

ABSTRACT

Title of dissertation: SIGNAL DENOISING FOR EXO-200
AND AN IMPROVED LIMIT ON
NEUTRINOLESS DOUBLE-BETA DECAY

Clayton G. Davis, Doctor of Philosophy, 2014

Dissertation directed by: Professor Carter Hall
Department of Physics

The EXO-200 detector is designed to measure the spectrum of Xenon-136 double-beta ($\beta\beta$) decay. This extremely low-rate radioactive process is interesting on its own, but more profound is its potential to reveal non-standard neutrinoless double-beta ($\beta\beta 0\nu$) decay. $\beta\beta 0\nu$ decay, if it occurs, would have important consequences for our understanding of the neutrino sector of the standard model. It would demonstrate the type of the neutrino mass term; set the mass scale of the neutrino sector; and demonstrate the first direct observation of lepton number non-conservation.

In the EXO-200 detector, we measure the summed energy of the two emitted electrons. The summed-electron-energy spectrum of standard $\beta\beta$ decay produces a smooth spectrum due to the "missing" energy of the two emitted neutrinos. In contrast, the non-standard $\beta\beta 0\nu$ decay has no neutrino products, and emits energy only in the form of electrons. Since all energy is captured by the electrons, the sum of electron energies will always be 2458 keV; thus, we can discover this process by looking for a peak in the summed electron energy spectrum at this energy.

The sensitivity of the search depends on two basic factors: accumulated Xenon-136 exposure, and expected background near 2458 keV from non- $\beta\beta 0\nu$ decays. The EXO-200 detector was designed to minimize background primarily through its construction. All materials used in the construction of the EXO-200 detector were carefully screened for low concentration of radioactive backgrounds, making the EXO-200 detector one of the lowest-background detectors in the world.

Backgrounds were also minimized in EXO-200 by ensuring that the energy resolution of the detector would be excellent. Good energy resolution ensures that all monoenergetic peaks (including both $\beta\beta 0\nu$ and background gamma lines) will be sharp. When background is integrated over the width of the $\beta\beta 0\nu$ expected peak, we wish for the integrated number of expected counts to be as small as possible; when peaks are sharper, we can expect that integral to be smaller.

When the EXO-200 detector turned on, its scintillation signals observed more electronic noise than had been expected based on design goals. The present work is concerned with improving the energy resolution of the EXO-200 detector by removing electronic noise by an offline denoising process. It will be concerned with a detailed characterization of the signal and noise behavior of the EXO-200 scintillation measurements; a novel algorithm for denoising such signals; and results from denoised EXO-200 data will be presented, showing an improved $\beta\beta 0\nu$ limit in Xenon-136. Since this scheme is still in its infancy, possible future improvements will be described throughout, along with discussion of motivations for these improvements.

SIGNAL DENOISING FOR EXO-200 AND AN IMPROVED
LIMIT ON NEUTRINOLESS DOUBLE-BETA DECAY

by

Clayton G. Davis

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Carter Hall, Chair/Advisor
Professor Radu Balan
Professor Betsy Beise
Professor Rabindra Mohapatra
Professor Peter Shawhan

© Copyright by
Clayton G. Davis
2014

Preface

If needed.

Foreword

If needed.

Dedication

If needed.

Acknowledgments

Thanks all for lots of help...

Contents

List of Figures	viii
List of Abbreviations	xi
1 Theory of $\beta\beta 0\nu$ Decay	1
1.1 Double-Beta Decay	1
1.2 Neutrinoless Double-Beta Decay	4
1.3 Neutrino Flavor Physics	9
1.4 Particle Physics Constraints	11
1.5 Nuclear Physics Constraints from $\beta\beta 0\nu$	14
2 The EXO-200 Detector	21
2.1 Overview of the EXO-200 Detector	21
2.2 Backgrounds to $\beta\beta 0\nu$ Decay	27
2.3 Passive Background Rejection	31
2.4 Active Background Rejection	40
2.5 Pulse Amplification and Waveform Readout	42
2.6 Calibration Systems	46
2.7 Summary	50
3 Denoising Theory	51
3.1 Setup	54
3.2 The Signal Model	56
3.2.1 APD Noise	58
3.2.2 APD Signals and Noise	62
3.3 Derivation	66
3.4 Matrix Version	73
3.5 Preconditioning	77
3.6 Solver	80
3.7 Implementation	81
3.8 Denoising in Practice	87
3.9 Future Plans	89
3.9.1 Anticorrelated Cluster Energies	89
3.9.2 Wire Denoising	93

4	Noise Measurements	94
5	The Lightmap	95
5.1	Four-Dimensional Lightmap	97
5.2	Algorithm Details	99
5.2.1	Event Selection	100
5.2.2	Function Parametrization	103
5.2.3	Convergence and Error Estimation	107
5.3	Implementation for this Analysis	110
5.4	Visualization	112
6	Denoising Results	120
7	Conclusions and Future Plans	121
	Bibliography	122

List of Figures

1.1	Feynman diagram for $\beta\beta 2\nu$ decay. The reaction products are equivalent to two β decays in succession, but this reaction can sometimes occur even if a single β decay would be energetically forbidden. . . .	2
1.2	Energy diagram of isotopes with atomic mass $A = 136$. The energies ΔE are the binding energies of the atom, compared to the bare masses of the same nuclei. Values are from [13].	3
1.3	Feynman diagram for $\beta\beta 0\nu$ decay. A virtual neutrino mediates the exchange. This is only possible if $\bar{\nu}_R$ can oscillate to ν_L , and the interaction that induces this oscillation also generates neutrino mass.	5
1.4	A generic $\beta\beta 0\nu$ process, with no assumptions of the underlying mechanism. If $\beta\beta 0\nu$ is observed, it requires that this reaction is possible even though it does not strictly imply that the oscillation in figure 1.3 is the cause.	6
1.5	Since $\beta\beta 0\nu$ requires that the reaction in Figure 1.4 occurs, we can insert that diagram into this larger one and assert that regardless of the underlying mechanism, $\beta\beta 0\nu$ implies an effective neutrino mass generated by $\bar{\nu}\text{-}\nu$ oscillation as a higher-order process.	7
1.6	$\beta\beta 0\nu$ half-life limits versus the year of the limit. Colors indicate which isotope is under study. Open circles indicate experiments which have not yet concluded data-taking. Figure from [42], with additional isotopes included, plus a couple of additional points for recent results which you preferred not to include – BUT MAYBE THAT’S MORE TROUBLE THAN IT’S WORTH?	8
1.7	The electron spectrum of Tritium (3H) β decay. The endpoint contains only a small fraction of the total statistics. Figure from [8]. . . .	13
1.8	The relationship between the effective Majorana mass $\langle m_{\beta\beta} \rangle$ and other fundamental neutrino quantities: the lightest neutrino mass eigenstate m_{min} , the sum of mass eigenstates $M = \sum m_i$, and the effective single-beta-decay neutrino mass $\langle m_{\beta} \rangle$. Black (magenta) lines indicate the allowed region for the inverted (normal) hierarchy; red (blue) hatches indicate uncertainty for the inverted (normal) hierarchy due to the unknown CP-violating and Majorana phases δ , α_1 , and α_2 . [20]	16

2.1	A two-dimensional energy spectrum of scintillation and ionization from a testbed liquid xenon experiment under an electric field of 4 kV/cm. The spectrum is from a ^{207}Bi source. Figure reproduced from [27].	22
2.2	Schematic of the inner EXO-200 TPC. Figure reproduced from [14]. .	24
2.3	Anode collection wires (u-wires) and induction wires (v-wires) from EXO-200. (1) and (5) indicate the wire support frame; (2) indicates the wires themselves, constructed as gangs of three wires; (3) illustrates the attachment between wires and the support frame; (4) shows the return cables from the wires to the data acquisition system. Figure reproduced from [14].	26
2.4	X-ray attenuation lengths in xenon. Compton (incoherent) scattering, pair production, and photoelectric absorption are shown, along with their combined attenuation length; coherent (Rayleigh) scattering is omitted because it produces no observable energy deposit. The vertical axis, in cm^2/g , can be multiplied by the density of the xenon to derive an attenuation factor per unit length. Figure produced by [19].	32
2.5	Wire triplet, read as one channel. Figure from [14].	35
2.6	APDs ganged together (bottom right). Wiring to the front-end electronics are visible as yellow “tape.” Figure from [14].	36
2.7	A cutaway schematic image of the TPC and surrounding materials in the cleanroom. HFE-7000 refrigerant is contained between the cryostat and the LXe vessel. Figure from [14].	37
2.8	Muon flux as a function of depth, in meters water equivalent. The WIPP site is indicated in relation to other underground science facilities. Figure from [29].	39
2.9	A cross-sectional schematic of an APD is shown in (a). The electric field as a function of depth is shown in (b). Figure from [51].	43
2.10	A guide tube permits gamma sources to be inserted to known locations near the TPC. Green dots indicate a few of the locations where the source may conventionally be placed. Figure from [14].	47
3.1	Coherent and incoherent noise power spectra for a sample set of APD channels without signal shaping. [61]	52
3.2	Shaped and unshaped APD waveforms. The normalization is shown to make the peak of the shaped waveform have a magnitude of one, and the time axis is shifted so that the unshaped waveform is a step function centered at $t = 0$	57
5.1	Lightmap position-dependence $R(\vec{x})$ for selected APD gangs.	112
5.2	Lightmap position-dependence $R(\vec{x})$ for selected APD gangs. Here extreme anode positions are omitted to permit better contrast for the lightmap in the fiducial volume.	113
5.3	Functions $S(t)$ for selected channels.	114
5.4	Functions $S(t)$ for selected channels.	115

5.5	Functions $S(t)$ for selected channels.	116
-----	---	-----

List of Abbreviations

EXO	Enriched Xenon Observatory (family of current or planned detectors)
EXO-200	Current (200-kg) iteration of the EXO family of detectors
nEXO	Planned next-generation EXO detector
APD	Avalanche Photo-diode
ADC	Analog-Digital Converter, or one of its discrete output units
DAQ	Data Acquisition System
TPC	Time Projection Chamber
WIPP	Waste Isolation Pilot Plant

Chapter 1: Theory of $\beta\beta 0\nu$ Decay

1.1 Double-Beta Decay

Standard-model double beta ($\beta\beta$) decay is the result of the particle interaction

$$2d \rightarrow 2u + 2e^- + 2\bar{\nu}_e \quad (1.1)$$

mediated by W^- -exchange, as depicted in Figure 1.1. It is effectively the simultaneous occurrence of two beta (β) decays from the same nucleus.

$\beta\beta$ decay is the result of a second-order weak interaction, so it has a remarkably slow rate compared to most β decay processes. As a result, although many nuclei are expected to decay by $\beta\beta$, the process is thoroughly masked by conventional β decay in most of them. We can only hope to detect $\beta\beta$ decay in isotopes where β decay is forbidden or highly suppressed.

In rare cases, β decay may be highly suppressed, enabling $\beta\beta$ decay to have a rate comparable to or exceeding single- β decay. This is, for example, the case in $^{48}_{20}\text{Ca}$: the ground state of $^{48}_{20}\text{Ca}$ has zero units of angular momentum, whereas its single- β -decay daughter product $^{48}_{21}\text{Sc}$ has six units of total angular momentum in its ground state. This β decay is thus highly suppressed by angular momentum

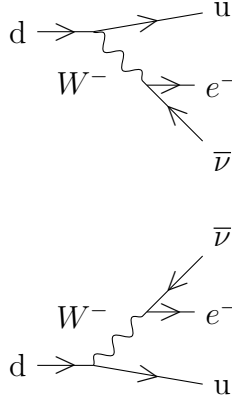


Figure 1.1: Feynman diagram for $\beta\beta 2\nu$ decay. The reaction products are equivalent to two β decays in succession, but this reaction can sometimes occur even if a single β decay would be energetically forbidden.

conservation. By contrast, $^{48}_{22}\text{Ti}$ has zero units of total angular momentum, making $\beta\beta$ decay of $^{48}_{20}\text{Ca}$ permitted by angular momentum considerations, and as a result the $\beta\beta$ decay mode of $^{48}_{20}\text{Ca}$ dominates. [45]

However, in most cases of interest β decay is forbidden by energy conservation. It is well-known that nuclei minimize their energy by arranging similar nucleons to have overlapping wavefunctions. [45] Thus, isotopes with an even number of protons and an even number of neutrons will have less nucleon-pairing potential energy than an isotope with either an odd number of protons or an odd number of neutrons, which in turn will have less nucleon-pairing potential energy than an isotope with an odd number of protons and an odd number of electrons.

A typical nuclear energy level diagram is shown in figure 1.2 for the $A = 136$ isobar: Xenon, Barium, Cerium, and Neodymium are even-even isotopes, and have systematically lower energies than the odd-odd isotopes Iodine, Cesium, Lanthanum,

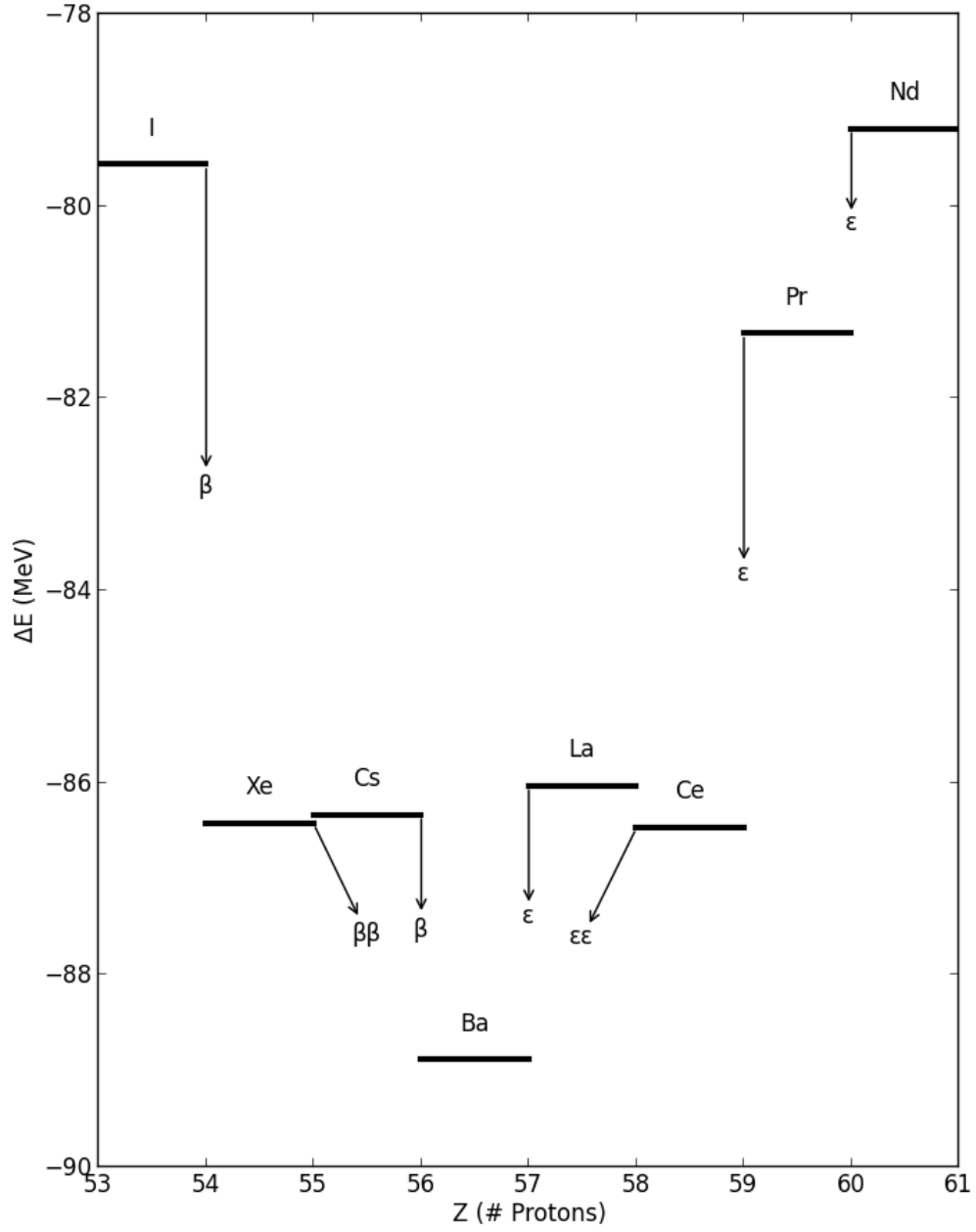


Figure 1.2: Energy diagram of isotopes with atomic mass $A = 136$. The energies ΔE are the binding energies of the atom, compared to the bare masses of the same nuclei. Values are from [13].

and Praseodymium. For this particular isobar, we can see that Xenon is energetically forbidden from single- β decaying to Cesium because the odd-odd isotope of Cesium has just slightly more potential energy than the even-even isotope of Xenon; as a result, the primary mode of decay of Xenon-136 will be by $\beta\beta$ decay to Barium-136. (We also note a similar feature, that the primary mode of decay of Cerium-136 will be by double-electron-capture ($\epsilon\epsilon$) to Barium-136. Many of the same implications of $\beta\beta$ decay also apply to $\epsilon\epsilon$ decay; however, in practice the rates for these decays will be lower than the rates for $\beta\beta$ decay, so we will not consider them further in this work.)

