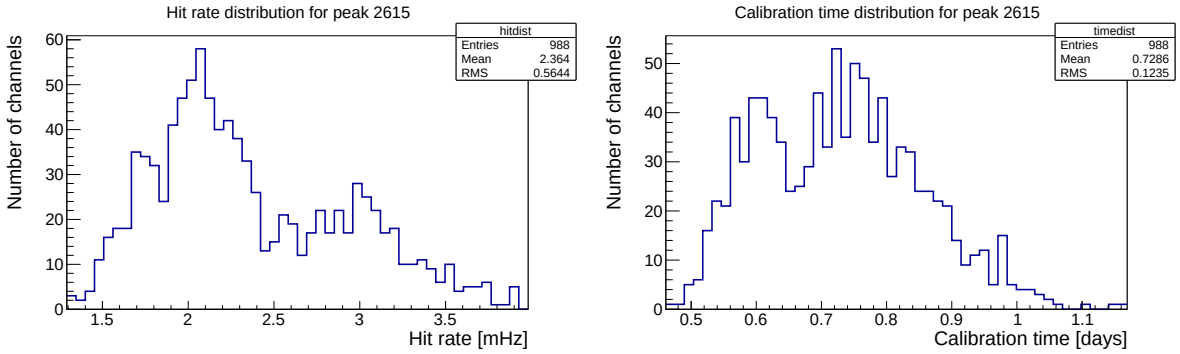


Figure 5.57: Hit rate (uncorrected for pileup) and calibration time (corrected for pileup) distributions for the 2614.5-keV peak.

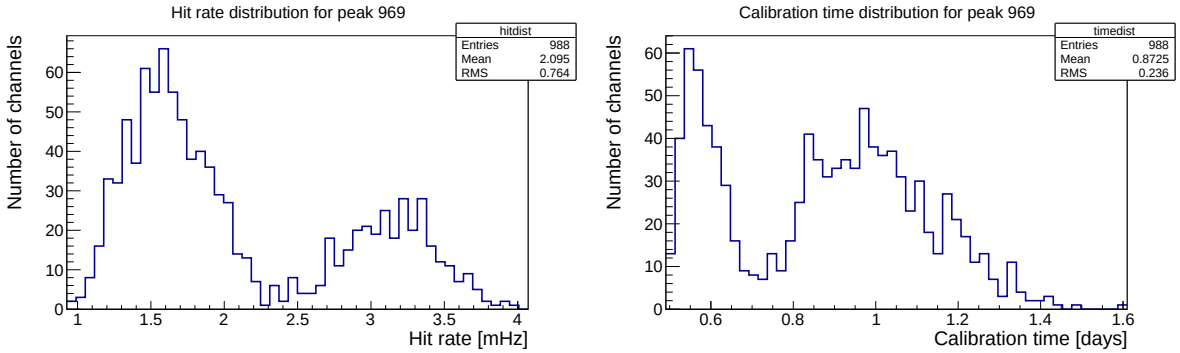


Figure 5.58: Hit rate (uncorrected for pileup) and calibration time (corrected for pileup) distributions for the 969-keV peak.

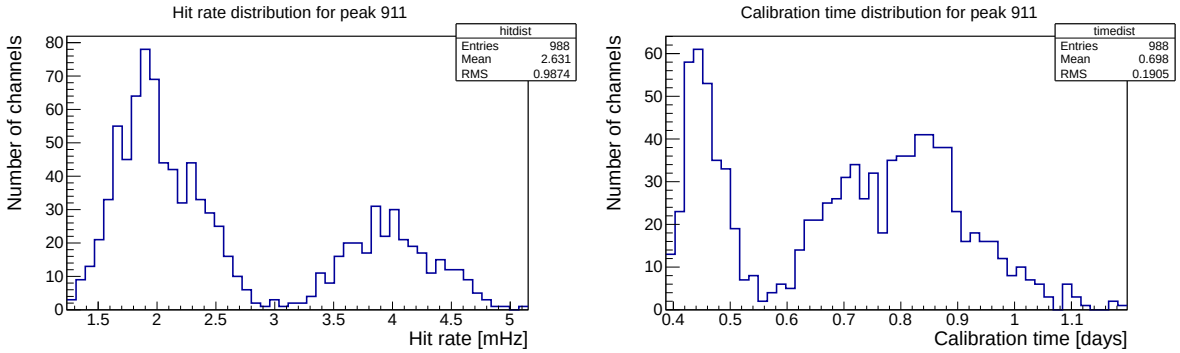


Figure 5.59: Hit rate (uncorrected for pileup) and calibration time (corrected for pileup) distributions for the 911-keV peak

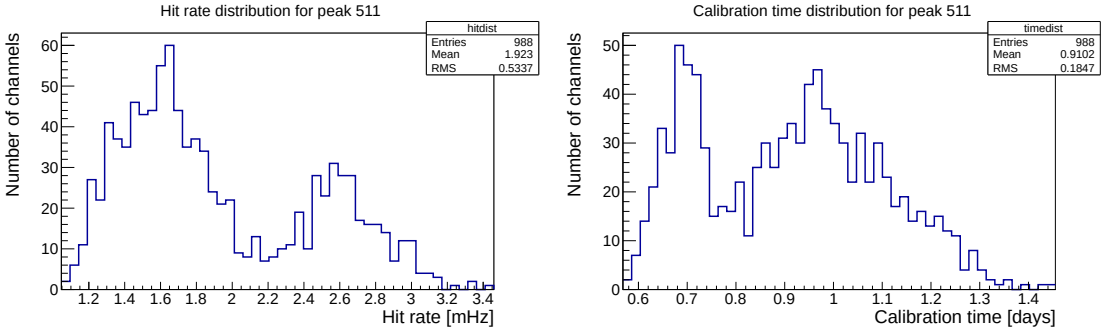


Figure 5.60: Hit rate (uncorrected for pileup) and calibration time (corrected for pileup) distributions for the 511-keV peak.

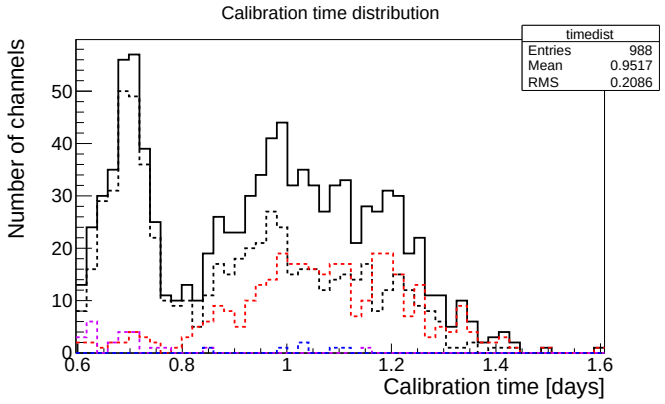


Figure 5.61: Overall calibration time distribution. Black solid line: all crystals. Black dashed line: crystals where calibration time is determined by 511 line. Blue dashed line: crystals where calibration time is determined by 911 line. Red dashed line: crystals where calibration time is determined by 969 line. Violet dashed line: crystals where calibration time is determined by 2615 line.

Table 5.8: Number of crystals for which the calibration time is determined by the specified calibration peak.

Peak	Number of Crystals
2615	23
969	335
911	6
511	624

From Figures 5.57-5.60, it can be seen that the calibration times for the hotter crystal group are, on average, the longest on the 511 line, followed by the 2615 and 969 lines, which have similar time distributions. For the colder crystal group, the 969 line and the 511 line have similar calibration time distributions that are, on average, longer than the calibration times on the 2615 or 911 lines. These observations lead to the expectation that the overall calibration time should be overwhelmingly determined by the 511 line on the hottest crystals, with some few instances in which it is determined by the 2615 or 969 peak, while the overall calibration time should be roughly equally likely to be determined by either the 969 or the 511 line on the colder crystals. This expectation is borne out by Figure 5.61, which shows the overall calibration time distribution broken down into its four component distributions corresponding to the calibration times determined by the four different primary calibration peaks. On a handful of crystals, the determining peak is the 911 line; this is probably due to statistical effects. Table 5.8 shows the number of crystals for which the calibration time is determined by each peak.

In the event that CUORE observes evidence of neutrinoless double beta decay, there may be interest in making a fiducial volume selection of those crystals with the best coincidence analysis to confirm that the observed signal is not caused by some confounding background process. In that case, it will be especially important that the calibration of the crystals within the fiducial volume is as reliable as possible. One possible way to construct this fiducial volume is to discount all crystals where at least one face is not facing another crystal; this requires discarding the top and bottom levels in their entirety as well as all crystals around the outside of each middle level (see Figure 5.62).

Figure 5.63 shows the result of including only those crystals that fall within this fiducial volume in the calibration time distribution. Applying the fiducial cut primarily has the effect of reducing the number of crystals in the colder group by about half; aside from this, the shape and range of the calibration time distribution remain essentially the same, although the fiducial cut appears to preferentially remove crystals on which the 969 line (as opposed to the 511) is the one that determines the calibration time. As a cautionary note, it should be considered that, because the crystals that would be discarded in this way constitute nearly half of the detector, an analysis using

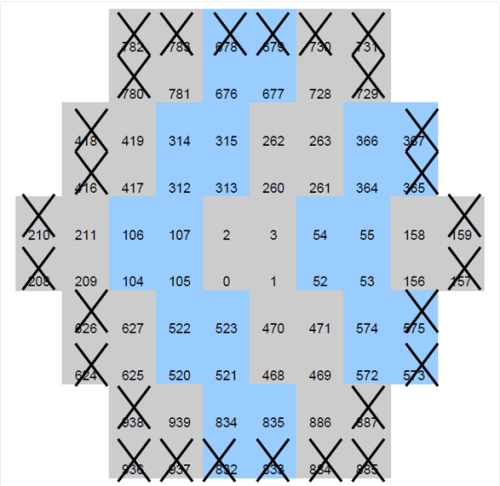


Figure 5.62: X's mark crystal positions that would be discarded on Layers 1-11 when constructing a fiducial detector volume.

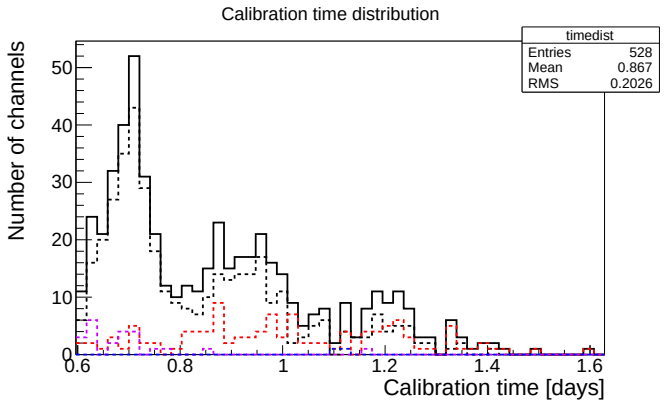


Figure 5.63: Overall calibration time distribution for only crystals that fall within the fiducial detector volume. Black solid line: all crystals. Black dashed line: crystals where calibration time is determined by 511 line. Blue dashed line: crystals where calibration time is determined by 911 line. Red dashed line: crystals where calibration time is determined by 969 line. Violet dashed line: crystals where calibration time is determined by 2615 line.

this sort of fiducial-volume cut would function primarily as a cross-check; it will still be necessary to successfully calibrate all of the crystals.

The calibration-time layer maps presented in Figure 5.64 illustrate that all crystals can be calibrated within two days, which was the nominal calibration time for Cuoricino. In fact, the crystals with the longest calibration times seem to be attributable to statistical fluctuations; once the weighted averages of the calibration times have been taken to produce Figure 5.65, the time vs. layer distribution appears to be quite flat and remain within 1.3 days. (Note: The weighted averages of the calibration times were not taken directly to construct Figure 5.65. Instead, the weighted averages of the calibration times on each peak were taken, then the longest of those was designated as the calibration time.)

The calibration hardware system for CUORE has been designed to enable the collection of high-quality calibration data by bringing radioactive calibration sources into close proximity of the detectors, illuminating the crystals as evenly as possible and allowing each detector to observe an energy spectrum with strong, identifiable gamma peaks. Extensive prototyping and testing of the elements of the system have demonstrated that a source carrier consisting of discrete capsules, which host radioactive material, attached to a flexible string can be deployed from a spool into a system of guide tubes to route calibration sources through a complicated path from storage at room temperature into calibration position in the detector region of the cryostat. The design and manufacturing of the spools and source carriers are carefully optimized to achieve reliable motion, and the activity distribution of the radioactive material in the source carriers is motivated by the desire to keep the time required for calibration low. With the hardware system appropriately adapted for the CUORE cryostat, simulations of the sources in the calibration geometry indicate that the calibration approach followed for Cuoricino can be expected to perform at least equally successfully in CUORE as it did in the previous experiment. At its core, after all, the detector technology remains essentially unchanged. Prospects for calibration performance in CUORE are explored in Chapter 6.

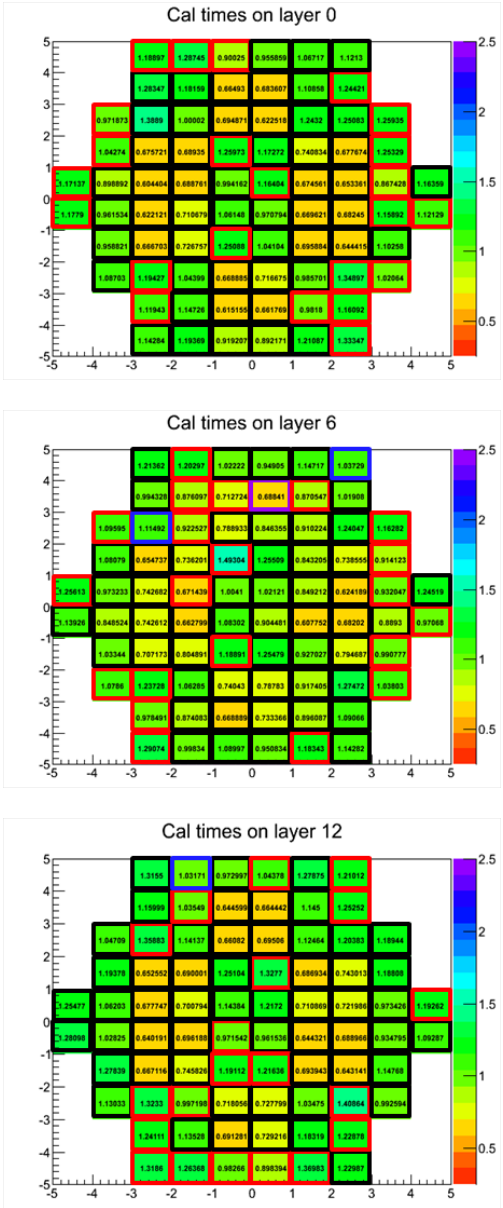


Figure 5.64: Calibration time maps. The colored borders correspond to the calibration peak that determines the calibration time on that crystal: 2615 (violet), 969 (red), 911 (blue), or 511 (black).

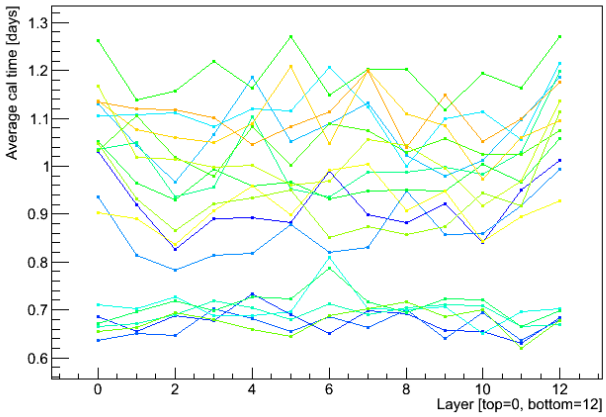


Figure 5.65: Calibration time vs. layer for the 18 unique crystal positions corresponding to one quadrant of the detector, calculated by taking the weighted average of the calibration time on each peak from all four quadrants and choosing the longest average peak calibration time

Chapter 6

Calibration Performance and Refinement in CUORE-Family Detectors

The thirty-two datasets of Cuoricino taken under normal operating conditions provide a large pool of calibration data with which to study the behavior and performance of the calibration software. This chapter discusses the calibration uncertainty achieved in Cuoricino and possibilities for extrapolating it to and/or improving it for CUORE.

6.1 Calibration Function

In Cuoricino, the energy calibration function fit was performed with two different functional forms:

1 A second-order log polynomial (logpol2)

$$\ln(E/(1\text{ keV})) = c_0 + c_1 \ln a_s + c_2 (\ln a_s)^2 \tag{6.1}$$

2 A third-order polynomial (pol3)

$$E = c_0 + c_1 a_s + c_2 a_s^2 + c_3 a_s^3 \tag{6.2}$$

Here, a_s is the amplitude of the event following stabilization with the stabilization heater as mentioned in Section 5.1.1.

The logpol2 function was used by the collaboration because it has historically proven to extrapolate well to the alpha region, although it was known to perform poorly at low energies because it is constrained to be unable to cross zero. While the analysis work for Cuoricino's final result

was being performed, however, it was found that there is a persistent trend to the residuals of the logpol2 calibration function throughout the gamma region.

The pol3 function was tried instead and proved to exhibit better results on Cuoricino data. For the majority of the channels and datasets, the final Cuoricino analysis used the pol3 calibration. However, the pol3 calibration function had one more parameter than the logpol2 because the intercept was not fixed to zero. Due to the fact that there are only four primary peaks and four parameters in the pol3, for a small subset of Cuoricino data, the pol3 calibration fit either failed or had zero degrees of freedom. For this data, the logpol2 calibration was used instead, as the performance of the two calibrations in the $0\nu\beta\beta$ R.O.I. is quite similar.

In CUORE, the pol3 function will be used with its intercept fixed to zero, reducing the number of free parameters to three, the same number as the logpol2[1]. The intercept is expected to be zero due to physical arguments (i.e., an 'event' of zero energy in the detector should cause a 'pulse' of zero amplitude), and a calibration function with intercept fixed to zero has shown to provide more reliable low-energy calibration in studies focused on the low-energy behavior of the detectors [17].

6.2 Secondary-Peak Finding and Choice of Interpolation Function

The peak-finding method used by the calibration module is briefly described in Section 5.1.2.1. During Cuoricino processing, the form of the interpolation function used to obtain the peak-location guesses for the secondary peaks (and any primary peaks missed on the first peak-finding pass) was the same as that of the calibration function itself; the result was that sometimes different peaks were identified as calibration peaks between the calibration performed with the logpol2 calibration function and the one performed with the pol3 calibration function on the same data.

In principle, of course, the 'correct' peak identification does not depend on the calibration function used, so in cases in which two different structures were identified as a particular calibration peak by the two different calibrations, at least one of those peak identifications must have

[1]This statement assumes that CUORE will continue to use the standard data processing used in Cuoricino. If the TR processing presented in [131] and tested in the Three Towers is adopted for CUORE, a second-order polynomial calibration function will be more appropriate. This was the functional form used for calibration in the high-rate calibration tests in the Three Towers discussed in Section 5.4, which were all processed with the TR processing.

been incorrect. Looking forward to CUORE, we would like to change the behavior of the calibration module such that the peak-finding is performed only once, saving the peak locations to the database for subsequent use any time we wish to use a different calibration function to process the data. This will require the choice of a single interpolation function, which will be used to perform the peak-finding no matter which calibration function is used for the first calibration performed on a given dataset. The fact that the Cuoricino data has been processed using two different functional forms for the interpolation means that there is a large sample of data on which the performance of each interpolation function can be evaluated to see if either demonstrates greater peak-finding reliability (either in finding peaks at all or in finding peaks more correctly). This kind of study can be used to drive the choice of the interpolation function that will be used in future.

Each time it is run on a dataset, the calibration module creates a supplementary output file containing fits, graphs, and histograms created as the calibration is performed. These files can be used to evaluate the calibration performance on Cuoricino; in particular, they can be used to determine

- whether a peak was found, by checking whether a corresponding peak fit function was saved to the file, and
- whether the same peak was found by both the logpol2 and the pol3 processing, by checking that

$$abs(\mu_1 - \mu_2) < min(\sigma_{\mu_1}, \sigma_{\mu_2}), \tag{6.3}$$

  where μ_i is the fitted position of the peak in the i^{th} processing and σ_{μ_i} is the corresponding error on the peak position from the fit,

in addition to a wealth of other information. In this way, it is possible to determine, for each peak,

- in how many calibrations the peak was found and the same structure was identified as the peak by both the logpol2 and the pol3 processing,
- in how many calibrations the peak was found by both the logpol2 and the pol3 processing, but different structures were identified as the peak in the two cases,

- in how many calibrations the peak was found by only the logpol2 processing,
- in how many calibrations the peak was found by only the pol3 processing, and
- in how many calibration the peak was found by neither processing.

When this determination is performed on all Cuoricino data, the total number of calibrations tested is (32 datasets)[2] $\times$ (57 channels)[3] $=$ 1824 calibrations.

Not all of these 1824 calibrations are representative of the automatic peak-finding results of the calibration module. If the module is unable to identify enough primary peaks on a channel to successfully interpolate, or if the module is unable to identify enough total peaks to successfully calibrate, the user is prompted to use a GUI tool to manually select the peak windows for fitting. This situation can be identified from the supplemental calibration output files because there will be no second-pass interpolation fit graph saved to the file for that channel. This manual GUI-based seeding was necessary on 62 of 1824 calibrations, or 3.4% of the time. In 5 of these 62 cases, one of the primary peaks on the channel could not be located even manually; in another 3 cases, no peaks could be successfully found on the channel at all.

Of the 1762 calibrations that did not trigger a request for manual GUI-based seeding, 37 (2%) contain fewer than four points in their corresponding interpolation fit graphs, indicating that they did not find all four primary peaks on the first peak-finding pass. However, only 8 of them have one primary peak that could not be eventually identified in at least one processing, and none is missing more than one primary peak in both processings.

Only 12 of the 37 calibrations that searched for at least one primary peak in the second peak-finding pass manifest a difference in peak-finding results between the logpol2 processing and the pol3 processing on a primary peak (1 511 peak found in different places; 3 511 peaks, 1 911 peak, and 1 969 peak found by logpol2 only; 3 911 peaks and 3 969 peaks found by pol3 only). None

[2]Dataset 30 consists entirely of test data and did not undergo normal processing, so it is skipped here.

[3]Channels 2 and 3 are discounted because they are deemed to be always bad for analysis due to their consistently poor performance throughout the history of Cuoricino; Channel 50 is also discounted because it did not have a functional heater, meaning could not be stabilized and therefore could not be properly calibrated. Other channels were sometimes determined to be bad for analysis on a dataset-by-dataset basis and excluded from the final Cuoricino analysis; however, only the three channels that are considered bad across all datasets are dropped here.

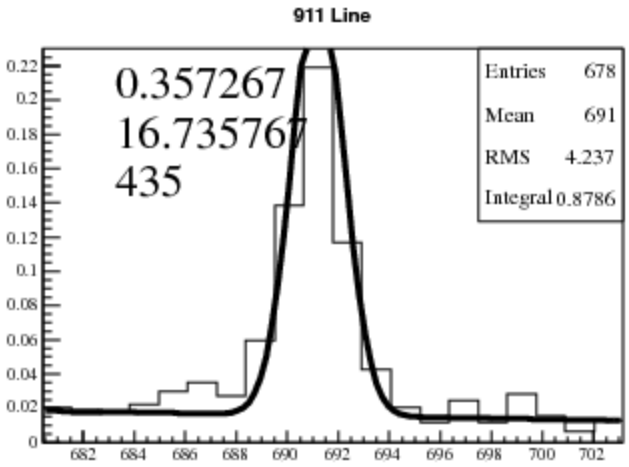


Figure 6.1: On Dataset 23, Channel 15, the 911 peak was not found on the first pass and, on the second pass, was found by the pol3 processing only. The peak location guesses used to define the 'wide' search window in each case were as follows: 691.551 (first pass), 691.477 (logpol2), 691.424 (pol3). The line displayed on the histogram shows the peak fit found by the pol3 processing; the values displayed, from top to bottom, are the background fraction in the window, the approximate significance of the peak above the background, and the approximate number of events in the peak.

differs in locating the 2615 peak, which is the most crucial peak for anchoring the calibration in the $0\nu\beta\beta$ region, and no more than one primary peak is found differently in any given calibration. There are no primary peaks that were identified differently on the *first* pass between the two different processings.

From these results, it can be seen that the peak-finding reliability of the calibration approach is quite high for the primary peaks, and in the rare cases that a primary peak is found on the second pass instead of the first, neither the logpol2 nor the pol3 interpolation function demonstrates an appreciably superior performance. In fact, when looking in detail at the primary peaks that were found by one processing and not the other, it seems that a difference in peak location guess of less than $1/10^{\text{th}}$ the resolution of the peak can mean the difference between finding the peak and not. An example of this behavior is presented in Figure 6.1. In a case like this, the success of one interpolation function over the other is evidently not due to the inherent superiority of the function for providing a peak location guess that is close to the true peak location but rather to 'luck.' It may therefore be worthwhile to add a provision to the calibration module that would, when a peak cannot initially be found, apply one or more tiny perturbations to the peak search window defined by the initial guess as the first attempt to relocate the peak.

As the logpol2 and pol3 processings are approximately equally reliable at finding primary peaks, as they should be, the choice between the two should be driven by their secondary-peak-finding performance when the peak-finding was entirely automatic and no manual GUI-based intervention was needed. To ensure that the two interpolation functions are tested in comparable circumstances, the calibration selection has been further reduced to include only the 1725 calibrations for which all four primary calibration peaks were found on the first pass and were therefore used for the interpolation fit. Figure 6.2 plots the number of calibrations in this subset that show the various pertinent peak-finding behaviors when comparing the logpol2 processing to the pol3 processing for each secondary peak.

It can be seen from Fig. 6.2 that the logpol2 and pol3 interpolation functions seem approximately equally capable of locating the 583 peak and the 1593 peak at all. The logpol2 processing located the 2104 peak slightly more often than did the pol3 processing; this may be due to the fact

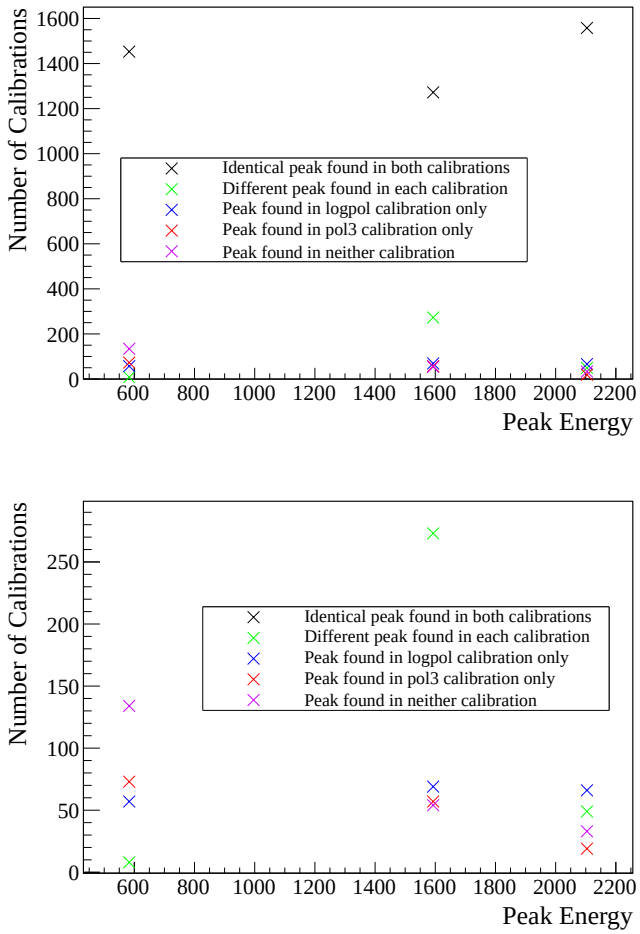


Figure 6.2: Calibration peak-finding behavior of secondary peaks in calibrations where all four primary peaks are found on the first pass and no manual intercession is necessary for peak-finding. The bottom figure is a zoom on the lower portion of the top figure for clarity.

that the pol3 function has one more free parameter than the logpol2 function and can therefore be expected to vary more greatly with small shifts in the primary-peak positions than the logpol2 in the energy region between the 969 and 2615 primary peaks. If true, the discrepancy should be addressed by fixing the pol3 intercept to zero, as the CUORE collaboration plans to do henceforth (see Sec. 6.1). In any case, the effect is small compared to the total number of calibrations.

It can also be seen from Figure 6.2 that it is very rare for the 583 peak to be found differently by the two different interpolation functions. This is expected, as the 583 peak is very close to the 511 peak, which is a primary peak. Also, the 583 peak will be much stronger in detectors calibrated with sources placed inside the cryostat shields as will be the case in CUORE, possibly strong enough to be treated as a primary peak instead of a secondary peak in CUORE (see Section 6.4), so the question of peak-finding reliability on the 583 peak as a secondary peak is probably relevant only to CUORE-0.

The most interesting peak to investigate for the comparison of the logpol2 interpolation function to the pol3 interpolation function appears to be the 1593 peak, which is identified in two different places by the two processings in 273 of the 1725 calibrations. We wish to construct a test that will allow us to determine which of the two interpolation functions more reliably identifies the *correct* structure as the peak; to do this, we begin by selecting the calibrations in which we believe the 1593 peak is *most likely* to have been correctly identified:

- Datasets affected by the high-radon periods identified by L. Kogler in her study of $2\nu\beta\beta$ in Cuoricino were removed [107].
- The pol3 calibration did not fail or have no free parameters.
- All four primary peaks were used for the interpolation fit for both processings, meaning that all four primary peaks were found on the first pass and found the same by both processings.
- The 1593 peak was identified the same by both the logpol2 and the pol3 processings.
- The 1593 peak was ultimately used for the final calibration function fit (i.e., not rejected for high deviation; see Section 5.1.2.3).

998 calibrations pass these selections. For the 1593 peak in each of the selected calibrations, we choose a few peak quality metrics that can be expected to show some identifiable characteristic behavior in correctly identified peaks:

Residual from calibration fit Ideally, the distribution of any peak's residual from the calibration fit ought to be a narrow peak centered on zero. If the functional form of the calibration function is not a perfect description of the detectors, it may be centered on a value slightly different from zero instead. Thus we expect that the residual distributions of the correctly identified peaks may differ between the logpol2 processing and the pol3 processing. The residual is calculated as

$$[\text{calibration function evaluated at the mean of the fitted peak}] - [\text{nominal peak energy}].$$

The residual is expressed in units of energy.

Deviation from calibration fit This is the quantity used by the calibration module to evaluate whether a peak should be rejected from the final calibration fit (see Section 5.1.2.3). As in the case of the residual, the distribution should ideally be a narrow peak centered on zero; the residual and the deviation are both measures of the distance of the point from the fit line, but they are calculated in perpendicular directions to one another. The deviation is calculated as

$$\frac{abs([\text{calibration function solved for peak mean at peak energy}] - [\text{peak mean}])}{[\text{peak mean error}]}.$$

The deviation is expressed in units of sigma referred to the fitted error on the peak mean.

Distance from interpolation peak location guess to mean of identified peak Unlike the residual and deviation, which judge the quality of the peak from the 'correct' calibration for the channel, this quantity measures the distance from the *initial* peak location guess to the location of the peak that was ultimately found. Ideally, if the interpolation function provides good guesses for the expected peak mean, the distribution of this quantity should also be a

narrow peak centered on zero, as in the case of the residual and the deviation. This quantity is calculated as

$$([\text{peak mean guess}] - [\text{peak mean}]) \times \frac{[\text{nominal peak energy}]}{[\text{peak mean}]}.$$

The scaling by $\frac{[\text{nominal peak energy}]}{[\text{peak mean}]}$ converts to units of approximate energy to roughly correct for detector responses so that the distributions taken from different detectors can be meaningfully combined.

Ratio of counts in 1593 peak to counts in 2615 peak The '1593' peak is actually composed of two gammas, one from the decay of ^{228}Ac at 1588.19 keV and one at 1592.533 keV, which is the double escape from the 2614.533-keV gamma of ^{208}Tl. The relative strength of the latter is geometry-dependent; the ratio of the amplitudes of the double-escape peak and the 2615 peak can be expected to be approximately fixed for the single tower of Cuoricino, although it will shift some depending on the exact relative orientation of the calibration source and the crystal. The peak-count ratio is estimated as

$$\frac{[\text{integral of 1593 peak-fit histogram}] \times (1 - f_{bkg_{1593}}) \times amp2_{1593}}{[\text{integral of 2615 peak-fit histogram}] \times (1 - f_{bkg_{2615}})}, \tag{6.4}$$

where f_{bkg} and $amp2$ are parameters of the functional forms used to fit the peaks as introduced in Section 5.1.2.2. For the 1593 peak, there is no simple *a priori* assumption that can be made for the expected relative amplitudes of the 1588.19-keV and 1592.533-keV peaks, so $amp2$ is merely constrained to the physical region and thus allowed to vary between 0 and 1. Unlike the previous three quantities, this quantity does not depend directly on the interpolation function or calibration function; the distribution from the logpol2 processing should therefore be identical to the distribution from the pol3 processing, aside from small differences in peak fits.

Histograms illustrating the distributions of the peak-quality metrics listed above are shown in Figures 6.3 – 6.6. It can be seen from Figures 6.3, 6.4, and 6.5 that the 'reference' distributions are approximately centered on zero, as expected, for the pol3 processing, while they are slightly offset

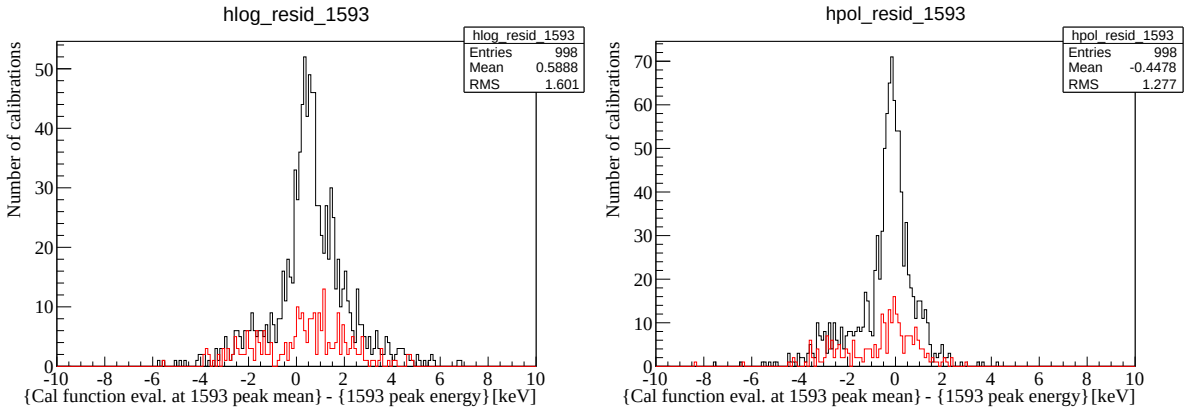


Figure 6.3: 'Residual from calibration fit' distributions for 1593 peak. Peaks identified differently by logpol2 and pol3 processings (red histogram) and 'reference' distribution (constructed from peaks that are believed most likely to be correctly identified; black histogram) are shown. Left: logpol2 processing. Right: pol3 processing.

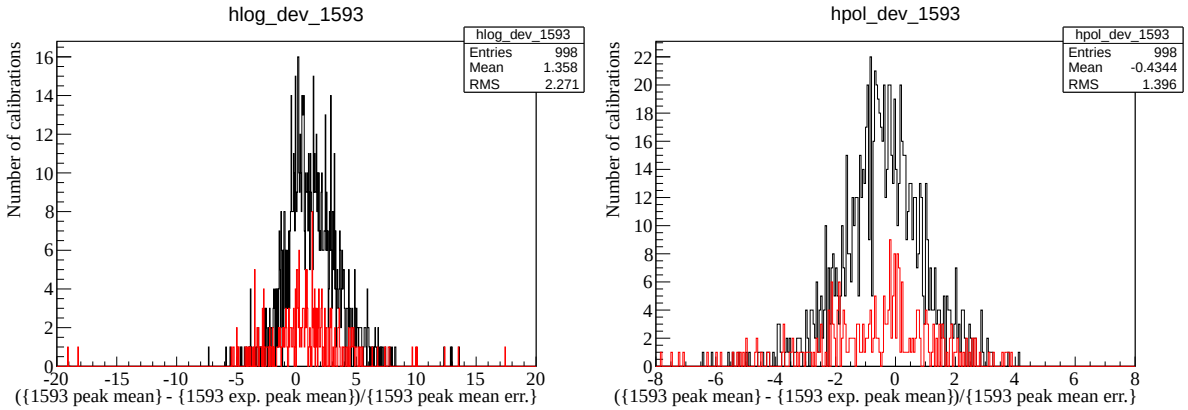


Figure 6.4: 'Deviation from calibration fit' distributions for 1593 peak. Peaks identified differently by logpol2 and pol3 processings (red histogram) and 'reference' distribution (constructed from peaks that are believed most likely to be correctly identified; black histogram) are shown. Left: logpol2 processing. Right: pol3 processing.