1.2 Neutrinoless Double-Beta Decay

The detection and study of $\beta\beta$ decay provides an opportunity to test a new class of nuclear matrix computations; however, it does not violate any fundamental symmetries and is, in this sense, a mundane prediction of the Standard Model. The primary appeal of $\beta\beta$ decay is through the opportunity it provides to probe the nature of neutrinos.

It has been suggested as early as 1937 that neutrinos could possess mass through a neutrino-antineutrino interaction, provided that the neutrino is its own antiparticle [49]. The theorized Majorana interaction comes from Lagrangian terms

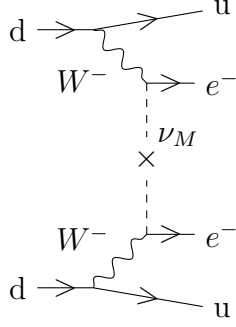


Figure 1.3: Feynman diagram for $\beta\beta 0\nu$ decay. A virtual neutrino mediates the exchange. This is only possible if $\bar{\nu}_R$ can oscillate to ν_L , and the interaction that induces this oscillation also generates neutrino mass.

of the form (for each of three neutrino eigenstates)

$$\begin{aligned} \mathcal{L}_{Maj} = & -\frac{m_L}{2} (\bar{\Psi}_L^c \Psi_L + \bar{\Psi}_L \Psi_L^c) \\ & -\frac{m_R}{2} (\bar{\Psi}_R^c \Psi_R + \bar{\Psi}_R \Psi_R^c), \end{aligned} \quad (1.2)$$

where the superscript- c represents charge conjugation. (This is the reason that Majorana mass terms are only possible for a chargeless lepton.) The masses m_L and m_R may be tuned independently; since there has never been an observation of right-handed neutrinos or left-handed anti-neutrinos, it is possible that $m_R = 0$ and that the fields Ψ_R and Ψ_R^c do not exist in nature [15].

If neutrinos do have Majorana mass interactions, then any isotope that undergoes $\beta\beta$ decay can also undergo the related process $2d \rightarrow 2u + 2e^-$, depicted in Figure 1.3, in which the two outgoing neutrinos in $\beta\beta$ decay are replaced by one virtual neutrino. We can interpret this as a mixing interaction between a left-handed neutrino and a right-handed antineutrino; this is only possible if neutrinos

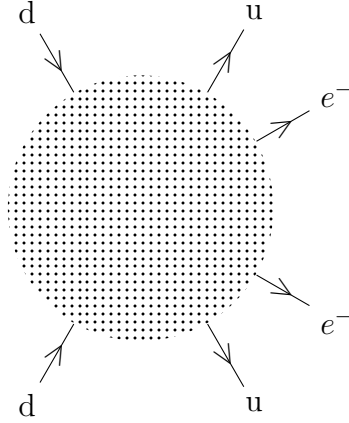


Figure 1.4: A generic $\beta\beta 0\nu$ process, with no assumptions of the underlying mechanism. If $\beta\beta 0\nu$ is observed, it requires that this reaction is possible even though it does not strictly imply that the oscillation in figure 1.3 is the cause.

are Majorana particles. We refer to the standard process as two-neutrino double beta ($\beta\beta 2\nu$) decay and the novel process as neutrinoless double beta ($\beta\beta 0\nu$) decay.

The tree-level diagram for $\beta\beta 0\nu$ has one additional interaction vertex compared to $\beta\beta 2\nu$ decay, and as a result we would expect it to occur at an even slower rate. However, the more immediate consequence of $\beta\beta 0\nu$ decay is that lepton number conservation would be violated. The lepton number changes by two units corresponding to the creation of two leptons with no balancing anti-leptons. Numerous theories have suggested other plausible modes of lepton number non-conservation [54][24], but none have yet reported a positive result. In the conventional Standard Model with massless neutrinos, lepton number conservation is an accidental symmetry [32], but in a model with massive neutrinos there may no longer be any reason *a priori* to expect conservation or non-conservation of lepton number.

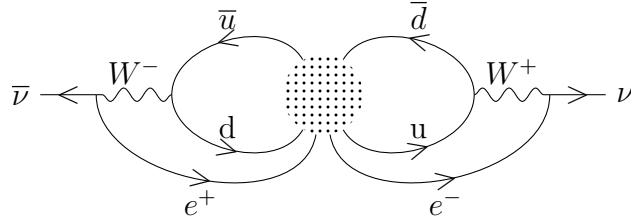


Figure 1.5: Since $\beta\beta 0\nu$ requires that the reaction in Figure 1.4 occurs, we can insert that diagram into this larger one and assert that regardless of the underlying mechanism, $\beta\beta 0\nu$ implies an effective neutrino mass generated by $\bar{\nu}\text{-}\nu$ oscillation as a higher-order process.

It is worth noting that the interaction term above assumes no mediating particles in the mechanism, whereas it is possible that $\beta\beta 0\nu$ could be mediated by some higher-order interaction terms. However, if $\beta\beta 0\nu$ decay is observed, it leads very generally to a conclusion that neutrinos have an *effective* Majorana mass interaction [56]. We can see this by examining the generic diagram for $\beta\beta 0\nu$ decay in figure 1.4. No matter what the mechanism of the decay is, we can insert this diagram into figure 1.5 and see that an effective neutrino-antineutrino mass vertex is implied by the existence of $\beta\beta 0\nu$ decay.

It has long been clear that appealing extensions to the Standard Model can be built around neutrinos with Majorana mass; see eg. the introduction in 1980 of the see-saw mechanism in [50], where the naturalness of the neutrino mass scale is addressed by hypothesizing m_R is inversely related to m_L and $m_L \ll m_R$. The motivation to search for $\beta\beta 0\nu$ decay became stronger when neutrinos were proven to have mass, particularly by the definitive proof of the oscillation of atmospheric

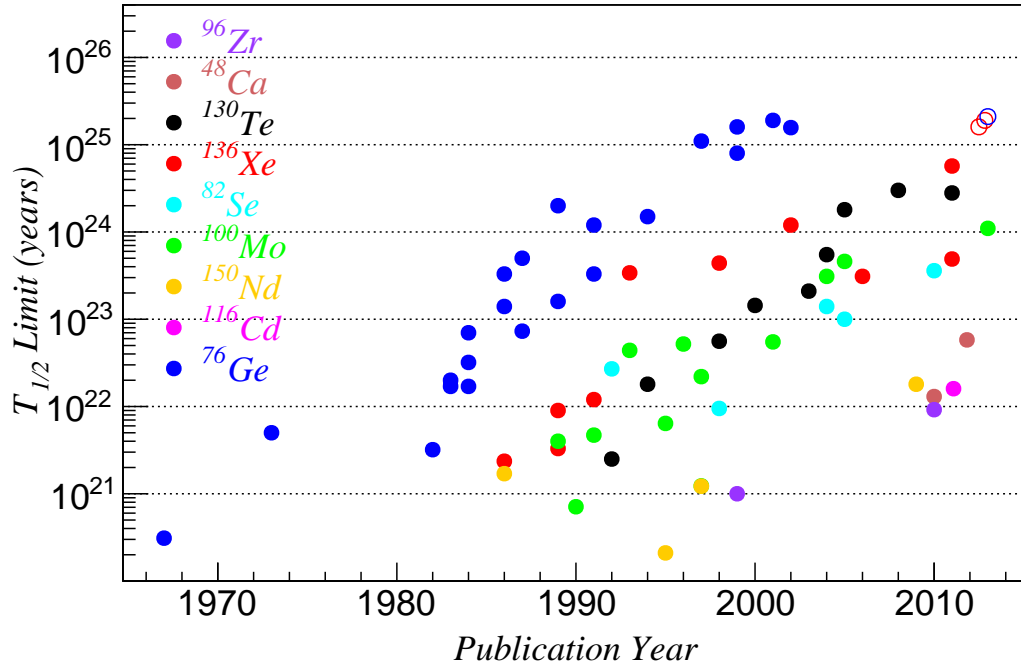


Figure 1.6: $\beta\beta 0\nu$ half-life limits versus the year of the limit. Colors indicate which isotope is under study. Open circles indicate experiments which have not yet concluded data-taking. Figure from [42], with additional isotopes included, plus a couple of additional points for recent results which you preferred not to include – BUT MAYBE THAT’S MORE TROUBLE THAN IT’S WORTH?.

neutrinos in [34]. As a result, the number of experiments searching for $\beta\beta$ decays exploded during this time period; figure 1.6 shows a collection of $\beta\beta 0\nu$ half-life limits obtained versus year, colored by the isotope under study. The search for $\beta\beta 0\nu$ has by now evolved into a highly competitive field, in which the isotopes ^{130}Te , ^{76}Ge , and ^{136}Xe are particularly dominant but searches are also being actively pursued in many other isotopes.

1.3 Neutrino Flavor Physics

Neutrinos are known to come in three flavors, or eigenstates which are diagonal with respect to the lepton interaction terms of the Standard Model. These flavors are ν_e , ν_μ , and ν_τ ; they interact, respectively, with e , μ , and τ leptons. We can hypothesize that neutrinos may also have a basis diagonalized with respect to rest mass which may not be identical to the flavor basis, and call the mass eigenstates ν_1 , ν_2 , and ν_3 with respective masses m_1 , m_2 , and m_3 . The unitary operator which translates between the two bases is specified by:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \mathbf{U} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \quad (1.3)$$

These formalities are uninteresting if the neutrino sector is massless, since the mass eigenstates are degenerate. However, in the case where neutrinos are massive, we can see that neutrinos which are created in one flavor eigenstate will oscillate between the flavor eigenstates with a period which depends on the differences be-

tween masses, as is typical in an N-state quantum system. In the neutrino sector, we can more specifically state that the probability for a transition from flavor α to flavor β will be: [37]

$$P_{\alpha\beta} = \delta_{\alpha\beta} - 4 \sum_{i=1}^2 \sum_{j=i+1}^3 \text{Re} [U_{\alpha i} U_{\beta i}^* U_{\alpha j}^* U_{\beta j}] \sin^2 \left(\frac{[m_i^2 - m_j^2] L}{4E} \right) \quad (1.4)$$

where L is the distance (or time, in $c = 1$ units) between neutrino source and destination, and E is the relativistic energy of the emitted neutrinos.

The transition probability is sensitive to the masses of the neutrinos, but only in the form $|m_i^2 - m_j^2|$. These measurements have now been performed in a variety of neutrino oscillation experiments, and the best current constraints are $|m_1^2 - m_2^2| = (7.50 \pm 0.20) \cdot 10^{-5} \text{eV}^2$ [35] and $|m_2^2 - m_3^2| = 2.32_{-0.08}^{+0.12} \cdot 10^{-3} \text{eV}^2$ [2].

However, oscillation experiments cannot constrain the overall mass scale of neutrinos; they can only set a conservative lower limit that $\max(m_2, m_3) \geq 0.048$ eV if we assume one of m_2 or m_3 is zero. Furthermore, they do not establish the sign of the difference. We can see that m_1 and m_2 are fairly close together, and m_3 is significantly different; but we cannot see whether m_3 is larger or smaller than the other two masses. We refer to the situation with $m_1 \simeq m_2 \ll m_3$ as the normal hierarchy, and $m_3 \ll m_1 \simeq m_2$ as the inverted hierarchy; the regime where $m_1 \simeq m_2 \simeq m_3 \gg |m_2^2 - m_3^2|$ is called the degenerate region. Distinguishing between these three situations is one of the significant open questions in neutrino physics because of its impact on observable quantities.

1.4 Particle Physics Constraints

To produce more detailed constraints on neutrino physics, it is generally useful to provide a parametrization of the mixing matrix \mathbf{U} from equation 1.3. The standard parametrization is:

$$\begin{aligned} \mathbf{U} &= \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_{23}) & \sin(\theta_{23}) \\ 0 & -\sin(\theta_{23}) & \cos(\theta_{23}) \end{pmatrix} \begin{pmatrix} \cos(\theta_{13}) & 0 & \sin(\theta_{13})e^{-i\delta} \\ 0 & 1 & 0 \\ -\sin(\theta_{13})e^{i\delta} & 0 & \cos(\theta_{13}) \end{pmatrix} \times \quad (1.5) \\ &\quad \begin{pmatrix} \cos(\theta_{12}) & \sin(\theta_{12}) & 0 \\ -\sin(\theta_{12}) & \cos(\theta_{12}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\alpha_1/2} & 0 & 0 \\ 0 & e^{i\alpha_2/2} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Out of the six parameters used in defining this matrix, the only ones which have been measured are the three mixing angles $\sin^2(2\theta_{12}) = 0.857^{+0.023}_{-0.025}$ [35], $\sin^2(2\theta_{13}) = 0.089 \pm 0.010(\text{stat}) \pm 0.005(\text{sys})$ [7], and $\sin^2(2\theta_{23}) > 0.95$ [1]. The Dirac phase δ is in principle observable from oscillation experiments, but no current experiments have achieved the sensitivity necessary to accomplish this. The Majorana phases α_1 and α_2 cannot be extracted from neutrino oscillations. [15]

The sum of the three mass eigenstates, $M = \sum m_i$, can be constrained by cosmological observations. This constraint (like all cosmologically-based constraints)

will be model-dependent; it relies on the expectation that low-mass, hot forms of dark matter similar to neutrinos promote the formation of large-scale structures in the early universe by allowing extremely remote regions of matter to remain in thermal equilibrium. Recent results from Planck combined with WMAP and baryon acoustic oscillations have restricted $M < 0.230$ eV with 95% confidence. [23] Considering the assertion in section 1.3 that the heaviest neutrino must have a mass no less than 0.048eV, we can see that this cosmological constraint pushes M to within a factor of five of its lower limit; taken at face value, the Planck measurement is the strongest existing constraint on the absolute mass scale of neutrinos.