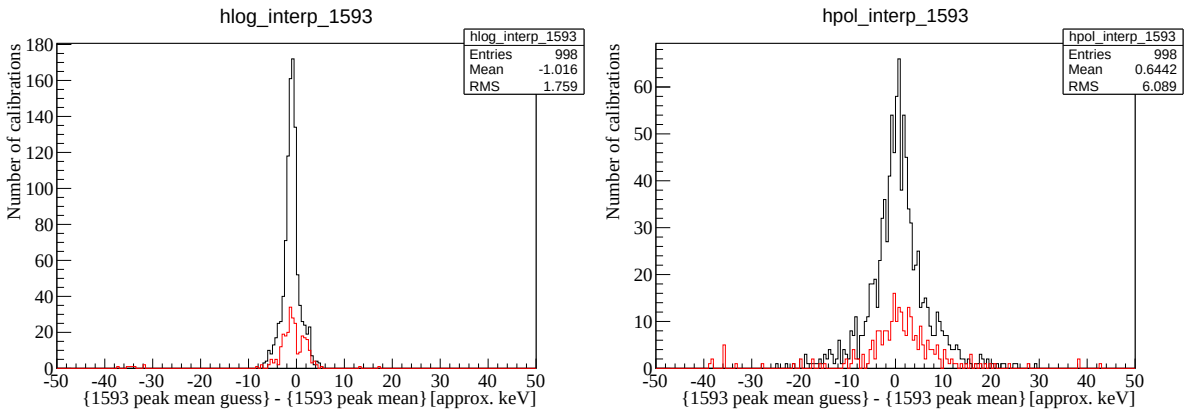


Figure 6.5: 'Distance from interpolation peak location guess to mean of identified peak' distributions for 1593 peak. Peaks identified differently by logpol2 and pol3 processings (red histogram) and 'reference' distribution (constructed from peaks that are believed most likely to be correctly identified; black histogram) are shown. Left: logpol2 processing. Right: pol3 processing.

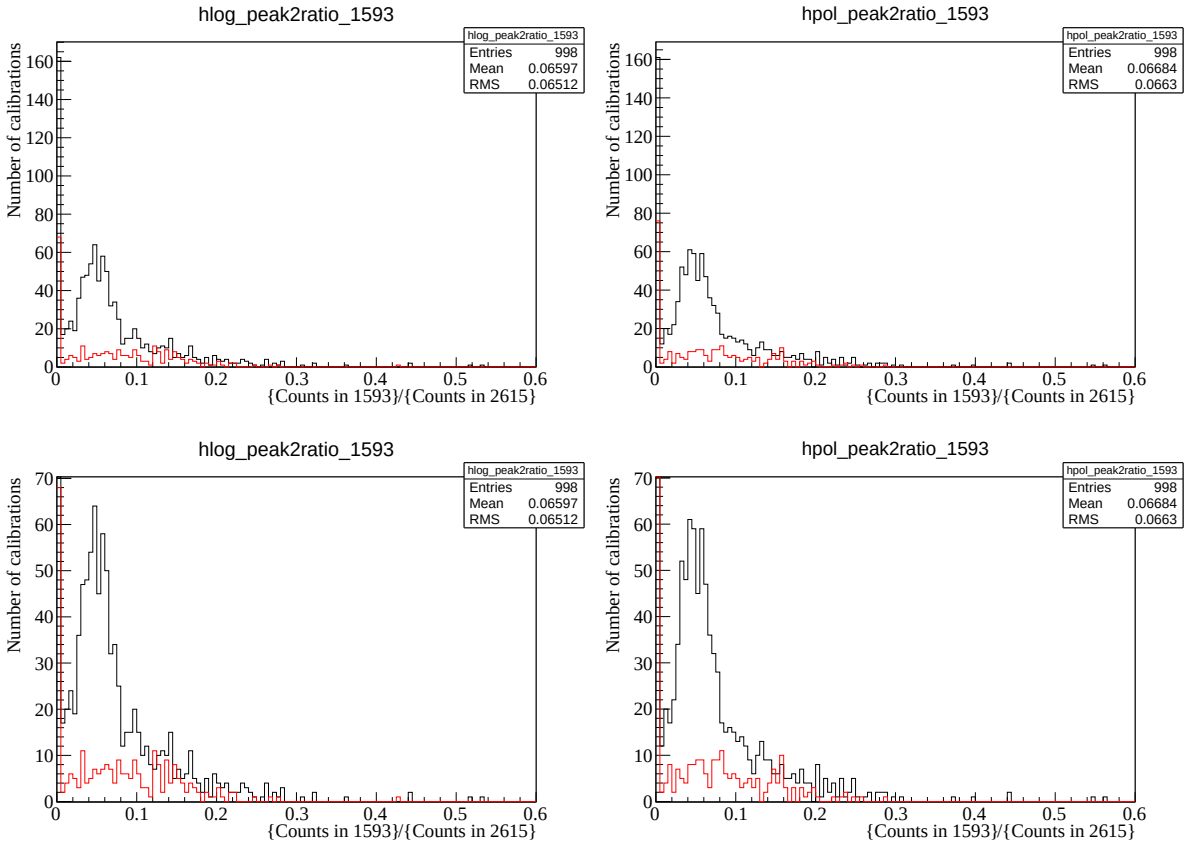


Figure 6.6: 'Ratio of counts in 1593 peak to counts in 2615 peak' distributions for 1593 peak. Peaks identified differently by logpol2 and pol3 processings (red histogram) and 'reference' distribution (constructed from peaks that are believed most likely to be correctly identified; black histogram) are shown. The bottom figures are zooms on the lower portions of the top figures for visibility. Left: logpol2 processing. Right: pol3 processing.

from zero for the logpol2 processing. This result is consistent with the residual trends observed from studying the gamma peaks in Cuoricino background data, as presented in Section6.3.2. It can also be seen from Figure 6.5 that the spread in peak-location guesses obtained from the pol3 interpolation function, compared to the position of the located peak, is much wider than that corresponding to the logpol2 interpolation function. This behavior is consistent with the expectation that the pol3 function can experience larger shifts due to having one more free parameter than the logpol2 function.

A conspicuous feature of Figures 6.3 – 6.6 is that many of the distributions appear to split into two distinct populations instead of manifesting only a single peak as expected, especially among the calibrations in which the 1593 peak was identified differently by the two processings. It is also evident from the large zero bin in Figure 6.6 that it is common for the peak fit to see only a single peak instead of a double peak shape; the zero bin is made of up calibrations in which the ^{228}Ac peak (amplitude proportional to $(1 - amp2)$) is 'chosen' by the fit as that single peak and the fit returns $amp2 \approx 0$. This behavior implies that, due to the lack of constraints on the $amp2$ parameter, the 1593 peak may be 'interfering with itself' — if one of the component peaks is commonly misidentified as the other, it would tend to form a second distinct peak in the peak-quality metric distributions, as seen. In particular, it seems that the lower-energy ^{228}Ac peak may be commonly misidentified as the higher-energy ^{208}Tl double-escape peak, pushing the identified peak location several keV below where it should be and giving rise to the structures that can be clearly seen on the low-energy tails of the residual distributions in Figure 6.3.

Histogramming the difference between the two identified peak means for the 273 'differently identified' peaks in units of approximate energy, as in Figure 6.7, shows that the peak identifications differ by $\lesssim 5$ keV in the majority of cases, roughly consistent with the nominal 4.34-keV energy difference between the ^{228}Ac peak and the double-escape peak. In 32 of these cases, the "1588" peak identified by the pol3 processing is consistent with being identical to the "1593" peak identified by the logpol2 processing; in 31 cases, the logpol2-identified "1588" peak is consistent with the pol3-identified "1593" peak.

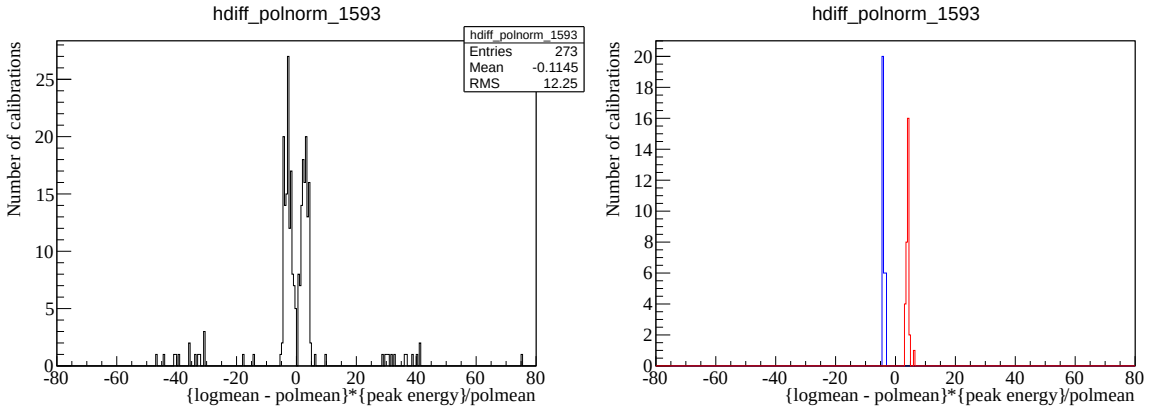


Figure 6.7: Left: Difference between mean of 1593 peak identified by logpol2 processing and mean of 1593 peak identified by pol3 processing, converted to approximate energy units. Right: Same as left-hand figure, for only those calibrations in which either the pol3-identified "1588" peak is identical to the logpol2-identified "1593" peak (blue) or the logpol2-identified "1588" peak is identical to the pol3-identified "1593" peak (red). The definition of peak identicalness used to generate the right-hand figure follows Equation 6.3. The free parameter used by the fit to determine the position of the 1593 peak is actually the mean of the "1588" ^{228}Ac peak, so the fitted mean and error of the "1593" double-escape peak is calculated from the fitted mean and error of the "1588" peak using the ratio of the nominal energies of the two peaks.

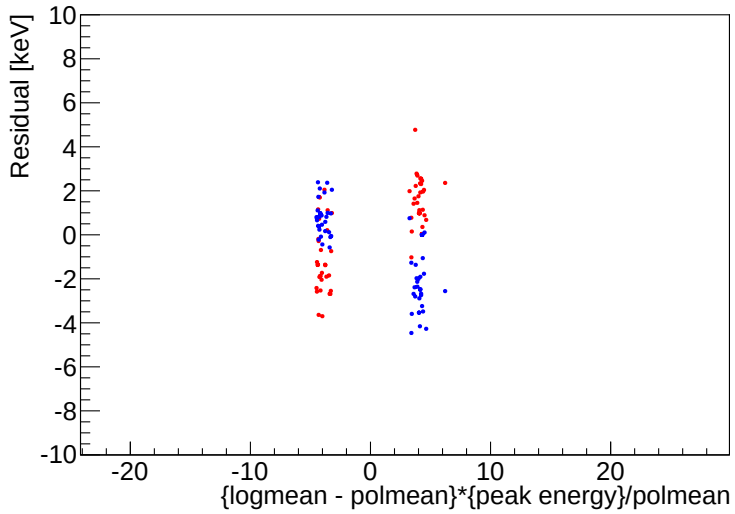


Figure 6.8: Residuals vs. the difference between the peak means identified by the two different processings of the calibration: logpol2-identified peak residuals (red) and pol3-identified peak residuals (blue). The x-axis serves only to separate the calibrations into 'pol3 "1588" = logpol2 "1593"' (left) and 'logpol2 "1588" = pol3 "1593"' (right), the same visual separation as in Figure 6.7. The behavior of the residuals is consistent with a '1588-as-1593' peak misidentification being the dominant occurrence, although a small population consistent with a '1593-as-1588' peak misidentification is also present.

In Figure 6.8, the residuals of the peaks found by the two processings for these 63 calibrations are plotted vs. the energy difference between the two identified peaks to preserve the visual separation of Figure 6.7; the left cluster corresponds to 'pol3-identified "1588" = logpol2-identified "1593"' and the right cluster corresponds to 'logpol2-identified "1588" = pol3-identified "1593"'. These residuals show behavior that is largely consistent with the hypothesis that the 1588 peak is stronger than the double-escape peak and often misidentified as the double-escape peak (i.e., the left cluster is mostly composed of correct pol3 peak identifications and incorrect logpol2 peak identifications, while the right cluster is mostly composed of correct logpol2 peak identifications and incorrect pol3 peak identifications):

- The majority of the blue points (pol3-identified peak residuals) in the left cluster are centered on zero, following the expected behavior of a correct peak identification, while the majority of the red points (logpol2-identified peak residuals) in the left cluster are too low.
- The majority of the red points in the right cluster are high, following the observed behavior of the logpol2 residuals in this energy region (see Section 6.3.2), while the majority of the blue points in the right cluster are too low.

This figure also shows some evidence of a small population of instances in which the alternate peak misidentification (i.e., 1593-as-1588) may have taken place: there is a grouping of 6 too-high blue points in the left cluster and a grouping of 5 near-zero blue points in the right cluster (no clean indicator of too-high red points can be found because of the logpol2 calibration's tendency to push residuals too high in this energy region).

The only direct indications of these 'one peak misidentified as the other' errors come from these calibrations in which the two processings identified two different peaks as the 1593. However, we would like to try to use this information to refine our selection of 'most likely to be correctly identified' peaks used to construct our reference distributions. Considering the apparent frequency of '1588-as-1593' peak misidentification and the fact that the "1588" is typically stronger than the "1593", it is likely that the peak misidentification is driven by relative peak strengths: the fit correctly finds the stronger peak but not the weaker one, then assigns the stronger peak to the wrong

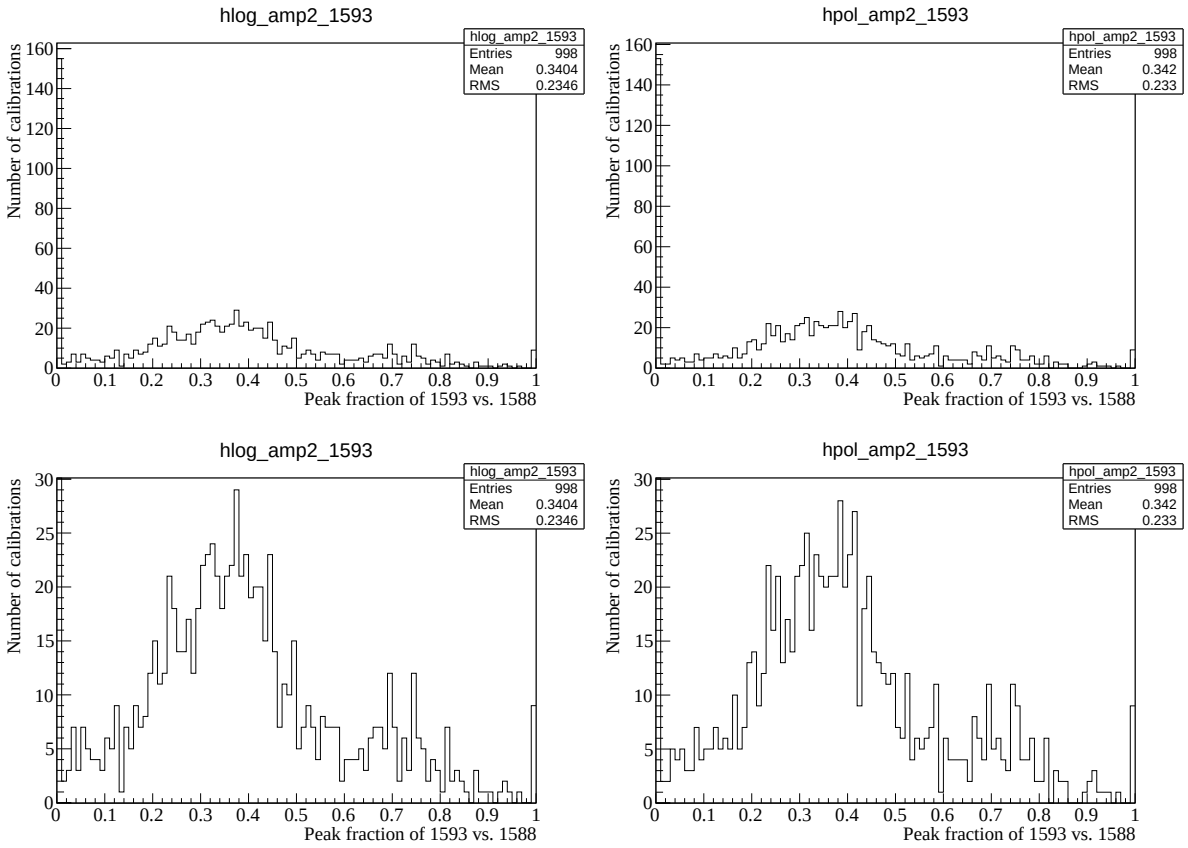


Figure 6.9: 'Reference' distributions (constructed from peaks that are believed most likely to be correctly identified) of $amp2$ for the 1593 peak. The distributions should be identical for both processings, differing only due to small shifts in peak fits; the Kolmogorov probability that they are the same is 0.999868. A double-peak structure is clearly evident. The bottom figures are zooms on the lower portions of the top figures for visibility. Left: logpol2 processing. Right: pol3 processing.

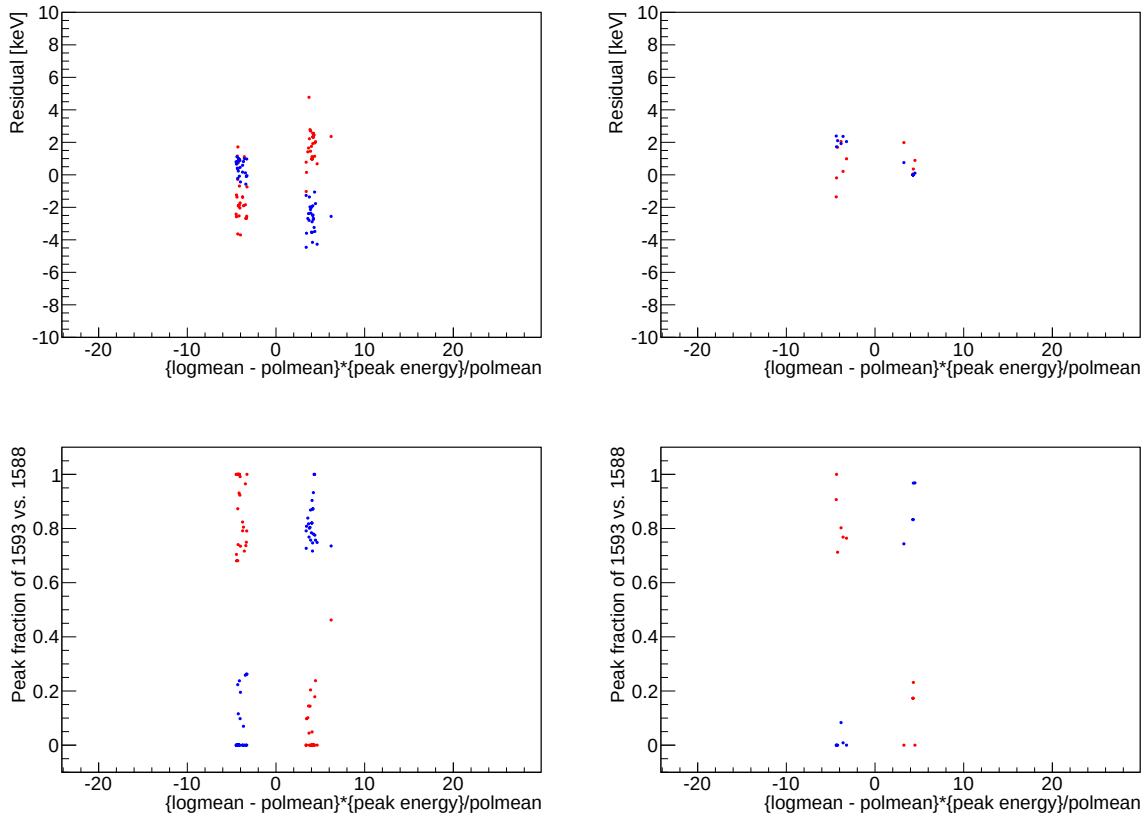


Figure 6.10: Top: Figure 6.8 split into the populations corresponding to inferred '1588-as-1593' misidentifications (left-hand figure) and inferred '1593-as-1588' misidentifications (right-hand figure). Bottom: Analogous scatter plots of $amp2$. In the '1588-as-1593' case (left-hand figure), the logpol2-identified (red) points in the left cluster and the pol3-identified (blue) points in the right cluster are believed to be incorrectly identified, and a cut of $amp2 < 0.65$ can be used to reject them; in the '1593-as-1588' case (right-hand figure), the pol3-identified (blue) points in the left cluster and the logpol2-identified (red) points in the right cluster are believed to be incorrectly identified, and a cut of $amp2 > 0.1$ can be used to mostly reject them.

energy. This is possible because of the lack of restriction on the $amp2$ fit parameter. Therefore, perhaps a cut on $amp2$ could preferentially select against this kind of identification mistake; this is a cut that can be straightforwardly made on the calibrations used to build the reference histograms. The $amp2$ reference distributions for the logpol2 and pol3 processings (which ought to be identical) and scatter plots of $amp2$ vs. peak-mean differences (only to maintain visual separation, as in Figure 6.8) are shown in Figures 6.9 and 6.10, respectively. From these figures, it can be seen that a cut of $0.1 < amp2 < 0.65$ rejects the great majority of peaks that we are confident are incorrectly identified while retaining reasonable statistics.

The effect of applying this $amp2$ cut to the reference distributions is illustrated in Figures 6.11 – 6.14. Both sides of the residual distributions are strongly reduced, making them narrower and also somewhat more symmetric. The shape of the logpol2 deviation distribution changes little, although most of the farthest outliers are removed; the pol3 deviation distribution, however, loses a significant portion of the high-energy side, becoming consistent with other indications that the pol3 calibration function has a mild tendency to push residuals low in this energy region (see Section 6.3.2). The logpol2 'distance from interpolation peak location guess to mean of identified peak' distribution is affected much as the residual distributions are, but the pol3 distribution, in contrast, experiences the greatest reduction toward the center of the distribution. Finally, the counts-in-peak ratio distributions are affected exactly as expected; the great majority of the cuts fall in the zero bin and the high tail because the counts-in-peak ratio is proportional to $amp2$.

To test the real discriminatory power of the reference distributions after the application of the $amp2$ cut, Table 6.1 summarizes the results of Kolmogorov shape-comparison tests between the reference distributions and the four distributions that can be extracted from the 52 calibrations that appear to be instances of one processing mistaking the "1588" peak as the "1593" peak; in 26 cases, the position of the logpol2-identified "1588" is the same as that of the pol3-identified "1593" and the logpol2-identified peaks are believed to be the correct ones (right cluster in the left-hand plots of Figure 6.10), while in the other 26 cases, the position of the pol3-identified "1588" is the same as that of the logpol2-identified "1593" and the pol3-identified peaks are believed to be the correct ones (left cluster in the left-hand plots of Figure 6.10).

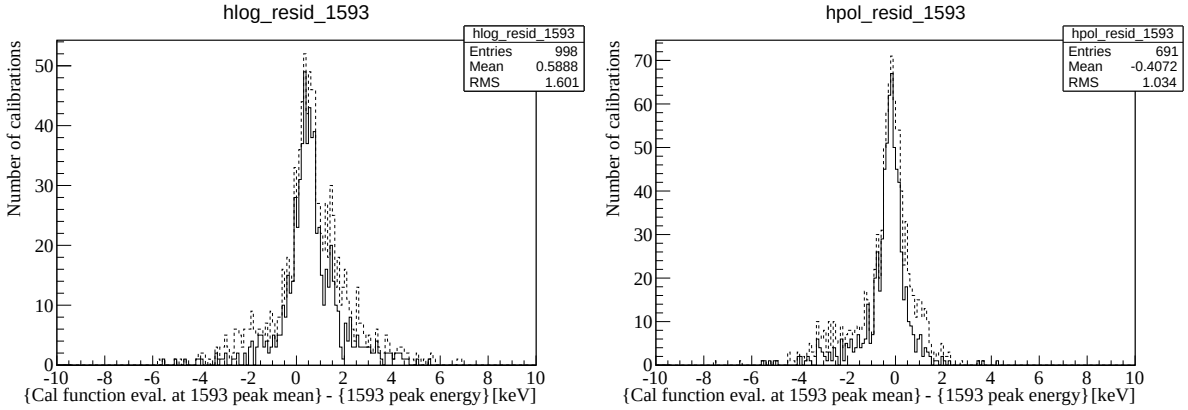


Figure 6.11: 'Residual from calibration fit' reference distributions for 1593 peak before (dotted line) and after (solid line) the application of the $0.1 < amp2 < 0.65$ cut. Left: logpol2 processing. Right: pol3 processing.

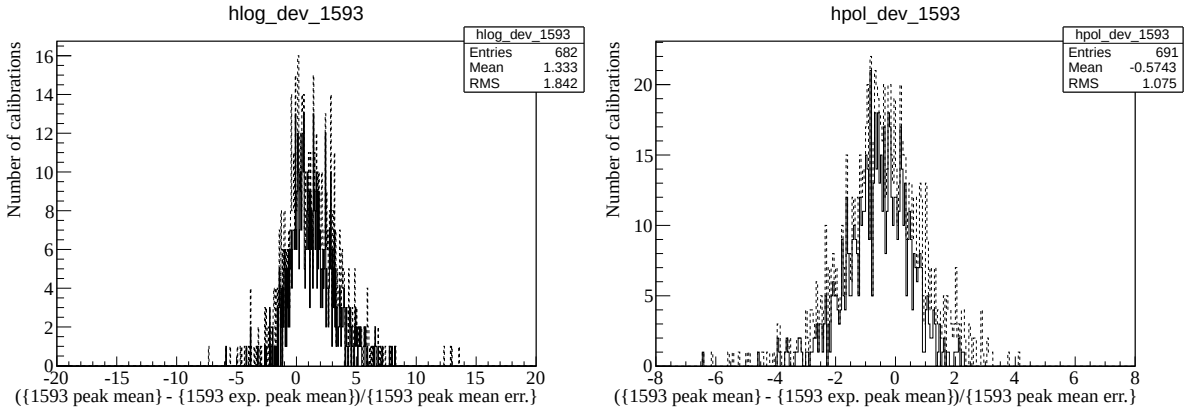


Figure 6.12: 'Deviation from calibration fit' reference distributions for 1593 peak before (dotted line) and after (solid line) the application of the $0.1 < amp2 < 0.65$ cut. Left: logpol2 processing. Right: pol3 processing.

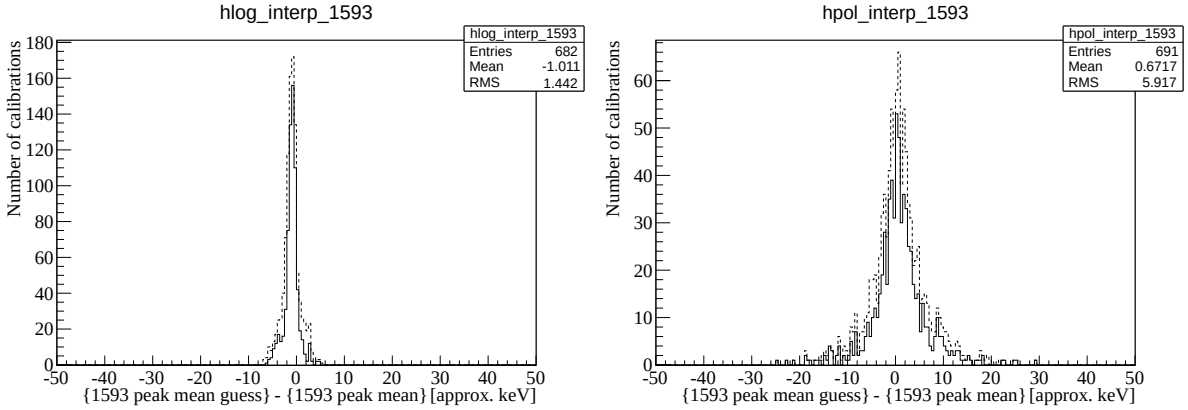


Figure 6.13: 'Distance from interpolation peak location guess to mean of identified peak' reference distributions for 1593 peak before (dotted line) and after (solid line) the application of the $0.1 < amp2 < 0.65$ cut. Left: logpol2 processing. Right: pol3 processing.

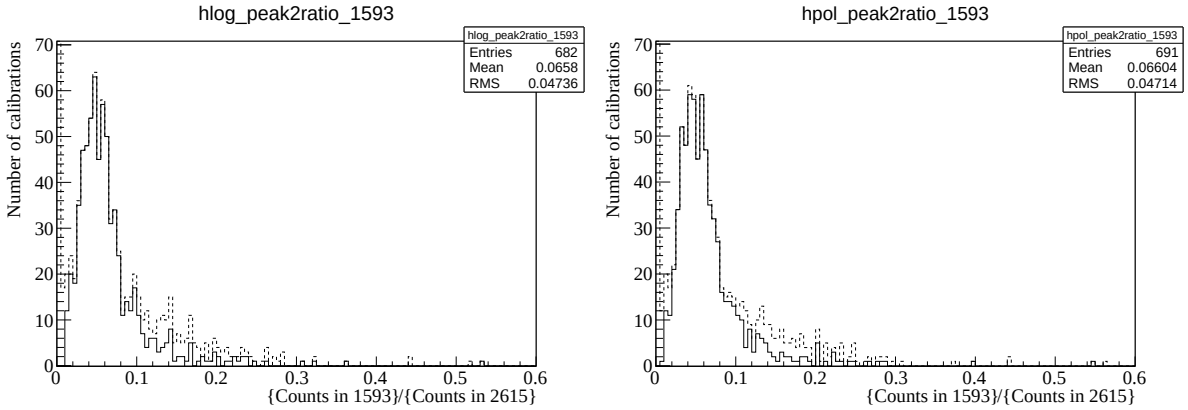


Figure 6.14: 'Ratio of counts in 1593 peak to counts in 2615 peak' reference distributions for 1593 peak before (dotted line) and after (solid line) the application of the $0.1 < amp2 < 0.65$ cut. Left: logpol2 processing. Right: pol3 processing.

Table 6.1: Results of Kolmogorov shape-comparison tests between reference distributions of peak-quality metrics and four test distributions: logpol2-identified peaks and pol3-identified peaks believed to be correctly or incorrectly identified. Test distributions drawn from peaks identified by the logpol2 processing are compared to reference distributions also drawn from peaks identified by the logpol2 processing, and similarly for the pol3 processing.

Test distribution			Kolmogorov prob. of compatibility w/ reference			
Peak (mis)ID	Processing	Believed correct?	Residual	Deviation	Distance from interp. to peak	Cts-in-peak ratio
"1588"$_{\text{logpol2}}$ = "1593"$_{\text{pol3}}$	logpol2	correct	2.13632e-07	0.0297263	1.77367e-07	2.70773e-14
"1588"$_{\text{pol3}}$ = "1593"$_{\text{logpol2}}$	logpol2	incorrect	1.47907e-11	1.44363e-11	1.2152e-12	6.87232e-14
"1588"$_{\text{pol3}}$ = "1593"$_{\text{logpol2}}$	pol3	correct	6.8929e-07	2.08324e-07	0.753539	1.9839e-13
"1588"$_{\text{logpol2}}$ = "1593"$_{\text{pol3}}$	pol3	incorrect	5.03309e-16	2.87189e-10	0.255393	8.47793e-13

The first conclusion that can be drawn from the results in Table 6.1 is that the 'distance from interpolation peak location guess to mean of identified peak' has no discrimination power for the pol3 processing. This is presumably caused by the wide spread of the distribution of this metric for the pol3 processing, which we have attributed to the pol3 function experiencing large shifts in the interpolation function in energy regions far from primary peaks due to having one more free parameter than the logpol2 function: a large distance between the interpolation guess for the peak location and the location of the peak that is ultimately found is not necessarily any indication that the peak may be incorrectly identified. The only value that can be drawn from looking at the distributions of this metric is to notice that the pol3 distribution is centered on zero and the logpol2 distribution is very narrow, both desirable characteristics. If the narrower spread of the logpol2 distribution can indeed be attributed to the fact that the logpol2 function has one fewer free parameter than the pol3, this is a promising indication that the pol3 function with the intercept fixed to zero may preserve both the lack of bias of the pol3 and the narrow spread of the logpol2.

The second conclusion that can be drawn from the results in Table 6.1 is that, because it is sensitive to the shapes of both the reference distributions and the test distributions, the Kolmogorov test cannot be used to directly compare the compatibilities of two test distributions to two differently shaped reference distributions. The clearest demonstration of this is the result of the deviation test: the comparison of the correctly logpol2-identified peaks to the logpol2 reference distribution yields a Kolmogorov probability of 0.0297263, while the comparison of the correctly pol3-identified peaks to the pol3 reference distribution yields a Kolmogorov probability of only 2.08324e-07.

A related observation is that none of the Kolmogorov probability values is very high; the test distributions, drawn as they are from the specific population of calibrations where the two processings identified the same structure and each assigned it to a different one of the two component peaks of the 1593 peak, are not representative of the distributions associated with the population of *all* 'probably correct' peaks. However, a clear preference for the correctly identified peaks over the incorrectly identified peaks *can* be seen in the residual and deviation metrics when comparing tests

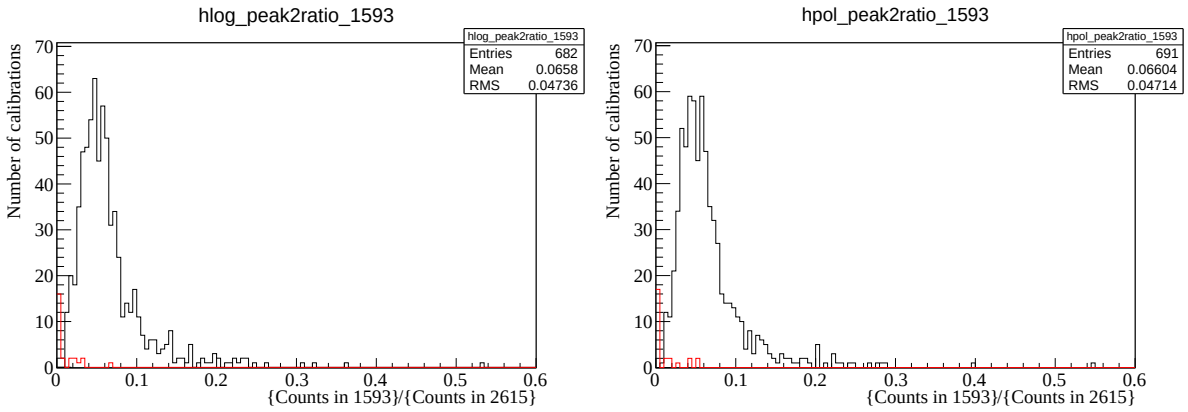


Figure 6.15: The counts-in-peak ratio reference distributions (black) and 'correctly identified' test distributions (red). The $amp2$ cut removed the zero bin from the reference distribution, significantly impacting the result of the Kolmogorov compatibility test. Left: logpol2 processing. Right: pol3 processing.

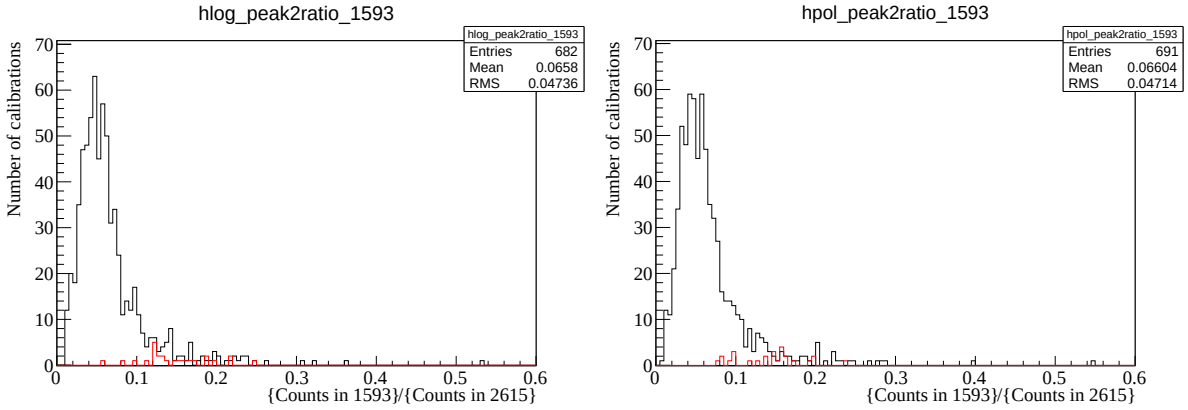


Figure 6.16: The counts-in-peak ratio reference distributions (black) and 'incorrectly identified' test distributions (red). The test distributions clearly fall in the far high tails of the reference distributions. Left: logpol2 processing. Right: pol3 processing.

of compatibility with the same reference distribution, so there *is* some discrimination power available that can be exploited. The exception is the counts-in-peak ratio, which just slightly prefers the incorrectly identified peaks. This is probably because the $amp2$ cut removed peaks where $amp2 \approx 0$, while a large percentage of the '1588-as-1593' peak identification mistakes seem to occur in cases where the "1593" peak cannot be resolved and thus the $amp2 \approx 0$ fit is likely to actually be the correct one (see Figures 6.15 and 6.16). With some care, however, it should still be possible to use this metric.

What we would like to do is construct an analysis method that will allow us to compare the logpol2 peak identification to the pol3 peak identification on a calibration-by-calibration basis, thus permitting us to determine which of the two is *more likely* to be the correct one in each case. It should be possible to define a properly normalized quality score:

1 First define a *residual score*. The residual and the deviation provide essentially the same information — distance from the final calibration function — and the discriminatory power of the residual appears to be more comparable between the logpol2 and the pol3.

   (a) Determine the error on the residual, σ_{residual}, from the fitted error on the mean, σ_μ, as

   $$\frac{1}{2}\times([\text{calibration function evaluated at}(\mu+\sigma_\mu)]+[\text{calibration function evaluated at}(\mu-\sigma_\mu)]).$$

   (b) Take the integral of the reference histogram from the bin associated with (residual $-$ σ_{residual}) to the bin associated with (residual $+$ σ_{residual}) and divide by the number of entries in the reference histogram.

   (c) Divide by the number of bins spanned or, if the error range spills over into the overflow and/or underflow bins, the number of bins that *would* be spanned if the histogram range extended far enough.

2 Then define a *counts-in-peak ratio score*.

   (a) Determine the error on the counts-in-peak ratio score, σ_{ratio}, using Equation 6.4 and the fitted errors on $amp2$, $f_{bkg_{1593}}$, and $f_{bkg_{2615}}$.