Closer to home, the mass of neutrinos is also reflected in β decay, $d \rightarrow u + e^- + \bar{\nu}_e$. The total energy of the daughter products is known, and will be shared between the electron and antineutrino; the minimum energy of the antineutrino will be its rest mass, so by searching for the maximum energy of the electron we can simultaneously measure the rest mass of the neutrino. The electron anti-neutrino emitted from beta decay has a mix of all three mass eigenstates; since no current or planned experiment has sufficiently good energy resolution to resolve the separate endpoints from the three neutrino masses, we can instead write an effective rest mass of an electron antineutrino as: [15]

$$\langle m_\beta \rangle^2 = \sum_i m_i^2 |U_{ei}|^2. \quad (1.6)$$

Attempts to measure $\langle m_\beta \rangle$ require experiments to measure the electron energy spectrum of beta decays at the endpoint, which is complicated by the fact that this portion of the electron spectrum contains only a small fraction of the total electron

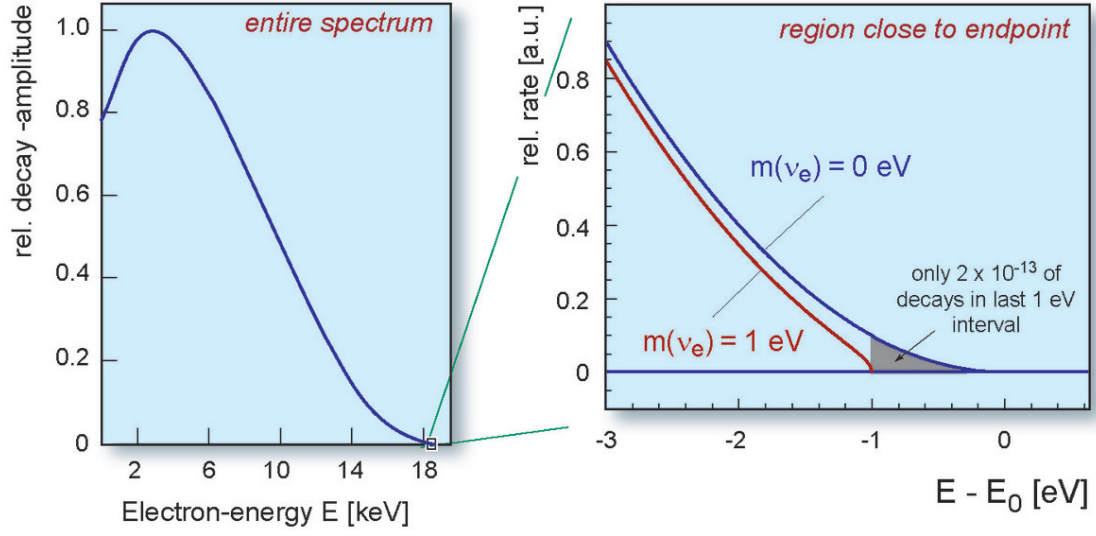


Figure 1.7: The electron spectrum of Tritium (3H) β decay. The endpoint contains only a small fraction of the total statistics. Figure from [8].

statistics. Tritium (3H) is commonly used for these experiments because it has a medium-length halflife of 12.3 years and an extremely low β decay endpoint energy of 18.6 keV, which maximizes the relative shift in endpoint energy due to $\langle m_\beta \rangle$. Figure 1.7 shows that for $\langle m_\beta \rangle = 1\text{eV}$ in Tritium, experiments must observe a shift in the spectrum which affects only about one decay in $5 \cdot 10^{12}$, making this level of sensitivity extremely difficult to achieve. The best existing limit from β decay is $\langle m_\beta \rangle < 2.05\text{ eV}$, with 95% confidence, from the Troitsk experiment which ran from 1994 to 2004. [12] The KATRIN experiment hopes to achieve a sensitivity of 0.2 eV, and is expected to begin taking data in 2015. [60][33]

1.5 Nuclear Physics Constraints from $\beta\beta 0\nu$

If the mechanism of $\beta\beta 0\nu$ decay really is, at tree level, a neutrino mass term as shown in figure 1.3, then the rate of $\beta\beta 0\nu$ decay will also reflect the magnitude of the neutrino mass parameters. We can write the partial half-life of an isotope which undergoes $\beta\beta 0\nu$ decay as

$$[T_{1/2}^{0\nu}]^{-1} = G_{0\nu}(Q_{\beta\beta}, Z) |M_{0\nu}|^2 \langle m_{\beta\beta} \rangle^2, \quad (1.7)$$

where $G_{0\nu}(Q_{\beta\beta}, Z)$ is a phase-space factor coming from the range of possible output states, $M_{0\nu}$ is a nuclear matrix element, and the last factor is the effective $\beta\beta$ neutrino mass, defined as:

$$\langle m_{\beta\beta} \rangle = \left| \sum_k m_k U_{ek}^2 \right|. \quad (1.8)$$

The phase-space factor $G_{0\nu}(Q_{\beta\beta}, Z)$ is difficult to compute exactly due to nuclear effects; for a recent computation, see [47] which estimates errors of 7% dominated by uncertainty in the nuclear radius. Its main feature is that it scales roughly like $Q_{\beta\beta}^5$, the fifth power of the decay energy. As a result, using $\beta\beta$ isotopes with high Q -values confers a significant advantage in neutrino mass sensitivity.

The dominant uncertainty in translating $\beta\beta 0\nu$ half-life limits into $\langle m_{\beta\beta} \rangle$ constraints comes from our knowledge of the matrix elements $M_{0\nu}$, which can have uncertainties as large as a factor of 2 – 3. [15] Calculations which attempt to estimate these values rely on a variety of matrix models, but often rely on approximations of some sort to make the calculations tractable. Efforts to improve the accuracy of

these values include a combination of more precise computations and experimental feedback to improve the accuracy of these models. See [62] for more discussion of attempts to improve this situation.

Unlike $\langle m_\beta \rangle$, which is an incoherent sum of strictly positive terms in equation 1.6, we can see that $\langle m_{\beta\beta} \rangle$ is a coherent sum of terms, each of which may have arbitrary complex phase which may increase or decrease the result of equation 1.8. In other words it is possible, even if neutrinos do have Majorana mass, for $\langle m_{\beta\beta} \rangle$ to be arbitrarily small if \mathbf{U} is tuned to produce cancellations between terms.

It is possible to relate the observables $\langle m_{\beta\beta} \rangle$, $\langle m_\beta \rangle$, and M within the model for \mathbf{U} of equation 1.5. These relations are shown in figure 1.8; we recall from section 1.4 that beta spectrum measurements constrain $\langle m_\beta \rangle < 2.05$ eV and cosmological observations constrain $M < 0.230$ eV, both with 90% confidence. The strongest constraints on $\langle m_{\beta\beta} \rangle$ from $\beta\beta 0\nu$ searches place $\langle m_{\beta\beta} \rangle < 0.15 - 0.4$ eV, depending on the choice of matrix element calculations. We can see that if the cosmological limits are to be trusted, they provide the strongest constraints on the neutrino mass parameters; however, all three observables are being probed competitively, and the wide range of experimental approaches means that systematic effects are unlikely to be shared by all three methods.

The sensitivity of an experiment for measuring $T_{1/2}^{0\nu}$ can be described by the approximate formula: [15]

$$T_{1/2}^{0\nu}(n_\sigma) = \frac{4.16 \cdot 10^{26} \text{ yrs}}{n_\sigma} \left(\frac{\epsilon a}{W} \right) \sqrt{\frac{Mt}{b\Delta E}}, \quad (1.9)$$

where M is the mass of material, a is the isotopic enrichment, W is the molecular

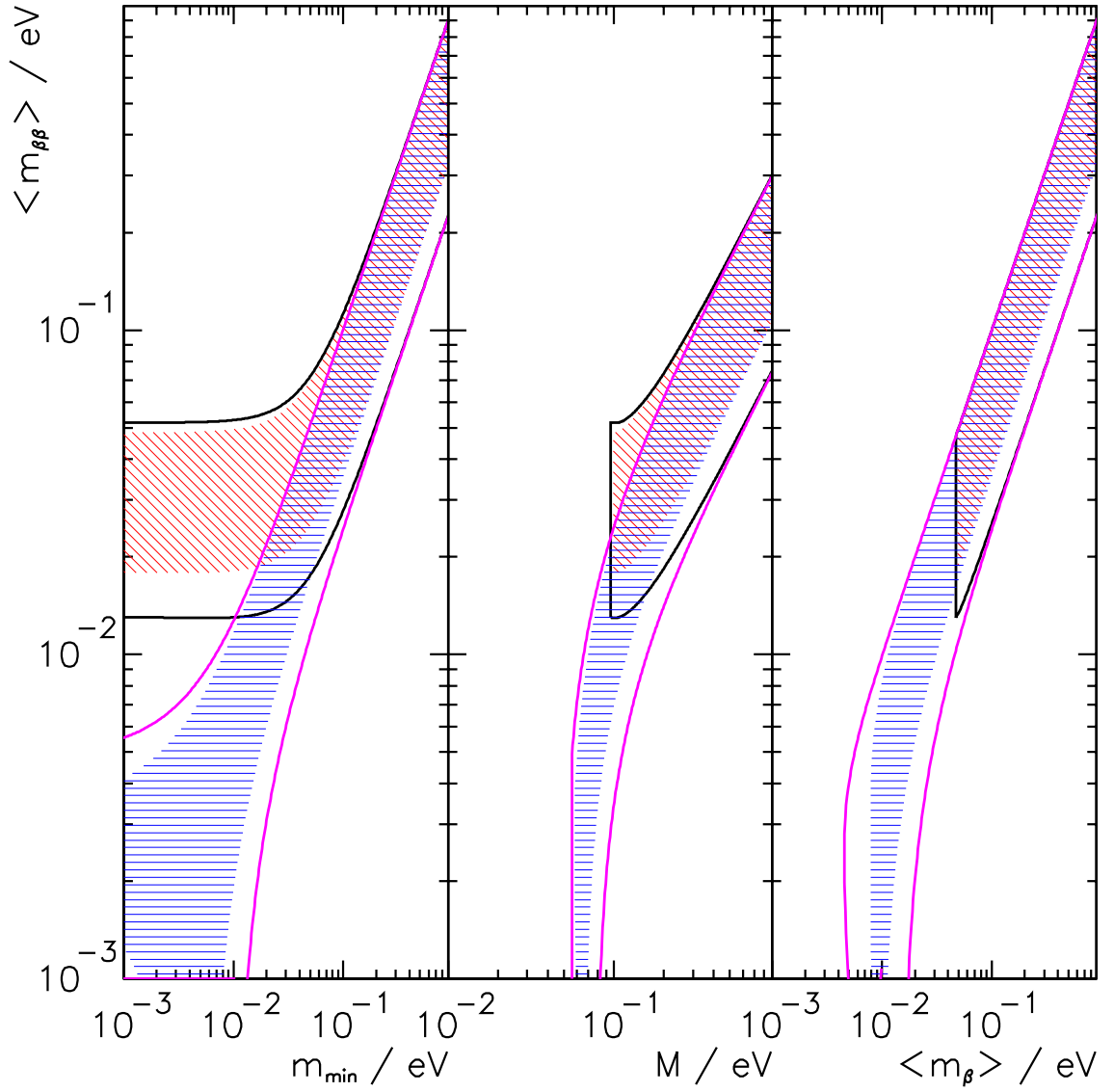


Figure 1.8: The relationship between the effective Majorana mass $\langle m_{\beta\beta} \rangle$ and other fundamental neutrino quantities: the lightest neutrino mass eigenstate m_{\min} , the sum of mass eigenstates $M = \sum m_i$, and the effective single-beta-decay neutrino mass $\langle m_{\beta} \rangle$. Black (magenta) lines indicate the allowed region for the inverted (normal) hierarchy; red (blue) hatches indicate uncertainty for the inverted (normal) hierarchy due to the unknown CP-violating and Majorana phases δ , α_1 , and α_2 . [20]

mass of the material in atomic units, and t is the live-time of the experiment; ϵ is the signal efficiency, b is the background rate (in counts per kg keV year, or some similar units), and ΔE is the energy resolution of the detector at the Q -value; and n_σ is the desired confidence limit, in sigmas, where the standard 90% confidence limit will require $n_\sigma = 1.64$. The scaling of this equation is most accurate when the energy resolution is much smaller than the Q -value and the background is uniformly distributed in energy; however as a rough guide it, combined with phase factor and matrix element estimates, allows us to compare the relative strengths of different $\beta\beta$ experiments.

What it tells us, broadly, is that we should prefer experiments for which:

- A large quantity of highly-enriched isotope can be obtained.
- Signal detection is highly efficient.
- Background contamination around the Q -value is small, and does not scale with detector mass.
- Good energy resolution can be achieved.

Currently the leading choices of materials for $\beta\beta$ searches are: [15]

^{76}Ge has by far the best energy resolution of any material, around 3 keV. It is also possible to grow Germanium crystals with fairly low background, with 0.06 counts/(kg keV year) having been demonstrated. However, it is difficult to grow large, clean crystals of Germanium, and as a result it may not be a scalable material. It also has a Q -value of only 2039 keV, giving it one of the

lowest phase factors among the popular isotopes. GERDA is currently the leading experiment, with data-taking underway.

^{130}Te also has an excellent energy resolution around 7 keV. Its Q -value is also higher than ^{76}Ge , at 2530 keV. TeO_2 crystals must be grown, though, so it may suffer from some of the same challenges of scalability as Germanium. The CUORE experiment is expected to begin data-taking in 2014.

^{136}Xe has a comparatively poor energy resolution around 35 – 40 keV. However, Xenon stands out for its scalability: it is easy to enrich in centrifuges, easy to purify chemically due to its status as a noble element, and it is easy to construct large monolithic detectors because it can be used in a liquid or gaseous state. Background rates of $1 - 2 \cdot 10^{-3}$ counts/(kg keV year) have been demonstrated. There are currently two leading experiments. KamLAND-Zen is the larger experiment, containing 180 kg of active ^{136}Xe ; however, it detects only scintillation, and as a result its resolution is roughly 100 keV. EXO-200 only contains roughly 80 kg of active ^{136}Xe , but it detects energy through both the scintillation and ionization modes, and as a result is able to achieve energy resolutions around 35 – 40 keV.

Table 1.1 tabulates the most current $T_{1/2}^{0\nu}$ limits in all $\beta\beta$ isotopes for which $\beta\beta 0\nu$ limits have been published. Representative limits on $\langle m_{\beta\beta} \rangle$ are also included; these come from one particular set of phase space factors and nuclear matrix element calculations chosen because they have been tabulated for a broad range of isotopes and permit comparisons across all available half-life limits. [47][18] How-

Isotope	$T_{1/2}^{0\nu}$ (years, 90% CL)	$\langle m_{\beta\beta} \rangle$ (eV, 90% CL)	Collaboration
^{48}Ca	$5.8 \cdot 10^{22}$	3.7	ELEGANT IV [59]
^{76}Ge	$2.1 \cdot 10^{25}$	0.24	GERDA [3]
^{82}Se	$3.6 \cdot 10^{23}$	1.1	NEMO-3 [17]
^{96}Zr	$9.2 \cdot 10^{21}$	8.0	NEMO-3 [10]
^{100}Mo	$1.1 \cdot 10^{24}$	0.56	NEMO-3 [30]
^{116}Cd	$1.6 \cdot 10^{22}$	6.1	NEMO-3 [17]
^{130}Te	$3.0 \cdot 10^{24}$	0.34	Cuoricino [11]
^{136}Xe	FILL IN		
^{136}Xe	$1.9 \cdot 10^{25}$	0.16	KamLAND-Zen [36]
^{150}Nd	$1.8 \cdot 10^{22}$	3.4	NEMO-3 [9]

Table 1.1: A listing of the strongest available $\beta\beta 0\nu$ limits; all limits are quoted at 90% confidence. Limits on $\langle m_{\beta\beta} \rangle$ are obtained using phase space factors from [47] and matrix elements from [18], both chosen for the completeness of their tabulations. For ^{136}Xe , this author has taken the liberty of including both Kamland-Zen’s published results and EXO-200 results described in this work.

ever, it should be remembered that errors in the matrix elements may propagate to errors in $\langle m_{\beta\beta} \rangle$ of as much as a factor of two for each isotope. We can see that although there are favoured isotopes, active and successful programs exist in a wide range of isotopes.

As the table indicates, ^{136}Xe has provided some of the strongest constraints on $\langle m_{\beta\beta} \rangle$ in spite of its relatively poor energy resolution. The present work will demonstrate that it is possible, in the EXO-200 detector, to produce significant improvements to the energy resolution through offline denoising of the scintillation signals. This denoising technique will be applied to data from the detector, and will result in a stronger limit on $T_{1/2}^{0\nu}$ from EXO-200 than could be obtained from the same data without denoising. Recent results will be presented which benefit from this and other improvements, as well as increased livetime.

Chapter 2: The EXO-200 Detector

This chapter will describe the physical apparatus of the EXO-200 detector. Section 2.1 will give a broad overview of the detector. Section 2.2 will identify the dominant backgrounds for $\beta\beta 0\nu$ decay, and sections 2.3 and 2.4 will describe methods used to mitigate these backgrounds. The pulse and waveform readout subsystems are described in section 2.5, where discussion of the scintillation readout will be particularly important for subsequent chapters. We conclude with a description of the calibration systems in section 2.6. Throughout, the reader is referred to the detailed description in [14] for more information.

2.1 Overview of the EXO-200 Detector

The EXO-200 detector is a cryogenic experiment containing 175 kg of liquid xenon enriched to 80.6% in ^{136}Xe . Of those 175 kg of liquid xenon, 110 kg are contained within the “active” volume where the detector is fully sensitive to deposited energy from $\beta\beta$ decay, [14] and 94.7 kg are contained within the fiducial volume where we believe the detector’s response is well-understood. Only fiducial xenon will be used in the search for $\beta\beta 0\nu$ decay; this means there are $3.39 \cdot 10^{26}$ atoms ^{136}Xe which will

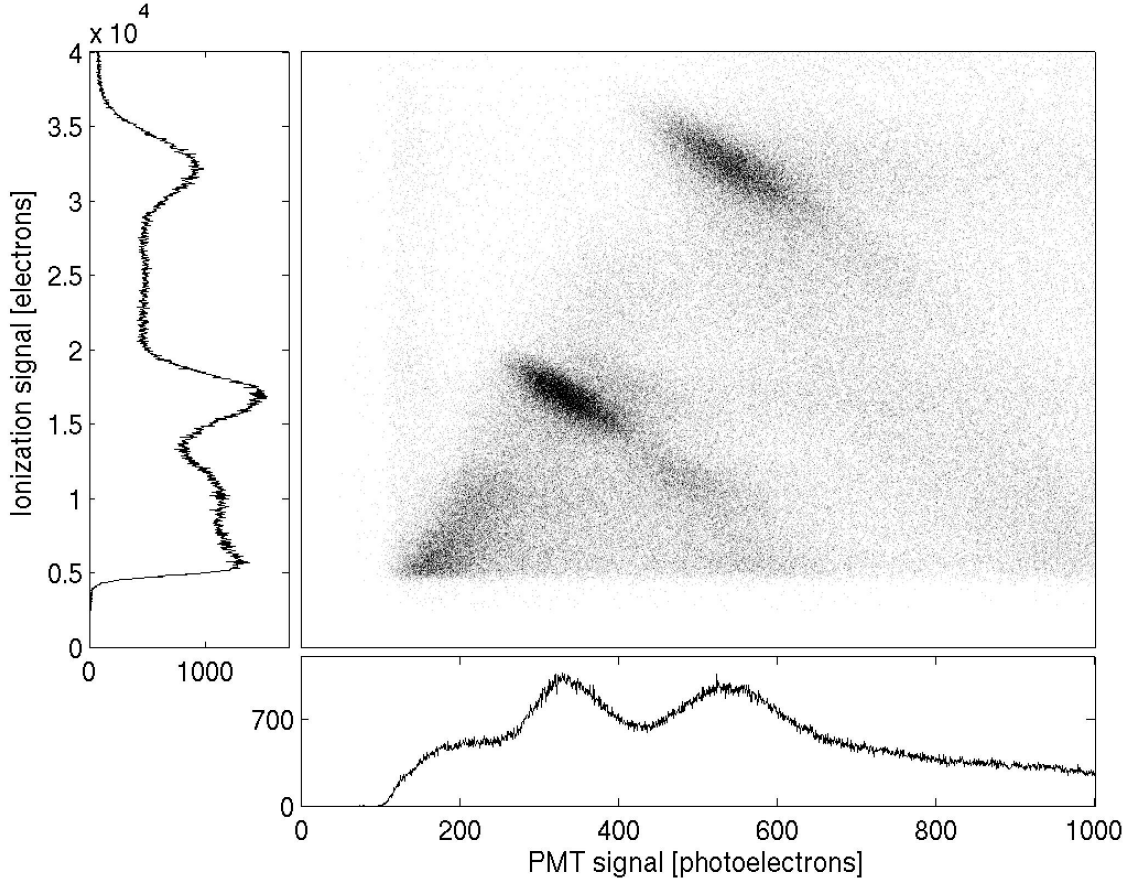


Figure 2.1: A two-dimensional energy spectrum of scintillation and ionization from a testbed liquid xenon experiment under an electric field of 4 kV/cm. The spectrum is from a ^{207}Bi source. Figure reproduced from [27].

be used for setting half-life limits. [31]

Energy deposits in xenon can be measured primarily in two ways: optical photons are emitted from the excitation and de-excitation of atomic electrons of xenon, and xenon atoms are ionized to produce free electrons. It is well-known [27] that liquid noble element calorimeters show significant fluctuations in their separate production of scintillation photons and free electrons, but that these separate quantities are strongly anticorrelated. As a result, is it possible to achieve far better

energy resolution if both light and charge are independently measured than if only one is detected; figure 2.1 illustrates this phenomenon in a testbed liquid xenon experiment, where it is apparent that using light and charge simultaneously lets us observe narrower gamma lines than either individually.

Scintillation will be produced from energy deposits in liquid xenon in all cases; [28] however, to observe free electrons we must exert an electric field on the xenon which drifts the electrons onto a collection anode. The EXO-200 detector is shaped as a cylinder, and the required electric field is produced by placing a cathode grid in the center of that cylinder and anode wires along each of the two endcaps of the detector; such a detector is called a time projection chamber, or TPC. The EXO TPC is shown in figure 2.2. The light is collected by avalanche photo-diodes (see section 2.5 for details) mounted on the endcaps behind the anode wires; the anode wires are thin, so the effect on light collection efficiency is minimal.

The cathode is maintained at a voltage of 8 kV above the anode wire voltage. Field shaping rings encircle the outside of the EXO cylinder to ensure that electric field lines are approximately parallel and the magnitude of the electric field is roughly constant in the bulk volume of xenon. (Near the edges, non-uniformities in the electric field are believed to exist; these are a continuing topic of research.) The cathode and anode wires are separated by roughly 20.4 cm; after accounting for edge effects, we believe that the bulk volume of xenon has an electric field of roughly 374 V/cm.

Electrons in liquid xenon drift at a velocity of $1.71 \text{ mm}/\mu\text{s}$ under an electric field of 375 V/cm. The maximum drift distance from the cathode to the anode

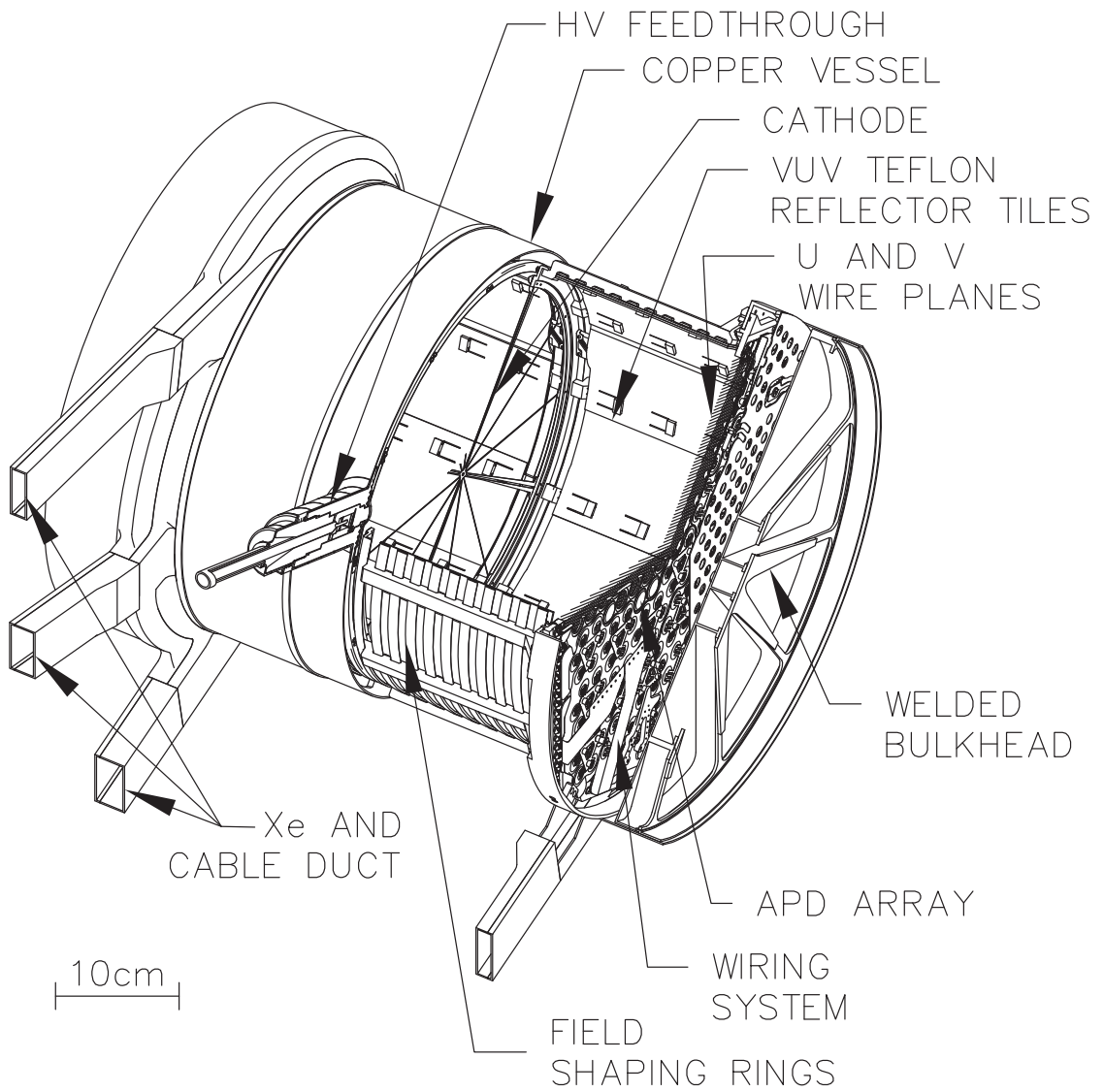


Figure 2.2: Schematic of the inner EXO-200 TPC. Figure reproduced from [14].

is 198.4 mm, resulting in a maximum electron drift time of roughly $116\ \mu\text{s}$. Free electrons are not absorbed by xenon – as for other noble elements, xenon has a low electronegativity, so in a pure xenon detector electrons could drift unimpeded for the time necessary to reach the anode. EXO-200 does have small quantities of electronegative impurities such as oxygen and methane; to minimize the concentration of these impurities, the xenon of EXO-200 is constantly circulated through a chemical purifier which extracts chemically active molecules and permits noble elements to pass through. [14]

It is necessary to associate charge pulses with their corresponding scintillation pulses in spite of their time separation. However, this is easily done provided the time between events is much longer than $116\ \mu\text{s}$. Thus, sources should be calibrated to produce a data rate in xenon no higher than roughly 1 kHz. By measuring the time difference between the observation of scintillation and the collection of charge, we can measure the position of the energy deposit along one dimension. Our time resolution permits us to reconstruct this position coordinate with an accuracy of 0.42 mm. [4]

At the anode, there are two sets of parallel wire planes. The first (closer to the cathode) parallel wire plane is called the “v-wire” plane, and the second plane is called the “u-wire” plane. The voltages of the two wire planes are set so that no electric field lines terminate on the v-wires, and all field lines instead penetrate through the v-wire plane and terminate on the u-wires. When charge is deposited within the liquid xenon and drifts toward the anode, it induces a current on the v-wires as it passes by, and then produces a current on the u-wires as it is collected.