(b) Take the integral of the reference histogram from the bin associated with (ratio $-\ \sigma_{\text{ratio}}$) to the bin associated with (ratio $+\ \sigma_{\text{ratio}}$) and divide by the number of entries in the reference histogram.

(c) Divide by the number of bins spanned or, if the error range spills over into the overflow and/or underflow bins, the number of bins that *would* be spanned if the histogram range extended far enough.

3 Finally, define the final quality score as (residual score)$\times$(counts-in-peak ratio score).

For a given calibration, the peak identification from the logpol2 processing is preferred if the quality score obtained for that peak is larger than the corresponding quality score obtained for the pol3-identified peak, and vice versa.

The results of performing this quality-score analysis on several subsets of the calibrations for which the logpol2 and pol3 processings identify different 1593 peaks are summarized in Table 6.2. The results, in large part, support the hypotheses of correct vs. incorrect peak identifications in the test populations inferred from Figure 6.8; this is especially true once any calibration for which *at least one* peak fit results in $amp2 \approx 0$, in which instance it is not clear that attempting to determine which peak is 'better' is meaningful, has been removed.

Overall, the performance of the pol3 interpolation function appears mildly superior to that of the logpol2 interpolation function, but the difference is so small as to be insignificant. In particular, the calibrations for which the peak identified by the logpol2 processing and the peak identified by the pol3 processing are separated by $\lesssim 5$ keV, which are of the greatest concern because they will have the greatest impact on the final calibration fit, show no significant preference for one function over the other at all.

A similar analysis can be performed for the 49 calibrations in which the single-escape peak (the 2104 peak) is identified differently by the two processings. This peak has only a single-peak shape, so it does not experience the same self-interference phenomenon or $amp2 \approx 0$ problem that the 1593 peak does; this probably accounts for its greater identification consistency, but it also means

Table 6.2: Results of quality-score analysis on the test distributions, for which the 'correct' peak identification can be independently inferred from Figure 6.8, and the remaining 210 calibrations in which the logpol2 and pol3 processings identify different 1593 peaks. While a mild preference for the pol3 interpolation function is evident, the significance of that preference is not compelling.

Calibration selection	Number of calibrations after the given cut for which the peak identified by the logpol2 / pol3 processing is the preferred one			
	No cut	$amp2 >\sim 0$	sep.$>\sim 5$ keV	sep.$>\sim 5$ keV & $amp2 >\sim 0$
"1588"-as-"1593" **logpol 'correct'**	**23** / 3	**9** / 0	**–** / –	**–** / –
"1588"-as-"1593" **pol3 'correct'**	5 / **21**	0 / **9**	– / **–**	– / **–**
"1593"-as-"1588" **logpol 'correct'**	**5** / 1	**2** / 0	**–** / –	**–** / –
"1593"-as-"1588" **pol3 'correct'**	2 / **3**	1 / **2**	– / **–**	– / **–**
Remaining 210 calibrations	95 / 115	48 / 65	15 / 16	7 / 10
Total	130 / 143	60 / 76	15 / 16	7 / 10

that there is no $amp2$ parameter to provide an additional handle on peak quality. Figures 6.17 – 6.20 show the distributions of the peak quality metrics for the 2104 peak. The shapes of the reference residual distributions are quite clean, especially that of the pol3 distribution, which seems to follow the same Crystal Ball shape seen on the stronger peaks in the energy spectrum of the detectors. There is a structure on the high-energy tail of the reference residual distribution for the logpol2 processing, however, that suggests that the logpol2 interpolation function may sometimes misidentify a higher-energy structure as the 2104 peak.

The quality-score analysis on the 2104 peak reveals that the peak identified by the logpol2 processing is preferred for 26 of the calibrations, while the peak identified by the pol3 processing is preferred for the remaining 23. This mild preference for the logpol2 interpolation function is no more compelling than the preference for the pol3 seen in the 1593 peak, especially considering the possible evidence of peak misidentification in the shape of the logpol2 residual distribution.

In summary, neither the logpol2 nor the pol3 shows a convincingly superior performance as the interpolation function used to search for secondary calibration peaks. This being the case, the recommended choice of interpolation function for use in CUORE-0 and CUORE is to use the default calibration function, which will be a pol3 with intercept set to zero. There are promising indications that this function may actually perform better as the interpolation function than either the logpol2 function or the pol3 function used for Cuoricino analysis.

The $amp2$ parameter's freedom to vary between 0 and 1 for the 1593 peak will continue to contribute to instability in the peak-finding behavior on this peak unless the range of the parameter can be reduced in a sensible way. Unfortunately, as there is no *a priori* relationship that can be assumed between the strengths of the double-escape peak and the ^{228}Ac peak, any $amp2$ range restriction would have to be motivated by simulation studies. The difficulty of this issue will be exacerbated in CUORE, where the more complicated geometry will mean that the expected relative amplitude between the two peaks may very quite significantly from one crystal to another; however, there is currently no infrastructure available in the calibration module for setting peak-fit parameters or parameter ranges differently for different channels. A distribution of $amp2$ for the CUORE array drawn from the same simulation presented in Section 5.5 is shown in Figure 6.21.

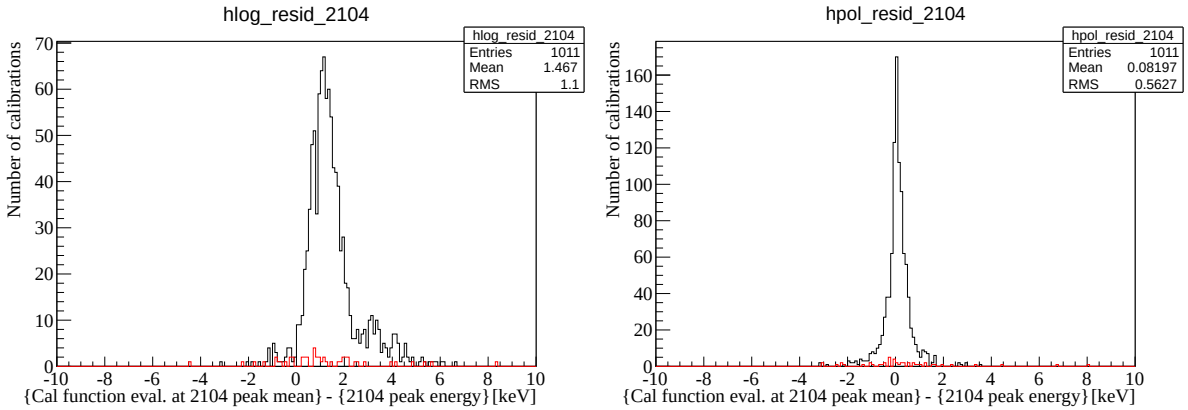


Figure 6.17: 'Residual from calibration fit' distributions for 2104 peak. Peaks identified differently by logpol2 and pol3 processings (red histogram) and 'reference' distribution (constructed from peaks that are 'most likely' to be correctly identified; black histogram) are shown. Left: logpol2 processing. Right: pol3 processing.

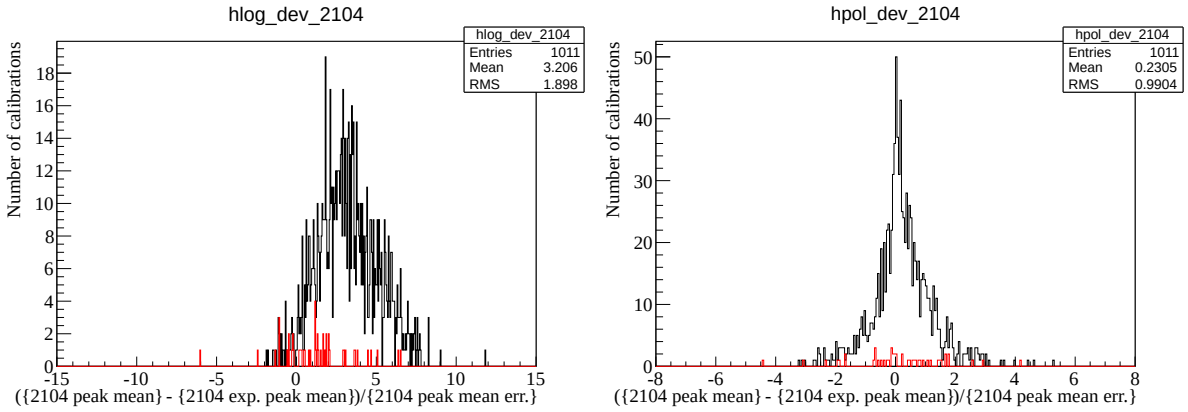


Figure 6.18: 'Deviation from calibration fit' distributions for 2104 peak. Peaks identified differently by logpol2 and pol3 processings (red histogram) and 'reference' distribution (constructed from peaks that are 'most likely' to be correctly identified; black histogram) are shown. Left: logpol2 processing. Right: pol3 processing.

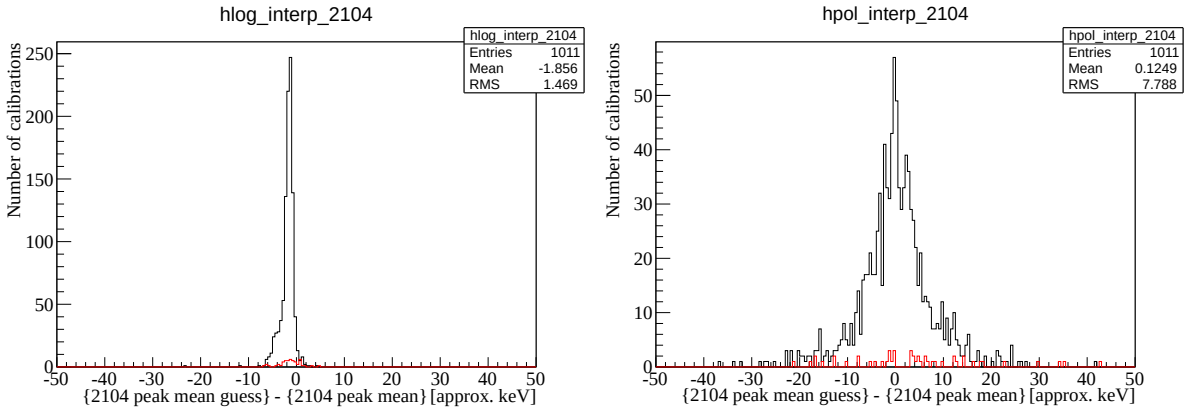


Figure 6.19: 'Distance from interpolation peak location guess to mean of identified peak' distributions for 2104 peak. Peaks identified differently by logpol2 and pol3 processings (red histogram) and 'reference' distribution (constructed from peaks that are 'most likely' to be correctly identified; black histogram) are shown. Left: logpol2 processing. Right: pol3 processing.

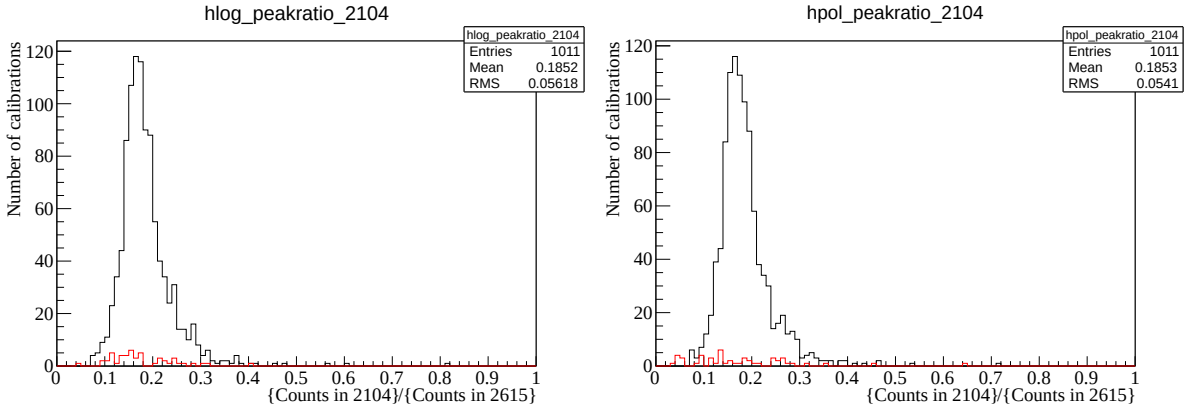


Figure 6.20: 'Ratio of counts in 2104 peak to counts in 2615 peak' distributions for 2104 peak. Peaks identified differently by logpol2 and pol3 processings (red histogram) and 'reference' distribution (constructed from peaks that are 'most likely' to be correctly identified; black histogram) are shown. The Kolmogorov probability of the two reference distributions being identical is 0.998817. Left: logpol2 processing. Right: pol3 processing.

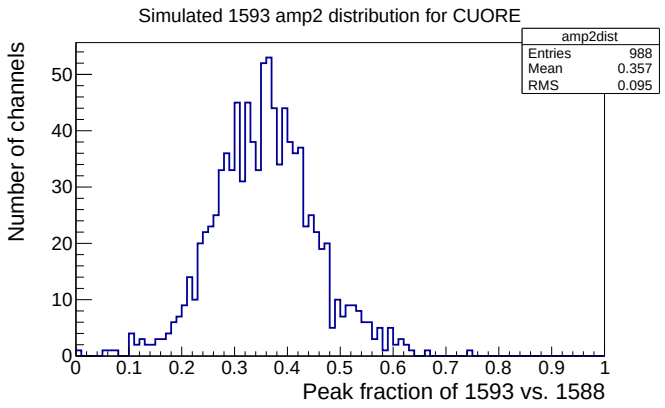


Figure 6.21: Simulated distribution of $amp2$ values for the 1593 peak in the CUORE array.

6.3 Evaluation of Calibration Uncertainty

A number of known background gamma peaks of varying strengths appear in the energy spectra collected by the Cuoricino detectors during normal data-taking. Once the Cuoricino data have been calibrated, the positions at which these lines appear in the calibrated sum spectrum can be compared to their nominal energies to gain an understanding of the effectiveness of the calibration procedure. The calibration performance in the region of interest (R.O.I.) surrounding the ^{130}Te $\beta\beta$ Q-value is of the most critical concern to the search for $0\nu\beta\beta$, while the investigation of lines throughout the gamma region of the spectrum can provide an understanding of the overall behavior of the detectors.

6.3.1 $0\nu\beta\beta$ R.O.I. Uncertainty

6.3.1.1 Energy Calibration Uncertainty in the Final Cuoricino Result

For the spectrum that was analyzed to yield the final Cuoricino result, the energy uncertainty arising from the calibration was evaluated by considering the residuals of peaks in the background spectrum near the region of interest. The following discussion is specifically aimed at determining the calibration uncertainty at the Q-value of double beta decay in ^{130}Te, but the approach can be easily generalized to estimate the uncertainty at other energies.

Table 6.3: Gamma lines used for extracting calibration uncertainty from residuals.

Source	Energy (keV)
^{208}Tl	2614.533
^{60}Co (sum peak)	2505.738
^{214}Bi	2447.86

The three peaks near the region of interest for ^{130}Te are listed in Table 6.3. The procedure for calculating the calibration uncertainty is as follows[4]:

1 In a spectrum constructed from the same cuts[5] used for the double-beta decay analysis, the background peaks are fit with the same response function used for the double-beta decay analysis with the mean left free (Figure 6.22). Two parallel approaches were taken to the final double-beta decay analysis in Cuoricino:

   (a) **'Berkeley type':** Each channel and data set is treated as a subset of the data, consisting of a Gaussian peak with its own fixed resolution and a flat continuum background; the fit is performed simultaneously to all channels and data sets at once, assuming the mean of the peak is the same for each.

   (b) **'Milano type':** The data is subdivided into only three categories — the big, small, and enriched crystals. The response function for each subset is a sum of Gaussians, one corresponding to each channel/dataset in the subset, using the associated fixed resolution and weighted by the appropriate exposure. Again, the three subsets are fit simultaneously, assuming the mean of the peak is the same for each.

2 The nominal energy of each peak is subtracted from the fitted mean of that peak, resulting in a set of three residuals. Each residual r_i is assigned a weight w_i proportional to $\frac{1}{\sigma_i^2}$, where σ_i is the error on the fitted mean; the weights are normalized such that $\sum_{i=1}^{3} w_i = 1$.

3 This set of three weighted points is used to calculate four values:

[4]The validity of this approach depends upon the validity of the assumption that the process of interest will 'look' like a gamma interaction to the detectors. The ^{60}Co residual is additionally problematic, being the result of the coincident interaction of two gammas (the 1173.237-keV and 1332.501-keV gammas from the decay of ^{60}Co) in the same detector, rather than a single gamma; it is possible that the fact that the ^{60}Co residual is notably larger than the other two is due to a real systematic effect that may arise from this physical difference. Thus, a conservative alternate choice for determining the calibration uncertainty for the double-beta decay analysis is to simply take it to be of the same magnitude as the ^{60}Co residual itself.

[5]The third-order polynomial calibration failed on a small number of channels and data sets. As the failed calibrations affect only about 1% of the data, the failed channels and data sets are simply excluded from this analysis for the third-order polynomial case.

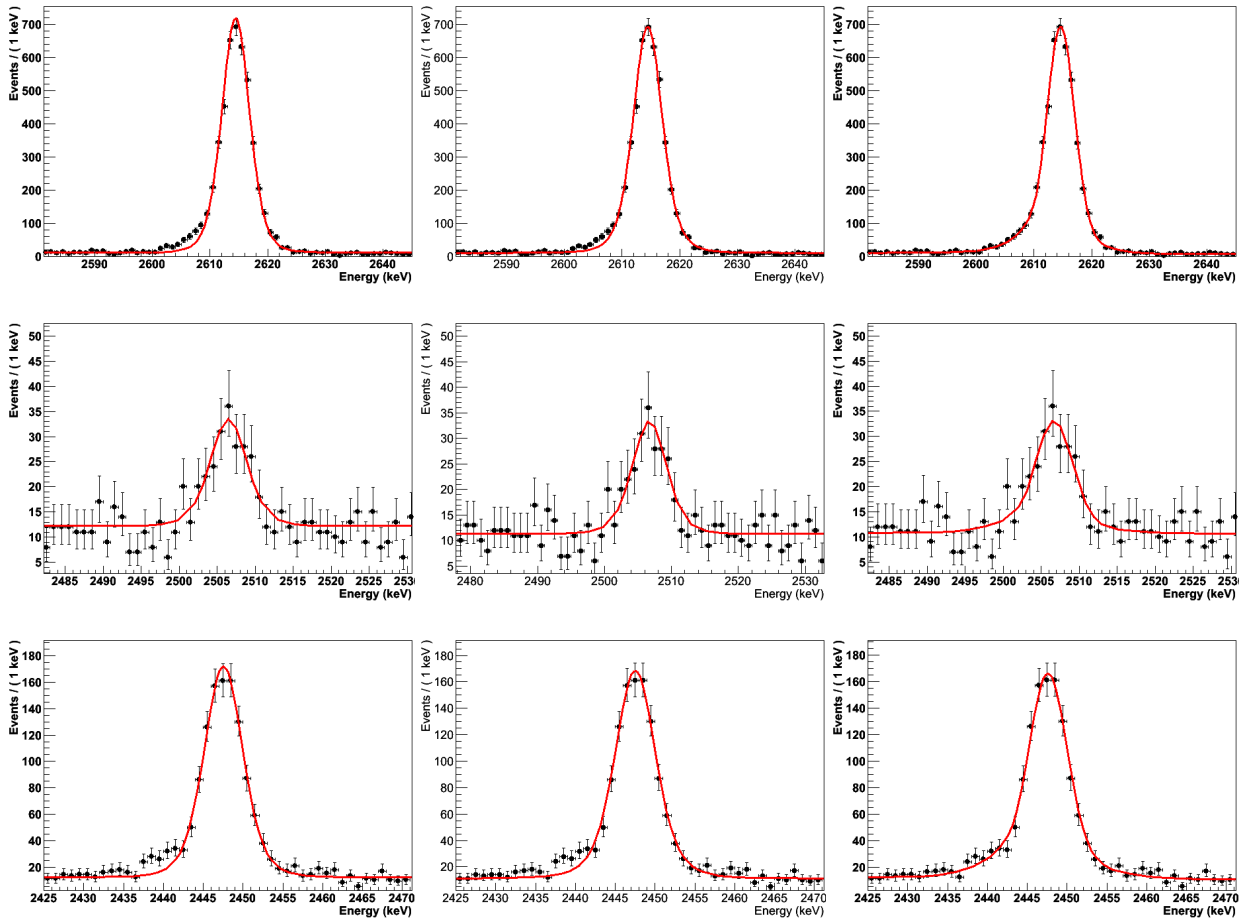


Figure 6.22: Left to right: Berkeley-type, Milano-type, and Crystal-Ball-based fits. Top to bottom: ^{208}Tl, ^{60}Co, and ^{214}Bi peaks. The data are calibrated with the third-order polynomial function and are selected using the same cuts as were used for the final Cuoricino result, except that channels/datasets on which the pol3 calibration failed are simply removed instead of being replaced by data calibrated with the logpol2 calibration.

(a) **Weighted mean of the residuals:** The usual weighted mean μ^* of a set of points, including the signs of the residuals. Ideally this should be consistent with zero.

(b) **Weighted standard deviation of the residuals:** Analogous to the usual sample standard deviation, taking into account the different weights.

$$s = \sqrt{\frac{1}{1 - \sum_{i=1}^{3} w_i^2} \sum_{i=1}^{3} w_i \left(r_i - \mu^*\right)^2} \tag{6.5}$$

(c) **Weighted mean of the absolute values of the residuals:** Intended as a measure of the average magnitude of the residuals.

(d) **Linear fit to the residuals, evaluated at the Q-value:** Estimate of the 'expected' residual at the Q-value, assuming the energy range is small enough for a linear trend to be a valid approximation.

4 We estimate the calibration uncertainty as 3b above if 3a is consistent with zero; otherwise, we estimate the calibration uncertainty as the sum of 3b and 3a.

5 3c and 3d, as well as the residuals themselves from the ^{208}Th and ^{214}Bi lines, should fall within the calibration uncertainty estimated in the previous step.

The results of step 3 above are tabulated in Table 6.4, and their relationship to the residuals themselves is presented graphically in Figure 6.23.

There are two aspects in which it is questionable whether it is really appropriate to extend the fit procedure for the double-beta decay analysis to the calculation of residuals. First, to fit all three peaks in a single window requires a large window width of 220 keV; the assumption of a flat continuum background may no longer hold over this wide an energy range. Second, the ^{208}Tl and ^{214}Bi peaks show a distinct asymmetry; both have a low-energy tail, implying that the peaks may not be fit well by a Gaussian. In either case, using the 'incorrect' model (for the continuum background or for the peak, respectively) may shift the fitted position of the peak.

Table 6.4: Results of residual analysis near the ^{130}Te double-beta decay region of interest for the determination of the calibration uncertainty on the final Cuoricino result.

Fit type	Functional form	Weighted mean	Weighted standard deviation	Weighted mean of absolute values	Linear fit evaluated at Q-value
Berkeley	logpol2	-0.13 ± 0.13	0.35	0.21 ± 0.07	0.01 ± 0.05
Berkeley	pol3	-0.13 ± 0.09	0.23	0.15 ± 0.07	-0.20 ± 0.05
Milano	logpol2	-0.11 ± 0.15	0.38	0.19 ± 0.09	0.03 ± 0.06
Milano	pol3	-0.12 ± 0.11	0.29	0.15 ± 0.09	-0.19 ± 0.06

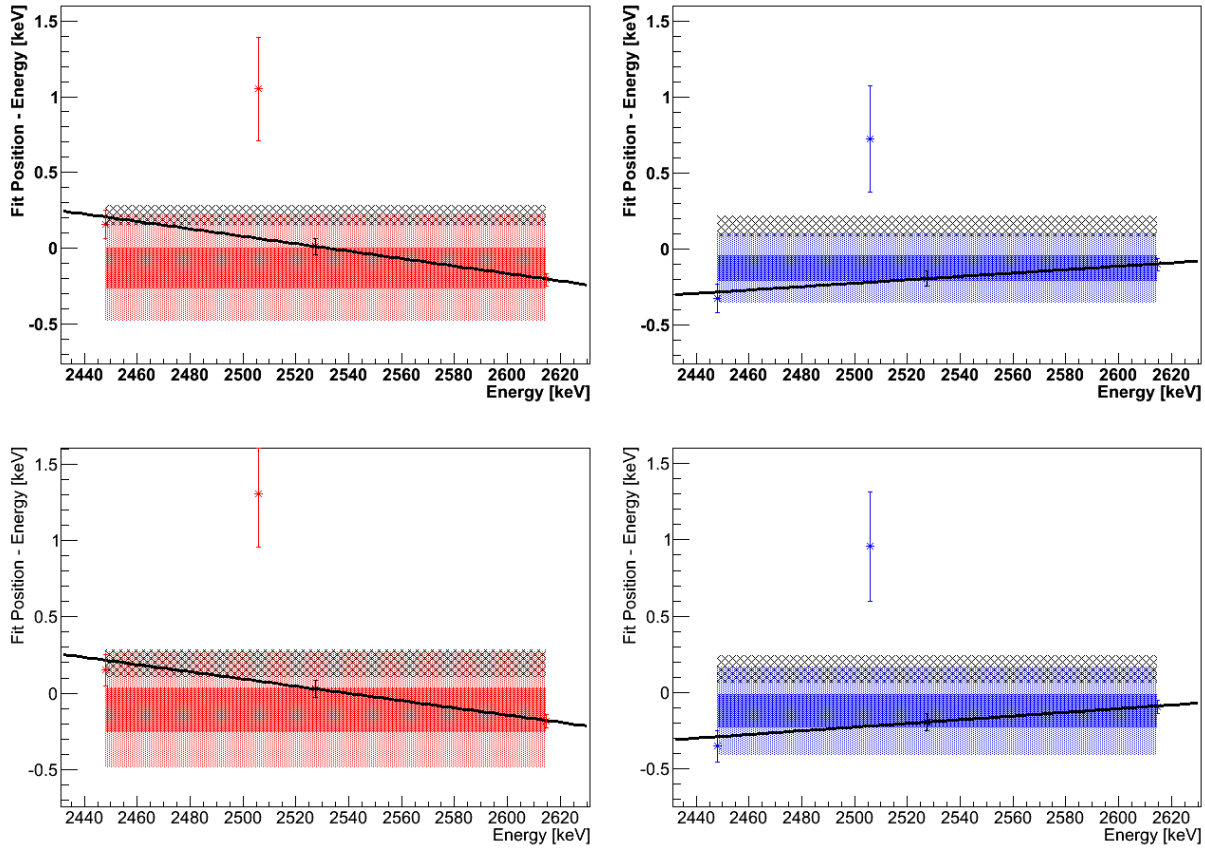


Figure 6.23: Graphical summary of the calibration uncertainty estimation for the final Cuoricino result. Residuals (colored points), weighted mean of residuals (dark colored band), weighted standard deviation of residuals centered on weighted mean (light colored band), weighted mean of absolute value of residuals (black hashed band), linear best fit to residuals (black line), and linear best fit to residuals evaluated at double-beta Q-value (black point) are shown. Red indicates the second-order log polynomial linearization; blue indicates the third-order polynomial linearization. Top: Berkeley-type fits. Bottom: Milano-type fits.

In order to investigate the question of the background shape, each peak was fitted in its own 80-keV window. This was done three times, each time fitting the background with a polynomial of order 0, 1, or 2, respectively, and the resulting residuals were compared with those obtained from the fit over a single 220-keV window. There was no significant effect.

To address the asymmetry of the peaks, the sum spectrum of all the data was fit, using a sum of two Crystal Ball functions with the same mean to fit each peak. The Crystal Ball peak shape consists of a Gaussian with a power-law low-energy tail; the shape parameters of the Crystal Ball are the Gaussian sigma (σ), the exponent defining the decay of the low-energy tail (N), and the threshold at which the Gaussian transitions into the power law (α, expressed as the number of σ below the mean). The fit was again performed over a single 220-keV window containing all three peaks. Although the shape parameters of the Crystal Ball functions were left free, they were constrained to be the same for each peak. Figure 6.22 compares this fit to the Berkeley-type fit and the Milano-type fit for each peak. Fitting the peaks with a Crystal Ball-based function instead of a Gaussian-based function shifted the fitted position of each peak approximately $0.1-0.2$ keV upward in energy (except for the case of the Milano-type fits to the ^{60}Co, which give residuals that are about the same as those from the Crystal Ball fits); however, the main effect of this shift was only to bring the weighted mean of the residuals closer to zero. Ultimately, it appears that the double-beta-analysis-based approach yields the more conservative results[6].

In the end, the analysis outlined above for either the Berkeley-type or Milano-type fit, and for either calibration function, confirms an energy uncertainty arising from the calibration of about ± 0.4 keV, as has been quoted for CUORICINO in the past [30]. The third-order polynomial, however, can be seen to be superior, both from the lack of a significant trend in the residuals over the gamma region (see Section 6.3.2) and from the somewhat smaller ^{60}Co residual compared to the second-order log polynomial.

[6]The exception to this is that in the case of taking the ^{60}Co residual as the calibration uncertainty, it would be more conservative to use the Crystal Ball result than the Berkeley-type result, as the ^{60}Co residual is positive in any case and the Crystal Ball fit only makes it more so. That said, the reliability of the Crystal Ball fit for the low-statistics ^{60}Co peak is arguably questionable, anyway, as the tail parameters are not well constrained.

In the final fit performed for the Cuoricino result [21], a Gaussian fluctuation of a size consistent with the more conservative calibration uncertainty estimate of ± 0.8 keV, obtained from the ^{60}Co sum peak residual, was allowed in the position of the $0\nu\beta\beta$ peak. In fact, the spectrum in the $0\nu\beta\beta$ region of interest was so flat that the difference between a ± 0.4-keV energy uncertainty and a ± 0.8-keV energy uncertainty had a negligible effect on the final analysis; because Cuoricino only set a limit on the $0\nu\beta\beta$ half-life instead of observing evidence of a peak, the more conservative uncertainty was considered more representative of the information that could ultimately be extracted from the experiment.

6.3.1.2 Comparing Uncertainties of the Two Calibration Functions on Data Allowing the Most Direct Comparison of the Calibration Functions Themselves

Although the study detailed above yields accurate results for the calibration uncertainty on the final published Cuoricino result, those results do not represent a true comparison of the performances of the two calibration functions. This is due to the fact the calibration module did not always find and use the same calibration peaks when performing the pol3 calibration as it did when performing the logpol2 calibration on the same channel and dataset (see Section 6.2).

Here, the calibration uncertainty study is repeated on a subset of data chosen to allow the most direct possible comparison of the performances of the two calibration functions. We include only channels/datasets for which all peaks were identified the same for both the logpol2 processing and the pol3 processing; this means that either the peak was not found by either processing, or it was found by both processings and its fitted location in one processing differed from its fitted location in the other processing by less than the smaller of the two fitted errors. We also remove all channels/datasets for which the pol3 calibration failed or fewer than 5 peaks were found.

The Milano-type fit is quite resource-intensive and its results are similar to those of the Berkeley-type fit, which corresponds to the analysis ultimately chosen for the publication of the final Cuoricino results, so the Milano-type fit was not performed for the calibration function comparison. However, we introduce a new fitting method that allows us to use a Crystal Ball peak shape in a

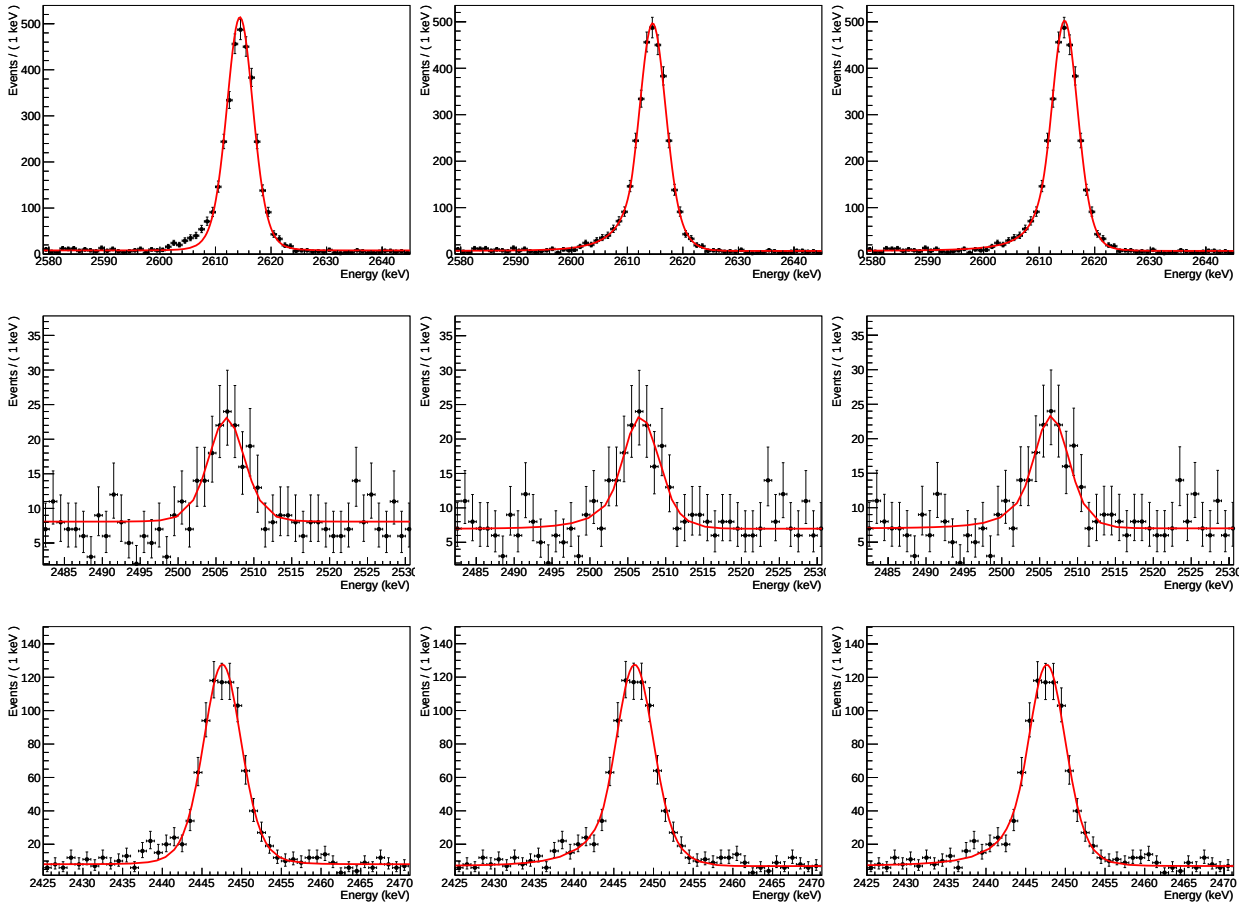


Figure 6.24: Left to right: Berkeley-type, sum-of-two-Crystal-Balls, and simultaneous-Crystal-Ball fits. Top to bottom: ^{208}Tl, ^{60}Co, and ^{214}Bi peaks. The data are selected for the most direct possible comparison of the two calibration functions and are calibrated with the third-order polynomial function.

Table 6.5: Results of residual analysis near the ^{130}Te double-beta decay region of interest for the comparison of the calibration uncertainties generated by the two calibration functions.

Fit type	Functional form	Weighted mean	Weighted standard deviation	Weighted mean of abs. values	Linear fit eval. at Q-value	Estimated calibration uncertainty
Berkeley	logpol2	-0.15 ± 0.15[7]	0.38	0.24 ± 0.06	0.02 ± 0.06	0.5
Berkeley	pol3	-0.15 ± 0.07	0.19	0.16 ± 0.06	-0.19 ± 0.06	0.3
2 CB	logpol2	0.01 ± 0.16	0.38	0.15 ± 0.12	0.15 ± 0.07	0.4
2 CB	pol3	-0.02 ± 0.11	0.27	0.08 ± 0.10	-0.08 ± 0.07	0.3
simul. CB	logpol2	-0.02 ± 0.14	0.34	0.15 ± 0.09	0.12 ± 0.06	0.3
simul. CB	pol3	0.01 ± 0.09	0.22	0.08 ± 0.07	-0.06 ± 0.06	0.2

simultaneous fit to all channels and datasets. Leaving the Crystal Ball shape parameters free for all of the data subsets would introduce far too many free parameters into the fit; however, it is possible to perform a successful fit if the Gaussian sigma on each channel is fixed to the resolution obtained from calibration data, as done for the Berkeley-type fit, and the tail shape parameters are constrained to be the same for all large crystals and for all small or enriched crystals.

The fits to the peaks on the data calibrated with the third-order polynomial using the Berkeley-type fit, the sum-of-two-Crystal-Balls fit with the peak shape parameters left free, and the simultaneous-Crystal-Ball fit with fixed Gaussian sigmas are shown in Figure 6.24. The results of the residual analysis are summarized in Table 6.5 and presented graphically in Figure 6.25.

As before, the main difference between the Berkeley-type fit and the sum-of-two-Crystal-Balls fit is that the Crystal-Ball-based fit pushes all residuals more positive, bringing the ^{208}Tl and ^{214}Bi residuals more consistent with zero but making the ^{60}Co residual worse.

[7] Although the weighted mean is just barely consistent with zero at two significant digits, when further decimal places are taken into account, the fit actually returns a value for the weighted mean that is slightly larger than its error.