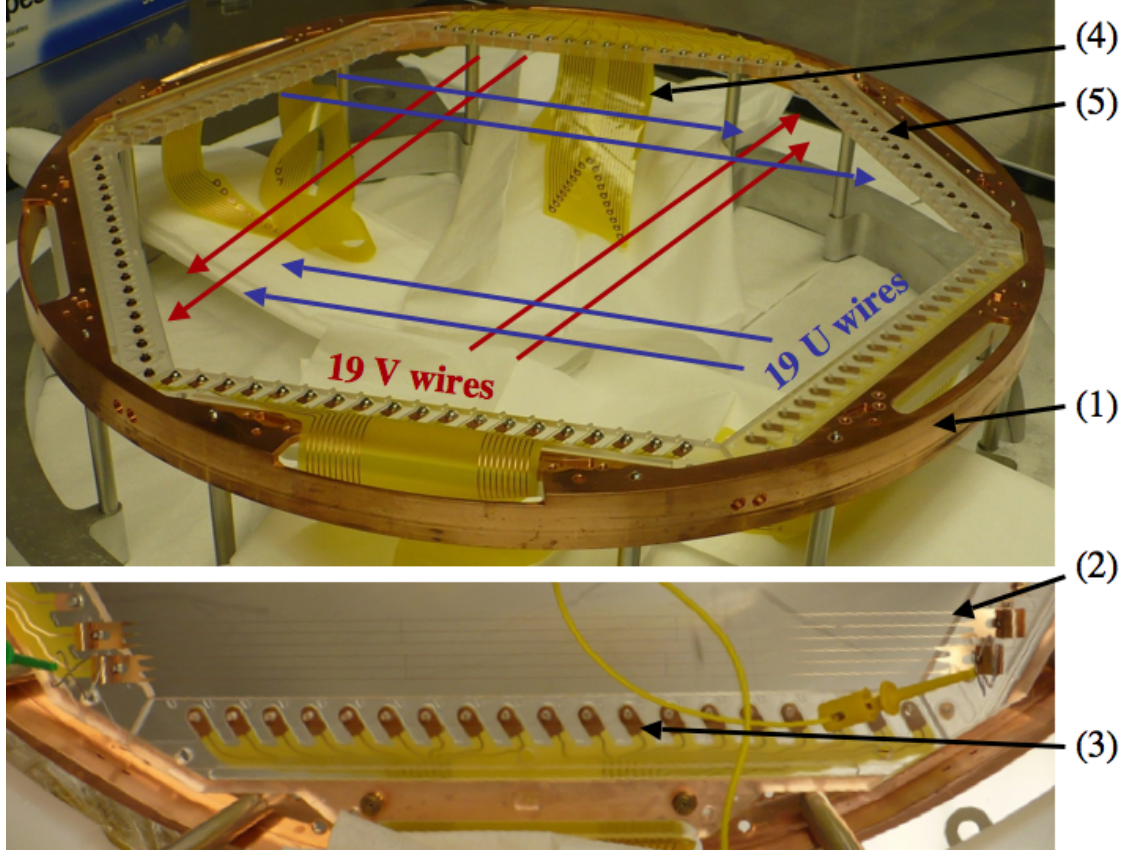


Figure 2.3: Anode collection wires (u-wires) and induction wires (v-wires) from EXO-200. (1) and (5) indicate the wire support frame; (2) indicates the wires themselves, constructed as gangs of three wires; (3) illustrates the attachment between wires and the support frame; (4) shows the return cables from the wires to the data acquisition system. Figure reproduced from [14].

Figure 2.3 indicates the relative orientation of the u-wires and v-wires. By detecting which (set of) wires observes a current pulse in both the u-wire and v-wire channels, it is possible to identify the two-dimensional location on the anode where the charge arrived. Channels in each plane are 9 mm wide; by taking advantage of signal sharing which sometimes occurs between u-wires and always occurs between v-wires, we are able to achieve position accuracy of 2.4 mm perpendicular to the u-wires and 1.2 mm perpendicular to the v-wires. [4]

Energy which appears to come from a single location is called a charge deposit “cluster” or “site”, and events are classified according to the number of sites they produce (their “multiplicity”) as either single-site or multi-site. We will see in section 2.4 that this provides a powerful tool for background rejection.

This section has described how energy deposits in the liquid xenon are transformed into scintillation and charge, which are then observed by APDs and anode wires, respectively. These observations are sufficient for us to reconstruct the positions and magnitudes of energy deposits in the liquid xenon.

2.2 Backgrounds to $\beta\beta 0\nu$ Decay

^{136}Xe has a relatively high Q -value compared to most other $\beta\beta$ decays; in the neutrinoless mode the two emitted electrons will share 2456.7 keV. [31] This means that the sensitivity of $T_{1/2}^{0\nu}$ to the mass of the neutrino is good in ^{136}Xe , as was described in section 1.5; it also means that energy spectrum around our Q -value will naturally be relatively free from most sources of background radiation. This section will

identify the types of background which can be expected to affect our $\beta\beta 0\nu$ search; subsequent sections will describe methods of mitigating those backgrounds.

It is common for alpha decays to have energies well in excess of our Q -value. As a result, we might expect all alpha decays to be possible backgrounds to $\beta\beta 0\nu$. However, alpha particles are stopped rapidly by even a small quantity of shielding, and as a result the only alpha decays which can be observed in the detector are those from sources which are dissolved into the xenon. Radon is the only alpha-emitting noble element, and only ^{222}Rn is sufficiently long-lived to diffuse into the xenon; as a result, we only expect ^{222}Rn and its daughter products to contribute a significant quantity of alpha decays to the detector.

Cosmic rays produce high-energy muons; these are another source of high-energy backgrounds. Muons will produce a streak of energy in the detector rather than discrete clusters; generally they will have sufficient energy to pass fully through the detector. We can expect that such an event will look substantially different from a $\beta\beta 0\nu$ event, and generally will deposit substantially more energy than our Q -value as well.

More interesting are the associated by-products from the passage of such a high-energy particle, called spallation products. Although high-energy muons can produce a range of fission products, the most significant spallation product of muons will be neutrons which can diffuse into the detector and activate materials there. ^{134}Xe and ^{136}Xe are both present in significant quantities in the TPC, and by activation can be converted into the radioactive isotopes ^{135}Xe and ^{137}Xe respectively. ^{135}Xe has too little energy to be a background to $\beta\beta 0\nu$ decay, but ^{137}Xe undergoes

β decay with a maximum energy of 4173 keV. We will see that indeed ^{137}Xe is a significant source of background in our detector. [31] [6]

Some sources of background are intrinsic to the $\beta\beta 0\nu$ search. Any isotope which can undergo $\beta\beta 0\nu$ decay can also undergo $\beta\beta 2\nu$ decay, and the endpoint of the $\beta\beta 2\nu$ spectrum is necessarily at our Q -value. The only means of reducing background from $\beta\beta 2\nu$ is to improve the energy resolution of the detector so that fewer such decays can mimic $\beta\beta 0\nu$ decay. Fortunately with the energy resolution exhibited by the EXO-200 detector, $\beta\beta 2\nu$ is subdominant to other backgrounds by many orders of magnitude. [31]

Similarly, it is hypothetically possible for neutrino absorption to stimulate a $\beta\beta 2\nu$ decay by the reaction

$$2d + \nu_e \rightarrow 2u + 2e^- + \bar{\nu}_e \quad (2.1)$$

which is obtained from the $\beta\beta 2\nu$ reaction by taking one neutrino from product to reactant in equation 1.1. The observable spectrum is similar to the standard $\beta\beta 2\nu$ spectrum, but with the endpoint shifted upward by the energy of the absorbed neutrino. The cross-section for this reaction is expected to be well below the cross-section for standard $\beta\beta 2\nu$, so this will be a negligible background well beyond the current generation of $\beta\beta$ detectors.

A gamma background which is particularly detrimental to EXO-200 comes from ^{214}Bi , a member of the radium decay chain. ^{214}Bi emits a gamma particle at 2448 keV, which with our energy resolution is indistinguishable from our Q -value. It will occur as a daughter product of ^{226}Ra , which has a half-life of 1600 years

and generally will be supported by ^{230}Th (75,000 years), ^{234}U (250,000 years), and ultimately by ^{238}U ($4.5 \cdot 10^9$ years) which has a primordial abundance in the Earth's crust and most natural materials. [21]

Most other backgrounds to $\beta\beta 0\nu$ will be gamma decays from ^{208}Tl , a member of the thorium decay chain. ^{208}Tl emits a gamma particle at 2614.5 keV; with our Q -value of 2456.7 keV, the two energies are separated by 157.8 keV. [21] The energy resolution (in σ/mean) of EXO-200 has typically been 1.5–2% at these energies, [31] meaning that separation between the central values of the gamma lines will be $3-4\sigma$. The thorium decay chain is supported by ^{232}Th , which has a half-life of 14 billion years and has a significant abundance in most natural materials; we may expect it to contribute a significant fraction of our radioactive backgrounds, and as a result we may expect that some events from the 2615 keV ^{208}Tl line will leak across those $3-4\sigma$ and act as backgrounds. Keeping our energy resolution at or below 1.5% may be of significant interest in reducing contamination from this gamma line.

The compton edge is a well-known feature of gamma decay spectra; it originates from gammas which enter a calorimeter, scatter once, and escape from the detector. The maximum energy which a gamma of incident energy E_{inc} can deposit in a single compton scatter is: [25]

$$E_{dep} = \frac{2E_{inc}^2}{m_e c^2 + 2E_{inc}}. \quad (2.2)$$

When the incident gamma has an energy of 2615 keV, the compton edge will lie at 2382 keV, or only 74.5 keV below our Q -value. Using the same estimate that our energy resolution is around $1.5-2\sigma$ at these energies, this separation is only around

$1.5 - 2\sigma$. It is clear that improving our energy resolution may have a strong impact on background contamination by improving separation from the Compton edge of the ^{208}Tl as well as its full deposit line.

We have here identified some of the primary backgrounds which can be expected in the EXO-200 detector, noting particularly those which can be significantly reduced through improvements to the energy resolution. The following sections will identify some of the mechanisms used to further minimize these backgrounds.

2.3 Passive Background Rejection

When we describe the methods by which backgrounds to $\beta\beta 0\nu$ are reduced, we can distinguish between two classes: passive methods in which backgrounds are reduced by reducing the amount of background reaching the xenon in the first place, and active methods in which backgrounds are observed by the detector but discriminated from $\beta\beta$ signal based on identifying characteristics. In this section, passive methods will be described; section 2.4 will describe the active methods of background rejection.

The simplest and most immediate method of reducing the presence of backgrounds is by exploiting the self-shielding properties of xenon. The xenon itself is extremely pure due to the ease of chemical purification; as a result most backgrounds will be external to the xenon. Since external gammas are attenuated by dense materials such as liquid xenon, we can expect that the xenon near the center of the detector will be exposed to less background than the xenon near the TPC

Xenon

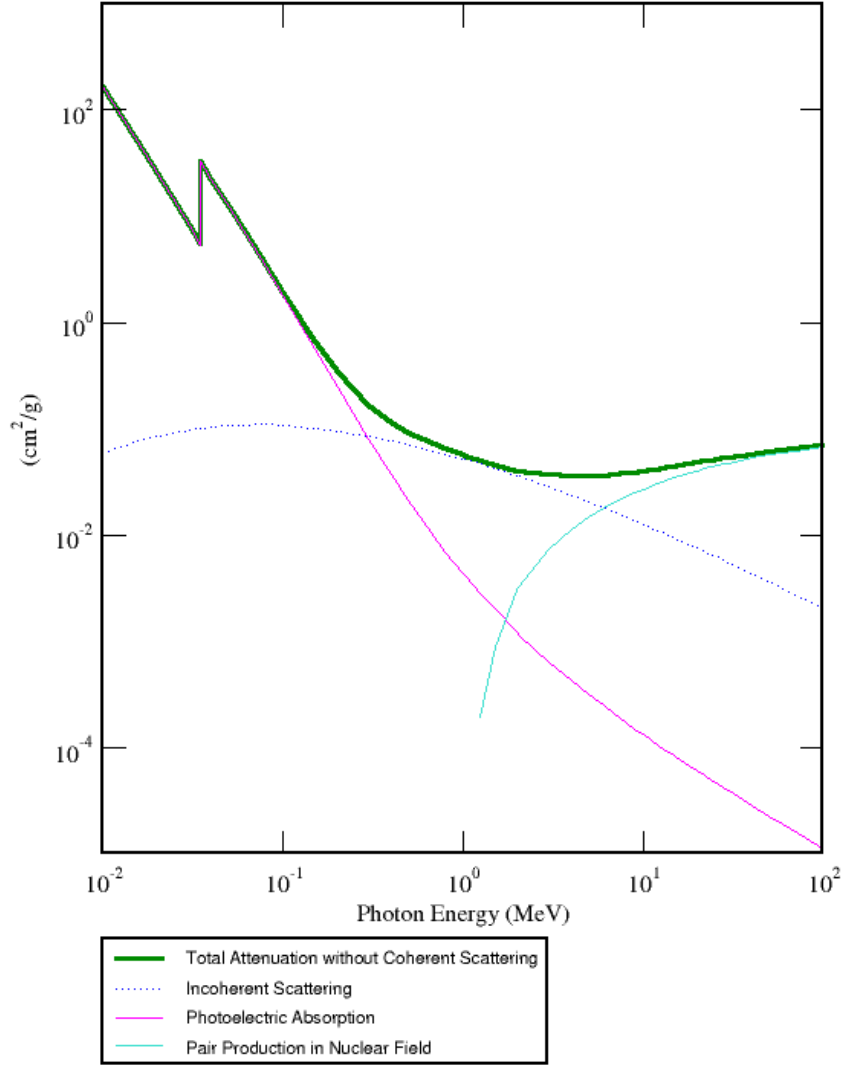


Figure 2.4: X-ray attenuation lengths in xenon. Compton (incoherent) scattering, pair production, and photoelectric absorption are shown, along with their combined attenuation length; coherent (Rayleigh) scattering is omitted because it produces no observable energy deposit. The vertical axis, in cm^2/g , can be multiplied by the density of the xenon to derive an attenuation factor per unit length. Figure produced by [19].

walls. This is one of the primary advantages of xenon as a source: it is easy to construct a large monolithic detector, maximizing the quantity of xenon which is shielded from the walls.

Figure 2.4 shows the attenuation lengths of gammas at a range of energies in xenon. Photoabsorption and pair production both convert the gamma entirely into short-ranged electron and positron carriers; compton (incoherent) scattering results in only some deposited energy, with the rest remaining in the gamma which will rebound and continue on its path. Liquid xenon has a density around 3 g/cm^3 , which means that the minimum attenuation factor is roughly $0.1/\text{cm}$ at energies around 4 MeV . Unfortunately, this is the same order of magnitude as our Q -value at 2456.7 keV , which means that at the energy of interest to us self-shielding is minimally effective. Nevertheless, there will still be some reduction in backgrounds deeper in the interior of the xenon, and lower-energy gammas will be attenuated more effectively. [19]

To reduce the quantity of background around the detector, all materials near the xenon were carefully screened for radioactive contamination. Backgrounds from ^{40}K , ^{232}Th , and ^{238}U were cataloged for all materials which have unshielded line-of-site access to the detector system. Requirements for ^{238}U were particularly stringent because its daughter products include the ^{214}Bi background which cannot be resolved from $\beta\beta 0\nu$ decay. For the copper of the TPC vessel and the lead shielding around the detector apparatus, samples from a range of companies and mines were tested to locate an optimal choice of material source. For the wires which were placed into the TPC, a range of manufacturing methods were tested; improvements to a

photo-etching scheme were identified which led to a reduction in background from these materials. This thorough material screening research is one of the distinctive processes which has enabled EXO-200 to fully reach its background targets. Details of the quantification of and constraints on material radioactivity can be found in [48].