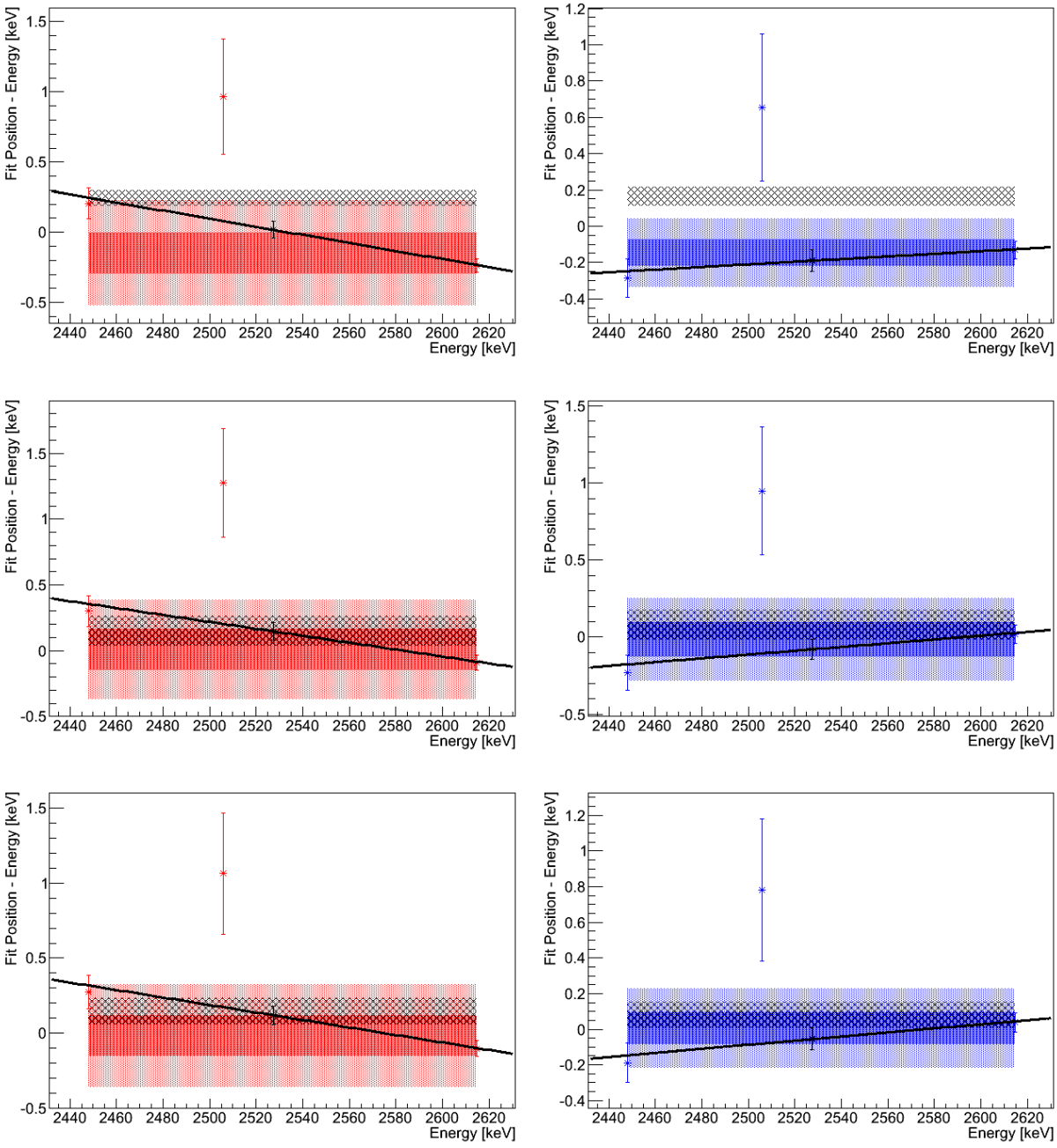


Figure 6.25: Graphical summary of the calibration uncertainty estimation for the comparison of the two calibration functions. Plot formatting is the same as in Figure 6.23. Top: Berkeley-type fits. Middle: Sum-of-two-Crystal-Balls fits. Bottom: Simultaneous-Crystal-Ball fits.

Interestingly, however, when compared to the sum-of-two-Crystal-Balls results, the simultaneous-Crystal-Ball fit brings the ^{60}Co and ^{214}Bi residuals closer to zero — bringing them up if they are low or down if they are high. The ^{60}Co residuals are shifted by about 0.2 keV, and the ^{214}Bi residuals are shifted by roughly $0.03 - 0.04$ keV. In contrast, the ^{208}Tl residuals are shifted farther from zero by the simultaneous-Crystal-Ball fit but by less than 0.02 keV, so the ultimate outcome is that the simultaneous-Crystal-Ball fit yields the smallest calibration uncertainties of the three fit methods. As there is no clear *a priori* explanation for this behavior, it seems reasonable to suppose that this may be a real effect of the calibration being good and the simultaneous-Crystal-Ball fit method being a better description of the peak shapes than the other two fit methods. Although the calibration module uses Gaussian peak shapes instead of Crystal Ball peak shapes, the statistics of the data used for calibration are also much lesser than the statistics of the summed background data of all channels over all datasets; the Crystal Ball tail may have a negligible effect on the fits to lower-statistics calibration data but a non-negligible one when fitting the higher-statistics sum spectrum.

Regardless of the fit method, the calibration uncertainty evaluated for data processed with the third-order polynomial calibration function is always smaller than the calibration uncertainty evaluated for the same data processed with the second-order log polynomial calibration function, indicating that the performance of the pol3 is superior.

6.3.1.3 Comparing Uncertainties of the Two Calibration Functions on Data Most Comparable to CUORE

CUORE will contain only $5 \times 5 \times 5$ cm^3 crystals, so an estimate of the expected calibration uncertainty in CUORE can be obtained by applying a further cut to include only data collected from the large crystals in Cuoricino. The simultaneous-Crystal-Ball fits to the R.O.I. peaks are shown in Figure 6.26, and the results of the simultaneous-Crystal-Ball residual analysis are shown in Table 6.6 and Figure 6.27. By far the majority of Cuoricino data is composed of data taken with large crystals, so the results are very similar to those in the preceding section.

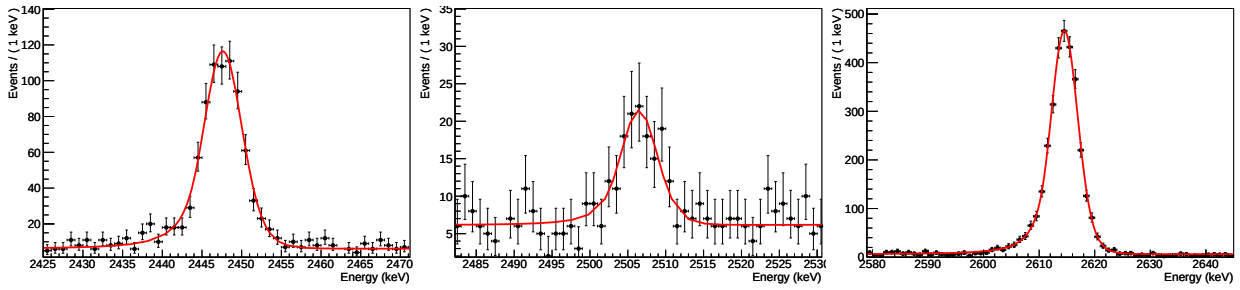


Figure 6.26: Simultaneous-Crystal-Ball fits to the R.O.I. peaks on data collected in Cuoricino by large crystals only. Left to right: ^{214}Bi, ^{60}Co, and ^{208}Tl peaks. The data are selected for the most direct possible comparison of the two calibration functions and are calibrated with the third-order polynomial function.

Table 6.6: Results of residual analysis near the ^{130}Te double-beta decay region of interest for the estimation of expected calibration uncertainty in CUORE.

Fit type	Functional form	Weighted mean	Weighted standard deviation	Weighted mean of abs. values	Linear fit eval. at Q-value	Estimated calibration uncertainty
simul. CB	logpol2	-0.01 ± 0.12	0.31	0.13 ± 0.09	0.10 ± 0.06	0.3
simul. CB	pol3	-0.03 ± 0.09	0.23	0.06 ± 0.09	-0.10 ± 0.06	0.2

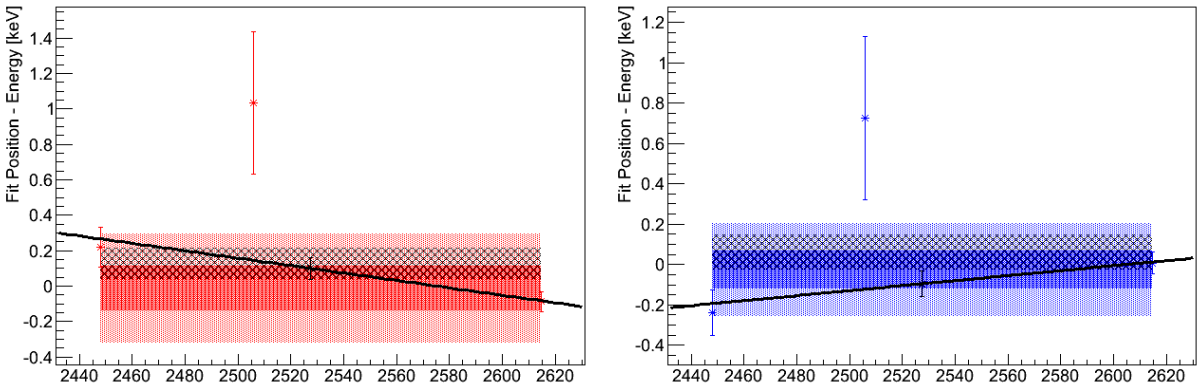


Figure 6.27: Graphical summary of the calibration uncertainty estimation for the expected calibration uncertainty in CUORE. Fits were performed with the simultaneous-Crystal-Ball fit method. Plot formatting is the same as in Figure 6.23 and Figure 6.25.

The standard calibration function for the processing of CUORE will be a third-order polynomial; however, it will differ from the pol3 calibration used here in that the intercept of the function will be fixed to zero. Nevertheless, the effect of fixing the intercept should be minimal at higher energies, such as in the region of interest for ^{130}Te $0\nu\beta\beta$ decay. Assuming similar calibration performances in CUORE as were seen in Cuoricino, then, the calibration uncertainty expected for CUORE is ± 0.2 keV (or ± 0.7 keV, following the more conservative choice of using the ^{60}Co sum peak residual as the calibration uncertainty).

In a way, a calibration uncertainty of ± 0.2 keV is a worst-case expectation for CUORE. If overall calibration performance can be improved — for example, by improving the peak-finding capabilities of the calibration module (see Section 6.2) — a smaller uncertainty may be achieved. If CUORE measures a spectrum that is flat in the region of interest for $0\nu\beta\beta$, as in the case of Cuoricino, the size of the energy uncertainty will not be very critical to the experimental result. However, if a peak structure begins to emerge in CUORE data, a reliable estimate of the energy uncertainty will be necessary to understand whether such an observation is indeed consistent with a $0\nu\beta\beta$ signal and to correctly evaluate the magnitude of that signal.

6.3.2 Residuals and Resolutions Across the Whole Gamma Region

High-resolution, low-background detectors like those of the CUORE-family experiments can provide interesting data for a number of different physics results, not just neutrinoless double-beta decay. For this reason, as well as to confirm our understanding of the detector's behavior, it is worthwhile to investigate peaks observed in Cuoricino across the entire gamma region of the energy spectrum. In the sum spectrum of all Cuoricino crystals, the statistics are sufficient to allow the identification and fitting of a number of peaks from a number of different known background sources, meaning that a residual study of the entire spectrum from a few hundred keV up to the 2614.5-keV peak can be performed in a similar manner to the study performed above for the region near the ^{130}Te $0\nu\beta\beta$ Q-value.

As the resolution of bolometers is somewhat energy dependent, it is not possible to fit peaks across the entire gamma spectrum in the same way as the $0\nu\beta\beta$ R.O.I. peaks can be fit, with a fixed resolution for each channel and dataset. We do not have a strong calibration peak like the 2614.5-keV peak in every energy regime from which such resolutions could be obtained. Thus, it is necessary to leave the resolution free for these peak fits, and it no longer makes sense to perform a simultaneous fit to all the individual channels and datasets. Instead, we will fit the sum spectrum of only those Cuoricino channels and datasets that a) provide the most direct comparison between the two calibration functions and b) are most representative of the expected CUORE experience:

- 'large' crystals only,
- only channels/datasets for which both calibrations identified all the same calibration peaks, and
- only channels/datasets for which both the logpol2 and pol3 calibrations were successful and did not have no free parameters in the calibration fit.

Fitting the peaks in the sum spectrum will yield an effective 'average' estimation of the crystals' resolution at each peak energy. Resolution should vary slowly with energy, so the fits are performed sequentially in order of decreasing energy and the initial value for the resolution parameter of each peak is set from the fit result of the previous one.

We would like to perform unbinned extended maximum likelihood fits on narrow energy windows (preferably containing only one peak) in which it is reasonable to approximate the background as linear. Performing extended fits allows us to use the relative branching ratios between similar-energy peaks from the decay of the same nuclide to restrict the amplitude in the fit of some peaks that are so close to one or more other peaks that they cannot be fit separately. Also, in keeping with the experience gained from fitting the R.O.I. peaks in Section 6.3.1, we wish to use Crystal Ball peak shapes.

The trouble with Crystal Ball fits is that, unless the peak statistics are quite high compared to the background, the fit can often fail if the tail shape parameters (α and N) are left free. To address this issue, instead of performing all of the peak fits strictly sequentially in order of decreasing energy, 8 fairly strong peaks distributed approximately evenly across the energy range were first fit simultaneously, requiring that α and N be the same for all of them. The remaining peaks were then fitted sequentially in order of decreasing energy with α and N set constant and equal to their fitted values as obtained from the strong peaks.

The background peaks fitted in this study are listed in Table 6.7. The 8 strong peaks used to establish the Crystal Ball tail parameters and cases in which multiple peaks were fitted in the same window are indicated.

The CUORE collaboration typically characterizes detector resolution in terms of the full width at half maximum (FWHM) of an observed gamma peak. In the case of a Gaussian peak shape, the FWHM can be straightforwardly calculated from the Gaussian σ:

$$\text{FWHM} = 2\sqrt{2\ln 2}\,\sigma \approx 2.3548200\,\sigma. \tag{6.6}$$

In other words, the height of a Gaussian function has dropped to half of its maximum value at a distance of $\pm\sqrt{2\ln 2}\,\sigma \approx 1.1774100\,\sigma$. The FWHM of a Crystal Ball peak, however, will depend on α, the particular point at which the Gaussian shape gives way to the power-law tail. If $\alpha \geq \sqrt{2\ln 2}$, the Crystal Ball FWHM will be identical to the Gaussian FWHM; otherwise, a more involved calculation is necessary.

Table 6.7: Gamma lines fitted to characterize calibration behavior over the gamma region. Starred peaks are strong peaks that are fit first and used to determine the Crystal Ball tail parameters used to fit the remaining peaks. Fit parameters (σ, mean, and/or peak yield) of daggered peaks are restricted with respect to the fitted parameters of stronger nearby peaks to ensure a sensible fit. Peak energies are from [68].

Label	Energy [keV]	Source	Label	Energy [keV]	Source
2615 *	2614.533	^{208}Tl	1064	1063.662	^{207}Bi
2506[8]	2505.738	^{60}Co	965†/969	968.971	^{228}Ac
2448	2447.86	^{214}Bi		965.766	^{228}Ac
2204 *	2204.21	^{214}Bi	934	934.061	^{214}Bi
2103/2119	2118.55	^{214}Bi	911 *	911.204	^{228}Ac
	2103.533	^{208}Tl 2615 sing. esc.	835/836†	835.710	^{228}Ac
1838†/1847	1847.420	^{214}Bi		834.848	^{54}Mn
	1838.36	^{214}Bi	768	768.356	^{214}Bi
1764 *	1764.494	^{214}Bi	601/609 [9] *	609.312	^{214}Bi
1730	1729.595	^{214}Bi		600.600	^{125}Sb
1661	1661.28	^{214}Bi	583	583.191	^{208}Tl
1509	1509.228	^{214}Bi	570	569.702	^{207}Bi
1461 *	1460.830	^{40}K	511†/511	511.0	annihilation γ
1402†/1408	1407.98	^{214}Bi		510.77	^{208}Tl
	1401.50	^{214}Bi	338/352	351.932	^{214}Pb
1333	1332.501	^{60}Co		338.320	^{228}Ac
1238	1238.110	^{214}Bi	295	295.224	^{214}Pb
1173 *	1173.237	^{60}Co	239/242† *	241.997	^{214}Pb
1120	1120.287	^{214}Bi		238.632	^{212}Pb

Table 6.8: Crystal Ball tail shape parameters obtained from initial fit to 8 'strong' peaks for data selected to be comparable to CUORE and to provide a direct comparison between the two calibration functions. The remaining peaks were fit with tail shape parameters fixed to these values.

Calibration function	α	N
logpol2	1.3277 ± 0.0371	6.30 ± 1.13
pol3	1.3702 ± 0.0488	6.07 ± 1.41

The results of the initial fit to the 'strong' peaks to determine α and N are summarized in Table 6.8. Because $\alpha > \sqrt{2 \ln 2}$, the FWHM resolutions can be directly calculated from the fitted σ of each peak in the same manner as for a Gaussian peak shape. The residuals of all fitted peaks in the gamma region are shown in Figure 6.28, while the FWHM resolutions are plotted in Figure 6.29 both in units of keV and as a percentage of the energy at which each is measured.

It is plain from Figure 6.28 that the superior performance of the pol3 calibration function holds throughout the gamma region of the energy spectrum. The poor behavior of the logpol2 function at low energies is expected; the functional form of the logpol2 cannot cross zero, so this calibration will always push events with energies below the lowest-energy calibration peak away from where they belong. The more telling behavior of the logpol2 function lies in the wide energy region between 969 and 2615 keV containing no primary calibration peaks. In this region, the fitted energy of almost every peak is higher than its nominal energy by up to around 1 keV [10]. While the pol3 function also demonstrates a bias (of the opposite sign) in this region, the resulting residuals have a magnitude of only up to about 0.5 keV.

[8] The ^{60}Co sum peak introduced in Section 6.3.1.

[9] There is also a tiny gamma line from ^{125}Sb at 606.718 keV present in this energy window, included in the fit as a Crystal Ball peak with a mean constrained to be at the expected location between the fitted means of the 609.312-keV and 600.600-keV peaks and a Gaussian σ constrained to be the mean of the two fitted σs.

[10] Three peaks have larger residuals. However, the 1838 line is a weak peak sitting on the tail of a considerably stronger one, and the sources of the ^{60}Co sum peak and the single-escape peak from the 2615 line are not single gammas, so these latter may have a slightly different peak shape from the majority of the background lines.

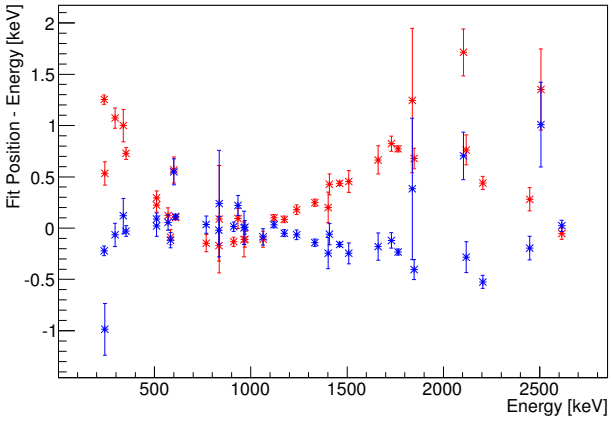


Figure 6.28: Residuals vs. energy for all background peaks listed in Table 6.8 in Cuoricino data selected to be comparable to CUORE and to provide a direct comparison between the two calibration functions used for Cuoricino processing. Red points are obtained from data calibrated with the second-order log polynomial calibration function; blue points are obtained from data calibrated with the third-order polynomial calibration function.

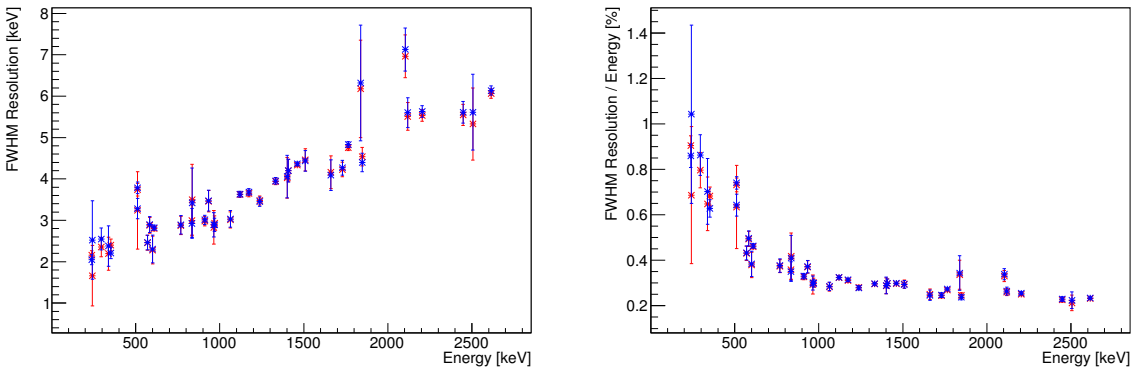


Figure 6.29: FWHM resolutions vs. energy for all background peaks listed in Table 6.8 in Cuoricino data selected to be comparable to CUORE and to provide a direct comparison between the two calibration functions used for Cuoricino processing. Red points are obtained from data calibrated with the second-order log polynomial calibration function; blue points are obtained from data calibrated with the third-order polynomial calibration function. Left: in units of energy. Right: in units of percent.

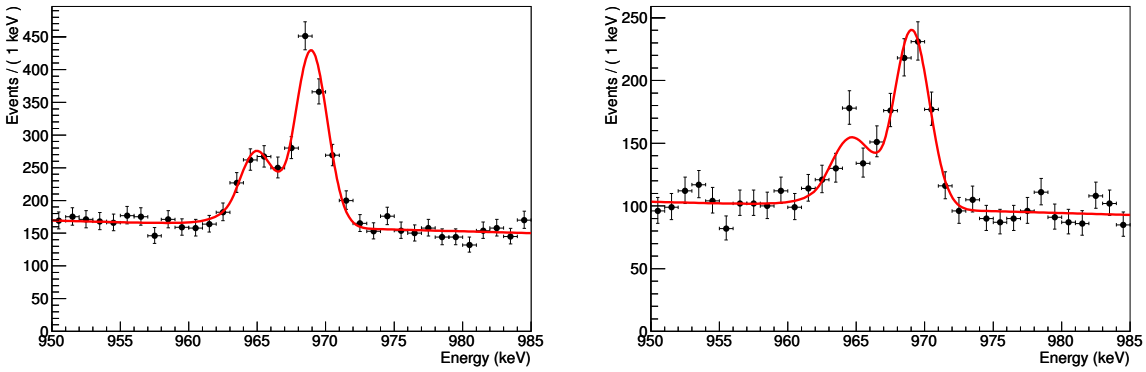


Figure 6.30: 969 calibration peak in Cuoricino sum spectrum of data selected to be comparable to CUORE and to provide a direct comparison between the two calibration functions used for Cuoricino processing. Left: only data calibrated with both initial and final calibration data. Right: only data calibrated with only initial or only final calibration data. Both peaks are from ^{228}Ac; the fit is obtained by restricting the fitted number of counts in the weaker peak to be within approximately 30% of the number expected by scaling from the number of counts in the stronger peak.

On the other hand, there is no significant difference in the fitted peak resolutions between the two calibration functions. This indicates that neither calibration function causes a significant deformation of peak shapes with respect to the other, a conclusion that is supported by the similarity between the fitted Crystal Ball tail shape parameters presented in Table 6.8. The resolutions follow a roughly linear trend with energy, especially at energies $\gtrsim 700$ keV; the only significant outlier is the single-escape peak from the 2615 line, which has a larger resolution than expected.

A factor that *can* lead to some deformation of peak shapes is whether the data was calibrated with both initial and final calibration data vs. only initial or only final calibration data. Although the baseline calibration approach calls for both initial and final calibration data to be taken for each dataset, various factors such as operational difficulties meant that approximately 1/3 of Cuoricino data was calibrated with only initial or only final calibration data. The difference in the shape of the 969 calibration peak between the Cuoricino sum spectrum of data calibrated with both initial and final calibration data and the sum spectrum of data calibrated with only initial or only final calibration data is illustrated in Figure 6.30.

Table 6.9: Fitted Crystal Ball tail shape parameters obtained from initial fit to 8 'strong' peaks for Cuoricino data calibrated with both initial and final calibration data and selected to be comparable to CUORE and to provide a direct comparison between the two calibration functions used for Cuoricino processing. Remaining peaks are fit with fixed tail shape parameters set to these values.

Calibration function	α	N
logpol2	1.2588 ± 0.0508	8.56 ± 2.67
pol3	1.2894 ± 0.0545	8.37 ± 2.69

It may be worthwhile, then, to consider the shape of the background residuals in 'ideal' calibration conditions, namely, to add a cut selecting only data calibrated with both initial and final calibration data. As it happens, this cut also removes the high-radon detector operation periods identified in [107] almost entirely; it therefore seems sensible to extend the definition of 'ideal' calibration conditions to also remove the datasets that contain the remaining high-radon data. As a result, the statistics in the peaks corresponding to gammas from ^{214}Bi, which is part of the decay chain fed by ^{222}Rn, are considerably reduced compared to other peaks in the spectrum. The results of the initial fit to the 'strong' peaks to determine α and N are summarized in Table 6.9, from which it can be seen that, again, $\alpha > \sqrt{2 \ln 2}$ and the FWHM resolutions can be found as for a Gaussian peak. The residuals in the gamma region for the data calibrated with both calibration functions is shown in Figure 6.31, while the FWHM resolutions are plotted in Figure 6.32.

With this additional cut, the residual trend of the pol3 calibration function is no longer clearly superior to that of the logpol2, except at low energy where this behavior is expected *a priori*. The magnitudes of the residuals in the $1000 - 2500$ keV energy region in Figure 6.31 are roughly comparable; the main difference is simply whether the bias is positive or negative. It is tempting to conclude that the third-order polynomial is more robust to 'non-ideal' calibration conditions,

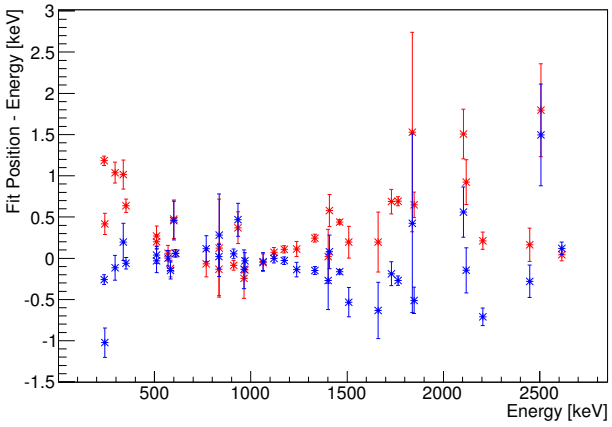


Figure 6.31: Residuals vs. energy for all background peaks listed in Tab. 6.8 in Cuoricino data calibrated with both initial and final calibration data and selected to be comparable to CUORE and to provide a direct comparison between the two calibration functions used for Cuoricino processing. Red points are obtained from data calibrated with the second-order log polynomial calibration function; blue points are obtained from data calibrated with the third-order polynomial calibration function.

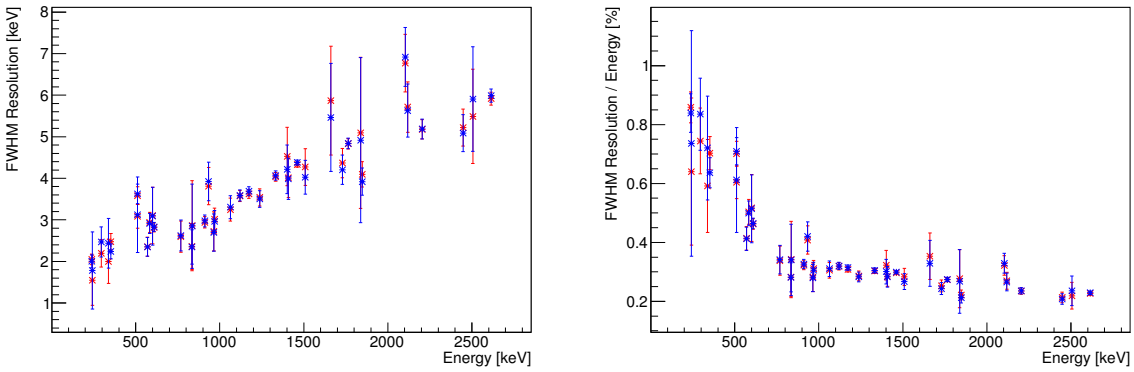


Figure 6.32: Full-width-at-half-maximum resolutions vs. energy for all background peaks listed in Tab. 6.8 in Cuoricino data calibrated with both initial and final calibration data and selected to be comparable to CUORE and to provide a direct comparison between the two calibration functions used for Cuoricino processing. Red points are obtained from data calibrated with the second-order log polynomial calibration function; blue points are obtained from data calibrated with the third-order polynomial calibration function. Left: in units of energy. Right: in units of percent.

but this interpretation seems unlikely, as the pol3 residuals worsen between Figure 6.28 and Figure 6.31 more than the logpol2 residuals improve.

However, it should be noted that all but three of the fitted peaks between 1200 and 2500 keV are from ^{214}Bi, as are three of the eight 'strong' peaks used to establish the Crystal Ball tail shape parameters for the peak fits. The considerable loss of statistics in these peaks is reflected in the larger error bars in Figure 6.31 compared to Figure 6.28 as well as in the larger errors on the values in Table 6.9 compared to Table 6.8. The removal of the high-radon data simply washes out much of the distinction that can be seen between the two calibration functions.

In any case, all comparisons of calibration uncertainty in data calibrated with the second-order log polynomial calibration function to calibration uncertainty in data calibrated with the third-order polynomial calibration function indicate that the pol3 is either comparable or superior to the logpol2 in the gamma energy region and specifically in the region of interest for $0\nu\beta\beta$ in ^{130}Te. The pol3, especially if its intercept is fixed to zero, can also be expected *a priori* to provide better behavior for any studies concerned with low-energy data. This serves as validation of the CUORE collaboration's decision to proceed to CUORE-0 and CUORE using the pol3 with intercept fixed to zero as the default calibration function. It will, of course, remain possible to calibrate with the logpol2 calibration function for any studies concerned with the alpha region.

The resolutions in Figure 6.32 remain comparable between the two calibration functions, and the single-escape peak from the 2615 line is no longer such a significant outlier. In fact, the same is true of its residual; the residual of the 2104 peak is quite a bit higher than those of most neighboring peaks in Figure 6.28 but much less so in Figure 6.31. Nevertheless, its residual and resolution both remain somewhat high. This fact is of particular interest because this peak is also a calibration peak. It has been proposed that the shape of the ^{60}Co sum peak at 2506 keV may be somewhat different from the single-gamma peaks, causing its energy to reconstruct unexpectedly high; a similar mechanism may be at work in the 2615 single-escape peak, which could ultimately distort the calibration function in that energy region to force nearby peaks to too-low energies, as we see in the Cuoricino pol3 residuals. In Cuoricino, it was rare for a well-calibrated channel/dataset to not find, or not use, the 2104 calibration peak, making it difficult to draw any meaningful conclusions

about the influence of the 2104 calibration peak by looking only at such data. It could be interesting to perform a dedicated set of calibrations that do not attempt to use the single-escape peak as a calibration peak, to see if this influence is indeed responsible for pushing the pol3 residuals away from zero.

6.4 Changes to Calibration Analysis Procedure Due to CUORE Geometry

All estimations of, and studies of possible methods to improve, the expected calibration performance in CUORE up to this point have been performed using the considerable amount of calibration data collected by Cuoricino. By necessity, then, the selection and fitting of calibration peaks was appropriate to the data collected by the Cuoricino detectors. This section will argue, however, that the behavior of the calibration module will need to be somewhat altered to reflect the calibration spectra that are expected to be collected by CUORE.

When the calibration module searches for calibration peaks, it does so in two phases: it first searches for the peaks that have been designated as 'primary' peaks, which are strong peaks that are expected to be found in every calibration, and then interpolates from those primary peaks to try to locate the weaker 'secondary' peaks. The relative strengths of calibration peaks are partially attributable to the branching ratios of the decay modes of the nuclides that give rise to the gammas that form the peaks, but the detector geometry also contributes in ways that can be evaluated only through simulations. For example, the strong 2614.5-keV peak, the primary peak that serves to anchor the calibration in the $0\nu\beta\beta$ region of interest, comes from a gamma produced in 99% of ^{208}Tl decays; the 583.191-keV gamma produced in 84.5% of decays of the same nuclide [68], however, resulted in a far weaker peak that was treated as a secondary peak in Cuoricino. This is because lower-energy gammas from the Cuoricino calibration sources were significantly attenuated by the inner lead shielding of the cryostat before reaching the detectors. In CUORE, however, the calibration sources will be inside the lead shields, and six of the twelve sources will be in between the detector towers themselves. Due to this closer placement of the sources to the detectors, a much greater percentage of the lower-energy gammas are expected to reach the detectors in CUORE than did so in Cuoricino.

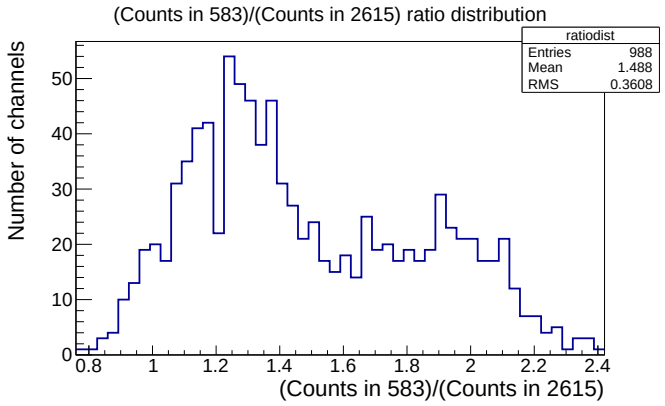


Figure 6.33: Simulated distribution of the ratio of the number of counts in the 583 peak to the number of counts in the 2615 peak in the CUORE array. The shape of the distribution reflects the two broad populations of detectors observed previously: 'hotter' (crystals adjacent to internal calibration sources; higher 583-to-2615 ratio) and 'colder' (crystals non-adjacent to internal calibration sources; lower 583-to-2615 ratio).

The effects of this change in detector geometry were reflected in the results of the simulation introduced in Section 5.5, and they suggest specific changes to the operation of the calibration module for the CUORE case that will be discussed in this section.

As can be seen from Figure 6.33, the simulation shows that the 583 peak is actually somewhat stronger than the 2615 peak on the majority of CUORE detectors. In the CUORE geometry, the shorter stopping distance of lower-energy gammas affects the relative strengths of the 583 and 2615 peaks in the opposite manner as it did in Cuoricino; the lower-energy gammas are now more likely to stop in the detector, while the higher-energy gammas are more likely to escape the detector region entirely. From these results, there seems to be no reason that the 583 peak could not be classified as a primary peak for CUORE calibration. Once CUORE begins taking data, it may be valuable to perform a study comparing calibration performances treating the 583 peak as a primary peak to calibration performances treating it as a secondary peak.

A related phenomenon is the composition of the 511 peak. There are two contributions to this peak: the 510.77-keV gamma from the decay of ^{208}Tl, which necessarily originates from the calibration sources, and the 511.0-keV annihilation gammas produced when an electron and a

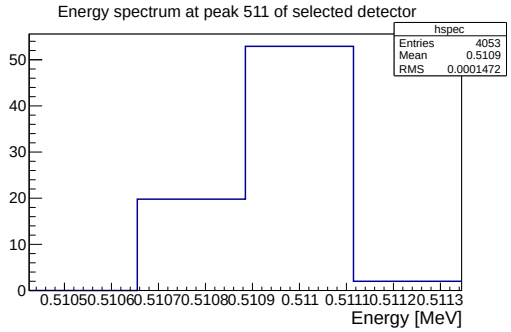


Figure 6.34: Illustration of the binning scheme used to estimate the ^{208}Tl contribution to the 511 peak from the CUORE simulation. The number of events attributed to ^{208}Tl is estimated by subtracting the leftmost bin from the second-left bin; the number of events attributed to the annihilation gamma is estimated by subtracting the rightmost bin from the second-right bin. The spectrum pictured is from Tower 12, Layer 3, Crystal 1, which is the detector with the smallest ^{208}Tl contribution to the 511 peak; the counts per bin are in arbitrary units.

positron annihilate, which may occur at some distance from the calibration sources. The energies of the two are separated by only 0.23 keV, which the bolometers are unable to resolve, so the calibration module fits the 511 peak with a single Gaussian. In the Cuoricino cryostat, simulations indicated that the ^{208}Tl gamma contributed only about 6% of the total intensity of the peak [107], so the calibration fit assigned an energy of 511.0 keV to the peak for the Cuoricino processing. In the CUORE cryostat, however, the likelihood of the 510.77-keV gammas interacting in the detectors should be enhanced with respect to Cuoricino just as it is for the 583.191-keV gammas.

To evaluate the relative composition of the 511 peak, it is necessary to use finer binning to resolve the 0.23-keV energy difference than the 1-keV-wide bins that have been used to analyze the simulations up to this point. Two adjacent 0.23-keV-wide bins, centered on the ^{208}Tl gamma energy and the annihilation gamma energy respectively, fully encompass the counts in the peak; the number of ^{208}Tl events is estimated by taking the number of counts in the bin centered on 510.77 keV and subtracting the number of counts in its left-by bin, while the number of annihilation-gamma events is estimated by taking the number of counts in the bin centered on 511.0 keV and subtracting the number of counts in its right-by bin. This binning scheme is illustrated in Figure 6.34.

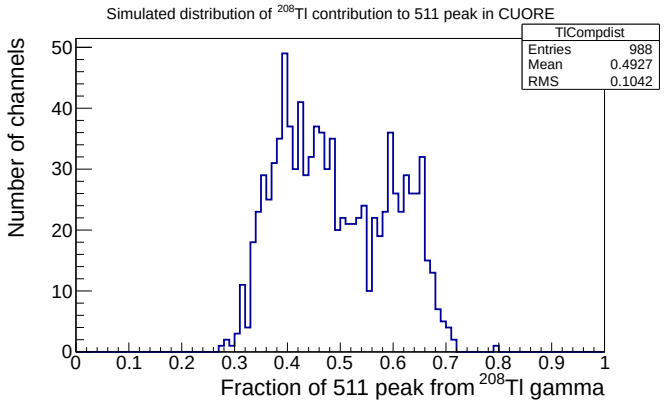


Figure 6.35: Simulated distribution of the approximate ^{208}Tl contribution to the 511 peak across all CUORE detectors. It is significantly enhanced compared to Cuoricino, in which the ^{208}Tl contribution to the 511 peak was only about 6%.

The fraction of the peak attributable to the ^{208}Tl gamma is histogrammed for all 988 CUORE detectors in Figure 6.35. The ^{208}Tl gamma can be seen to contribute between 30% and 70% of the 511 peak. To first approximation, then, it may be appropriate to assign the mean energy of the two source gammas to the calibration peak for CUORE processing.

Evidence of two broad populations of detectors can once again be seen, however, suggesting that in principle the energy that should be assigned to the 511 peak may differ from one crystal to another, depending on their locations relative to the calibration sources. There is currently no framework in the calibration module for treating different detectors differently in this manner. It would be advisable to apply a realistic detector resolution to the simulated spectra and fit the 511 peak with a Gaussian plus linear background, as the calibration module does, to determine whether the possible detector-to-detector differences in the relative composition of the 511 peak will be expected to cause enough of a shift in the fitted mean of the peak between detectors that such a mechanism will be necessary.

Even if the effective position of the 511 peak does drift from detector to detector, in the end, it may be expected that the promotion of the 583 peak to primary status should help to compensate any resulting uncertainty in the calibration function in that energy region. Additionally, the

mere fact of having one more primary peak should improve the calibration module's ability to successfully obtain a calibration function fit. The calibration uncertainty achieved by CUORE may therefore be even better than indicated by the Cuoricino-based estimations of Section 6.3.1.3.

Chapter 7

Experimental Sensitivity to $0\nu\beta\beta$

7.1 The Meaning of Sensitivity

In essence, 'experimental sensitivity' is a metric that distills the capabilities of an experiment into a single value. This can allow for meaningful comparisons among highly disparate experiments, as long as the sensitivity parameter is constructed in a consistent way for every experiment considered. The sensitivity is typically expressed in terms of the physics quantities the experiment aims to explore in order to convey the scale of its experimental reach. In this chapter, two possible approaches for deriving such a sensitivity that is appropriate for an experiment like CUORE will be discussed.

In Section 7.1.1, we will develop a simplified, formula-based calculation that uses several basic experimental parameters (e.g., resolution, background rate, mass) to express sensitivity in terms of expected background fluctuations. We will refer to this calculation as the "background-fluctuation" sensitivity. The background-fluctuation sensitivity cannot be extended to the ideal zero-background case, so we develop an analytical expression for zero-background sensitivity in Section 7.1.2.

In Section 7.1.3, we will discuss a procedure to express sensitivity as the average limit that a particular experiment can expect to set in the case that the true $0\nu\beta\beta$ rate is zero, by applying the experiment's analysis tools to a suite of Monte Carlo trials. We will refer to this calculation as the "average-limit" sensitivity.

As we will show, the results of the two approaches are compatible, although the philosophies of their construction differ. Section 7.1.4 offers a discussion of the philosophies and interpretations of the two sensitivity methods.

Finally, Section 7.1.5 extends the sensitivity framework to the calculation of the discovery potential, another way to express the physics reach of an experiment.

7.1.1 Sensitivity with Respect to Background Fluctuation

Double-beta decay is a second-order weak process, so half-lives are typically long: two-neutrino double-beta decay half-lives are at least of order 10^{18} years, while current limits on $0\nu\beta\beta$ half-lives are on the order of 10^{24} years or greater. With such long half-lives, the radioactive decay law can be approximated as

$$N(t) \simeq N_0 \left(1 - \ln(2) \cdot \frac{t}{T_{1/2}}\right), \tag{7.1}$$

where $T_{1/2}$ is the half-life, N_0 is the initial number of atoms and $N(t)$ is the number of atoms left after time t has passed.

The mean value S_0 of the $0\nu\beta\beta$ signal, i.e., the expected number of $0\nu\beta\beta$ decays observed during the live time t is

$$S_0 = \frac{M \cdot N_A \cdot a \cdot \eta}{W} \cdot \ln(2) \cdot \frac{t}{T^{0\nu}_{1/2}} \cdot \varepsilon, \tag{7.2}$$

where M is the total active mass, η is the stoichiometric coefficient of the $0\nu\beta\beta$ candidate (i.e., the number of nuclei of the candidate $0\nu\beta\beta$ element per molecule of the active mass), W is the molecular weight of the active mass, N_A is the Avogadro constant, a is the isotopic abundance of the candidate $0\nu\beta\beta$ nuclide, and ε is the physical detector efficiency.

In Equation (7.2), $T^{0\nu}_{1/2}$ refers to the (unknown) *true value* of the $0\nu\beta\beta$ half-life, and S_0 is therefore also unknown. The background-fluctuation sensitivity formulates the sensitivity in reference to the magnitude of the observed-count fluctuations due to background expected in an experiment. In our derivation, we first determine the sensitivity in terms of a number of counts (analogous to

S_0) and then use the form of Equation (7.2) to convert to a half-life sensitivity (analogous to $T^{0\nu}_{1/2}$). In order to prevent confusion between sensitivities and true values, hatted quantities (e.g., $\widehat{T^{0\nu}_{1/2}}$, $\widehat{S_0}$) will be used to represent the sensitivities corresponding to the unhatted true values.

An experiment can expect to see a background contribution to the counts acquired in the energy window of interest for the $0\nu\beta\beta$ signal. For any experiment in which the source is embedded in the detector (common though not universal for $0\nu\beta\beta$ experiments), we can express the mean number of background counts $B(\delta E)$ in an energy window δE as

$$B(\delta E) = b \cdot M \cdot \delta E \cdot t, \tag{7.3}$$

where b is the background rate per unit detector mass per energy interval (units: cts/(keV kg y)).

Usually, b is independently measured by a fit over an energy range much larger than the energy window of interest δE. The background in δE follows a Poisson distribution with a mean value of $B(\delta E)$.

Equation (7.3) assumes that the number of background events scales linearly with the absorber mass of the detector. We will use this simplified model for our background-fluctuation sensitivity calculations. However, other cases, most notably surface contaminations, are in fact possible wherein the background might not scale with M. Therefore, the final analysis of experimental data requires a detailed understanding of the physical distribution of the background sources and Monte Carlo simulations of the specific detector geometry under consideration.

We will use Equations (7.2) and (7.3) as analytic expressions for the expected numbers of signal and background counts assuming an experimental configuration in which the source is embedded in the detector, but an analogous estimation is possible for any detector configuration.

With the background $B(\delta E)$ as defined in Equation (7.3), we can calculate the number of counts that would represent an upward background fluctuation of a chosen significance level. For simplicity, we consider a single-bin counting experiment wherein the width of the bin is equal to the energy window δE; this way we need to consider only a single measured value, sampling a count distribution with mean $B(\delta E)$, and we can decouple the sensitivity calculation from the specific analysis approach used by the experiment.

In this case, the background-fluctuation sensitivity is the smallest mean signal $\widehat{S}_0$ that is *greater than or equal to* a background fluctuation of a chosen significance level. If $B(\delta E)$ is large enough, the background fluctuation will be Gaussian, and the significance level can be expressed in terms of a number of Gaussian standard deviations n_σ. Then $\widehat{S}(\delta E)$ is given by

$$\widehat{S}(\delta E) = \widehat{S}_0 \cdot f(\delta E) = n_\sigma \cdot \sqrt{B(\delta E)}, \tag{7.4}$$

where $\sigma = \sqrt{B(\delta E)}$ and $f(\delta E)$ is the fraction of signal events that fall in the energy window cut δE around the Q-value. The signal fraction $f(\delta E)$ is a simple estimate of the analysis efficiency.

For a signal that is Gaussian-distributed in energy around the Q-value, $f(\delta E)$ is

$$f(\delta E) = \text{erf}\left(\frac{\delta E}{\Delta E} \cdot \sqrt{\ln(2)}\right), \tag{7.5}$$

where ΔE is the detector FWHM energy resolution. The value of δE can be chosen to maximize the $\widehat{S}(\delta E)$-to-$\sqrt{B(\delta E)}$ ratio in the energy window of interest, which in turn optimizes the sensitivity criterion expressed by Equation (7.4); this optimal choice corresponds to $\delta E \approx 1.2\Delta E$. It is, however, common to take $\delta E = \Delta E$. In this case, the sensitivity differs by less than 1% from the one calculated at the optimal cut.

By using the expressions for $\widehat{S}_0$ and $B(\delta E)$ from Equations (7.2) and (7.3), we obtain the Gaussian-regime expression for the background-fluctuation sensitivity of $0\nu\beta\beta$ experiments in the following form:

$$\widehat{T^{0\nu}_{1/2}}(n_\sigma) = \frac{\ln(2)}{n_\sigma} \frac{N_A \cdot a \cdot \eta \cdot \varepsilon}{W} \sqrt{\frac{M \cdot t}{b \cdot \delta E}} \cdot f(\delta E). \tag{7.6}$$

This equation is useful in evaluating the expected performances of prospective experiments, as it analytically links the experimental sensitivity with the detector parameters. Aside from the inclusion of the signal fraction, it is similar to the familiar 'factor of merit' expression used within the $0\nu\beta\beta$ experimental community [129] (see also Chapter 3).

For small numbers of observed events, i.e., extremely low backgrounds, the Gaussian approximation of Equations (7.4) and (7.6) does not provide the correct probability coverage, and therefore

the meaning of the significance level is not preserved. If $B(\delta E)$ is less than ≈ 24 counts, the Gaussian calculation of a 1σ sensitivity will differ from its Poissonian counterpart (developed below) by 10% or more.

Although the Gaussian limit will possibly still be sufficient for CUORE (see Section 7.3), a more careful calculation might be necessary in the case of a lower background or smaller exposure, or for more sensitive experiments in the future. We therefore compute the sensitivity by assuming a Poisson distribution of the background counts.

In terms of Poisson-distributed variables, the concept expressed by Equation (7.4) becomes [95]

$$\sum_{k=\mathcal{C}}^{\infty} p_B\left(k\right) = \alpha, \tag{7.7}$$

where α is the Poisson integrated probability that the background distribution alone will cause a given experiment to observe a total number of counts larger than $\mathcal{C} = \widehat{S}(\delta E) + B(\delta E)$. Equation (7.7) can be solved only for certain values of α because the left-hand side is a discrete sum. To obtain a continuous representation that preserves the Poisson interpretation of Equation (7.7), we exploit the fact that the (discrete) left-hand side of Equation (7.7) coincides with the (continuous) normalized lower incomplete gamma function $P(a, x)$ (see page 260 of [7] for details):

$$P(\widehat{S}(\delta E) + B(\delta E), B(\delta E)) = \alpha. \tag{7.8}$$

The computation of $\widehat{S}_0$ from Equation (7.8), for given values of $B(\delta E)$ and α, can be achieved numerically. Once $\widehat{S}_0$ is computed in this way, the corresponding Poisson-regime background-fluctuation sensitivity to the half-life $T^{0\nu}_{1/2}$ for neutrinoless double-beta decay is simply calculated by reversing Equation (7.2).

Henceforth, we will use the Poisson-regime calculation based on Equation (7.8) to evaluate our background-fluctuation sensitivity. However, to indicate the significance level with the familiar n_σ notation instead of the less-intuitive α, we will label our sensitivities with the n_σ corresponding to a Gaussian upper-tail probability of the α value used for the calculation (for example, we will call a background-fluctuation sensitivity calculated with $\alpha = 0.159$ in Equation (7.8) a '1σ sensitivity').

7.1.2 Analytical Expression for Zero-Background Sensitivity

It is meaningless to define sensitivity in terms of background fluctuations when $B(\delta E) = 0$; therefore, the background-fluctuation sensitivity calculation cannot be extended to the ideal 'zero-background' case. To develop a formula-based, analysis-decoupled zero-background sensitivity calculation, we consider again a single-bin counting experiment in the same way as we did for the background-fluctuation sensitivity; however, we must adopt a new method of constructing our sensitivity parameter.

To construct the zero-background sensitivity, we choose to follow the Bayesian limit-setting procedure. Instead of comparing the mean signal value $S(\delta E)$ to the mean background value $B(\delta E)$, we are now obliged to consider $S_{max}(\delta E)$, the upper limit on $S(\delta E)$ in the case that the experiment observes zero counts (i.e., no background *or* signal) in δE during its live time. $S_{\max}(\delta E)$ can be evaluated using a Bayesian calculation with a flat signal prior (see Equations (36.44) – (36.46) of [48]):

$$\frac{\int_{S=0}^{S_{max}(\delta E)} p_S(0)dS}{\int_{S=0}^{\infty} p_S(0)dS} = \frac{\int_{S=0}^{S_{max}(\delta E)} S^0 e^{-S} dS}{\int_{S=0}^{\infty} S^0 e^{-S} dS} = \frac{\text{C.L.}}{100}, \tag{7.9}$$

where $p_S(k)$ is the Poisson distribution $p_\mu(k)$ with mean $\mu = S$ and the credibility level C.L. is expressed as a percent. Equation (7.9) can be solved analytically for $S_{max}(\delta E)$:

$$S_{max}(\delta E) = S_{max} \cdot f(\delta E) = -\ln(1 - \frac{\text{C.L.}}{100}), \tag{7.10}$$

where S_{max} is the inferred upper limit on S_0. Using S_{max} in place of S_0 in Equation (7.2), we obtain

$$\widehat{T_{1/2}^{0\nu}}(\text{C.L.}) = -\frac{\ln(2)}{\ln(1 - \frac{\text{C.L.}}{100})} \frac{N_A \cdot a \cdot \eta \cdot \varepsilon}{W} M \cdot t \cdot f(\delta E). \tag{7.11}$$

Depending upon the resolution of the experiment, it may be advantageous to consider a wider window than $\delta E = \Delta E$ in the zero-background case, as there is no longer the need to optimize the signal-to-background ratio; the only concern is that the window remain sufficiently narrow that the irreducible background from the $2\nu\beta\beta$ continuum remains negligible if possible.

For practical purposes, this background-free approximation becomes necessary when the background-fluctuation sensitivity in units of counts is of the order of unity or less, $\widehat{S}_0 \lesssim 1$ count, as it is not possible to observe a fraction of an event. By definition, the interpretations of the zero-background sensitivity and the background-fluctuation sensitivity do not entirely coincide.

7.1.3 Sensitivity with Respect to the Average Expected Limit

In the finite-background case, an alternative approach is to use a Monte-Carlo-based procedure to evaluate the experimental sensitivity in terms of the limit that will be set in the case that the observation is consistent with background. Following what the CUORE collaboration has done in [21], the method requires generating a large number of toy Monte Carlo spectra assuming zero $0\nu\beta\beta$ signal in the fit window (much wider than the $\delta E = \Delta E$ window used for the background-fluctuation sensitivity, in order to utilize the available shape information in the fit). For each Monte Carlo spectrum, a binned maximum likelihood fit to the spectrum is performed and used to extract the associated Bayesian limit with a flat signal prior by integrating the posterior probability density (the same analysis technique used in [26, 30]). Finally, the distribution of the limits calculated from the Monte Carlo spectra is constructed, and its median is taken to be the sensitivity.

The average-limit sensitivity method is, in a way, more powerful than the analytical background-fluctuation method because it can in principle take into account subtle and detector-dependent experimental effects, which can be difficult or sometimes impossible to model with analytical formulas. However, because the average-limit approach relies on analysis of statistical ensembles, it lacks the great advantages of clarity and simplicity offered by the analytical approach of the background-fluctuation sensitivity formulas. The two methods are, as will be shown, essentially equivalent given the same input parameters, though a minor systematic difference arises because the probability distribution of the limits is not symmetric and the median found with the MC does not coincide with the $\widehat{S}(\delta E)$ computed with Equation (7.8).

For a completed experiment like Cuoricino, the experimental parameters (e.g., background rate(s) and shape(s), resolution(s), exposure) have been directly measured and are used as inputs to the Monte Carlo. To adapt the approach for an upcoming experiment, it is of course necessary

to instead use the expected experimental parameters to generate the Monte Carlo spectra. Calculating the average-limit sensitivity in this way allows for the direct comparison of an upcoming experiment with previously reported experimental limits. The average-limit sensitivity is often considered in specific $0\nu\beta\beta$ experiments; for example, the GERDA experiment reports a sensitivity calculated in essentially this manner [53], although they choose to report the mean expected limit instead of the median.

7.1.4 Philosophical Interpretation of Sensitivity and the Choice of Significance or Credibility Level

Although the quantitative results of the background-fluctuation sensitivity calculation and the average-limit sensitivity calculation are ultimately similar, the motivation behind the way each is constructed is significantly different. In essence, the background-fluctuation sensitivity corresponds to the longest half-life of the decay under study that can produce a mean signal that would look like an upward background fluctuation at or above the chosen significance level. In contrast, the average-limit approach defines the sensitivity as the average limit at the chosen credibility level expected to be set by the experiment in the case that the true signal is *zero*. This being the case, the natural choice of significance/credibility level differs between the two. Therefore, although the background-fluctuation sensitivity and the average-limit sensitivity are comparable if computed at corresponding significance/credibility levels, an as-reported background-fluctuation sensitivity will typically *not* be comparable to an as-reported average-limit sensitivity because the significance/credibility levels will typically not coincide.

In the case of the background-fluctuation sensitivity, the natural choice of significance level is 1σ. A 1σ criterion corresponds to the expected error on the experiment's measurement that will arise due to background fluctuations; setting a background-fluctuation sensitivity at 1σ would thus define the sensitivity as the signal for which the experiment's expected measurement would be of the same magnitude as the expected error.

It is important to be clear that the significance level of the background-fluctuation sensitivity is defined in reference to $B(\delta E)$, the *mean* expected number of background counts, and that the

$\widehat{S_0}$ extracted using Equation (7.4) or Equation (7.7) is similarly considered to be the *mean* number of signal events that would be produced assuming a $0\nu\beta\beta$ half-life of $\widehat{T^{0\nu}_{1/2}}$. Said otherwise, if the true value $T^{0\nu}_{1/2}$ of the half-life is equal to the background-fluctuation sensitivity $\widehat{T^{0\nu}_{1/2}}$, then the experiment is not guaranteed to actually have an outcome with a number of observed events equal to or greater than $\widehat{S}(\delta E)+B(\delta E) = S(\delta E)+B(\delta E)$ falling into the energy window of interest (i.e., it is not guaranteed to achieve an observation at the desired significance level). In fact, considering that we have no prior knowledge of the true magnitude of the signal, the experiment's probability of achieving a particular result is not well defined. Attempts have been made to overcome this limitation in the definition of the statistical significance of a signal above background [50, 86]. Nevertheless, given its simplicity and intuitiveness, the definition of Equations (7.4) and (7.7) tends to persist, although adhering to this definition requires the understanding that the usual probabilistic interpretation associated with the use of a Gaussian σ cannot be applied to expectations for what the experiment will actually observe.

The average-limit sensitivity is constructed following a markedly different philosophy than the background-fluctuation sensitivity. Instead of being found by supposing that the $0\nu\beta\beta$ rate may be non-zero and determining how small a signal the experiment can observe without the signal becoming obscured by background fluctuations, the average-limit sensitivity is found by supposing that the $0\nu\beta\beta$ rate is zero and determining the average limit that the experiment is expected to set in that case. The natural choice for the average-limit sensitivity is therefore to report it at the percent credibility level at which the experiment chooses to set its limit; 90% C.L. is the most common choice. By nature, the average-limit sensitivity is intimately related to the experiment's analysis tools and specific method for setting the limit; it is therefore appropriate to evaluate it with a Monte Carlo study, in which the pertinent experimental parameters and a $0\nu\beta\beta$ rate of zero are used as inputs to simulate a suite of spectra, the experiment's real analysis tools are applied to each spectrum to extract a limit, and the average of the resulting distribution of limits is taken to be the average-limit sensitivity. There are several possible methods of defining this 'average' (e.g., mean, median, most probable value); the particular one chosen should be reported alongside the average-limit sensitivity value.

The CUORE collaboration prefers the use of the background-fluctuation sensitivity for upcoming experiments. The goal of $0\nu\beta\beta$ experiments is to discover and measure neutrinoless double beta decay; thus, from a philosophical perspective, it is more attractive to characterize an experiment's capabilities in terms of what it is able to see in the case that the $0\nu\beta\beta$ signal is non-zero than it is to focus on what the experiment will see if there is no signal to find. We consider the average-limit sensitivity to be most valuable for completed experiments; the average-limit sensitivity is meaningful for a completed experiment that has not seen evidence of a signal because it provides an understanding of how fortunate the experiment was in the limit it was able to set.

Although upcoming CUORE-family experiments have historically shown 1σ background-fluctuation sensitivities [23, 28], which quantitatively roughly coincide with 68% C.L. average-limit sensitivities, other upcoming $0\nu\beta\beta$ experiments commonly report 90% C.L. sensitivities. To prevent confusion between our sensitivity approach and that commonly used by other $0\nu\beta\beta$ experiments, it is instructive to compare 1.64σ background-fluctuation sensitivities to 90% C.L. average-limit sensitivities, which will be done in the following sections.

As noted in Section 7.1.2, the background-fluctuation sensitivity calculation cannot be extended to the ideal 'zero-background' case because it is meaningless to define sensitivity in terms of background fluctuations when $B(\delta E) = 0$. The zero-background sensitivity is thus, by necessity, analogous to the average-limit sensitivity.

7.1.5 Experimental Potential to Discover $0\nu\beta\beta$

In the case of experiments like those searching for $0\nu\beta\beta$, it may be desirable to frame the experiment's capabilities in terms of discovery potential rather than sensitivity; in other words, one may wish to report the maximum $0\nu\beta\beta$ half-life for which the experiment can be reasonably expected to be able to truly claim discovery of the decay. For the formulation of discovery potential, two criteria must be established: the requirement to claim discovery given a particular experimental observation, and the requirement to 'reasonably expect' to obtain a particular experimental

observation (in particular, one that satisfies the discovery criterion) given a particular true $0\nu\beta\beta$-signal-plus-background magnitude. The discovery potential then corresponds to the minimum true $0\nu\beta\beta$ signal magnitude that would satisfy these requirements.

For sufficiently large expected backgrounds $B(\delta E) >> 1$, the requirement to claim discovery can be straightforwardly expressed in the framework of the background-fluctuation sensitivity. When a finite background is present, it can never be entirely certain that a given observation is due to the presence of a signal, as there is always some possibility that the observation may arise from the background count distribution alone; however, the convention is that discovery may be claimed if an upward Gaussian background fluctuation of 5σ or greater would be required to explain the observation with the background distribution alone, corresponding to a probability of 2.87×10^{-7}. If we state that the requirement to 'reasonably expect' to be able to claim discovery is that the mean of the true (signal plus background) count distribution is at least large enough to fulfill this requirement, then the finite-background Gaussian-regime discovery potential is defined by Equation (7.4) or, equivalently, Equation (7.6), for $n_\sigma = 5$; in essence, it is the '5σ sensitivity.' In the Poisson regime, then, the finite-background discovery potential may be similarly considered to be the $0\nu\beta\beta$ half-life that would give rise to the mean signal $\widehat{S_0}$ found from Equation (7.8) for $\alpha = 2.87 \times 10^{-7}$ and the appropriate value of $B(\delta E)$.

For very small expected background levels $B(\delta E) \simeq 0$, however, we cannot continuously extrapolate the ability to claim that the experimental observation is inconsistent with the background-only hypothesis at a certain significance level; it is not possible to observe a fraction of an event, so the minimum requirement to be able to claim discovery is the observation of a single signal event. One way to ensure that the discovery potential represents the minimum possible signal for which the experiment can be reasonably expected to be able to claim discovery is to consider the true zero-background case. Unlike in the finite-background case, if the background is truly zero, the observation of a single event will satisfy the requirement to claim discovery. However, it is still necessary to set the requirement to 'reasonably expect' to be able to claim discovery (i.e., observe more than zero events). This can be done by requiring that the true expected signal distribution

corresponding to $\widehat{S}(\delta E)$ must yield at least a certain probability $\mathcal{P}$ of observing more than zero counts:

$$1 - p_{\widehat{S}(\delta E)}(0) = 1 - e^{-\widehat{S}(\delta E)} \geq \mathcal{P}. \tag{7.12}$$

This is mathematically equivalent to the upper limit in the case of zero observed counts that would be found from Equation (7.10) with a credibility level of $\mathcal{P}$.

Unlike the conventional requirement that an experimental observation must correspond to at least a 5σ background fluctuation to claim discovery, the choice of $\mathcal{P}$ is arbitrary. It defines a flat minimum threshold in $\widehat{S}(\delta E)$ depending upon how certain one wishes to be that an experiment will observe at least one signal event in its region of interest.

For a sufficiently small expected number of background counts, depending on the choice of $\mathcal{P}$, the requirement for the expected observation to be inconsistent with background becomes less stringent than the requirement that the experiment be reasonably likely to observe any signal event at all. A simple formulation of discovery potential can therefore be established by setting two criteria:

- $\widehat{S}(\delta E) \geq -\ln(1 - \mathcal{P})$ and
- $P(\widehat{S}(\delta E) + B(\delta E), B(\delta E)) \leq 2.87 \times 10^{-7}$,

where $P(a, x)$ is the lower normalized incomplete gamma function, as discussed in reference to Equation (7.8). The discovery potential curve is then defined by the minimum value of $\widehat{S}(\delta E)$ that satisfies both criteria.

Alternatively, a Monte-Carlo-based discovery potential can be constructed in an analogous manner to the average-limit sensitivity using the specific analysis mechanisms and choice of discovery criteria defined by the particular experiment. The sensitivity tools of [53] provide a prescription for a discovery potential calculated in this way.

7.2 Cuoricino

As Cuoricino is a completed experiment, it can serve well as a concrete illustration of the relationships between the two sensitivity calculations presented in Sections 7.1.1 and 7.1.3 and between an experiment's sensitivity and its observed result.

7.2.1 Detector Parameters and Backgrounds

There were three different types of crystals operated in Cuoricino: big natural crystals, small natural crystals, and small enriched crystals, typically abbreviated as big, small, and enriched (see Section 4.2.1 !). The physical detection efficiencies, average FWHM resolutions at 2614.5 keV, exposures, and flat background rates for each crystal type can be found tabulated in Table 4.1, subdivided by crystal type and Run as appropriate.

Some care must be taken in using the values in Table 4.1 to calculate the background-fluctuation sensitivity as it has been formulated in Section 7.1.1. First, the exposures are given in units of kg(^{130}Te)·y; as they will be used in place of $M \cdot t$ in Equations (7.2) and (7.3), they must be converted to total-mass exposures instead using the effective abundance of ^{130}Te in each crystal type. Second, the flat background rates have not been corrected for instrumental efficiency (see Section 4.2.3.2). For simplicity, this efficiency is neglected in the formulation of the background-fluctuation sensitivity (i.e., it is implicitly assumed to be 100%), so all background rates reported in Table 4.1 must be divided by the instrumental efficiency, 0.947, before being used to calculate Cuoricino's background-fluctuation sensitivity.

7.2.2 Sensitivity vs. Reported Limit

Cuoricino did not see any evidence for $0\nu\beta\beta$ and published a limit based on its observed spectrum, which was presented alongside an average-limit sensitivity. This sensitivity was evaluated as the median of the distribution of 90% C.L. limits extracted from toy Monte Carlo simulations that used the measured detector parameters as inputs, and it was determined to be $\widehat{T^{0\nu}_{1/2}}(90\%\ \text{C.L.}) = 2.6 \times 10^{24}$ y.

Because of the different crystal types present in Cuoricino, if we wish to calculate a background-fluctuation sensitivity for Cuoricino to compare with this average-limit sensitivity, we need to slightly adjust the background-fluctuation calculation presented in Section 7.1.1 to accommodate different parameter values for the different crystal types. Cuoricino can be considered as the sum of virtual detectors, each representing one of the crystal types during one of the two major data-taking periods, Run I and Run II. With the caveats discussed in Section 7.2.1 above, we can use the values given in Table 4.1 to calculate both our expected signal $\widehat{S}(\delta E)$ and expected background $B(\delta E)$ as sums of the contributions from these virtual detectors, then follow the Poisson-regime background-fluctuation sensitivity procedure. To quantitatively compare to a 90% C.L. average-limit sensitivity, we must choose to calculate the background-fluctuation sensitivity at 1.64σ ($\alpha = 0.051$); indeed, doing so yields $\widehat{T^{0\nu}_{1/2}}(1.64\sigma) = 2.6 \times 10^{24}$ y, in perfect agreement with the average-limit sensitivity.

Due to a tiny downward fluctuation in the observed flat background rate in the region of the Cuoricino spectrum right around the ^{130}Te $0\nu\beta\beta$ Q-value, Cuoricino set a limit of $T^{0\nu}_{1/2} > 2.8 \times 10^{24}$ y (90% C.L.) [21], which is slightly better than its sensitivity.

Following previously established convention for past bolometric experiments [23, 28], we choose to report background-fluctuation sensitivities at 1σ ($\alpha = 0.159$) for upcoming experiments. For the purpose of illustration, the corresponding background-fluctuation sensitivity for Cuoricino would be $\widehat{T^{0\nu}_{1/2}}(1\sigma) = 4.2 \times 10^{24}$ y.

7.3 CUORE-0 and CUORE

7.3.1 Detector Parameters and Background Estimates

The experimental parameters of CUORE-0 and CUORE that are used in the sensitivity calculations are summarized in Table 7.1.

Of these experimental parameters, the background rate is probably the one with the most impact on the sensitivity evaluation, as it and the exposure together determine the number of background counts $B(\delta E)$ expected to be observed in the region of interest, from which the sensitivity is then determined. Figure 7.1 illustrates both the 1σ sensitivity and the 5σ discovery potential in units of signal counts in the region of interest, $\widehat{S}(\delta E)$, as a function of $B(\delta E)$. For the sensitivity, curves for

Table 7.1: Values used in the estimation of the sensitivity of CUORE-0 and CUORE. Symbols are defined in Equations (7.2) – (7.4). See Section 4.2.3.2 for a discussion of the background values.

Experiment	a (%)	η	ε (%)	W (g/mol)	M (kg)	ΔE (keV)	$f(\Delta E)$ (%)	b (cts/(keV kg y))
CUORE-0	34.167	1	87.4	159.6	39	5	76	0.05
CUORE	34.167	1	87.4	159.6	741	5	76	0.01

the zero-background (zero-count limit), small-background (Poisson background-fluctuation sensitivity), and large-background (Gaussian background-fluctuation sensitivity) regimes are all shown at equivalent significance/credibility levels. For the discovery potential, only the Poisson and Gaussian curves are shown; in this case, the zero-background criterion will depend upon the threshold criterion chosen to define the desired probability of observing at least one signal event. Regions indicating the background regimes into which CUORE-0 and CUORE are expected to fall for a range of live times are shown. A brief discussion of the background estimates follows.

Because the geometry of CUORE-0 is similar to that of Cuoricino (Section 4.3.1, the surface contamination reduction factors reported in Section 4.2.3.2 scale almost directly to the background we expect to observe in the region of interest. The total amount of copper facing the crystals is only slightly reduced with respect to Cuoricino, but its surface has been treated with the new procedure developed for CUORE. CUORE-0 is housed in the Cuoricino cryostat, so the gamma background from contamination in the cryostat shields remains approximately the same as in Cuoricino. We consider that the irreducible background for CUORE-0 comes from the 2614.5-keV ^{208}Tl line due to ^{232}Th contaminations in the cryostat, in the case that all other background sources (i.e., surface contaminations) have been rendered negligible; this would imply a lower limit of ≈ 0.05 cts/(keV kg y) on the expected background in CUORE-0. Similarly, an upper limit of 0.11 cts/(keV kg y) follows from scaling the Cuoricino background in the conservative case, described above, of a factor of 2 improvement in crystal and copper contamination.

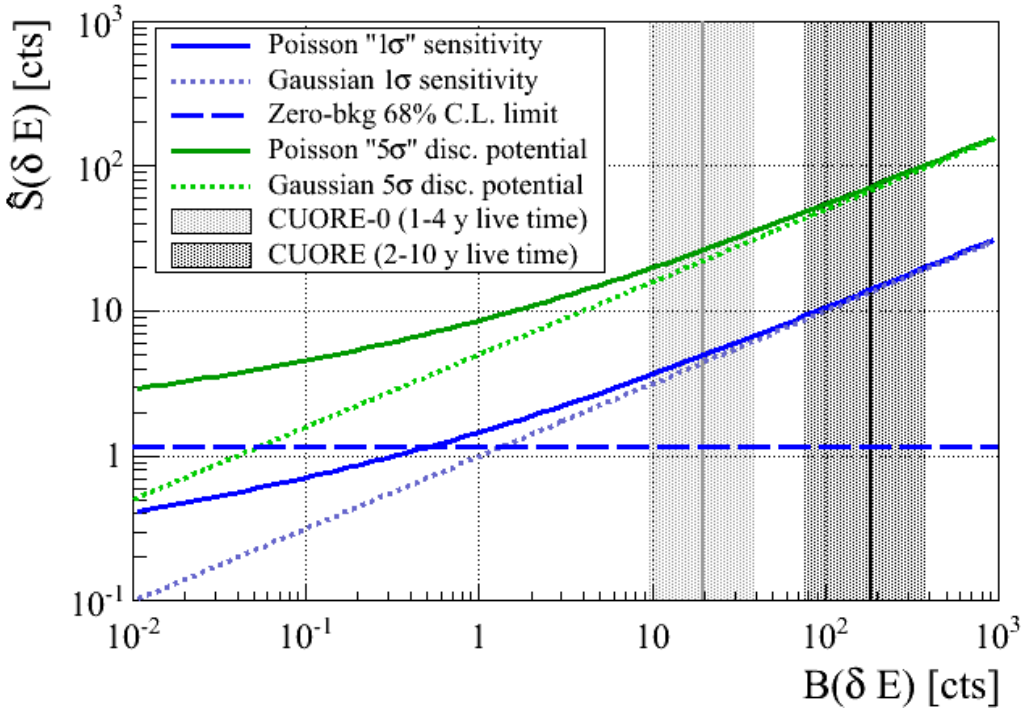


Figure 7.1: Background-fluctuation sensitivity and discovery potential curves in units of counts in δE. The Poisson curve approaches the Gaussian curve at the same significance level for $B(\delta E) >> 1$. For very small $B(\delta E)$, the discovery potential will follow whatever flat minimum $\widehat{S}(\delta E)$ threshold is chosen until that threshold crosses the Poisson 5σ curve. The shaded regions indicate the regimes into which CUORE-0 and CUORE are expected to fall for $\delta E = \Delta E$, given their anticipated exposures; the vertical lines indicate the values of $B(\delta E)$ corresponding to 2 y of CUORE-0 live time (with $b = 0.05$ cts/(keV kg y)) and 5 y of CUORE live time (with $b = 0.01$ cts/(keV kg y)), respectively.

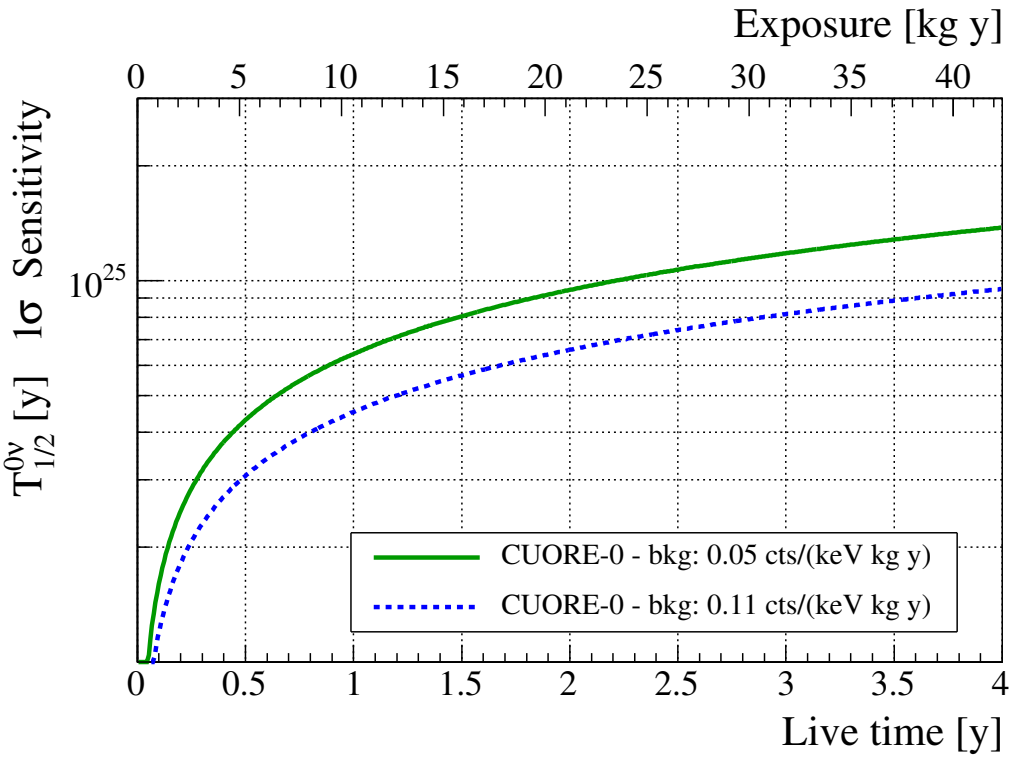


Figure 7.2: CUORE-0 background-fluctuation sensitivity at 1σ for two different values of the background rate in the region of interest, 0.05 cts/(keV kg y) (solid line) and 0.11 cts/(keV kg y) (dotted line), representing the range into which the CUORE-0 background is expected to fall.