Beyond selecting extremely clean materials, it is possible to reduce backgrounds by minimizing the mass of these external materials. Most notably, EXO-200 is constructed with a copper TPC which in most places is only 1.37 mm thick; to maintain structural integrity, supporting structures are welded to the TPC where needed, and it was possible to keep the total mass of copper below 30 kg. [14]

Some materials were dispensed with entirely. Typically, silicon APDs are encapsulated with ceramic to isolate them from water contamination and provide electrical insulation. However, this ceramic material would have contributed backgrounds. Instead, the APDs were delivered “bare,” without any encapsulation, and protected from water by storing them in a dry-nitrogen container. Liquid xenon itself serves as an excellent electrical insulator, ensuring the APDs would function properly during detector operations. [53]

Another example of material avoidance comes from the cabling from wires and APDs to the electrical amplifiers and digitizers. These electronics contain many high-background plastics and other complex materials, so they were placed outside the lead shielding rather than placing them close to the TPC. Wires connect the sensors in the TPC to these electronics; ordinarily such long wires would need to be shielded by coaxial cabling to minimize cross-talk noise. However, the coaxial cabling was expected to contribute a high quantity of radioactive background, and

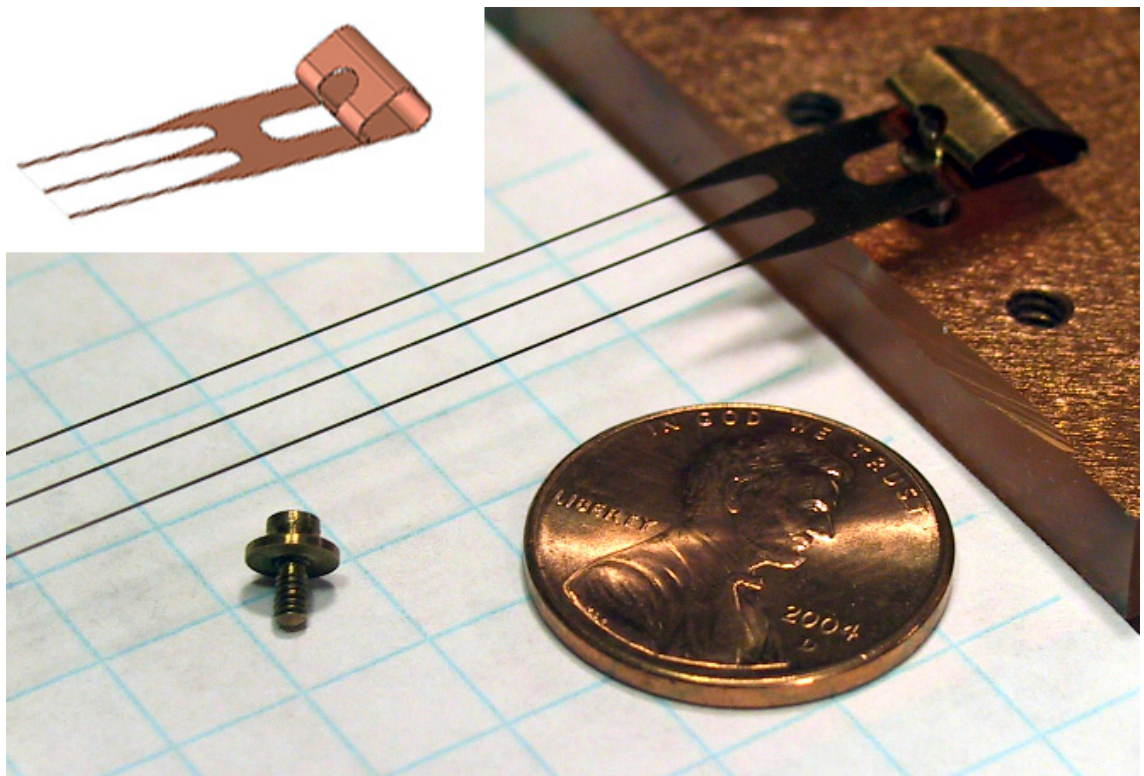


Figure 2.5: Wire triplet, read as one channel. Figure from [14].



Figure 2.6: APDs ganged together (bottom right). Wiring to the front-end electronics are visible as yellow “tape.” Figure from [14].

was omitted; instead wires are partially shielded by surrounding them with inactive wires which reduce cross-talk noise. Furthermore, the total number of wires needed was reduced by ganging together triplets of u- and v-wires and groups of six or seven APDs into single channels; this reduced the fineness of event information available in analysis, but reduced the quantity and complexity of material placed near the detector. See the wire gangs in figure 2.5; APD gangs and wiring to external electronics are visible in figure 2.6. [14]

We can imagine the TPC being shielded by nested layers of clean material designed to prevent backgrounds from reaching the xenon. In the innermost layer, HFE-7000 refrigerant (see figure 2.7) is used to maintain the detector at liquid xenon temperatures and maintain similar pressures inside and outside of the copper vessel walls. However, the HFE has a significant quantity of hydrogen, which makes it an

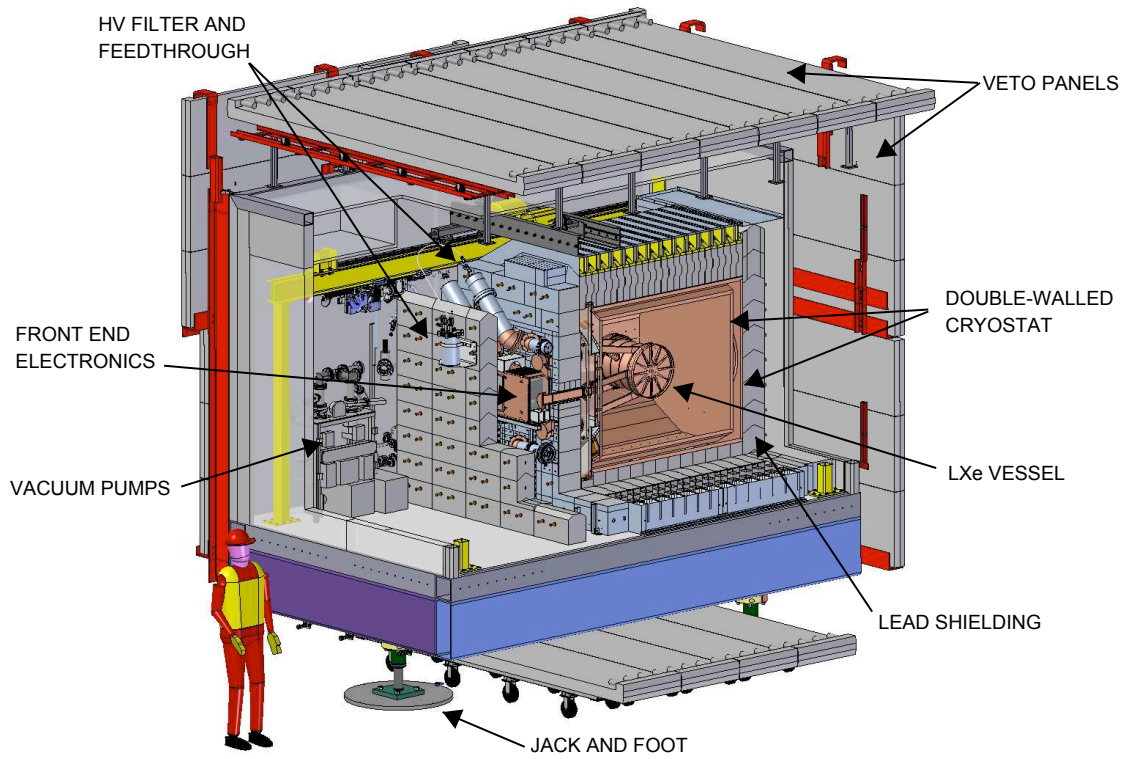


Figure 2.7: A cutaway schematic image of the TPC and surrounding materials in the cleanroom. HFE-7000 refrigerant is contained between the cryostat and the LXe vessel. Figure from [14].

excellent stopping agent for thermal neutrons. The refrigerant layer is more than 50 cm thick, so most neutron products of muons have time to thermalize and absorb onto hydrogen before reaching the TPC and producing ^{137}Xe . [14]

The following layer of shielding is lead, shown in figure 2.7. This is a common approach for low-background experiments due to the high density of lead, which enables it to stop gammas in a short distance. Lead bricks surround the detector on all sides; the bricks were designed with an interlocking shape to ensure no seams between bricks left a line-of-sight path from external sources into the TPC. In the front of the detector, where plumbing must enter and exit the detector through the first lead wall, a second lead wall is assembled with the purpose of blocking any line-of-sight paths into the TPC through the plumbing gaps of the first wall. [14]

Outside of the lead shielding, restrictions on materials are less stringent; however, control over the presence of materials is still desired. The entire apparatus is contained inside a class-1000 clean room facility. Entire facility is located in the Waste Isolation Pilot Plant (WIPP) facility, a DOE-owned waste repository located in a salt mine in Carlsbad, NM. The overburden of the facility is 1585 m water-equivalent, leading to a significant reduction in muon rate compared to the rate observed on the Earth’s surface. Figure 2.8 demonstrated the reduction in muon flux due to depth at the WIPP site and a selection of other underground science facilities.

These passive approaches have all been demonstrated to reduce the rate of backgrounds depositing energy in the liquid xenon. The following section will describe an active set of background-rejection approaches.

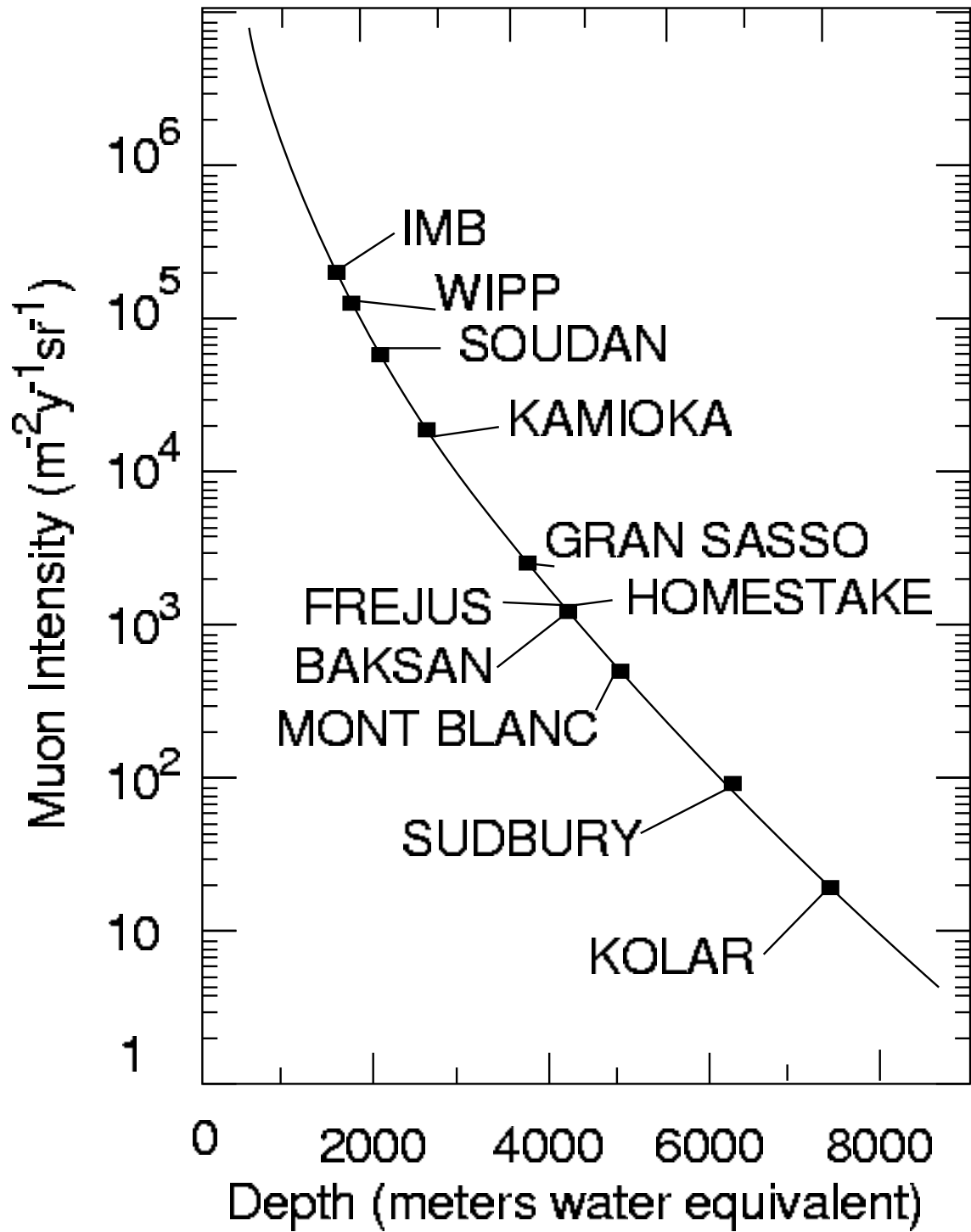


Figure 2.8: Muon flux as a function of depth, in meters water equivalent. The WIPP site is indicated in relation to other underground science facilities. Figure from [29].

2.4 Active Background Rejection

This section will describe the “active” forms of background reduction employed by the EXO-200 detector, in which events are observed but discriminated from $\beta\beta$ events based on their characteristics. We neglect the energy resolution as a background rejection tool here, since section 2.2 has already described how the impact of certain backgrounds may be reduced by improvement to the energy resolution.

A number of cosmogenic backgrounds have been described: neutrons and fission products can be produced from muons penetrating down to the depth of the WIPP facility. To reject many of these events, it is sufficient to detect the passage of a muon anywhere near the detector. This is accomplished by means of a set of veto panels surrounding the cleanroom on four of six sides. Muons are detected by coincident panel hits on opposite sides of the detector, with $(96.0 \pm 0.5)\%$ efficiency for detecting muons which traverse the TPC. [14] Further reduction in the impact of muon-induced backgrounds is achieved by searching for muon tracks observed within the TPC itself, and also by applying a one-second coincident event cut as a catch-all for spallation products of muons which may decay multiple times in rapid succession.

Alpha decays within the xenon were identified as another possible source of high-energy backgrounds. However, it is possible to discriminate alpha decays from beta and gamma decays based on the properties of their energy measurement. Alphas produce a far more dense energy deposit in liquid xenon than betas and gam-

mas; the dense collection of free electrons and ionized xenon has a higher rate of recombination than is observed with betas and gammas, which translates into a higher ratio of scintillation to ionization for alphas than for betas and gammas. This feature can be used to eliminate nearly all alpha background.

Discriminating between gamma, beta, and double-beta decays is more difficult. Gammas generally deposit their energy by ionizing atomic electrons of xenon, so in this way their behavior may closely mimic that of betas. However, most gammas within our energy window (and particularly gammas near our Q -value) will be attenuated by compton (incoherent) scattering, as can be seen in figure 2.4. By this process the gamma will deposit energy in more than one discrete location, which beta and double-beta decay will not do. We are generally able to resolve the positions of these energy deposits and classify the event as multi-site; more than half of external gamma backgrounds are classified as multi-site, whereas the fraction of $\beta\beta$ events classified as multi-site is only 5%. [4]

These are some of the basic techniques employed to reduce backgrounds in offline analysis, and the detector has been designed with the goal of facilitating their application. Although the passive methods of background reduction cannot be improved once construction of the detector is complete, it may be hoped that these active methods may be improved or extended through further analysis. These improvements could have the potential to reduce backgrounds in data which has already been collected, making them a particularly attractive target now that EXO-200 has collected a significant dataset.

2.5 Pulse Amplification and Waveform Readout

After an APD or wire collects a signal, how is it converted into a waveform on disk? This section will provide a brief overview of the EXO electronics and waveform readout. Particular attention will be paid to subsystems which are relevant to the denoising algorithm described in chapter 3, including the APDs and front-end electronics.

The APDs are constructed from a silicon semiconductor which is doped and biased so that: [51]

1. Photons arriving at the active surface of the APD can deposit their energy by generating an electron-hole pair in the silicon semiconductor.
2. The electron-hole pair is drifted apart by an electric field. Here the electric field is low and the silicon is doped to be of p-type (electron-poor), ensuring a similar number of electron-hole pairs are produced by all photons before any amplification can occur.
3. The electron enters a high-field n-type (electron-rich) region of silicon, and liberates additional electron-hole pairs in an avalanche effect.
4. Electrons reach another p-type region of silicon, and then are collected on a cathode as an output current pulse.

These phases of the APD are illustrated in figure 2.9.

EXO-200 had 851 APDs specially made with pure aluminum and no casing;

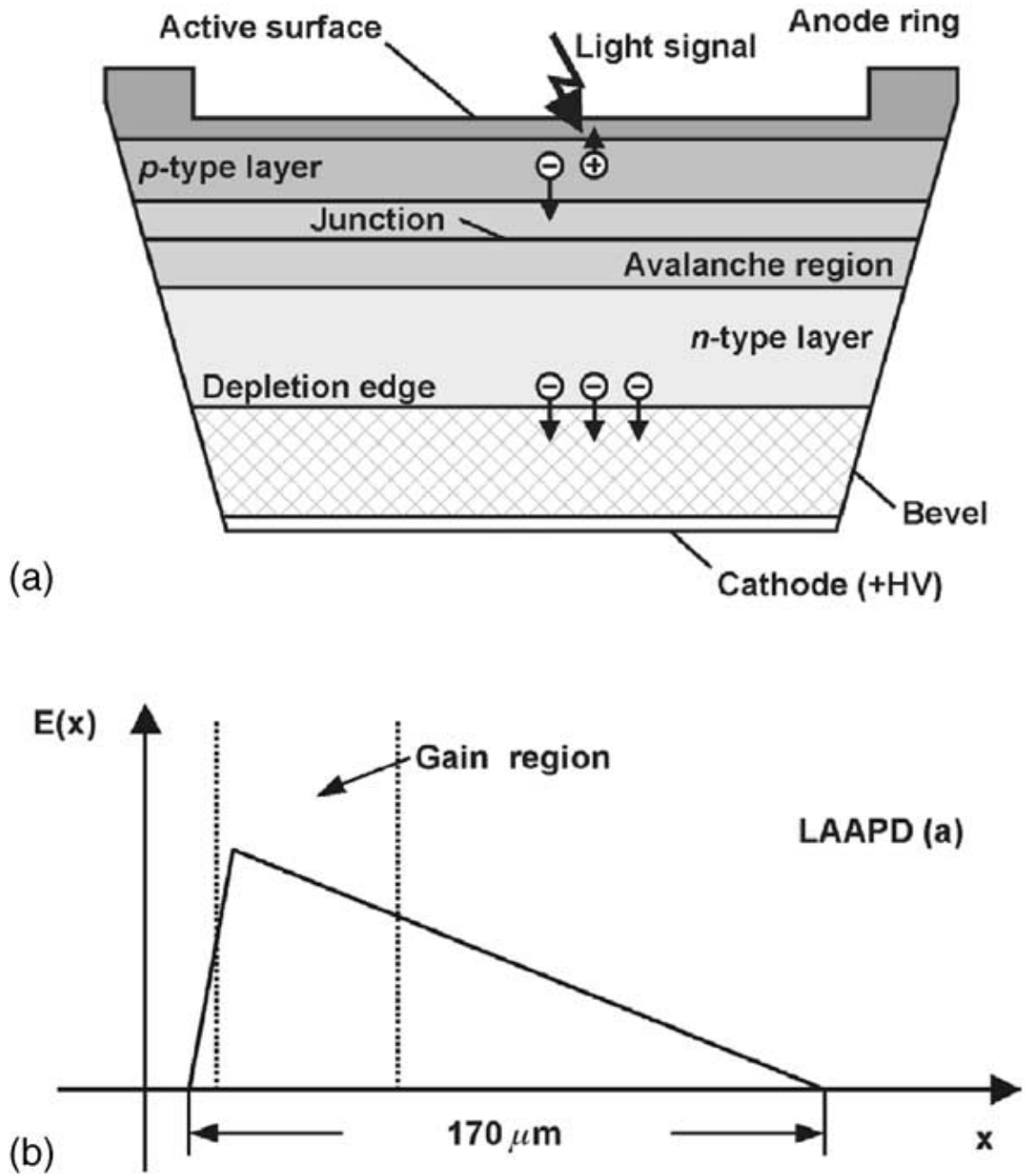


Figure 2.9: A cross-sectional schematic of an APD is shown in (a). The electric field as a function of depth is shown in (b). Figure from [51].

from those, 468 were selected for use based on their excellent noise and gain properties. [53] For each APD, the relative quantum efficiency for electron-hole production and the bias voltage needed to ensure the product of gain and quantum efficiency equals one hundred were both measured. APDs were assigned to gangs and pies so that their products of gain and quantum efficiency could be set to one hundred as uniformly as possible. [52] It is impossible to make the gain of all APDs perfectly uniform because of limitations in the configurability of the APD bias voltages in EXO-200. [14] We note that the current operating conditions of EXO-200 establish a higher operating gain of around 200 – 300; this may have the side effect of making the product of gain and quantum efficiency less uniform in each APD gang.

The means by which wires deliver charge is somewhat simpler than for APDs. In the u-wires, charge is collected and delivered directly to the electronics as current. In the v-wires, no net charge is delivered, but transient currents are induced in such a way that the v-wire remains at a constant voltage; these transient currents produce a pulse with a different shape than the u-wire pulses, such that the integral (before shaping) is zero. We note that the pulse shape depends on the path of electrons in the xenon – both sets of wires deliver a current signal which is induced from approaching charge before the charge deposits on the u-wires, and the path of the electrons can affect the timing of that induced current. Furthermore, the pulse magnitude will depend on the point of origin of the electrons because at that location positively-charged ions will remain long after the electrons have been collected, and the ions will hold a net charge on the u- and v-wires which is released only on much longer timescales than we can observe in our waveforms. (Ions, being much more massive

than electrons, drift much more slowly through liquid xenon.) [5] [39]

The front-end electronics, located between the two lead walls in front of the detector, consist of four stages: [14] [40]

1. A transimpedance amplifier which converts the current signals from the detector to voltage signals. This stage applies amplification and a differentiator with a long time constant (roughly $60\mu\text{s}$ for wires and $300\mu\text{s}$ for APDs).
2. Two differentiators and two integrators, to improve signal-to-noise ratio for real-time triggering. For the u-wires, each integrator has a time constant of $1.5\mu\text{s}$ and each differentiator has a time constant of $40\mu\text{s}$; for the APDs and v-wires, each integrator has a time constant of $3\mu\text{s}$ and each differentiator has a time constant of $10\mu\text{s}$. Some signal amplification is applied at this stage as well.
3. An analog-to-digital converter (ADC) converts voltage pulses into 12-bit digital waveforms. The full 12-bit scale is equivalent to 2.5V. Digitization is performed at a rate of 1MHz.
4. A triggering module reads the digitized waveforms, and issues a trigger to the data acquisition (DAQ) when signals exceed programmable thresholds. The triggering module can also accept external requests for a “solicited” trigger.

In most cases, when a trigger is issued, the DAQ will write out 2048 samples (just over 2ms) from all waveforms, centered on the first sample responsible for the trigger.

In cases where the data rate is high, waveforms can be truncated to a length less

than 2048 samples.

2.6 Calibration Systems

Energy resolution has been discussed as a general feature of the detector which can have a significant impact on the reduction of backgrounds. A significant factor in the resolution is the quality of the detector's absolute energy calibration. This is particularly important because of the long lifetime of the experiment. If the calibration of the detector drifts over time, this will be perceived in the cumulative low-background spectrum as a smearing out of energy peaks, worsening the effective resolution. This section will describe some of the techniques used to produce an accurate time-dependent calibration for the detector.

The most critical type of calibration comes from external gamma sources. The EXO experiment was constructed with a guide tube passing through the HFE refrigerant and wrapping loosely around the TPC, as shown in figure 2.10. Currently available to the experiment are sources containing ^{228}Th (primary gamma line at 2615 keV), ^{60}Co (simultaneous emission of gammas at 1173 and 1332 keV), and ^{137}Cs (gamma line at 662 keV); additionally, a new ^{226}Ra source became available in the summer of 2013 (numerous gammas, including 2448 keV). Together these sources allow us to observe the detector response to external gammas with known energies, permitting energy calibrations, energy resolution measurements, and verification of monte carlo simulations.

All of these sources have been used periodically (roughly every three months)

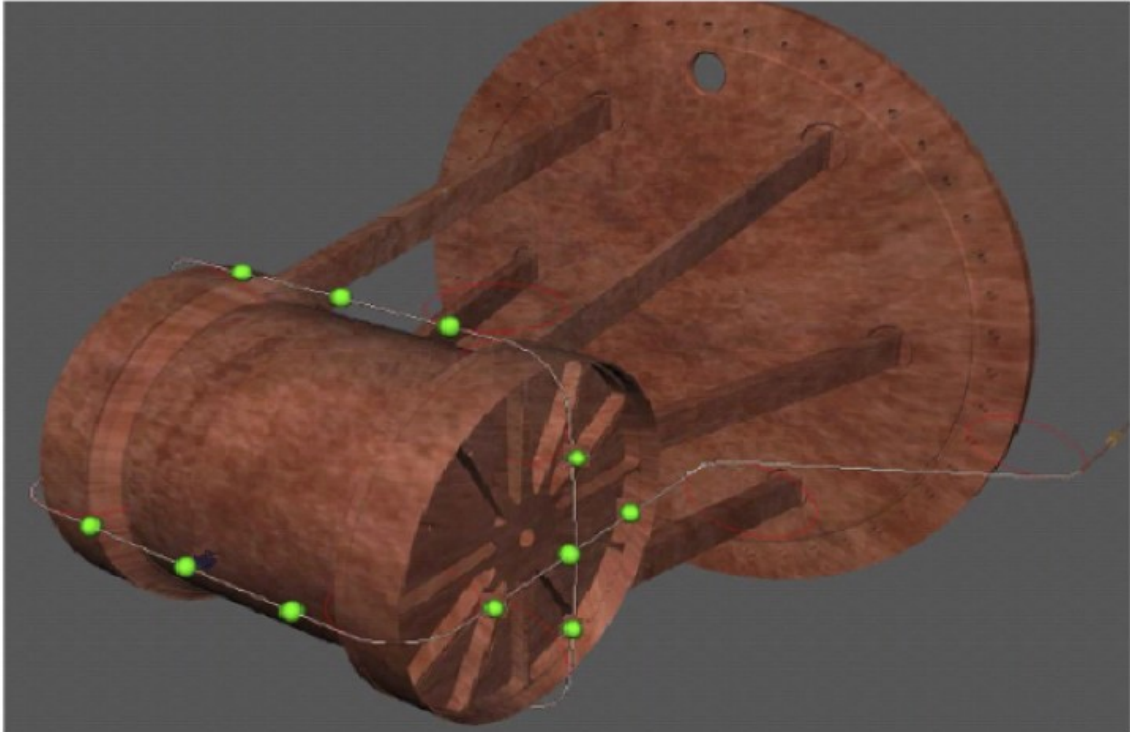


Figure 2.10: A guide tube permits gamma sources to be inserted to known locations near the TPC. Green dots indicate a few of the locations where the source may conventionally be placed. Figure from [14].

in calibration campaigns intended to permit a full characterization of the detector response across a wide energy range at a particular moment. Additionally, the ^{228}Th source has been deployed regularly, typically three times every week for 2 – 3 hours, for the entire history of the dataset. The fullness of the ^{228}Th source dataset, along with the benefit that it contains a relatively clean high-energy gamma line, make it a particularly convenient dataset to work with; it will play a prominent role in the generation of our lightmap described in chapter 5.

Other calibrations are also performed on the detector, and may in some instances supplement the information provided by the source data. To perform direct calibrations of the APDs, one APD mount contains a light diffuser rather than an APD device; the diffuser can be illuminated by a laser through optical fibers. [14] The APDs can be calibrated by pulsing the laser while varying the voltages of the APDs to adjust their gain, and in this way it is possible to measure the gain of the APDs in situ. A laser system was available at the beginning of the dataset used in this work; however, there were certain deficiencies to the laser system which made its data difficult to use, and as a result a new and more accurate laser system was commissioned in August 2012. Laser data in some form has been collected roughly once per week for much of the time containing this dataset. Composing the laser data into a coherent history of the APD gains is an ongoing project, but it has been possible to use APD gain information from specific moments to inform the APD noise model described in section 3.2.1.

Just as it is possible and desirable to learn the gain of the APDs separately from other components using laser runs, it is also desirable to learn the gain of the

electronics amplifiers separately from other detector function. This is accomplished using “internal calibrations:” simultaneously with software triggers solicited by the data acquisition, a switched capacitor injects charge into the input of the electronics cards of individual or groups of channels. [43] The resulting waveform is then read normally, and the size of the output pulse reflects the gain of the electronics. This system is employed approximately daily on all channels; it provides an excellent measure of relative changes in the electronics gains over time. However, the absolute gains cannot be extracted accurately from this data, and must be obtained by other means.

To measure the absolute gains of the electronics, the front-end cards must be physically removed from their enclosures; when this is done, it is possible to use a more accurately calibrated capacitor to make an absolute gain measurement in the same style as the internal calibrations. These “external calibrations” can be performed in two flavors: manual, in which a capacitor calibrated to 1% is used to inject charge, and automatic, in which the calibration is automatically performed at a much faster rate but accuracy is limited to 2 – 3%. Due to the inconvenience of removing electronics cards from their enclosures, this process has only been performed once in February 2011. Both styles of external calibration were performed on the u-wire channels. Less demanding needs for resolution on the v-wires and APD channels were foreseen, so only the less accurate automatic calibration was performed on those channels. The data from these runs has not been fully analyzed, but may be revived in future analysis to permit a better calibrations of the subsequent internal calibrations which were taken. [5]

2.7 Summary

This chapter has described some of the key characteristics of the construction and operation of the EXO-200 detector. Background reduction has been a theme throughout – EXO-200 has employed both passive and active methods to reject background and improve its sensitivity to $\beta\beta 0\nu$ decay. Energy resolution has also been described as a key element in background rejection because with sufficiently precise energy measurements, many backgrounds will no longer resemble $\beta\beta 0\nu$ decay. Chapter 3 will begin our discussion of a new scheme to extract more accurate energy measurements from the APD channels, thereby improving our overall energy resolution and the sensitivity of the limit on the $\beta\beta 0\nu$ half-life we will be able to demonstrate.