Table 7.2: Several estimators of the experimental capabilities of CUORE-0 under different background estimations after one, two, and four years of live time. The boldfaced column corresponds to the anticipated total live time of two years. The background-fluctuation half-life sensitivities at 1σ are the official sensitivity values reported by the collaboration. 1.64σ background-fluctuation sensitivities and 90% C.L. average-limit sensitivities, in italics, are provided to illustrate the similarity of the two values. The 5σ discovery potentials for $\mathcal{P} = 0.90$ are also given.

			half-life sensitivity		
b	ΔE	Method	(10^{25} y)		
(cts/(keV kg y))	(keV)	(sig./conf. level)	1 y	**2 y**	4 y
0.11	5	1σ	0.45	**0.66**	0.95
		1.64σ	*0.28*	*0.40*	*0.58*
		90% C.L.	*0.29*	*0.41*	*0.59*
		5σ	0.085	**0.13**	0.18
0.05	5	1σ	0.64	**0.94**	1.4
		1.64σ	*0.39*	*0.58*	*0.84*
		90% C.L.	*0.39*	*0.59*	*0.83*
		5σ	0.12	**0.18**	0.26

A plot of the expected 1σ background-fluctuation sensitivity of CUORE-0 as a function of live time in these two bounding cases is shown in Figure 7.2. Table 7.2 provides a quantitative comparison among 1σ background-fluctuation sensitivities (as shown in Figure 7.2), 1.64σ background-fluctuation sensitivities, 90% C.L. average-limit sensitivities, and 5σ discovery potentials for CUORE-0 at several representative live times. The anticipated total live time of CUORE-0 is approximately two years; for this live time at the 0.05 cts/(keV kg y) background level, $B(\delta E) \approx 20$ cts, meaning that the Poisson-regime calculation is necessary (see Section 7.1.1). The pale shaded region in Figure 7.1 illustrates where the CUORE-0 live time range considered in Table 7.2 lies with respect to the statistical regime of the sensitivity calculations for the 0.05 cts/(keV kg y) background level.

CUORE, in addition to the new crystals and frames already present in CUORE-0, will be assembled as a 19-tower array in a newly constructed cryostat. The change in detector geometry will have two effects. First, the large, close-packed array will enable significant improvement in the anticoincidence analysis, further reducing crystal-related backgrounds. Second, the fraction of the total crystal surface area facing the outer copper shields will be reduced by approximately a factor of 3. In addition to these considerations, the new cryostat will contain thicker lead shielding and be constructed of cleaner material, which should result in a gamma background approximately an order of magnitude lower than that in the Cuoricino cryostat. Based on the above considerations and the Cuoricino results, CUORE is expected to achieve its design background value of 0.01 cts/(keV kg y).

An overview of the 1σ background-fluctuation sensitivities of the Cuoricino, CUORE-0, and CUORE TeO_2 bolometric experiments is shown in Figure 7.3. The Cuoricino 1σ sensitivity calculated in Section 7.2.2 is shown for reference. A 1σ half-life sensitivity close to 10^{25} years is expected from 2 years' live time of CUORE-0. Once CUORE starts data-taking, another order of magnitude improvement in sensitivity is expected in another two years.

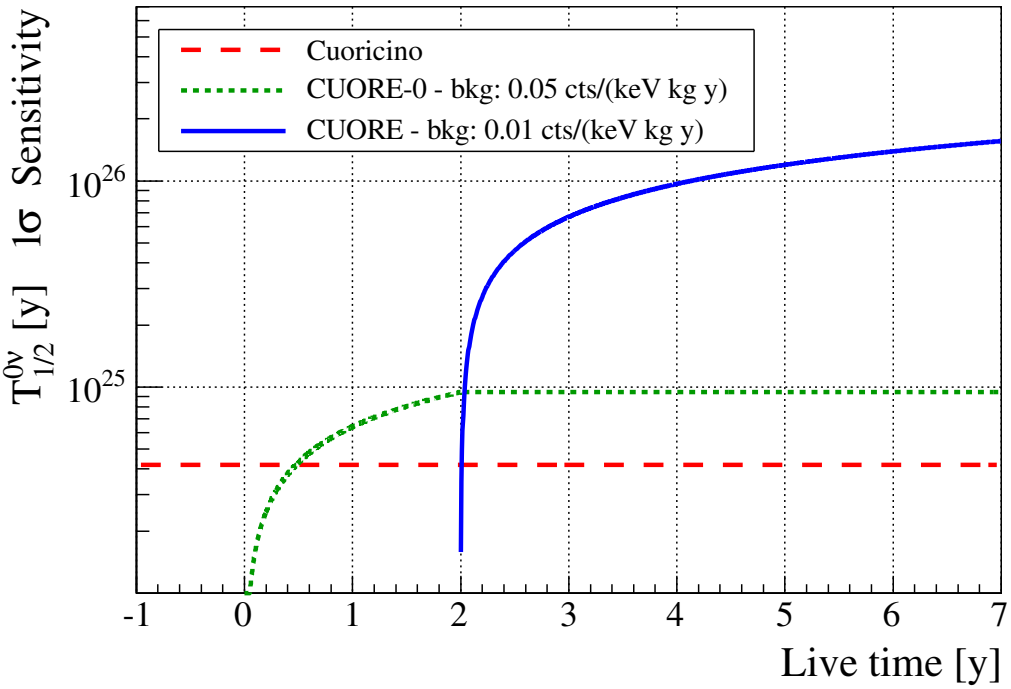


Figure 7.3: 1σ expected background-fluctuation sensitivities for the CUORE-0 (dotted line) and CUORE (solid line) experiments, calculated from Equations (7.8) and (7.2) with the experimental parameters shown in Table 7.1. The Cuoricino 1σ sensitivity calculation (dashed line) is discussed in Section 7.2.2.

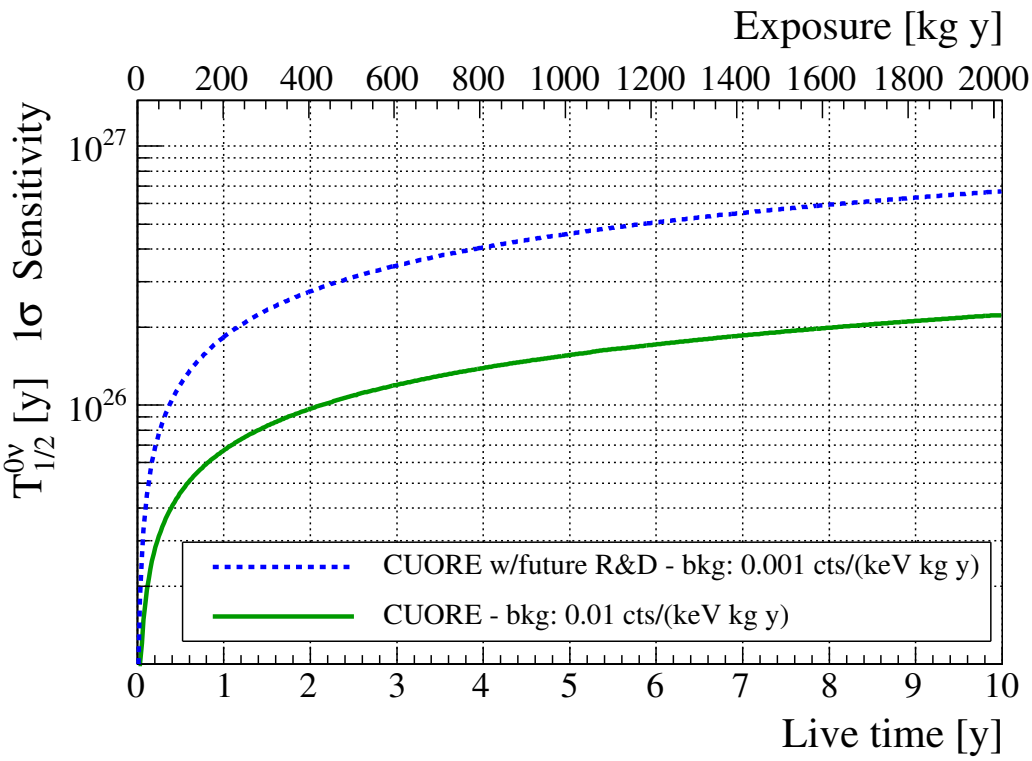


Figure 7.4: Background-fluctuation sensitivity of the CUORE experiment at 1σ (solid line) for the design goal background level. The sensitivity for an order-of-magnitude improvement over the baseline background is also shown (dotted line).

Table 7.3: Several estimators of the experimental capabilities of CUORE after two, five, and ten years of live time. The boldfaced column corresponds to the anticipated total live time of five years. The values are reported for the design goal background level, as well as for an order-of-magnitude improvement over the design goal. The background-fluctuation half-life sensitivities at 1σ are the official sensitivity values reported by the collaboration. 1.64σ background-fluctuation sensitivities and 90% C.L. average-limit sensitivities, in italics, are provided to illustrate the similarity of the two values. The 5σ discovery potentials for $\mathcal{P} = 0.90$ are also given.

			half-life sensitivity		
b	ΔE	Method	(10^{26} y)		
(cts/(keV kg y))	(keV)	(sig./conf. level)	2 y	**5 y**	10 y
0.01	5	1σ	0.97	**1.6**	2.2
		1.64σ	*0.59*	*0.95*	*1.4*
		90% C.L.	*0.59*	*0.97*	*1.4*
		5σ	0.19	**0.30**	0.44
0.001	5	1σ	2.7	**4.6**	6.7
		1.64σ	*1.7*	*2.8*	*4.1*
		90% C.L.	*1.6*	*2.8*	*4.2*
		5σ	0.50	**0.86**	1.3

A plot of the CUORE experiment's sensitivity as a function of live time and exposure is shown in Figure 7.4. Table 7.3 provides a quantitative comparison among 1σ background-fluctuation sensitivities (as shown in Figure 7.4), 1.64σ background-fluctuation sensitivities, 90% C.L. average-limit sensitivities, and 5σ discovery potentials for CUORE at several representative live times. The anticipated total live time of CUORE is approximately five years; for this live time at the design goal background level, $B(\delta E) \approx 190$ cts, meaning that the Gaussian approximation would still be valid in this case. The sensitivity values shown here nevertheless differ from those previously reported by the experiment [29, 23] by about 25%. This difference can be attributed to the inclusion of the signal fraction $f(\delta E)$, which was not previously considered. The dark shaded region in Figure 7.1 illustrates where the CUORE live time range considered in Table 7.3 lies with respect to the statistical regime of the sensitivity calculations for the 0.01 cts/(keV kg y) background level.

While it is unlikely that CUORE itself will reach a background rate of 0.001 cts/(keV kg y) or below, R&D activities are already underway pursuing ideas for further reduction of the background in a possible future experiment. Techniques for active background rejection are being investigated [79, 78, 56]) that could provide substantial reduction of the background. Sensitivities for a scenario with 0.001 cts/(keV kg y) in a CUORE-like experiment are given in Figure 7.4 and Table 7.3.

7.3.2 Comparison with the claim in ^{76}Ge

It is interesting to compare the CUORE-0 and CUORE sensitivities with the claim for observation of $0\nu\beta\beta$ in ^{76}Ge [102, 103, 105]. The authors of this claim have reported several different values for the half-life of ^{76}Ge, depending upon the specifics of the analysis; the longest of these, and thus the one requiring the greatest sensitivity to probe, is $T^{0\nu}_{1/2}\,(^{76}\text{Ge}) = 2.23^{+0.44}_{-0.31} \times 10^{25}$ y [105]. From Equation (2.3), it follows that

$$T^{0\nu}_{1/2}\,(^{130}\text{Te}) = \frac{F_N(^{76}\text{Ge})}{F_N(^{130}\text{Te})} \cdot T^{0\nu}_{1/2}\,(^{76}\text{Ge}).$$

However, simply directly using the F_N values summarized in Table 2.1 and this equation to estimate the expected half-life for the $0\nu\beta\beta$ of ^{130}Te can be misleading. Instead, correlations between the F_N calculations for the two nuclides should be taken into account.

A method of treating NME uncertainties based on the QRPA-T calculations is suggested, and shown to be roughly consistent with the QRPA-J and ISM calculations, in [71]; the authors linearize Equation (2.3) with logarithms to allow for simpler treatment of asymmetric errors, extract conservative 1σ error contours from the QRPA-T calculations, and show that the QRPA-J and ISM calculations are consistent with the 3σ contours obtained in this way. Although the values have not been updated to utilize the most recent QRPA-T calculations, the authors argue in a recent addendum to the original article that they remain a valid estimate of the spread of NME calculations [72]. Following this method and applying the updated phase space factors calculated in [108] (with an appropriate correction for $r_0 = 1.1$ fm instead of 1.2 fm; see Section 2.2), the expected 1σ range of $T^{0\nu}_{1/2}$ (^{130}Te) is $(0.49-1.0) \times 10^{25}$ y (including the 1σ uncertainty on the ^{76}Ge claim as done in [71]).

The mathematical framework of the background-fluctuation sensitivity calculation can be inverted to determine the magnitude of the mean signal in terms of n_σ that an assumed 'true' half-life value will produce in an experiment. Figure 7.5 shows the n_σ significance level at which CUORE-0 can probe the ^{76}Ge claim as it accrues statistics over its anticipated live time. The band is bounded by curves corresponding to the maximum and minimum $T^{0\nu}_{1/2}$ (^{130}Te) of the range given above. As can be deduced from the plot, CUORE-0 will achieve at least a 1σ sensitivity to any signal within the expected 1σ range of $T^{0\nu}_{1/2}$ (^{130}Te) within two years. By combining data from CUORE-0 and Cuoricino, the claim could be verified in a shorter time with higher sensitivity.

Thanks to the increased size and lower background, if the ^{130}Te $0\nu\beta\beta$ half-life indeed falls in the 1σ range implied by the claim in ^{76}Ge, CUORE will already be able to achieve a 5σ expected signal above background within about eight months of live time.

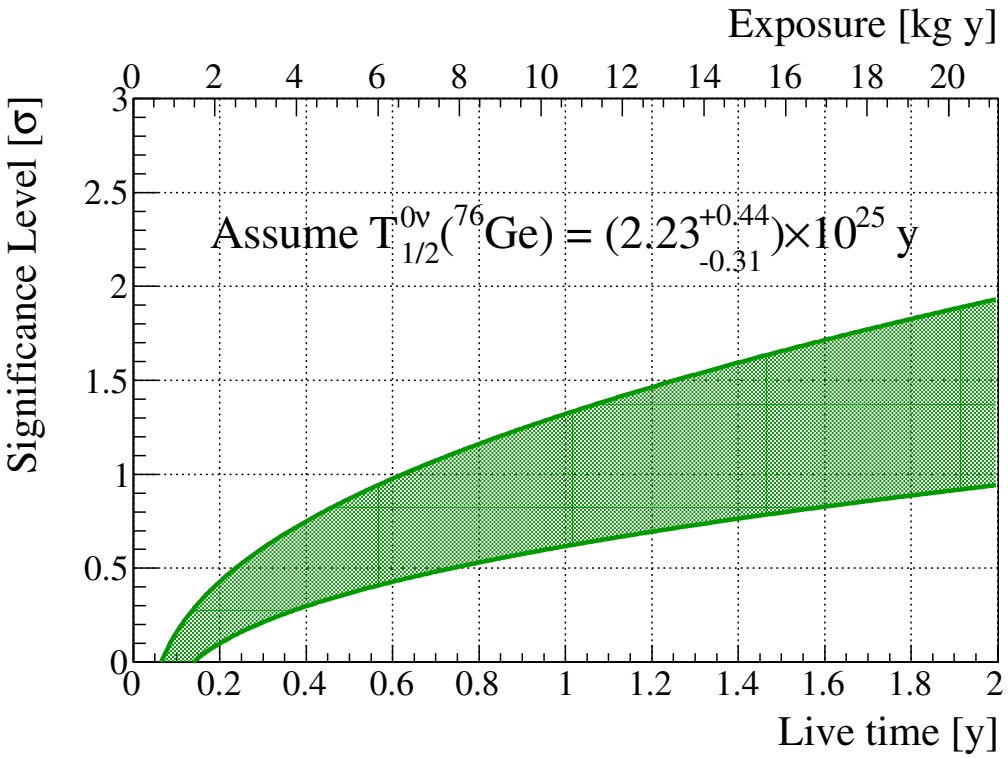


Figure 7.5: Significance level at which CUORE-0 can observe a signal corresponding to the ^{76}Ge claim, assuming the best expected background of 0.05 cts/(kev kg y). The width of the band accounts for both the 1σ uncertainty on the ^{76}Ge claim and the 1σ range of QRPA-T NMEs calculated in [71], but it is by far dominated by the NME spread.

7.3.3 Comparison with $\langle m_{\beta\beta} \rangle$ Allowed Regions

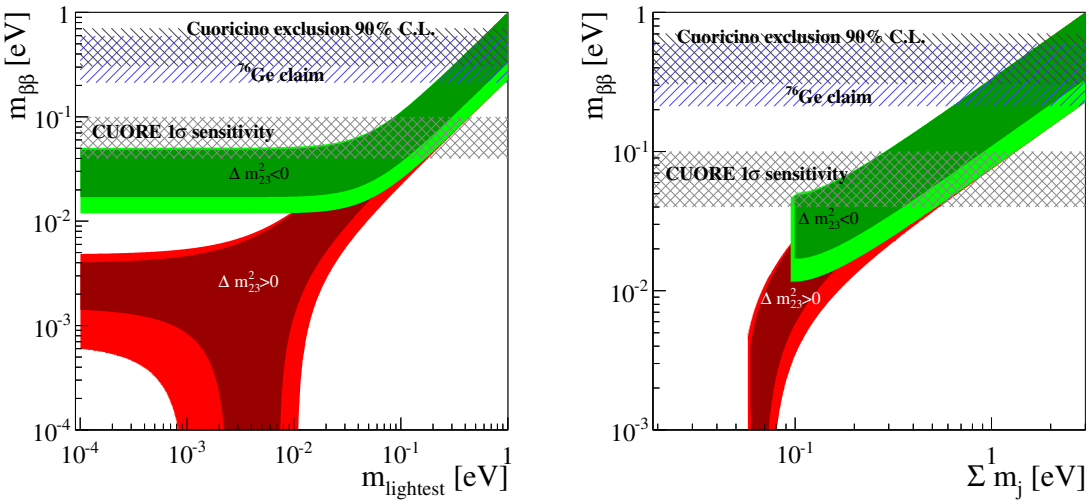
As discussed in Section 2.2, the preferred values of the neutrino mass parameters obtained from neutrino oscillation experiments define a certain allowed region of parameter space into which the effective $0\nu\beta\beta$ mass $\langle m_{\beta\beta} \rangle$ may fall if the neutrino is indeed a Majorana particle. Ultimately, $\langle m_{\beta\beta} \rangle$ is the physical parameter that $0\nu\beta\beta$ experiments wish to measure, but the theoretical uncertainty in NME values makes such a determination difficult. If we wish to discuss experimental sensitivity in terms of $\langle m_{\beta\beta} \rangle$, it is cleanest to consider the single half-life sensitivity value of the experiment at its expected total live time.

Table 7.4 contains a summary of 1σ background-fluctuation sensitivities to the neutrino Majorana mass at the expected total live times of CUORE and CUORE-0. These values are considered the official sensitivity values for CUORE-family experiments. The $\langle m_{\beta\beta} \rangle$ sensitivities are determined from the half-life sensitivities according to several different NME calculations, assuming that the exchange of a light Majorana neutrino is the dominant $0\nu\beta\beta$ mechanism, as discussed in Section 2.2; the corresponding nuclear factors of merit are summarized in Table 2.1.

For illustrative purposes, Table 7.4 also shows the limiting "zero-background" case for both CUORE-0 and CUORE. The calculation is performed at 68% C.L. so that the values can be considered as zero-background extrapolations of the finite-background 1σ background-fluctuation sensitivities. CUORE-0 and CUORE will both have sufficiently good resolution that the considered counting window may be expanded with the background contribution from the $2\nu\beta\beta$ continuum remaining negligible; therefore, the zero-background sensitivities in Table 7.4 are calculated for $\delta E = 2.5\Delta E$. As discussed in Section 7.1.2, the zero-background approximation applies when $B(\delta E) \lesssim 1$ count; the background rate that each experiment would have to achieve to fulfill this requirement can be determined from Equation (7.3). CUORE-0 would require $b \lesssim 1.0 \times 10^{-3}$ cts/(keV kg y); CUORE would require $b \lesssim 2.2 \times 10^{-5}$ cts/(keV kg y), nearly three orders of magnitude better than the baseline background rate.

Table 7.4: Summary table of expected parameters and 1σ background-fluctuation sensitivity in half-life and effective Majorana neutrino mass. The different values of $\langle m_{\beta\beta}\rangle$ depend on the different NME calculations; see Section 2.2 and Table 2.1. Zero-background sensitivities for a window of $\delta E = 2.5\Delta E$, in italics, are also provided as an estimation of the ideal limit of the detectors' capabilities; they are presented at 68% C.L. so that they can be considered as approximate extrapolations of the 1σ background-fluctuation sensitivities.

				$\langle m_{\beta\beta}\rangle$ (meV)					
Setup	t (y)	b (cts/(keV kg y))	$\widehat{T^{0\nu}_{1/2}}(1\sigma)$ (y)	QRPA-T	QRPA-J	ISM	IBM-2	PHFB	GCM
CUORE-0	2	0.05	9.4×10^{24}	160 – 280	170 – 290	340 – 420	190 – 210	170 – 300	170
	zero-bkg. case at 68% C.L.:		*5.3×10^{25}*	*68 – 120*	*73 – 120*	*140 – 180*	*79 – 91*	*74 – 130*	*74*
CUORE baseline	5	0.01	1.6×10^{26}	40 – 69	42 – 71	83 – 100	46 – 53	43 – 74	43
	zero-bkg. case at 68% C.L.:		*2.5×10^{27}*	*9.9 – 17*	*11 – 18*	*21 – 26*	*12 – 13*	*11 – 18*	*11*



(a) (b)

Figure 7.6: The Cuoricino result and the expected CUORE 1σ background-fluctuation sensitivity overlaid on plots that show the bands preferred by neutrino oscillation data (inner bands represent best-fit data; outer bands represent data allowing 3σ errors) [81]. Both normal ($\Delta m^2_{23} > 0$) and inverted ($\Delta m^2_{23} < 0$) neutrino mass hierarchies are shown. (a) The coordinate plane represents the parameter space of $\langle m_{\beta\beta} \rangle$ and $m_{lightest}$, following the plotting convention of [128]. (b) The coordinate plane represents the parameter space of $\langle m_{\beta\beta} \rangle$ and Σm_j, following the plotting convention of [80]. The width of the CUORE band is determined by the maximum and minimum values of $\langle m_{\beta\beta} \rangle$ obtained from the six NME calculations discussed in Section 2.2; the widths of the Cuoricino and ^{76}Ge bands are drawn from the $\langle m_{\beta\beta} \rangle$ ranges reported in the publications of the respective results.

In Figure 7.6, the expected sensitivity of CUORE is compared with the preferred values of the neutrino mass parameters obtained from neutrino oscillation experiments. The sensitivity of CUORE will allow the investigation of the upper region of the effective Majorana neutrino mass phase space corresponding to the inverted hierarchy of neutrino masses.

Note that the Cuoricino result, the sensitivity of CUORE, and the ^{76}Ge claim are shown as bands in Figure 7.6. In the Cuoricino case, the 'true' limit lies somewhere within the band, and the experiment excludes all greater values of $\langle m_{\beta\beta} \rangle$. Similarly, the true sensitivity of CUORE lies somewhere within its band, and the experiment is considered to be sensitive to all greater values of $\langle m_{\beta\beta} \rangle$. In contrast, as the ^{76}Ge claim is a claim of observation, it corresponds to a claim that the *true value* of $\langle m_{\beta\beta} \rangle$ lies somewhere within its corresponding band. Although the Cuoricino band almost entirely overlays the ^{76}Ge band, Cuoricino does not exclude the Klapdor-Kleingrothaus claim for any given NME calculation.

7.3.4 Sensitivity Comparison with Other Leading $0\nu\beta\beta$ Experiments

The main value of a sensitivity calculation is to allow for a direct, transparent comparison of the capabilities of experiments that may have very different systematics and approaches to analysis. The sensitivity formulation presented in this chapter allows us to perform such a comparison between CUORE-0 and CUORE and the other current and upcoming $0\nu\beta\beta$ experiments introduced in Section 3.2; the one exception is SuperNEMO, because it will be operating in a regime in which its background in the $0\nu\beta\beta$ region is dominated by the tail of the $2\nu\beta\beta$ spectrum, a situation that is not well modeled by our sensitivity calculation.

To obtain a full picture of how the sensitivities of a selection of contemporaneous experiments compare to each other, it is valuable to plot sensitivity as a function of calendar time, with the curve for each experiment beginning at the date on which it is expected to turn on. To do this, it is necessary to estimate the live time fraction for each experiment. The baseline expected live time fraction for CUORE-0 and CUORE is 75%, driven by the expectation of one week of calibration activities per month of detector operation.

Table 7.5: Values used in the estimation of the sensitivity of all $0\nu\beta\beta$ experiments considered in this comparison. Symbols are defined in Equations (7.2) – (7.4). For each experiment, either $\eta = 1$ or an effective molecular weight W_{eff} is chosen such that η_{eff}=1. In each case, ΔE is the FWHM resolution of the detector, so $f(\Delta E) = 76\%$. Two different background estimates are considered for SNO+ due to an ambiguity in the presentation of the expected ^{214}Bi contribution (see text).

Experiment	a (%)	ε (%)	W (g/mol)	M (kg)	ΔE (keV)	b (cts/(keV kg y))	Live time frac. (%)
CUORE-0	34.167	87.4	159.6	39	5	0.05	75
CUORE	34.167	87.4	159.6	741	5	0.01	75
GERDA Phase I	86	87	75.6	12.7	4.5	0.02	95
GERDA Phase II	87	87	75.7	20	3	0.001	95
Majorana demonstrator 1	86	90	75.6	10	4	0.00075	100
Majorana demonstrator 2	86	90	75.6	30	4	0.00075	100
EXO-200 first data	80.6	71	135.5	63	260	0.0015	60
EXO-200	80.6	71	135.5	98.5	96.7	0.0015	60
NEXT-100	90	100	135.7	100	25	0.0008	50
KamLAND-Zen first data	90.93	100	5388 (eff.)	5450	244	0.0001	93
KamLAND-Zen restart	90.93	100	5388 (eff.)	5450	244	0.000003	93
SNO+	34.167	100	42536 (eff.)	156000	270	0.0000008	80

A discussion of estimates of the experimental parameters needed for the sensitivity calculation for each of the considered experiments follows. The parameters are summarized in Table 7.5.

GERDA Phase I has been operating since Nov. 1, 2011, and its experimental parameters have been well established. Two of the enriched detectors are problematic and are not used for physics analysis, leaving a total mass of 14.6 kg, which must be corrected by the active mass fraction remaining after discounting the outer dead layer of the detectors (86.7%) [11]. These detectors are enriched to 86% in ^{76}Ge; as the centrifuge used for enrichment preferentially selects heavier isotopes, the atomic weight can be estimated by assuming that the remainder is composed of ^{74}Ge. The physical efficiency of the detectors has been evaluated to be 87% by simulations [53]. The mass-weighted average FWHM resolution of the detectors in the Phase I run is 4.5 keV, the observed background rate is 0.02 cts/(keV kg y), and the live time percentage has proven to be $\approx 95\%$ [11, 88]. Phase II data-taking will begin in early 2013; around 20 kg of new detectors will be added, but the Phase I detectors will also continue to take data. The live time percentage and physical efficiency for the new detectors can be considered the same as for Phase I. However, the enrichment level for these detectors will be 87% [110], and a conservative extrapolation of the resolutions obtained for the new detectors [88] yields a FWHM of approximately 3 keV. The background level goal in these detectors is 0.001 cts/(keV kg y) [53].

As the Majorana demonstrator has not yet begun data-taking, its detector parameters are as yet less concrete than those of GERDA. Approximately 10 kg will be deployed in the first phase, followed by an additional 20 kg to bring the total to 30 kg for the second phase. The detectors will be enriched to 86% and have a physical efficiency of 90%; the FWHM resolution is expected to be 4 keV, and the live time percentage is expected to be near 100% [76, 64]. The background level goal is 3 c/(tonne y) in the 4-keV region of interest [76], corresponding to 0.00075 cts/(keV kg y).

EXO-200 began taking data on May 21, 2011. The detector contains 175 kg of liquid xenon, enriched to 80.6% in ^{136}Xe with the remainder being ^{134}Xe [9]. The reconstruction efficiency for $0\nu\beta\beta$ events is 71%, and the observed background level is 0.0015 cts/(keV kg y) [34]. The live time percentage can be deduced to be $\approx 60\%$ based on the operating dates and collected live times reported in [9] and [34]. For the first data collected by EXO-200, the fiducial mass of the

analysis was 63 kg of the 175 kg total, and the energy resolution was found to be $\sigma(\mathrm{E})/\mathrm{E} = 4.5\%$, which corresponds to a FWHM of 260 keV at the ^{136}Xe Q-value of 2458 keV [9]. However, an analysis of data collected starting September 22, 2011, after the optimization of electron lifetime in the detector, achieved a fiducial mass of 98.5 kg and a FWHM energy resolution of 96.7 keV ($\sigma(\mathrm{E})/\mathrm{E} = 1.67\%$) [34]; these parameters are likely to be more representative of the performance of the detector going forward.

NEXT-100 will contain $100-150$ kg of xenon gas [87]. As we have been unable to find an efficiency estimate for the detector, we will use a mass of 100 kg and assume that this applies to a fiducial volume selected such that the efficiency is $\approx 100\%$. The background level goal is 0.0008 cts/(keV kg y); we will follow the 'conservative' resolution and live time percentage values used for the sensitivity estimate in [87] and use 25 keV and 50%, respectively. Based on the composition of NEXT-DEMO, we will assume an isotopic abundance of 90% ^{136}Xe [19], with the remainder assumed to consist of ^{134}Xe.

The KamLAND-Zen setup is different from those we have considered thus far in that the target mass is a mixture of several materials of which one is the source, instead of a compound containing the source (CUORE-0/CUORE) or a pure element in some form (GERDA/Majorana, EXO, NEXT). The scintillator is loaded with 2.52% xenon by mass; the xenon is enriched to 90.93% in ^{136}Xe, and the fiducial mass is reported to be 125 kg ^{136}Xe [83]. From this, it can be deduced that the total fiducial target mass is 5450 kg. For the purposes of the sensitivity calculation, we can use the same values to construct an effective molecular weight for the target mass by assuming a 1:1 ratio between one mole of xenon and one 'unit' of the scintillator mixture, allowing us to use an effective stoichiometric coefficient of $\eta_{eff} = 1$; this results in $W_{eff} = 5388$ g/mol. We assume that the fiducial volume is chosen such that the physical efficiency is $\approx 100\%$, as other analysis cuts are stated to reduce the efficiency by less than 0.1%; the $6.6\%/\sqrt{E(\mathrm{MeV})}$ σ resolution corresponds to a 244 keV FWHM resolution, and a live time percentage of approximately 93% can be deduced from reported data collection statistics [83, 82]. Due in part to the comparatively poor energy resolution, background modeling for KamLAND-Zen is more complex than for the other experiments we have considered up to this point; spectral shapes corresponding to a variety

of background sources are included in the fit for the $0\nu\beta\beta$ analysis. However, the experiment has published a figure showing the overall background rate in the 2.2 MeV $< E/, < 3.0$ MeV energy region over time [82] from which an approximate background rate of 0.0001 cts/(keV kg y) can be extracted. When data taking resumes after the purification of the xenon-loaded scintillator, the rate due to contaminants attributed to the Fukushima reactor accident is expected to be reduced by approximately a factor of 100 [98]. We assume that this will recover a background rate similar to that reported in [106], where 9.35 background events are expected for 400 kg of xenon in 2.5%-loaded scintillator in a roughly 230-keV-wide signal region, corresponding to 0.000003 cts/(keV kg y).

SNO+ is similar to KamLAND-Zen in many respects, except that it will be loaded with natural tellurium, which has an atomic weight of 127.60833 g/mol [74], instead of xenon. The $0\nu\beta\beta$ candidate of interest is ^{130}Te, and the anticipated loading is 0.3% by weight [1, 89]; this yields $W_{eff} = 42536$ g/mol calculated in a similar manner as for KamLAND-Zen. The SNO+ collaboration's understanding of the sensitivity expectations for the tellurium-loaded configuration is still in a preliminary stage; however, the available estimates of the pertinent detector parameters [89] can be used to obtain a sensitivity estimate as follows. The expected sigma energy resolution is 4.5% at the ^{130}Te Q-value, which corresponds to a FWHM resolution of about 270 keV. The background modeling for SNO+ is similarly complicated to that of KamLAND-Zen; a rough background estimate of 0.0000008 cts/(keV kg y) can be deduced from the spectrum shown in [89], which considers a 3.5-m-radius fiducial volume, encompassing 20% of the 780 tonnes of total target mass, to reduce the contribution of external backgrounds. We assume that the fiducial volume also ensures a physical efficiency of near 100%, as in KamLAND-Zen. Sensitivity calculations performed by the SNO+ collaboration for the ^{150}Nd configuration estimated a live time percentage of 80% [93], and there is no reason to suppose that this parameter does not apply equally well to the ^{130}Te case.

Sensitivity curves as a function of calendar time for the full range of NMEs considered previously in this chapter in Table 7.4 and Figure 7.6 are shown in Figure 7.7; the nuclear factors of merit used for the calculation are listed in Table 2.1. Sensitivities for multi-stage experiments and experimental programs are shown as continuous curves with kinks appearing when subsequent stages turn on. For the case of CUORE-0 and CUORE, the curve after the kink represents CUORE

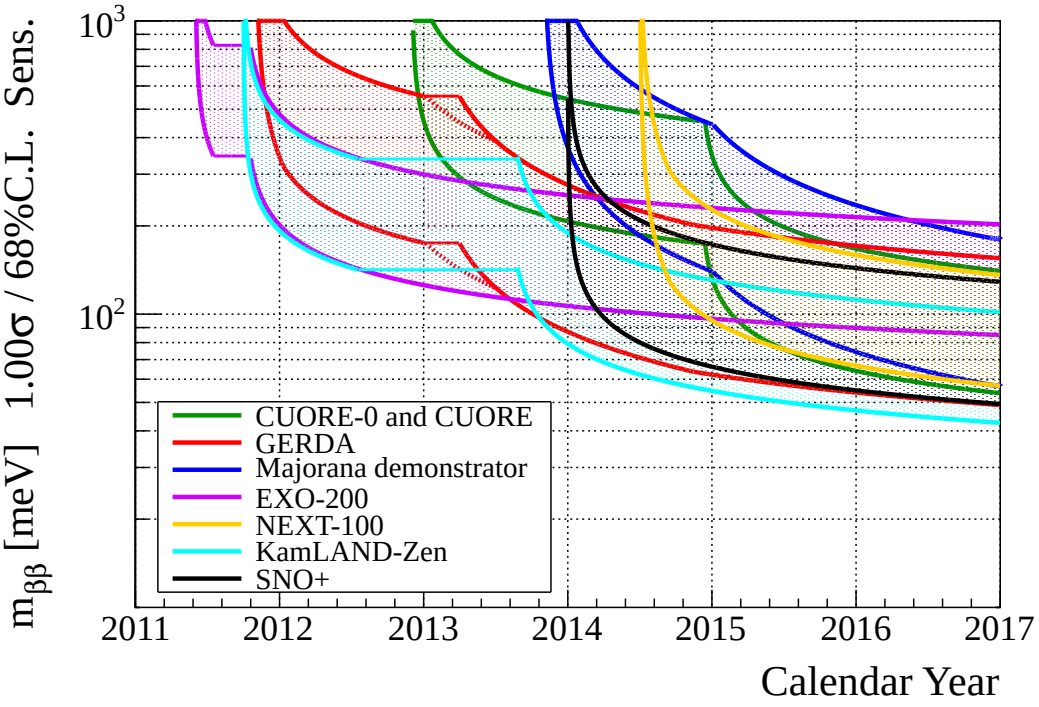


Figure 7.7: Sensitivity curves vs. calendar time for a selection of leading $0\nu\beta\beta$ experiments. Multi-stage experimental programs are represented by continuous curves with kinks when subsequent stages turn on. The NME range represented corresponds to that which appears in Table 7.4 and Figure 7.6; the nuclear factors of merit used for the calculation are listed in Table 2.1. The dotted lines appended to the GERDA curve illustrate the boundaries of the sensitivity band arising from combining data from Phase I and Phase II during the time when it is advantageous to do so.

data only; the background expected in CUORE is so much better than that expected in CUORE-0 that it is not advantageous to combine data. Similarly, after the flat portions of the EXO-200 and KamLAND-Zen curves (corresponding to the time spent tuning the electron lifetime in EXO-200 and the time spent purifying the scintillator in KamLAND-Zen), the curves represent only data taken after the experiments' respective maintenance periods. In the case of GERDA, however, there will be a brief period of Phase II data-taking in which Phase-II-only sensitivity will not have caught up to Phase I sensitivity; here, the boundaries of the sensitivity band obtained by combining the data sets is indicated with a dotted line. As the only change in experimental parameters between the first and second stages of the Majorana demonstrator is the addition of more active mass, unlike the other experiments, the curve after the kink for Majorana corresponds to all data collected from the start of the experiment.

All six NME calculations considered here provide values for ^{130}Te. The PHFB reference does not provide values for ^{76}Ge or ^{136}Xe, but F_N range for ^{130}Te that arises from the PHFB NME calculations does not determine either the maximum or minimum of the full F_N range for ^{130}Te considered here, so the bands in Figure 7.7 are all fully comparable to one another. See Section 2.2 for a brief discussion of nuclear matrix elements and a comparison of values for the nuclides considered here.

The inherent relative sensitivities of different candidate isotopes to $\langle m_{\beta\beta} \rangle$ shift slightly in relation to one another depending upon the NME calculation. QRPA-T values are used as a standard for reference by many $0\nu\beta\beta$ experiments; this calculation tends to be somewhat less favorable toward ^{136}Xe by comparison to other $0\nu\beta\beta$ candidates than the other methods. Figure 7.8 shows the sensitivity bands of the experiments, using the higher and lower SNO+ background estimates respectively, calculated only with the QRPA-T values with CCM short-range correlations of [73].

The CUORE experiment and the SNO+ experiment will both investigate the same $0\nu\beta\beta$ candidate nuclide, ^{130}Te, using two very different choices of detector apparatus. Interestingly, if both experiments are assigned the same arbitrary start date for the purposes of direct comparison, the detector parameter estimates considered here yield nearly identical sensitivity curves for the two. Despite the very large scintillator mass of the SNO+ detector, once the 20% fiducial volume cut

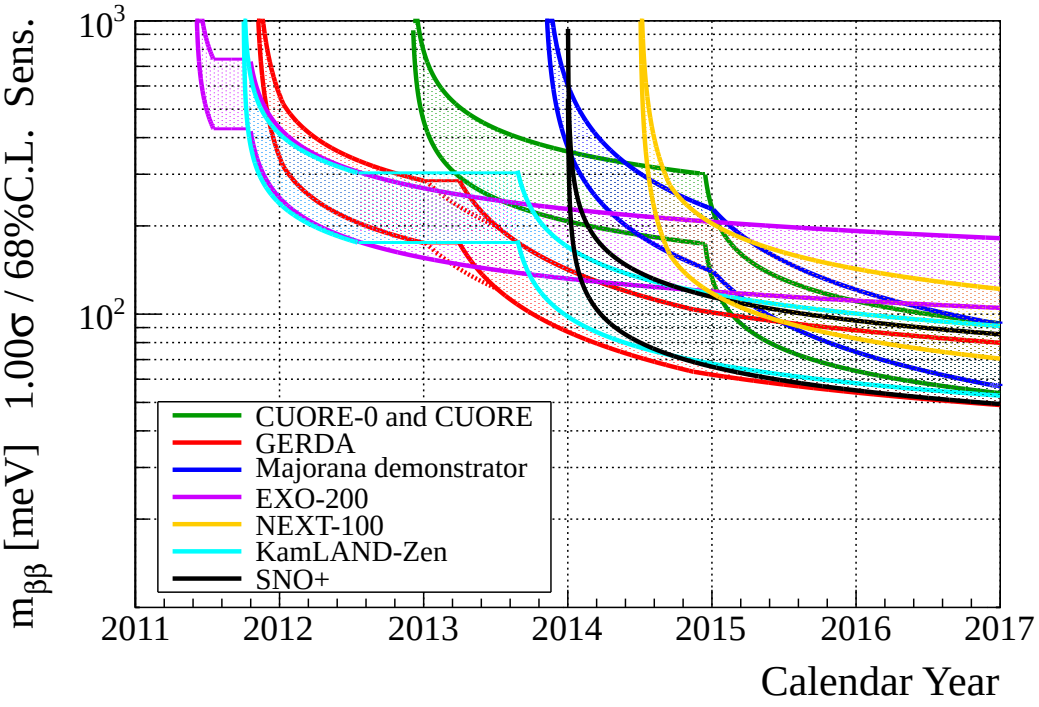


Figure 7.8: Sensitivity curves vs. calendar time for a selection of leading $0\nu\beta\beta$ experiments. Multi-stage experimental programs are represented by continuous curves with kinks when subsequent stages turn on. The dotted lines appended to the GERDA curve illustrate the boundaries of the sensitivity band arising from combining data from Phase I and Phase II during the time when it is advantageous to do so. The width of each sensitivity band arises from the QRPA-T NME range with CCM short-range correlations reported in [73].

has been applied, the source isotope masses of the two experiments are on a comparable scale. If the background rates in the two experiments are expressed in units of counts per source isotope mass instead of in units of counts per total detector mass, the background rate in SNO+ remains superior to that in CUORE by a factor of nearly 50; on the other hand, CUORE's energy resolution is far superior to that of SNO+, by a factor of a bit more than 50. The relative advantages of the two experiments therefore effectively cancel one another out, and in the end, both will explore a very similar space.

SNO+ is expected to begin real $0\nu\beta\beta$ data taking approximately one year ahead of CUORE, so it is likely that SNO+ will be the first to see indications of any ^{130}Te $0\nu\beta\beta$ signal that may eventually be measured by CUORE. However, the resolution of SNO+ will be insufficient to clearly resolve the $0\nu\beta\beta$ peak from the upper end of the $2\nu\beta\beta$ continuum; instead, SNO+ will rely upon careful modeling and fitting of the various components of the spectrum to distinguish the deformation of the $2\nu\beta\beta$ continuum caused by the $0\nu\beta\beta$ signal. Even if SNO+ observes an excess of events near the ^{130}Te Q-value, it may not be possible to unambiguously identify the source of that excess as a signal shape consistent with $0\nu\beta\beta$. In contrast, CUORE's very narrow energy resolution renders the $2\nu\beta\beta$ continuum negligible in the energy region where the $0\nu\beta\beta$ signal is expected to appear. If CUORE observes evidence of $0\nu\beta\beta$, it will appear as a clear, sharp Gaussian peak. Thus, if the first year or so of SNO+ data reveals an indication of a non-zero $0\nu\beta\beta$ rate in ^{130}Te, CUORE will be well poised to confirm that measurement with an observation of the $0\nu\beta\beta$ signal peak.

By taking into account sensitivity curves drawn with various nuclear matrix elements, it can be seen that there is a large number of $0\nu\beta\beta$ experiments, both currently operating and anticipated for the near future, all investigating a very similar range of $\langle m_{\beta\beta}\rangle$ space within the next few years. This is an excellent situation for the neutrino physics community. Experiments investigating the same candidate isotope can provide nearly immediate checks on one another's results, and in the exciting event that any experiment observes evidence of neutrinoless double-beta decay, it can be expected that other experiments should shortly follow. Observation of $0\nu\beta\beta$ in a number of different isotopes will help to refine nuclear matrix element calculations and narrow the theoretical

uncertainty on the determination of $\langle m_{\beta\beta} \rangle$, not to mention providing confirmation of the discovery of the Majorana nature of neutrinos in several detectors with very different systematics. Even if the discovery of $0\nu\beta\beta$ is not within the reach of the current generation of detectors, this rich, broad experimental program represents a healthy store of knowledge and expertise that can propel the search for $0\nu\beta\beta$ into the next generation of experiments and beyond.

Chapter 8

Conclusion

In 1930, Pauli proposed the existence of a particle that had never been observed to resolve a conundrum that puzzled the physics community. More than eighty years later, neutrinos remain difficult to detect due to how rarely they interact with matter; nevertheless, measurements have unambiguously demonstrated that different flavors of neutrinos mix with one another, indicating that these particles, previously believed to be massless, must indeed have some non-zero mass and raising the question of how this mass can be characterized. The mechanism of neutrino mass is not yet understood, but the modifications to the theoretical framework of the Standard Model that allow it to accommodate neutrino mass hint that neutrinos may hold the key to not only the theory of beta decay, for which they were originally proposed, but to the very matter composition of the universe.

Although neutrinos are now known to have mass, the nature and absolute scale of that mass have not yet been determined. This is crucial information to acquire before the neutrino's true place in the history of the universe can be understood and calculated. While the neutrino was instrumental in understanding beta decay, the nuclear process that was the focus of investigations that laid the foundation for the theory of the weak force, the related second-order weak process of double-beta decay offers the most promising avenue for reaching a greater understanding of the neutrino. The observation of the lepton-number-violating neutrinoless double-beta decay in which two electrons (positrons) are ejected from a nucleus with no accompanying antineutrinos (neutrinos) would establish that the neutrino is a Majorana particle, that is, its own antiparticle; a measurement of the rate of this process would also provide some constraints on the overall neutrino mass scale. The search for $0\nu\beta\beta$ is presently the only experimentally feasible method of

attempting to establish the Majorana nature of the neutrino, which would have implications for the understanding of the mechanism by which the neutrino gains its mass. Although the neutrino is the only known fermion with the potential to be a Majorana particle because it carries no strictly conserved charge, such as electrical or color charge, the discovery of a Majorana mass mechanism for the neutrino could help to guide the search for possible exotic particles that could gain their mass from the same mechanism.

Although feasible, the search for $0\nu\beta\beta$ is nevertheless challenging. The decay is very rare, if it occurs at all; the half-lives for the decay in various nuclides are expected to be $10^{25}-10^{28}$ y for an effective double-beta mass of 50 meV. It is possible that the effective double-beta mass may be much smaller (leading to much longer half-lives), especially if the neutrino mass hierarchy follows the 'normal' hierarchy pattern, in which case the effective double-beta mass may vanish even if neutrinos are indeed Majorana particles. In contrast, the longest $2\nu\beta\beta$ half-life ever measured is that of ^{128}Te, only 7×10^{24} y [48], and this measurement was accomplished only geochemically; the longest half-life that has been measured directly is the 2×10^{21}-y half-life of ^{136}Xe [9, 83]. Experiments require large source masses, long measuring times, excellent energy resolutions, and stringent control of backgrounds to achieve sensitivities that can probe thus-far-unexplored regions of the $0\nu\beta\beta$ parameter space. A number of experimental groups are pursuing various detector technologies to achieve these requirements; many adopt the approach in which the source is embedded in the detector to take advantage of the high efficiencies and good energy resolutions typically offered by such a configuration, but this places additional restrictions on the types of detectors that are suitable for the investigation of particular nuclides. While consistent observation of $0\nu\beta\beta$ in several nuclides with different nuclear structures and systematic concerns will be necessary to confirm any discovery of the decay, the comparison of experimental results for different nuclides is difficult due to theoretical uncertainties in the calculation of the nuclear matrix elements that describe the effects of nuclear structure on the rate of the decay; such calculations can differ by factors of $2-3$ for the same nuclide.

One claim of observation of $0\nu\beta\beta$ in ^{76}Ge has been reported for a dataset collected with germanium diodes. There is some skepticism regarding the strength of the claim within the physics

community, due to dissatisfaction with some elements of the analysis applied to the data and doubts with respect to the origins of some of the other peaks observed in the spectrum, but no experiment to date has fully excluded the claim. One of the major goals of the present generation of $0\nu\beta\beta$ experiments is to operate detectors capable of directly confirming or directly refuting this claim. Two ^{136}Xe experiments have recently achieved limits that exclude at 90% C.L. a majority of the parameter space implied by the claim, but it is not yet fully excluded, and in any case, there exists a possibility of some as-yet-unknown effect that could cause $0\nu\beta\beta$ to manifest in ^{76}Ge in a manner consistent with the claim while still not being present in ^{136}Xe. Only another germanium experiment will be able to directly refute the observed signal in ^{76}Ge. By the same token, however, only an observed signal in a different nuclide will constitute an unambiguous confirmation of the $0\nu\beta\beta$ claim, as a signal in another germanium experiment could possibly be attributable to some as-yet-unknown process unique to the setup of the germanium detectors instead of $0\nu\beta\beta$. Phase I of the GERDA experiment is the earliest germanium experiment expected to begin to see evidence of a signal if the claim is accurate — using the sensitivity approach and parameter values presented in Chapter 7, GERDA Phase I has a 1σ half-life sensitivity of 2.6×10^{25} y for one year of live time, implying at least a 1σ sensitivity to the majority of the claimed $2.23^{+0.44}_{-0.31} \times 10^{25}$ y half-life range — but GERDA will not be capable of fully excluding the claim until Phase II. Five of the Phase II detectors have already been deployed, and Phase II data taking is expected to begin this year.

The CUORE family of $0\nu\beta\beta$ detectors represents an experimental program with two decades of experience operating TeO_2 crystals as bolometric detectors searching for the neutrinoless double-beta decay of ^{130}Te. Bolometric detectors can be built from $0\nu\beta\beta$ candidate materials that are not suitable for other event detection methods, as long as these materials can be grown into high-purity dielectric and diamagnetic crystals to ensure that they have a very small heat capacity under cryogenic conditions, and they offer excellent energy resolution for the separation of the $0\nu\beta\beta$ peak from the $2\nu\beta\beta$ continuum as well as mass scalability. As a candidate nuclide, ^{130}Te is attractive due to its reasonably high Q-value and to its considerable natural abundance, the highest of all nuclides commonly considered for experimental double-beta searches. Following many years of operating small arrays of crystals, Cuoricino was the first large-scale implementation of the TeO_2

bolometer technology with a total crystal mass of 40.7 kg. It collected data from 2003 until 2008, measuring a total exposure of 19.75 kg y of ^{130}Te to set a lower limit of 2.8×10^{24} y (90% C.L.) on the $0\nu\beta\beta$ half-life in ^{130}Te, the most stringent limit to date on $0\nu\beta\beta$ in that isotope. In the process, the collaboration gained considerable experience regarding detector design, background suppression, and analysis techniques that can now be applied to the ton-scale experiment CUORE, which is presently under construction. CUORE will be hosted in the largest cryostat of its kind ever built, and it is not expected to be feasible to construct a bolometric array of larger physical dimensions. However, CUORE itself will provide valuable experience operating a detector on this scale, and options are already under evaluation that will allow the next stage of evolution of the technology within these physical confines, including isotopic enrichment of the crystal material, surface-sensitive bolometers based on panels of differing heat capacity, and the simultaneous collection of either scintillation or Cherenkov light (depending on the characteristics of the particular crystal). Thus, CUORE is the present state of the art in an ever-evolving experimental program with a long history of success that is anticipated to continue well into the future, supported by the knowledge and expertise that will be gained from the design, construction, and operation of CUORE itself.

While techniques such as pulse-shape analysis and coincidence analysis can be applied to close-packed arrays of CUORE-like bolometers to reject some non-$0\nu\beta\beta$ events such as those caused by electronic noise and penetrating environmental particles, respectively, the only real identifying information acquired for each event is its energy. CUORE-family experiments rely upon their excellent energy resolution to allow the $0\nu\beta\beta$ peak to be distinguished from the background in the region of interest near the Q-value. For this reason, it is crucial that the true energy of each event is precisely determined by calibrating the voltage signals obtained from the detectors to gamma lines of known energies from a radioactive source. Reliable energy calibration involves two major tasks: first, the design and implementation of calibration hardware suitable to deliver the calibration sources to the detectors in a manner that will allow the collection of quality calibration data without compromising the operation of the detectors, and second, analysis techniques for

fitting the calibration spectra to obtain voltage-to-energy conversions to be applied to the $0\nu\beta\beta$-search spectra. Because no cryostat can achieve perfectly stable operating conditions, calibrations must be performed periodically throughout the lifetime of the experiment to confirm the detector responses over time, so the calibration hardware must be suitable to operate for multiple calibration cycles over a number of years and the analysis should ideally be automated as much as possible to minimize the effort required to repeatedly calibrate each of 988 individual channels. Like Cuoricino before it, CUORE will be calibrated using the decay chain of ^{232}Th because it is well suited to both the hardware and analysis requirements for the experiment: thoriated tungsten wire can be used as solid source material to be loaded into a calibration source delivery system; the 10^{10} y half-life of the isotope means that the calibration sources should never need to be replaced during the lifetime of the experiment; and the decay chain produces a number of strong gamma peaks between 511 keV and 2615 keV, including a very strong peak near the ^{130}Te $\beta\beta$ decay Q-value, providing a number of calibration points to ensure that the calibration function fit is well constrained in the gamma region of the energy spectrum and a strong anchor near the energy at which a good understanding of the detector calibration is most crucial.

The calibration hardware requirements in a CUORE-scale experiment pose a much greater challenge than in any previous iteration of bolometric TeO_2 experiments. Previously, all of the detector crystals faced the outer edge of the cryostat; it was therefore possible to evenly illuminate the detectors with sources placed symmetrically outside the outer cryostat shield, so long as the source activity was sufficiently high to penetrate the shields and reach the detectors. In a multi-tower crystal array such as CUORE, however, the inner towers will be shielded by the outer towers; it will therefore be necessary to insert calibration sources into the interior of the array to achieve the most even possible illumination of all crystals in order to calibrate every detector in as little time as possible without exceeding the maximum safe event rate on any detector, which could cause signal-spoiling pileup or even an overall increase in operating temperature. To address these concerns, a calibration system has been designed for CUORE based on a source carrier comprising a continuous, flexible string to which individual capsules housing radioactive source material are attached. The distribution of mass along the active length of the source carrier allows

it to travel down into the cryostat under its own weight when unspooled into a system of guide tubes that defines its complicated route through the upper stages of the cryostat into the detector region; with only one attachment point above the 300-K flange of the cryostat, it can be fully retracted from the cryostat during normal data-taking, allowing the minimal possible disruption of the low-background environment of the cryostat. Extensive motion tests have shown that with careful quality control of the components and manufacture of each source carrier, the computer-controlled string-and-guide-tube calibration system developed for CUORE provides reliable and well-characterized motion for the delivery of the source to the detector region that should allow it to be operated with confidence in the sealed environment of the cryostat.

With the sources inside the cryostat shields, the physical space available differs considerably from the situation in which the sources are placed outside the cryostat, with implications for the optimum activity that should be loaded into the source carriers. The source attenuation due to shielding is considerably lower, so the source activities must be accordingly lower to avoid causing excessively high event rates in the detectors; the CUORE source carrier design is in fact most appropriate for low source activities, constructed as it is of small source capsules placed at some distance from one another along a string. The vertical space in which the source strings can extend above and below the height of the crystal towers is restricted by the interior height of the cryostat vessels, creating a solid-angle effect that results in lower event rates in detectors at the tops and bottoms of the towers when the source activity distribution is constant along the active length of the source carrier. The CUORE source carrier design allows this effect to be easily compensated by loading additional source material into the capsules at the top and bottom of each source string. High-rate calibration tests were performed on CUORE-like detectors to determine the maximum safe event rate that can be tolerated on the detectors before the calibration performance of the analysis software deteriorates, which was found to be 150 mHz; when compared with the results of these tests and calibration performances in Cuoricino, simulations of the final activity distribution defined for the CUORE source strings in the real cryostat geometry demonstrate that all CUORE detectors can be successfully calibrated within two days of calibration data-taking.

Cuoricino produced a high-statistics collection of data taken with TeO_2 bolometers in the same manner as will be done for CUORE and also processed with the same analysis tools, the Diana software suite. Cuoricino data thus provides an opportunity to evaluate the performance of the calibration software on a large number of calibrations over a long data-taking period and refine its behavior for future application to CUORE. A number of known gamma lines of varying strengths appear in the calibrated energy spectra collected by Cuoricino detectors; the effectiveness of the calibration procedure can be evaluated by constructing the calibrated sum spectrum of all channels and data sets that fit given criteria of interest and then fitting these gamma lines to determine their resolutions and their deviations from their nominal energies. Both throughout the gamma region and within only the $0\nu\beta\beta$ region of interest, this kind of residual analysis provides strong indications that, of the two calibration functions used to process Cuoricino data, a third-order polynomial performs better in the gamma region than a second-order log polynomial. A study of the peak-finding performance of the interpolation function used to search for the weaker secondary calibration peaks after identifying the stronger primary calibration peaks was less conclusive, showing only a mild preference for the third-order polynomial, though it also indicated that secondary-peak-finding performance could be improved independently of the interpolation function by tightening the constraints on one of the fit parameters of the 1588-keV/1593-keV secondary calibration peak. Based on the superior performance of the third-order polynomial as the calibration function for the Cuoricino detectors and on physical arguments that an energy deposition of zero should give rise to a pulse of zero amplitude, the calibration function intended for use in CUORE is a third-order polynomial with the intercept fixed to zero; this conclusion is also supported by calibrations performed in a study focused on the low-energy portion of the spectrum [17]. The third-order polynomial with intercept fixed to zero appears to be the best description of the response of the detectors, so because there is no clear indication that a different functional form exhibits superior performance when used to interpolate from the primary calibration peaks to find the secondary calibration peaks during automated calibration, we recommend that the third-order polynomial with intercept fixed

to zero be used for interpolation as well as for calibration; there are some indications that this function may in fact show better interpolation performance than either of the interpolation functions that were directly investigated.

When the three peaks in the $0\nu\beta\beta$ region of interest are used to estimate the calibration uncertainy at the Q-value, the uncertainty associated with the final Cuoricino result is ± 0.4 keV; using only the subset of Cuoricino data that is most comparable to the anticipated CUORE data, this uncertainty already improves to ± 0.2 keV even with no additional improvements to the calibration procedure. Cuoricino calibration data is not directly comparable to CUORE due to the placement of the calibration sources; with the sources inside the shields, simulations indicate that the 510.77-keV gamma from ^{208}Tl no longer contributes negligibly in comparison to the 511-keV annihilation gamma, as it did in the Cuoricino cryostat, meaning that the nominal energy assigned to the peak should be shifted. To first order, it appears that it may be appropriate to simply assign the mean of the two gamma energies to the peak, but the relative contribution of the two gammas is somewhat dependent upon the location of the crystal in the detector array; once CUORE has collected calibration data, an empirical calibration study may be necessary to determine whether the calibration software should be modified to allow channel-dependent modifications to calibration peak fit parameters. However, simulations also indicate that the nearby 583.191-keV line, also from ^{208}Tl, may be able to be treated as a primary peak instead of a secondary peak in the CUORE cryostat, which should alleviate the influence of the increased uncertainty in the position of the 511 peak. A calibration uncertainty of ± 0.2 keV, or possibly better, should therefore be achievable in CUORE. While this is on a similar scale to the errors achieved on two recent measurements of the ^{130}Te Q-value [126, 118], it is more than an order of magnitude larger than the error on another recent measurement [120], meaning that the energy-scale uncertainty in the analysis of CUORE data will likely be dominated by the calibration uncertainty. If CUORE observes a flat spectrum in the $0\nu\beta\beta$ region of interest and sets a limit on the half-life of the decay, the energy-scale uncertainty will not be a critical factor in the experimental results; however, if a peak structure appears in the CUORE spectrum, a well-understood energy calibration uncertainty will be necessary for the

reliable identification of the peak as consistent with a $0\nu\beta\beta$ signal and for the accurate evaluation of the magnitude of such a signal.

While CUORE strives to measure a precise energy spectrum and controls backgrounds mostly through passive methods such as shielding and stringent control of contaminations in detector materials, some other $0\nu\beta\beta$ experiments pursue detector technologies that prioritize the various factors that affect sensitivity to $0\nu\beta\beta$ somewhat differently. A simple sensitivity calculation can be performed by comparing the number of signal and background counts expected to be observed in a particular detector configuration within an energy window of a width equal to the detector's resolution, centered on the Q-value of the isotope under study. When the sensitivity of CUORE is considered in this way, it can be seen that CUORE expects to see a signal with a significance of at least 5σ above background within about eight months of live time (i.e., less than a calendar year of operating time) if $0\nu\beta\beta$ occurs in ^{130}Te at a rate consistent with the range implied by the Klapdor-Kleingrothaus claim of observation in ^{76}Ge; the 1σ half-life sensitivity of CUORE for its anticipated 5-year total live time is 1.6×10^{26} y, corresponding to a range of $42-100$ meV in the effective neutrino mass $\langle m_{\beta\beta} \rangle$. This simple sensitivity approach allows the relative physics reach of highly varied detector designs to be compared on the basis of the performance they can achieve in terms of various simple detector parameters such as mass, resolution, isotope choice, and background rate. When such a comparison is performed for many of the leading present $0\nu\beta\beta$ experiments and those that will begin to take data within the next few years, it is evident that there is currently a rich experimental program across many $0\nu\beta\beta$ candidate nuclides, with many experiments all probing a very similar space. The Klapdor-Kleingrothaus claim will be thoroughly tested within the next few years; following the method discussed in Section 7.3.2 and using the experimental parameters considered in Section 7.3.4, GERDA and SNO+ are expected to confirm or exclude it at a level of 3σ by early 2014 at the latest. Depending on the true values of the nuclear matrix elements, the $\langle m_{\beta\beta} \rangle$ range corresponding to the inverted neutrino mass hierarchy may also begin to come within reach.

This is an exciting time for neutrino physics. With so many experiments of similar capabilities, any discovery of $0\nu\beta\beta$ is likely to be quickly followed by confirmation in one or more other nuclei,

providing a wealth of information leading the way toward a greater understanding of new physics beyond the Standard Model. Even if the current generation of experiments is not fortunate enough to reach a discovery, the present $0\nu\beta\beta$ program represents an extensive community of physicists developing the expertise on designing and operating large-scale, low-background experiments that will allow the evolution of present detectors to continue to pursue the search for new frontiers of rare physics well into the future.

BIBLIOGRAPHY

[1] About SNO+: Neutrinoless Double Beta Decay. SNO+ Experiment webpage. http://snoplus.phy.queensu.ca/Double_Beta.html. Accessed 18 Apr 2013.

[2] Advanced Spectra Processing: Documentation of TSpectrum class. ROOT Reference Guide on ROOT webpage. http://root.cern.ch/root/html534/TSpectrum.html. Accessed 24 Apr 2013.

[3] Press Release: The 1979 Nobel Prize in Physics. Nobelprize.org. http://www.nobelprize.org/nobel_prizes/physics/laureates/1979/press.html. Accessed 22 Mar 2013.

[4] C. Aalseth et al. Neutrinoless double beta decay and direct searches for neutrino mass. Full text of the report of the working group for the APS Multidivisional Study on the Future of Neutrino Physics. arXiv:hep-ph/0412300, Dec 2004.

[5] C. E. Aalseth et al. Comment on 'Evidence for neutrinoless double beta decay'. *Mod. Phys. Lett. A*, 17(22):1475–1478, Jul 2002.

[6] C. E. Aalseth et al. (IGEX Collaboration). IGEX ^{76}Ge neutrinoless double-beta decay experiment: Prospects for next generation experiments. *Phys. Rev. D*, 65(9):092007, May 2002.

[7] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Courier Dover Publications, 1972.

[8] I. Abt et al. Background reduction in neutrinoless double beta decay experiments using segmented detectors — A Monte Carlo study for the GERDA setup. *Nucl. Instrum. Meth. A*, 570(3):479–486, Jan 2007.

[9] N. Ackerman et al. (EXO Collaboration). Observation of Two-Neutrino Double-Beta Decay in ^{136}Xe with the EXO-200 Detector. *Phys. Rev. Lett.*, 107(21):212501, Nov 2011.

[10] P. A. R. Ade et al. (Planck Collaboration). Planck 2013 results. XVI. Cosmological parameters. Submitted to Astronomy & Astrophysics. arXiv:1303.5076 [astro-ph.CO], Mar 2013.

[11] M. Agostini et al. (GERDA Collaboration). Measurement of the half-life of the two-neutrino double beta decay of ^{76}Ge with the GERDA experiment. *J. Phys. G: Nucl. Part. Phys.*, 40(3):035110, Mar 2013.

[12] A. Alessandrello et al. A search for neutrinoless double beta decay of ^{130}Te with a thermal detector. *Phys. Lett. B*, 285(1-2):176–182, Jul 1992.

[13] A. Alessandrello et al. First tests on a large mass, low temperature array detector. *Nucl. Instrum. Meth. A*, 360(1-2):363–366, Jun 1995. Proceedings of the 6th Pisa Meeting on Advanced Detectors, 22–28 May 1994, La Biodola, Italy.

[14] A. Alessandrello et al. Preliminary results on double beta decay of ^{130}Te with an array of twenty cryogenic detectors. *Phys. Lett. B*, 433(1-2):156–162, Aug 1998.

[15] A. Alessandrello et al. The first step toward CUORE: Cuoricino, a thermal detector array to search for rare events. *Nucl. Phys. B Proc. Suppl.*, 87(1-3):78–80, Jun 2000. Proceedings of the 6th International Workshop on Topics in Astroparticle and Underground Physics, 6–10 Sep 1999, Paris, France.

[16] F. Alessandria et al. (CUORE Collaboration). CUORE crystal validation runs: results on radioactive contamination and extrapolation to CUORE background. *Astropart. Phys.*, 35(12):839–849, Jul 2012.

[17] F. Alessandria et al. (CUORE Collaboration). The low energy spectrum of TeO_2 bolometers: results and dark matter perspectives for the CUORE-0 and CUORE experiments. *J. Cosmol. Astropart. Phys.*, 2013(01):038, Jan 2013.

[18] F. Alessandria et al. (CUORE Collaboration). Validation of techniques to mitigate copper surface contamination in CUORE. *Astropart. Phys.*, 2013. Article in press, doi:10.1016/j.astropartphys.2013.02.005.

[19] V. Álvarez et al. (NEXT Collaboration). Initial results of NEXT-DEMO, a large-scale prototype of the NEXT-100 experiment. *J. Instrum.*, 8(04):P04002, Apr 2013.

[20] E. Ambler, R. W. Hayward, D. D. Hoppes, R. P. Hudson, and C. S. Wu. Further Experiments on β Decay of Polarized Nuclei. *Phys. Rev.*, 106(6):1361–1363, Jun 1957.

[21] E. Andreotti et al. (CUORE Collaboration). ^{130}Te Neutrinoless Double-Beta Decay with CUORICINO. *Astropart. Phys.*, 34(11):822–831, Jun 2011.

[22] E. Andreotti et al. (CUORICINO Collaboration). Muon-induced backgrounds in the CUORICINO experiment. *Astropart. Phys.*, 34(1):18–24, Aug 2010.

[23] R. Ardito et al. (CUORE Collaboration). CUORE: A cryogenic underground observatory for rare events. CUORE experiment proposal. arXiv:hep-ex/0501010, Jan 2005.

[24] C. Arnaboldi et al. A calorimetric search on double beta decay of ^{130}Te. *Phys. Lett. B*, 557(3-4):167–175, Apr 2003.

[25] C. Arnaboldi et al. First results on neutrinoless double beta decay of ^{130}Te with the calorimetric CUORICINO experiment. *Phys. Lett. B*, 584(3-4):260–268, Apr 2004.

[26] C. Arnaboldi et al. New Limit on the Neutrinoless $\beta\beta$ Decay of ^{130}Te. *Phys. Rev. Lett.*, 95(14):142501, Sep 2005.

[27] C. Arnaboldi et al. Production of high purity TeO_2 single crystals for the study of neutrinoless double beta decay. *J. Cryst. Growth*, 312(20):2999–3008, Oct 2010.

[28] C. Arnaboldi et al. (CUORE Collaboration). Physics potential and prospects for the CUORICINO and CUORE experiments. *Astropart. Phys.*, 20(2):91–110, Nov 2003.

[29] C. Arnaboldi et al. (CUORE Collaboration). CUORE: A cryogenic underground observatory for rare events. *Nucl. Instrum. Meth. A*, 518(3):775–798, Feb 2004.

[30] C. Arnaboldi et al. (CUORICINO Collaboration). Results from a search for the $0\nu\beta\beta-$decay of ^{130}Te. *Phys. Rev. C*, 78(3):035502, Sep 2008.

[31] R. Arnold et al. First Results of the Search for Neutrinoless Double-Beta Decay with the NEMO 3 Detector. *Phys. Rev. Lett.*, 95(18):182302, Oct 2005.

[32] V. N. Aseev et al. Upper limit on the electron antineutrino mass from the Troitsk experiment. *Phys. Rev. D*, 84(11):112003, Dec 2011.

[33] G. Audi, O. Bersillon, J. Blachot, and A. H. Wapstra. The NUBASE evaluation of nuclear and decay properties. *Nucl. Phys. A*, 729(1):3–128, Dec 2003. *The 2003* NUBASE *and Atomic Mass Evaluations*.

[34] M. Auger et al. (EXO Collaboration). Search for Neutrinoless Double-Beta Decay in ^{136}Xe with EXO-200. *Phys. Rev. Lett.*, 109(3):032505, Jul 2012.

[35] F. T. Avignone, S. R. Elliott, and J. Engel. Double beta decay, Majorana neutrinos, and neutrino mass. *Rev. Mod. Phys.*, 80(2):481–516, Apr 2008.

[36] A. M. Bakalyarov, A. Ya. Balysh, S. T. Belyaev, V. I. Lebedev, and S. V. Zhukov. Results of the experiment on investigation of ^{76}Ge double beta decay. *Phys. Part. Nucl. Lett.*, 2:77–81, 2005.

[37] A. B. Balantekin and W. C. Haxton. Neutrino Oscillations. arXiv:1303.2272 [nucl-th], Mar 2013.

[38] A. S. Barabash. Double-beta decay: Present status. *Phys. Atom. Nucl.*, 73(1):162–178, Jan 2010.

[39] A. Barabash (for the SuperNEMO Collaboration). SuperNEMO double beta decay experiment. *J. Phys. Conf. Ser.*, 375(4):042012, Jul 2012. Proceedings of the 12th International Conference on Topics in Astroparticle and Underground Physics, 5–9 Sep 2011, Munich, Germany.

[40] P. S. Barbeau, J. I. Collar, and O. Tench. Large-mass ultralow noise germanium detectors: performance and applications in neutrino and astroparticle physics. *J. Cosm. Astro. Phys.*, 2007(09):009, Sep 2007.

[41] M. Bardon, P. Franzini, and J. Lee. Helicity of μ^- Mesons; Mott Scattering of Polarized Muons. *Phys. Rev. Lett.*, 7(1):23–25, Jul 1961.

[42] J. Barea and F. Iachello. Neutrinoless double-β decay in the microscopic interacting boson model. *Phys. Rev. C*, 79(4):044301, Apr 2009.

[43] J. Barea, J. Kotila, and F. Iachello. Nuclear matrix elements for double-β decay. *Phys. Rev. C*, 87(1):014315, Jan 2013.

[44] R. Becker-Szendy et al. Electron- and muon-neutrino content of the atmospheric flux. *Phys. Rev. D*, 46(9):3720–3724, Nov 1992.

[45] F. Bellini et al. CUORICINO first-level data analysis and official n-tuple production. CUORE Internal Note 001, Version 2, May 2010.

[46] M. Berglund and M. E. Wieser. Isotopic compositions of the elements 2009 (IUPAC Technical Report). *Pure Appl. Chem.*, 83(2):397–410, 2011.

[47] J. Bergström. Combining and comparing neutrinoless double beta decay experiments using different nuclei. *J. High Energy Phys.*, 2013(2):93, Feb 2013.

[48] J. Beringer et al. (Particle Data Group). The Review of Particle Physics. *Phys. Rev. D*, 86:010001, 2012.

[49] R. Bernabei et al. Investigation of $\beta\beta$ decay modes in ^{134}Xe and ^{136}Xe. *Phys. Lett. B*, 546(1-2):23–28, Oct 2002.

[50] S. I. Bityukov and N. V. Krasnikov. Uncertainties and discovery potential in planned experiments. Talk presented by S. I. Bityukov at Advanced Statistical Techniques in Particle Physics, Mar 18–22, 2002, Durham, UK. arXiv:hep-ph/0204326, Apr 2002.

[51] L. M. Brown. The idea of the neutrino. *Phys. Today*, 31(9):23–28, Sep 1978.

[52] C. Bucci et al. Background study and Monte Carlo simulations for large-mass bolometers. *Eur. Phys. J. A*, 41(2):155–168, Aug 2009.

[53] A. Caldwell and K. Kröninger. Signal discovery in sparse spectra: A Bayesian analysis. *Phys. Rev. D*, 74(9):092003, Nov 2006.

[54] J. Clark and L. Ejzak. CUORE DCS Source String Technical Design and Production Summary. UW-CUORE internal note, Oct 2012.

[55] B. T. Cleveland et al. Measurement of the Solar Electron Neutrino Flux with the Homestake Chlorine Detector. *Astrophys. J.*, 496(1):505, Mar 1998.

[56] CUORE Collaboration. The frontier of neutrinoless double beta decay with bolometers. To be submitted for publication, 2013.

[57] C. L. Cowan Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the Free Neutrino: A Confirmation. *Science*, 124(3212):103–104, Jul 1956.

[58] S. T. Cowell. Scaling factor inconsistencies in neutrinoless double beta decay. *Phys. Rev. C*, 73(2):028501, Feb 2006.

[59] G. Danby et al. Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos. *Phys. Rev. Lett.*, 9(1):36–44, Jul 1962.

[60] Z. Daraktchieva (for the NEMO Collaboration). Search for neutrinoless double beta decay with NEMO 3 experiment. *Nucl. Phys. A*, 827(1-4):495c–497c, Aug 2009. Proceedings of the 18th Particles and Nuclei International Conference, 9–14 Nov 2008, Eilat, Israel.

[61] S. Davidson, E. Nardi, and Y. Nir. Leptogenesis. *Phys. Rept.*, 466(4-5):105–177, Sep 2008.

[62] A. de Gouvêa, B. Kayser, and R. N. Mohapatra. Manifest *CP* violation from Majorana phases. *Phys. Rev. D*, 67(5):053004, Mar 2003.

[63] R. de Putter et al. New Neutrino Mass Bounds from Sloan Digital Sky Survey III Data Release 8 Photometric Luminous Galaxies. *Astrophys. J.*, 761(1):12, Dec 2012.

[64] J. Detwiler. Private communication, 2 Dec 2011.

[65] M. Dine and A. Kusenko. The Origin of the Matter - Antimatter Asymmetry. *Rev. Mod. Phys.*, 76(1):1–30, Dec 2003.

[66] G. Drexlin, V. Hannen, S. Mertens, and C. Weinheimer. Current Direct Neutrino Mass Experiments. *Adv. High Energy Phys.*, 2013:293986, 2013.

[67] DuPont (E. I. du Pont de Nemours and Company). *KEVLAR Aramid Fiber Technical Guide*. http://www2.dupont.com/Kevlar/en_US/assets/downloads/KEVLAR_Technical_Guide.pdf. Accessed 8 Mar 2013.

[68] L. P. Ekström and R. B. Firestone. *WWW Table of Radioactive Isotopes*. Database version 2/28/99 from URL http://ie.lbl.gov/toi/index.htm. Accessed 8 Apr 2013.

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

[70] S. R. Elliott and P. Vogel. Double beta decay. *Ann. Rev. Nucl. Part. Sci.*, 52:115–151, Dec 2002.

[71] A. Faessler et al. Quasiparticle random phase approximation uncertainties and their correlations in the analysis of $0\nu\beta\beta$ decay. *Phys. Rev. D*, 79(5):053001, Mar 2009.

[72] A. Faessler et al. Addendum to: QRPA uncertainties and their correlations in the analysis of neutrinoless double beta decay. *Phys. Rev. D*, 87(5):053002, Mar 2013.

[73] A. Faessler, V. Rodin, and F. Šimkovic. Nuclear matrix elements for neutrinoless double-beta decay and double-electron capture. *J. Phys. G: Nucl. Part. Phys.*, 39(12):124006, Dec 2012.

[74] M. A. Fehr, M. Rehkämper, and A. N. Halliday. Application of MC-ICPMS to the precise determination of tellurium isotope compositions in chondrites, iron meteorites and sulfides. *Int. J. Mass Spectrom.*, 232(1):83–94, Mar 2004.

[75] F. Feruglio, A. Strumia, and F. Vissani. Neutrino oscillations and signals in β and $0\nu2\beta$ experiments. *Nucl. Phys. B*, 637(1-3):345–377, Aug 2002.

[76] P. Finnerty et al. (MAJORANA Collaboration). The MAJORANA DEMONSTRATOR: Progress towards showing the feasibility of a tonne-scale ^{76}Ge neutrinoless double-beta decay experiment. Proceedings of the 18th International Symposium on Particles, Strings and Cosmology, 3–8 Jun 2012, Merida, Mexico. To be published in IOP Conference Series. arXiv:1210.2678 [nucl-ex], 2012.

[77] E. Fiorini and T.O. Niinikoski. Low-temperature calorimetry for rare decays. *Nucl. Instrum. Meth.*, 224(1-2):83–88, Jul 1984.

[78] L. Foggetta et al. Composite macro-bolometers for the rejection of surface radioactive background in rare-event experiments. *Astropart. Phys.*, 34(11):809–821, Jun 2011.

[79] L. Foggetta, A. Giuliani, C. Nones, M. Pedretti, and S. Sangiorgio. Surface-sensitive macrobolometers for the identification of external charged particles. *Appl. Phys. Lett.*, 86(13):134106, Mar 2005.

[80] G. L. Fogli et al. Observables sensitive to absolute neutrino masses. II. *Phys. Rev. D*, 78(3):033010, Aug 2008.

[81] D. V. Forero, M. Tórtola, and J. W. F. Valle. Global status of neutrino oscillation parameters after Neutrino-2012. *Phys. Rev. D*, 86(7):073012, Oct 2012.

[82] A. Gando et al. (KamLAND-Zen Collaboration). Limits on Majoron-emitting double-beta decays of ^{136}Xe in the KamLAND-Zen experiment. *Phys. Rev. C*, 86(2):021601, Aug 2012.

[83] A. Gando et al. (KamLAND-Zen Collaboration). Measurement of the double-β decay half-life of ^{136}Xe with the KamLAND-Zen experiment. *Phys. Rev. C*, 85(4):045504, Apr 2012.

[84] M. Goeppert-Mayer. Double Beta-Disintegration. *Phys. Rev.*, 48(6):512–516, Sep 1935.

[85] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of Neutrinos. *Phys. Rev.*, 109(3):1015–1017, Feb 1958.

[86] J. J. Gómez-Cadenas et al. Sense and sensitivity of double beta decay experiments. *J. Cosmol. Astropart. Phys.*, 2011(06):007, Jun 2011.

[87] J. J. Gómez-Cadenas, J. Martín-Albo, and F. Monrabal. NEXT, high-pressure xenon gas experiments for ultimate sensitivity to Majorana neutrinos. *J. Instrum.*, 7(11):C11007, Nov 2012. Proceedings of the 14th International Workshop on Radiation Imaging Detectors, 20–25 Jun 2012, Figueira da Foz, Portugal.

[88] P. Grabmayr. Results from GERDA. Talk presented at the XXV International Conference on Neutrino Physics and Astrophysics, 3–9 Jun 2012, Kyoto, Japan.

[89] S. Grullon. Liquid Scintillator Light Yield Measurements for the SNO+ Experiment. Talk presented at the American Physical Society April Meeting 2013, 13–16 Apr 2013, Denver, Colorado, USA.

[90] T. D. Gutierrez. Distinguishing between Dirac and Majorana Neutrinos with Two-Particle Interferometry. *Phys. Rev. Lett.*, 96(12):121802, Mar 2006.

[91] E. F. Hansen and W. S. Ginell. The Conservation of Silk with Parylene-C. In *Historic Textile and Paper Materials II*, volume 410, chapter 8, pages 108–133. American Chemical Society, 1989.

[92] H. L. Harney. Reply to the comment on 'Evidence for neutrinoless double beta decay'. [Mod. Phys. Lett. A 16(2001)2409]. arXiv:hep-ph/0205293, May 2002.

[93] J. Hartnell (for the SNO+ Collaboration). Neutrinoless Double Beta Decay with SNO+. *J. Phys. Conf. Ser.*, 375(4):042015, Jul 2012. Proceedings of the 12th International Conference on Topics in Astroparticle and Underground Physics, 5–9 Sep 2011, Munich, Germany.

[94] J. Hellmig and H. V. Klapdor-Kleingrothaus. Identification of single-site events in germanium detectors by digital pulse shape analysis. *Nucl. Instrum. Meth. A*, 455(3):638–644, Dec 2000.

[95] J. J. Hernández, S. Navas, and P. Rebecchi. Discovery limits in prospective studies. *Nucl. Instrum. Meth. A*, 372(1-2):293–296, Mar 1996.

[96] K. S. Hirata et al. Observation of a small atmospheric ν_μ/ν_e ratio in Kamiokande. *Phys. Lett. B*, 280(1-2):146–152, Apr 1992.

[97] M. G. Inghram and J. H. Reynolds. Double Beta-Decay of Te^{130}. *Phys. Rev.*, 78(6):822–823, Jun 1950.

[98] K. Inoue. Results from KamLAND-Zen. Talk presented at the XXV International Conference on Neutrino Physics and Astrophysics, 3–9 Jun 2012, Kyoto, Japan.

[99] K. Inoue. Private communication, 19 Dec 2012.

[100] B. Kayser. Are neutrinos their own antiparticles? *J. Phys. Conf. Ser.*, 173(1):012013, Jun 2009. Proceedings of the Carolina International Symposium on Neutrino Physics, 15–17 May 2008, Columbia, South Carolina, USA.

[101] I. V. Kirpichnikov. Klapdor's claim for the observation of the neutrinoless double beta-decay of Ge-76. Analysis and corrections. arXiv:1006.2025 [hep-ph], 2010.

[102] H. V. Klapdor-Kleingrothaus, A. Dietz, H. L. Harney, and I. V. Krivosheina. Evidence for neutrinoless double beta decay. *Mod. Phys. Lett. A*, 16(37):2409–2420, Dec 2001.

[103] H. V. Klapdor-Kleingrothaus, A. Dietz, I. V. Krivosheina, and O. Chkvorets. Data Acquisition and Analysis of the ^{76}Ge Double Beta Experiment in Gran Sasso 1990-2003. *Nucl. Instrum. Meth. A*, 522(3):371–406, Apr 2004.

[104] H. V. Klapdor-Kleingrothaus et al. Latest Results from the Heidelberg-Moscow Double Beta Decay Experiment. *Eur. Phys. J. A*, 12(2):147–154, Oct 2001.

[105] H. V. Klapdor-Kleingrothaus and I. V. Krivosheina. The evidence for the observation of $0\nu\beta\beta$ decay: The identification of $0\nu\beta\beta$ events from the full spectra. *Mod. Phys. Lett. A*, 21(20):1547–1566, Jun 2006.

[106] M. Koga. KamLAND double beta decay experiment using ^{136}Xe. Talk presented at the 35th International Conference on High Energy Physics, 22–28 Jul 2010, Paris, France.

[107] L. Kogler. *A measurement of the 2 neutrino double beta decay rate of ^{130}Te in the CUORICINO experiment*. PhD thesis, University of California, Berkeley, Nov 2011.

[108] J. Kotila and F. Iachello. Phase space factors for double-β decay. *Phys. Rev. C*, 85(3):034316, Mar 2012.

[109] T. D. Lee and C. N. Yang. Parity Nonconservation and a Two-Component Theory of the Neutrino. *Phys. Rev.*, 105(5):1671–1675, Mar 1957.

[110] B. Majorovits. Status of the GERDA Experiment. Talk presented at the International Workshop on Double Beta Decay and Neutrinos, 14–17 Nov 2011, Osaka, Japan.

[111] R. D. McKeown and P. Vogel. Neutrino masses and oscillations: triumphs and challenges. *Phys. Rept.*, 394(6):315–356, May 2004.

[112] J. Menéndez, A. Poves, E. Caurier, and F. Nowacki. Disassembling the nuclear matrix elements of the neutrinoless $\beta\beta$ decay. *Nucl. Phys. A*, 818(3-4):139–151, Mar 2009.

[113] R. N. Mohapatra. Physics of neutrino mass. *eConf*, C040802:L011, 2004. Proceedings of the 32nd SLAC Summer Institute on Particle Physics, 2–13 Aug 2004, Menlo Park, California, USA.

[114] S. H. Moseley, J. C. Mather, and D. McCammon. Thermal detectors as x-ray spectrometers. *J. Appl. Phys.*, 56(5):1257–1262, Sep 1984.

[115] F. Piquemal. Future double beta decay experiments. Talk presented at the XXV International Conference on Neutrino Physics and Astrophysics, 3–9 Jun 2012, Kyoto, Japan.

[116] F. Pobell. *Matter and Methods at Low Temperatures*. Springer, 3rd, revised and expanded edition, 2007.

[117] H. Primakoff and S. P. Rosen. Double beta decay. *Rep. Prog. Phys.*, 22:121, 1959.

[118] S. Rahaman et al. Double-beta decay Q values of ^{116}Cd and ^{130}Te. *Phys. Lett. B*, 703(4):412–416, Sep 2011.

[119] P. K. Rath, R. Chandra, K. Chaturvedi, P. K. Raina, and J. G. Hirsch. Uncertainties in nuclear transition matrix elements for neutrinoless $\beta\beta$ decay within the projected-Hartree-Fock-Bogoliubov model. *Phys. Rev. C*, 82(6):064310, Dec 2010.

[120] M. Redshaw, B. J. Mount, E. G. Myers, and F. T. Avignone, III. Masses of ^{130}Te and ^{130}Xe and double-β-decay Q-value of ^{130}Te. *Phys. Rev. Lett.*, 102(21):212502, May 2009.

[121] F. Reines and C. L. Cowan. Detection of the Free Neutrino. *Phys. Rev.*, 92(3):830–831, Nov 1953.

[122] R. G. H. Robertson. Empirical Survey of Neutrinoless Double Beta Decay Matrix Elements. *Mod. Phys. Lett. A*, 28(8):1350021, Mar 2013.

[123] T. R. Rodríguez and G. Martínez-Pinedo. Energy Density Functional Study of Nuclear Matrix Elements for Neutrinoless $\beta\beta$ Decay. *Phys. Rev. Lett.*, 105(25):252503, Dec 2010.

[124] S. Schael et al. (ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, and SLD Electroweak and Heavy Flavour Groups). Precision electroweak measurements on the Z resonance. *Phys. Rept.*, 427(5-6):257–454, May 2006.

[125] J. Schechter and J. W. F. Valle. Neutrinoless double-β decay in $\mathrm{SU}(2) \times \mathrm{U}(1)$ theories. *Phys. Rev. D*, 25(11):2951–2954, Jun 1982.

[126] N. D. Scielzo et al. Double-β-decay Q values of ^{130}Te, ^{128}Te, and ^{120}Te. *Phys. Rev. C*, 80(2):025501, Aug 2009.

[127] R. Slansky, S. Raby, T. Goldman, and G. Garvey (as told to N. G. Cooper). The Oscillating Neutrino — An Introduction to Neutrino Masses and Mixing. *Los Alamos Science*, 25:28–63, 1997.

[128] A. Strumia and F. Vissani. Implications of neutrino data circa 2005. *Nucl. Phys. B*, 726(1-2):294–316, Oct 2005.

[129] A. Strumia and F. Vissani. Neutrino masses and mixings and... arXiv:hep-ph/0606054v3 [hep-ph], Apr 2010.

[130] J. Suhonen and O. Civitarese. Review of the properties of the $0\nu\beta^-\beta^-$ nuclear matrix elements. *J. Phys. G: Nucl. Part. Phys.*, 39(12):124005, Dec 2012.

[131] M. Vignati. Model of the response function of large mass bolometric detectors. *J. Appl. Phys.*, 108(8):084903, Oct 2010.

[132] P. Vogel. Nuclear physics aspects of double beta decay. In F. Ferroni, F. Vissani, and B. Brofferio, editors, *Measurements of Neutrino Mass*, pages 49–103. Società Italiana di Fisica, 2009. Proceedings of the International School of Physics "Enrico Fermi", Vol. 170.

[133] F. L. Wilson. Fermi's Theory of Beta Decay. *Am. J. Phys.*, 36:1150, Dec 1968. English translation of *Z. Physik*, 88:161, Feb 1934.

[134] M. Yoshimura. Neutrino pair emission from excited atoms. *Phys. Rev. D*, 75(11):113007, Jun 2007.

Appendix A: Future Improvements to the Diana Calibration Module

Investigations of the behavior of the Diana calibration module when applied to Cuoricino data have revealed a number of changes to the calibration module that can be anticipated to improve the performance of the calibration analysis. Judging by the calibration uncertainty results of Section 6.3.1, the harmful impact of the behaviors that these changes would correct was minimal in the processing of Cuoricino data, but it may be advantageous to implement them in preparation for CUORE, especially as the number of channels in CUORE will be more than an order of magnitude greater than the number of channels in Cuoricino, rendering manual cross-checking of calibration performance far less feasible.

This appendix summarizes changes that should prevent several non-optimal behaviors in the calibration module's peak-rejection algorithm (Section A.1) as well as the changes to the peak-finding process (Section A.2) and changes due to CUORE detector geometry (Section A.3) proposed in Chapter 6. Any changes in Sections A.1 and Section A.2 that are to be implemented for CUORE should preferably be thoroughly tested on CUORE-0 data in calibrations using a third-order polynomial with intercept fixed to zero as the calibration function.

A.1 Non-Optimal Handling of Peak Rejection During Calibration Function Fitting

Section 5.1.2.3 introduced the information that if the χ^2 of the first attempt at the calibration function fit is too high, the calibration module will try to throw out the worst point to see if it improves the fit under the assumption that said peak may have been incorrectly identified. There are a few behaviors in this feature of the calibration module as it is currently implemented that could compromise the reliability of automated calibration performance.

First, there is currently a requirement that the number of peaks successfully fit by the calibration module must be greater than the number of free parameters in the calibration function for the calibration module to attempt to throw out a peak due to a high-χ^2 calibration function fit. However, this requirement allows a situation in which the attempt to discard a peak may lead to a

calibration function fit with no free parameters. The requirement ought to be that the number of peaks successfully fit by the calibration module must be greater than the number of free parameters in the calibration function *plus one*.

Second, there is currently no provision preventing the calibration module from discarding the 2614.5-keV calibration peak. In at least two instances in Cuoricino, this peak was removed from the calibration function fit: Dataset 19, Channel 68 and Dataset 23, Channel 15 (both in the processing using the third-order log polynomial calibration function). However, this peak is crucial for anchoring the calibration in the ^{130}Te $0\nu\beta\beta$ region; a channel is automatically marked bad for analysis due to excessive calibration uncertainty if the 2614.5-keV peak cannot be found, and there should be no case in which the calibration is performed without it. If the 2614.5-keV peak would be thrown out, the calibration module should instead print an explicit warning that the calibration fit is not sufficiently consistent with the 2614.5-keV peak and recommend a manual check of the channel.

Third, the reduced-χ^2 warning threshold was optimized assuming the second-order log polynomial to be the default calibration function. The warning threshold is currently hard-coded to $\frac{\chi^2}{n.d.f.} > 25$. However, as has been demonstrated in Section 6.3, the pol3 calibration function overall describes the energy calibration of the detectors in the gamma region better than the logpol2 calibration function does; therefore, we would expect the 'characteristic' χ^2 of this functional form to be smaller than that of the logpol2. Indeed, this behavior can be seen by investigating the calibration fit χ^2s of the best-quality Cuoricino calibrations:

- Datasets affected by the high-radon periods identified by L. Kogler in her thesis were removed [107].
- The pol3 calibration did not fail or have no free parameters.
- All four primary peaks were used for the interpolation fit for both processings, meaning that all four primary peaks were found on the first pass and found the same by both processings.

Figure A.1 shows the distributions of the calibration fit χ^2s for both the logpol2 and pol3 processings using this data selection.

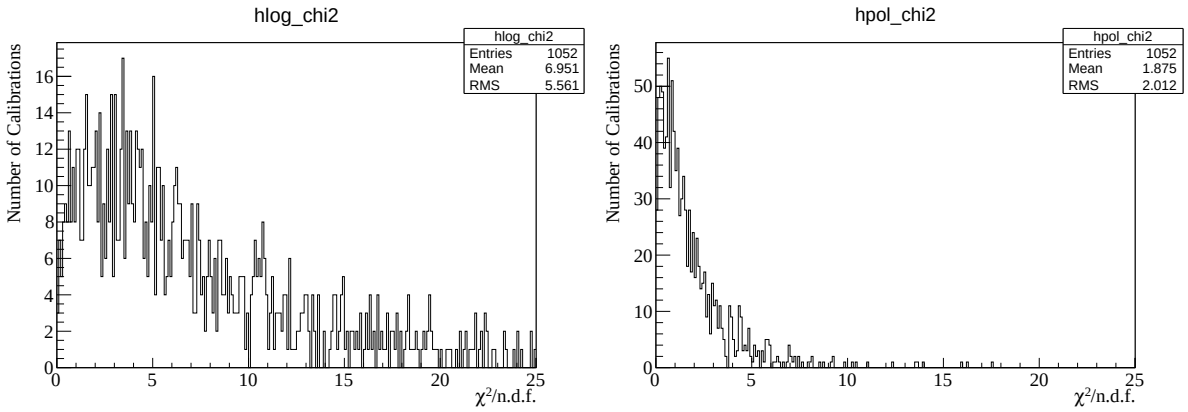


Figure A.1: Distributions of final calibration fit χ^2 in 'best-quality' Cuoricino calibrations. Left: logpol2 processing. Right: pol3 processing.

Even a brief visual inspection of Figure A.1 reveals that a lower reduced-χ^2 warning threshold, perhaps around 9 or 10, would be appropriate for the pol3 calibration function. In fact, 16 of the 1052 calibrations lie above $\frac{\chi^2}{n.d.f} = 25$ in the logpol2 case, while 16 of the 1052 calibrations lie above $\frac{\chi^2}{n.d.f.} = 8.5$ in the pol3 case.

It is clear that, instead of setting the χ^2 warning with a hard-coded parameter, it would be preferable to assign to each functional form used for the calibration function a χ^2 threshold that will trigger a warning and/or peak rejection on only the worst $\sim$1.5% of 'good-quality' calibrations. The drawback is that a large number of calibrations will be required to find this threshold empirically for a given calibration function; in the case of the default calibration function for CUORE, the third-order polynomial with intercept set to zero, however, we can in principle expect its behavior to be very similar to that of the pol3 function investigated here on Cuoricino data.

A.2 Refinement of Peak-Finding Behavior

A study of the automated peak-finding performance of the Diana calibration module using Cuoricino data is presented in Section 6.2. Specifically, this study focuses on the secondary calibration peaks, which are found by interpolating from the primary peaks. Several recommendations for changes to the calibration module have arisen as a result of this study and are summarized here.

First, the calibration module should be rewritten to perform the peak finding (both primary and secondary peaks) only once on a given channel and dataset. Currently, the peak finding is performed every time the calibration module is run on a set of calibration data, and a different interpolation function is used to search for the secondary peaks depending on the form of the calibration function that is used. This means that it is possible for different peaks to be found and used for calibration each time a calibration is performed on a given set of data, although in principle there can be only one correctly identified collection of peaks. Instead, the calibration peaks found for a given channel and dataset should be stored in the database. When the calibration module is run, it should first check the database to see if calibration peaks are stored for that channel and dataset; if they are, it should load the stored peaks and fit them to the calibration function. Only if no calibration peaks are stored in the database should the calibration module perform the peak finding and subsequently store the located peaks to the database.

Second, the way the peak finding is performed should not depend upon the form of the calibration function. If the peak finding is to be performed only once for a given set of calibration data, it should be performed the same no matter the details of the processing that happens to be the first processing applied to the data. The goal of the study in Section 6.2 was to determine the optimal functional form to be used for the interpolation function for the search for the secondary calibration peaks. The conclusion of that study was to recommend a third-order polynomial with the intercept fixed to zero for the interpolation function. This is the same functional form that is recommended for the calibration function, but it should be implemented separately in the calibration module so that no matter what functional form is used for the calibration function, the peak finding will be performed with a third-order polynomial with the intercept fixed to zero[1].

Third, it appears that the successful identification of a peak can sometimes be unexpectedly sensitive to small changes in the peak mean guess. The identification of each peak begins with a single value, the 'peak mean guess'; this value can be obtained in several different ways depending

[1]It is possible that decoupling the interpolation function from the calibration function will make it unnecessary to actually save the calibration peaks to the database; in principle, if the peak finding is performed the same way regardless of the other settings of the module, even if the peak finding is performed multiple times, the peak-finding results should be the same each time.

on the peak, but in each case it is used to define a window of the detector spectrum in which to search for a structure consistent with the peak (see Sections 5.1.2.1 and 5.1.2.2 for details). It seems that a very small shift in the original peak mean guess used to define this window can sometimes cause the module to be unable to locate a structure that it recognizes as having a sufficiently large significance over background to qualify as a calibration peak, even when such a peak is indeed present. Therefore, it may be advantageous to implement a functionality in the calibration module that, in the case that no significant structure is found for a given peak using the original peak mean guess, would apply a small perturbation to the original peak mean guess and attempt the peak finding again with this new value. Such a feature would require careful testing on data to determine the optimal number of trials that should be attempted on each peak before determining that said peak cannot be located.

Fourth, the allowed range of the $amp2$ parameter of the fit to the calibration structure consisting of two peaks at 1592.533 keV and 1588.19 keV should be restricted to reduce the present unstable behavior of the fit. Currently, the parameter is allowed to vary between 0 and 1 because there is no well-defined *a priori* value that can be assigned to the expected relative strength of the two peaks. However, allowing both $amp2 = 0$ and $amp2 = 1$ to be valid parameter values allows a 4.3-keV ambiguity in the energy assignment of the peak in the case that it is fitted as a single-peak structure. A preliminary calibration simulation with the CUORE geometry indicates that a range of $0.1 \leq amp2 \leq 0.7$ should allow the expected peak shape to be well fit in the majority of CUORE detectors. A similar simulation of the CUORE-0 cryostat would be necessary to define the appropriate $amp2$ range for this peak in CUORE-0 calibrations.

A.3 Changes for CUORE Due to Detector Geometry

Section 6.4 identifies two changes to the calibration procedure that will be required for CUORE (but not for CUORE-0) due to the location of the calibration sources inside the cryostat shields. They are summarized here for convenience.

First, the 583.191-keV gamma from ^{208}Tl will no longer be significantly be attenuated by the shields before reaching the CUORE detectors. In fact, simulations indicate that this gamma line

can be expected to be even stronger than the 2614.5-keV gamma line on the majority of CUORE detectors. For CUORE calibration, then, it should be possible to define the 583.191-keV calibration peak as a primary peak instead of a secondary peak.

Second, the 510.77-keV gamma from ^{208}Tl will similarly be considerably less attenuated with respect to the 511.0-keV annihilation gamma. The detectors cannot separately resolve the two lines, so the nominal energy assigned to the peak will need to shift to accommodate the contribution from both structures. The expected contribution of the 510.77-keV gamma is somewhat dependent on the position of the crystal in relation to the calibration sources; however, the calibration module does not presently allow a channel-dependent assignment of peak energy. A preliminary simulation indicates that to first order, it may be appropriate to assign the calibration peak an energy of 510.885 keV, the mean of the two energies, on all channels. Once CUORE calibration data has been collected, a dedicated study can be performed to determine whether it will be necessary to adapt the calibration module to allow channel-dependent peak energies.

The assignment of nominal energy and primary or secondary status for each peak is implemented in the calibration source definition in the calibration module. The simplest method of implementing these geometrically motivated changes is to define a new calibration source that can be passed to the module for the calibration of CUORE data.

Appendix B: Detailed Procedure Documents for String Production

Presented here are the procedure documents used by J. Clark during the production of the final CUORE calibration source carriers. The initial drafts of both the cleaning procedure document and the source string production document were written by L. Ejzak; the final version of the cleaning procedure document incorporates experience gained from tests performed by J. Clark, and the final version of the source string production document incorporates the jig operation experience of L. Ejzak, D. Lenz, I. Guinn, and J. Clark. The cleaning procedure for the heat gun was verbally communicated to J. Clark by the manufacturer.

The source string production document references Table 8 in the internal note regarding the source carrier production [54] for the precut lengths of the PTFE heat shrink sleeves, which differ for different PTFE batches and were empirically determined in a battery of shrink tests performed by J. Clark prior to final source carrier production. It also references the Spacing Document, which is a list of predefined position settings for the main reference bar on the source manufacturing jig used to properly place the capsules on the string, and the Activity Document, a list of the wire insert specifications (number, size, and ThO_2 content) for each capsule on the string. All of these documents are stored in the clean room near the jig for easy reference during manufacturing.

The source string production document also references several spacers used to position the tool support arms for the various operations involved in building the capsules. The 'height' of the spacer refers to the vertical space it enforces between the reference bar and the tool support arm. Spacer X was produced with a nominal height of 0.858" (2.18 cm); it is used to position the crimp tool support arm for the bottom crimp, and its height is technically arbitrary[1]. Spacer Y was produced with a nominal height of 1.161" (2.95 cm), 7.7 mm greater than Spacer X; it was originally intended to be used to position the crimp tool support arm for the top crimp for the source capsules, allowing only 0.15 mm of the length of the copper tube to be taken up by each crimp. Spacer Z was produced with a nominal height of 1.138" (2.89 cm), slightly shorter than

[1] Its height was originally defined to allow clearance for the C-clamp that helps anchor the crimp tool to the support arm, but when the clamp was replaced with a new stainless steel clamp that is appropriate for clean room use, the top bar of the new clamp was larger than that of the old clamp and had to be cut down.

Spacer Y; it was originally intended to be used only to position the pliers arm for shrinking the PTFE sleeve, but it was found to provide a more reliably secure top crimp placement than Spacer Y. As a result, Spacer Y is no longer used for any of the source-string-assembly operations. Spacer AA was accordingly produced at a nominal height of 1.0787" (2.74 cm), 1.65 mm shorter than Spacer Z; this spacer positions the top crimp for the weight capsules, which are 1.65 mm shorter than the source capsules.

B.1 Cleaning Procedure Document

All production materials are cleaned as specified below before string production. This includes all lengths of copper and PTFE, PTFE beads, copper bead inserts, and Kevlar. The DI water used in our cleaning procedures has a measured conductivity of less than 1.0 μS.

1 Initial Cleaning — Use degassed solution of 20% acetone, 80% alcohol in beakers (5 minutes degas). Then add pieces and run in ultrasonic for 20 minutes.

2 Drain pieces in sieve and rough rinse with DI water.

3 Cleaning — Use degassed solution of 5% citric acid, 1% H_2O_2, 94% DI water in beakers (5 minutes degas). Then add pieces and run in ultrasonic for 20 minutes.

4 Drain pieces in sieve, then rinse thoroughly with DI water.

5 Rinsing — Use degassed 100% alcohol in beakers (5 minutes degas). Then add pieces and run in ultrasonic for 20 minutes.

6 Remove from ultrasonic and dry outer surface of beakers. Carry into clean room.

7 Drain pieces in sieve (catching drain alcohol in bowl). This should be done in clean room only.

8 Dry pieces under nitrogen flux in desiccator box within clean room.

9 Remove drain bowl from clean room for disposal.

10 Store pieces under nitrogen until string production.

B.2 Cleaning Procedure for Heat Gun

1 Perform an external alcohol wipe-down of the outside of the heat gun, removing nozzles and attachments to clean separately.

2 Use a small brush to wipe the side and back vents. Do not push the bristles into the gun; use a sweeping motion of brush.

3 Make sure that side and back mesh vents are clear and open, with nothing blocking them.

4 Use compressed air with no additives or bitterants to clear out the interior of the heat gun; clean pressurized nitrogen works well. Aim air directly into the nozzle (with any external nozzle or attachments removed) and use short, quick blasts. Debris/particles will exit through the back mesh vents. If using commercial 'canned' air, be careful not to allow it to overcool; follow manufacturer instructions to avoid excess liquid production.

5 Raise the heat gun to temperature (or slightly above) and shoot continuous blasts with the gun to remove any possible particles. This step should be performed in a clean room environment so that intake air is clean, if possible. Run the gun continuously for approximately 2 minutes, then allow gun to cool completely.

6 After the gun has cooled again, bring it to temperature and run continuously for an additional $3-4$ minutes. This should be done in a clean room environment. Allow the gun to cool completely.

7 Perform a final complete external alcohol wipe-down of the gun itself, and clean removable nozzles and attachments with the full clean room procedures defined in the previous document.

B.3 Source String Production Document

To ensure that each string is produced identically, we include detailed instructions for final production.

Materials per Source Carrier

1 PTFE-coated Kevlar string: 4.5 m

2 25 + 2 pieces of 8-mm copper sleeves

3 8 + 3 pieces 6.35-mm copper sleeves

4 25 + 5 pieces of small heat shrink, $11-13$ mm (See Table 8 for cut lengths)

5 8 + 3 pieces of large heat shrink, $10-12$ mm (See Table 8 for cut lengths)

6 1 PTFE ball

7 1 copper core insert

8 4 + 2 pieces of 1% Th wire, 0.38-mm OD, 2.46-mm length (INTERNAL string)

9 21 + 6 pieces of 1% Th wire, 0.25-mm OD, 3.96-mm length (INTERNAL string)

10 8 + 2 pieces of 1% Th wire, 0.25-mm OD, 2.85-mm length (INTERNAL string)

11 12 + 4 pieces of 2% Th wire, 0.35-mm OD, 3.90-mm length (EXTERNAL string)

12 42 +10 pieces of 2% Th wire, 0.35-mm OD, 2.54-mm length (EXTERNAL string)

13 16 + 4 pieces of 2% Th wire, 0.35-mm OD, 2.92-mm length (EXTERNAL string)

Tools Used

1 Masking tape

2 Spacing Document (Detailed list of set spacing)

3 Activity Document (Detailed list of activity per capsule)

4 Source jig (includes frame, crimping tool, custom pliers, stop platform, spacers)

5 Heat gun (Steinel Electric Heat Gun, Model HL 2010 E, Type #3482)

6 Paper clips

7 Tweezers

Step-by-Step Instructions

1 Cut 4.5 m of Kevlar string.

2 Secure the string to the tube on top of the jig with masking tape.

3 Wrap the remaining string on the tube and insert the free end of the string through the center hole on the top crossbar of the jig.

4 Slide 7 small PTFE pieces followed by 5 regular copper tubes onto string; move to top and secure with paper clip below lowest copper tube.

5 Slide 11 small PTFE pieces followed by 10 regular copper tubes onto string; move to top and secure with paper clip below lowest copper tube.

6 Slide 12 small PTFE pieces followed by 10 regular copper tubes onto string; move to top and secure with paper clip below lowest copper tube.

7 Slide 12 large PTFE pieces followed by 8 weight copper tubes onto string; move to top and secure with paper clip below lowest copper tube.

8 Insert the free end of the string through the holes in the crimper and jig.

9 Attach string at bottom clamp point, tying a knot.

10 Tension the string by loosening the mounts of the top tube, rotating the tube, and retightening the mounts.

11 Ensure that no heat gun nozzle is attached; the bare heat gun directed at an angle toward capsules takes longer than using the nozzle but has a more consistently even result at top/bottom of capsules. Start heat gun, set to $750\,^{\circ}$F, so that it will be warm by the time it is needed.

12 Position reference arm at the bottom of the jig (its carriage will rest against the bottom crosspiece) and ensure that this position corresponds to the 0.1-cm point on the scale on each side rail. (Follow Spacing Document for position of each subsequent capsule.)

13 Position crimper arm with respect to reference arm using spacer X. The spacers should always be placed in a consistent position, as far *forward* toward the free end of the crimper arm as possible.

14 Position plier/stop arm above crimper arm.

15 Extend sliding stop platform on plier/stop arm so that the scored line on the platform lines up with the Kevlar.

16 Move one copper tube down to rest on sliding stop platform.

17 Slide pliers out to grab copper piece; align the front of the pliers carriage with the end of the plier arm and tighten the set screw on the carriage fully before finalizing the positioning of the copper piece in the pliers. Verify position is centered and will fit in crimper (look down string).

18 Use tweezers to make sure copper is against sliding stop (pull down if necessary with tweezers), and verify that copper is vertically straight. Then retract sliding stop platform.

19 Move plier/stop arm down until the pliers rest on the crimper (between the screws). Be careful not to hit copper on crimper; RESET position if there is any contact between capsule and crimper, as this will bump your capsule out of position. Hold pliers firmly closed while moving arm to prevent friction against the string from causing the copper to slip in the pliers (especially likely for weight capsules).

20 Visually confirm that the copper piece is properly centered in the crimper — this is necessary both to prevent warping the pliers and to prevent a poor-quality crimp. Check spacing position one more time (position can slip).

21 Hold pliers in left hand and crimper in right hand. Gently begin closing crimpers until contact is made with capsule, then release pliers. Hold pliers open so that crimper does not damage them; the capsule is now held by the contact with the crimper. Now finish crimping by fully squeezing crimper (bottom crimp). Note that crimper will not release/open until it has reached its fully closed position.

22 Loosen plier/stop arm so that it is free to slide on the vertical rail, and loosen the set screw on the pliers carriage so that it is free to slide on the arm — this is necessary in order to be able to open and remove the pliers.

23 Keep pliers open and slide them back on the arm, while simultaneously lifting the arm (do not catch on string).

24 Insert active wires. Follow Activity Document for number and length of active wires per capsule. Use wire pushing tool (stripped twist tie end) to fully insert all active wires.

25 Position crimper arm with respect to reference arm using spacer AA (weight capsules) or Z (regular capsule). Be careful not to push capsule down with edge of crimper. The spacer should always be placed in a consistent position, as far *forward* toward the free end of the crimper arm as possible.

26 Visually confirm that the copper piece is properly centered in the crimper. It helps to hold the string to ensure that the copper is centered in the crimper. Check spacing position one more time (position can slip), THEN crimp (top crimp).

27 Remove spacer and slide crimper arm up out of the way.

28 Position plier/stop arm with spacer Z (always Z for any capsules). The spacers should always be placed in a consistent position, as far *forward* toward the free end of the crimper arm as possible.

29 Extend sliding stop platform on plier/stop arm so that the scored line on the platform lines up with the Kevlar.

30 Move one heat shrink piece down to rest on sliding stop platform. Large PTFE pieces should be gently folded and pushed through crimper, then unfolded and re-shaped with tweezers. If there are any snags or PTFE is deformed, use a new piece. Regular PTFE pieces should be able to pass through crimp tool without manipulation.

31 Verify PTFE piece is centered around copper capsule. If platform prevents accurate centering, instead use paper clips above/below PTFE piece to hold in centered position. (Note: Platform may put 'extra' PTFE on top edge of capsule.) For weight capsules centering is more difficult; use combination of paper clips/tweezers to maintain centering. Platform may be adequate for centered placement in some capsules.

32 Use heat gun directed at an angle to shrink sleeve around copper. PTFE has a tendency to slide upward when heat is applied; use a paper clip above capsule to keep PTFE in proper position.

33 Retract sliding stop platform.

34 Slide plier/stop arm up out of the way.

35 Move reference arm to next spacing. Follow set spacing as indicated on the Spacing Document.

36 Repeat steps 13 to 35 for all 8 weight capsules (top crimp using AA spacer) and all 25 regular capsules (top crimp using Z spacer). Verify each capsule position with the Spacing Document.

37 Unwind string from top cross bar (carefully so string does not become tangled).

38 Remove the source string from the jig.

39 Cut away bottom of string where it was attached at the clamp point, as this can damage the string. (Enough length should remain for the bottom PTFE ball weight.)

40 Cut any remaining additional length from the top of the string.

41 Remove any remaining spare PTFE pieces carefully from string (cut small pieces if needed).

42 Thread PTFE bead and copper core insert from bottom of string.

43 Using free-held crimper, crimp size 2, lightly grip copper core insert so that crimper teeth are holding it. Verify that the larger-diameter flat 'step' at bottom end of insert is free of crimper teeth. Use small copper wire to adjust position of insert in crimper.

44 Using small copper wire, pull PTFE bead as close to bottom weight capsule as possible. Slide crimper with held insert up until touching wire holding the PTFE bead. (This procedure may require two people, one to hold tension on the string and the PTFE bead away from the crimper teeth and one to crimp the insert.)

45 Verify crimper position is as close to weight capsule as possible while not catching the PTFE bead. Crimp insert.

46 Check for placement and quality of crimp. Tie knot under insert.

47 Press-fit PTFE bead onto copper core insert.

48 Cut additional string from bottom.

49 After QA testing, a large (weight) copper capsule is added to the top of the string for anchoring in the drive spool. This capsule can be crimped off the jig with a free-held crimper, crimp size 4.