Chapter 3: Denoising Theory

As has been discussed, the performance of double-beta experiments like EXO-200 is partially determined by their energy resolution. In EXO-200, the energy resolution is limited by the scintillation energy resolution, which in turn is dominated by electronic noise in the APDs. Significant effort must therefore be expended to understand and reduce the noise in the scintillation signals.

When the noise in the APDs is studied, it was found that the individual APD channels met targetted root-mean-square noise levels of 2000 electrons. However, rather than observing the noise on summed APD channels increase proportionally to the square root of the number of channels, the summed APD noise is roughly two to three times higher than projected. This worse-than-expected scaling of the noise with channels indicates that noise across different channels is correlated, and further analysis confirms that the bulk of the noise on an unshaped APD waveform is correlated with other channels, as shown in Figure 3.1. There are many possible sources of coherent noise in the hardware, and reducing the amount of coherent noise is a topic for further research. [61]

However, the observation that coherent noise is the dominant source of noise in

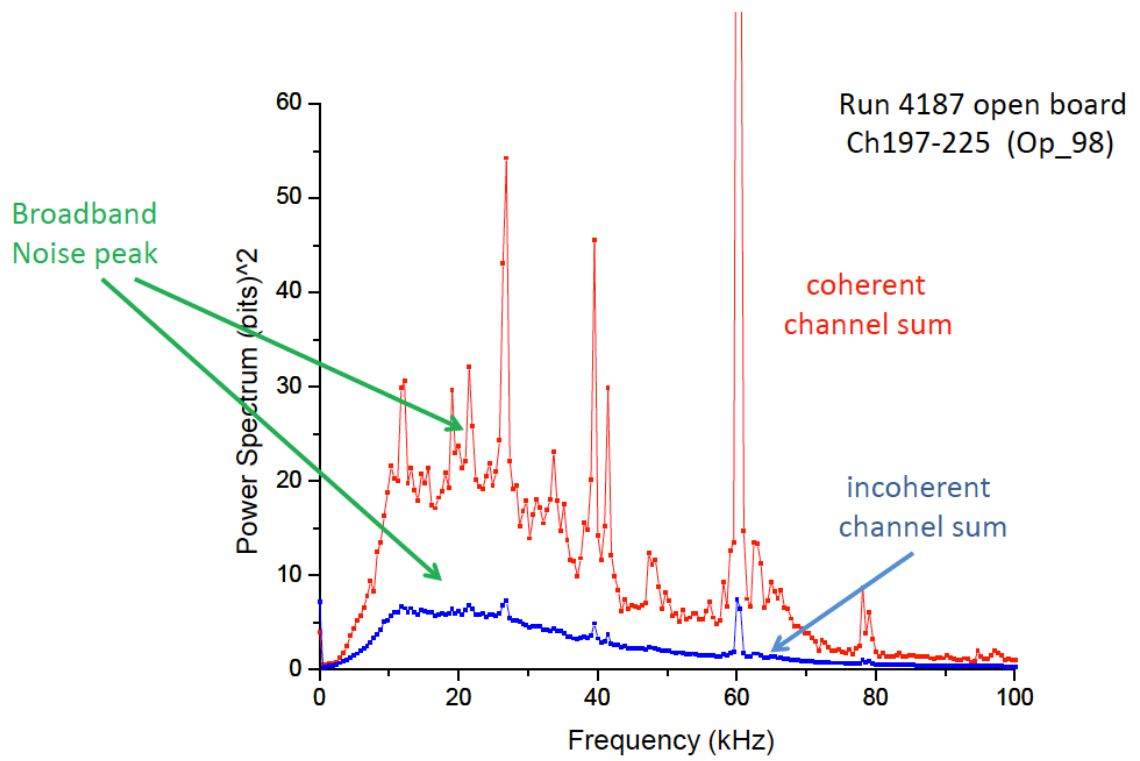


Figure 3.1: Coherent and incoherent noise power spectra for a sample set of APD channels without signal shaping. [61]

the APDs, and thus also the limiting factor in EXO-200’s energy resolution, means that it should be possible to exploit these correlations and reduce the level of noise in offline analysis. This chapter will describe a scheme to accomplish that goal and produce an optimal estimate of scintillation energy which takes noise correlations into account.

It is worth noting that in fact there are a number of different qualitative approaches to reducing the noise levels in the scintillation channel. In casual terms, we will refer to “passive” approaches as those in which components of our signals are weighted more or less heavily based on their relative signal-to-noise content. “Active” approaches, by contrast, will be classified as those which attempt to improve the signal-to-noise content of signal components. We have identified the following three types of denoising:

Frequency weighting

On a given channel signal, weight more heavily the frequency components which contain larger signal-to-noise ratio. This passive denoising scheme requires knowledge of the shapes in fourier space of a signal and the power spectrum of the noise.

Channel weighting

Different channels may have different levels of noise, so some may generally have higher-quality signals. More importantly, though, the amount of signal on a given channel depends strongly on the proximity of the APD gang to the source of scintillation within the detector. This passive denoising scheme

therefore allows us to weight more heavily the channels which have more scintillation, provided we can identify these weights on an event-by-event basis. This requires knowledge of the magnitude of noise on each channel and of the correspondence between event position and signal magnitude on each APD gang; the latter is described by the lightmap, described earlier.

Noise subtraction

This active form of denoising consists of using correlations between noise on different channels to produce a better estimate of the noise component of signals than each signal taken independently could provide. To accomplish this, we will require detailed information about the pairwise noise correlations across channels at each frequency.

In the following, we will describe a general linear operator on the APD signals which in principle can accomplish each of these forms of denoising, and we will presume that all of the described inputs are available; by identifying the parameters for that linear operator which are optimal, we can be confident that the denoising operator we produce will accomplish all three forms of denoising described above.

3.1 Setup

We first establish a number of notational conventions:

- i, j will represent indices over APD channels.
- a, b, c will represent indices of signals in an event.

- τ will represent the time indices of a discrete waveform; t_a represents the calendar time of a signal a .
- f, g will represent the frequency indices of Fourier-transformed waveforms.
- For a waveform $*[\tau]$, we will represent the discrete Fourier transform of that waveform with $\widetilde{*}[f]$, where the particular convention used to evaluate the Fourier transform is not significant.
- For a Fourier-transformed waveform $\widetilde{*}[f]$, we denote the real and imaginary parts of that waveform by $\widetilde{*}^R[f]$ and $\widetilde{*}^I[f]$, respectively.
- For an unknown parameter $*$, the symbol $\widehat{*}$ will identify an estimator for $*$.
- For an expression $*$ containing random variables, $\langle * \rangle$ is the expectation value of $*$.

We describe the data as a collection of discretely sampled waveforms, $X_i[\tau]$. We assume that all signal times and shapes are already known, so we can model the waveforms by $X_i[\tau] = \sum_a M_{ia} Y_{ia}[\tau] + N_i[\tau] + b_i$, where Y is the shape of signal a on channel i , M is the magnitude of that signal, and N and b represent the electronic noise and baseline, respectively, of the channel.

To break the degeneracy between M and Y , we choose to fix the magnitude of the template signal Y . We choose to require that the signal Y has a magnitude of one, as described in figure 3.2. We assume that the expected magnitude of a signal on channel i from a single-site 2615-keV deposit is known, and is characterized as a function $L_i(\vec{x}, t)$ of the deposit position \vec{x} and calendar time t . The exact methods

for measuring $L_i(\vec{x}, t)$ will be described in chapter 5. In cases where a multi-site event deposits energy in multiple locations, we will use the charge information to estimate the energy deposited in each location; the estimated lightmap yield from such an event will be a weighted sum:

$$L_i^{MS}(\vec{x}_1, \dots, \vec{x}_{n_{max}}, t) = \frac{\sum_n E_n^{charge} L_i(\vec{x}_n, t)}{\sum_n E_n^{charge}}. \quad (3.1)$$

For notational simplicity, we will still characterize the expected yield on a gang i from a multi-site event as $L_i(\vec{x}, t)$; in practice it will always be clear how to compute the expected yields for any event from this function. We can thereby characterize the expected yields M_{ia} from an event with energy E_a as:

$$\langle M_{ia} \rangle = L_i(\vec{x}_a, t_a) E_a. \quad (3.2)$$

Noise correlations will be much simpler in frequency space, so we take the Fourier transform and drop the zero-frequency component to obtain

$$\tilde{X}_i[f] = \sum_a M_{ia} \tilde{Y}_{ia}[f] + \tilde{N}_i[f]. \quad (3.3)$$

As with the lightmap, we will assume that all correlations in the noise, of the form

$$\begin{aligned} & \langle \tilde{N}_i^R[f] \tilde{N}_j^R[g] \rangle \langle \tilde{N}_i^R[f] \tilde{N}_j^I[g] \rangle \\ & \langle \tilde{N}_i^I[f] \tilde{N}_j^R[g] \rangle \langle \tilde{N}_i^I[f] \tilde{N}_j^I[g] \rangle, \end{aligned} \quad (3.4)$$

are known; our means of measuring these correlations will be described in chapter 4.

3.2 The Signal Model

It is first important to characterize the response of the APDs to energy deposits in EXO-200. This will require us to understand the signal amplification of the detector

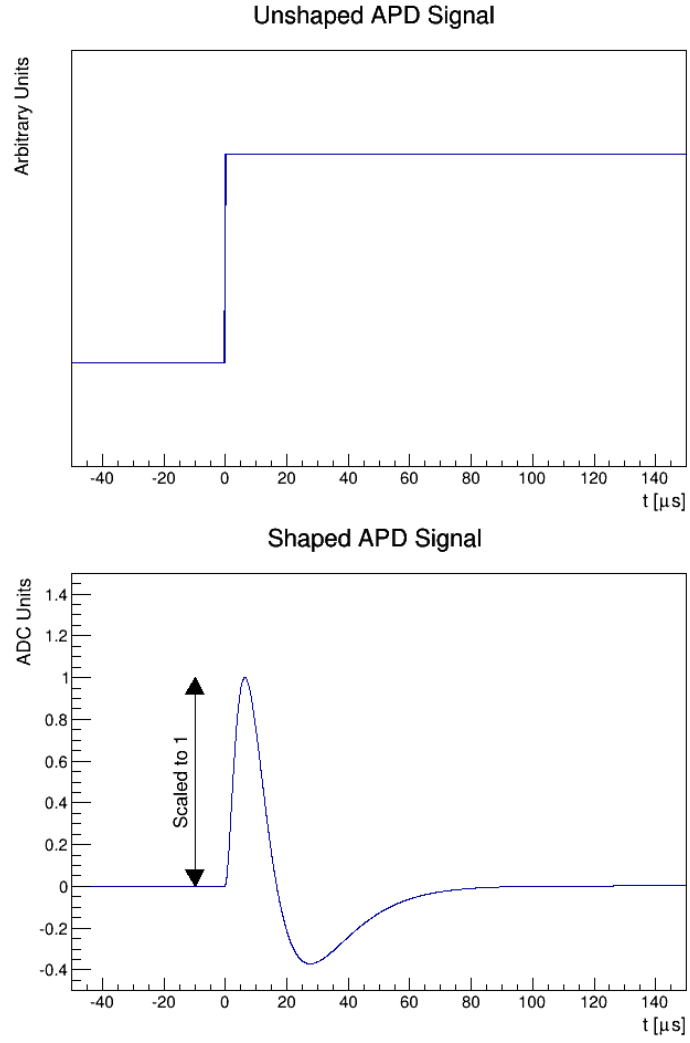


Figure 3.2: Shaped and unshaped APD waveforms. The normalization is shown to make the peak of the shaped waveform have a magnitude of one, and the time axis is shifted so that the unshaped waveform is a step function centered at $t = 0$.

at each physical stage, as well as the noise introduced by each of these processes.

We will describe both in detail here.

3.2.1 APD Noise

We will assume that there are two sources of noise in this model. First, the electronic noise $N_i[\tau]$ is assumed to be random. We require that noise with different frequencies is uncorrelated:

$$\langle \tilde{N}_i[f] \tilde{N}_j[g] \rangle = 0 \text{ when } f \neq g.$$

The noise correlations $\langle \tilde{N}_i[f] \tilde{N}_j[f] \rangle$ are assumed to be known; our means for measuring them are described in chapter 4.

The second random variable will be the magnitude of the signal itself, M_{ia} . When an energetic decay occurs in bulk of the detector, it will release some number of photons. We will denote the number of photons released by signal a by $P_a^{(0)}$, and this is the parameter we wish to measure. However, the magnitude signal actually observed on APD channels number of photons actually collected by sensors is a random variable whose distribution depends on $P_a^{(0)}$, and we will now identify the chain of events which produce it.

First, each APD gang i has some number $P_{ia}^{(1)}$ of photons which reach them. The mean fraction of photons reaching a particular gang i from position \vec{x}_a , $f_i(\vec{x}_a)$, is assumed to be known; however, the initial paths of the optical photons emitted from the source, their trajectories through the Xenon, and their success in reflecting off of teflon are all random, so we treat $P_{ia}^{(1)}$ as a Poisson-distributed random variable

with mean $f_i(\vec{x}_a)P_a^{(0)}$. A typical deposit from a $\beta\beta 0\nu$ event may deposit up to roughly 1000 photons on a nearby APD gang, resulting in roughly 3% uncertainty from photon statistics; in the bulk of the Xenon, it may be that the largest single APD signal is only one-tenth of this size, resulting in closer to 10% uncertainty due to photon statistics on the nearest APD gang. Thus, we find that the Poisson noise on a single APD channel may be quite significant.

Additionally, the number of photons reaching different gangs are not uncorrelated; since a photon which reaches gang i cannot deposit on a different gang j , $P_{ia}^{(1)}$ and $P_{ja}^{(1)}$ are anticorrelated for different gangs $i \neq j$. This process is described by a multinomial distribution. Explicitly, we can identify the expectation values of a multinomial distribution:

$$\langle P_{ia}^{(1)} \rangle = f_i(\vec{x}_a)P_a^{(0)} \quad (3.5)$$

$$\langle P_{ia}^{(1)} P_{jb}^{(1)} \rangle = \langle P_{ia}^{(1)} \rangle \langle P_{jb}^{(1)} \rangle + [f_i(\vec{x}_a)\delta_{ij} - f_i(\vec{x}_a)f_j(\vec{x}_a)] P_a^{(0)}\delta_{ab} \quad (3.6)$$

(As a detail, we should note that the multinomial distribution is an incorrect model in one important respect: photons may reflect off of teflon one or more times on its way to the APD. Photons incident on the teflon may be absorbed, leading to a reflection efficiency which may be less than one. Furthermore, it is possible that teflon may be mildly fluorescent at the wavelength of our scintillation, leading to an increase in efficiency for reflection. As a result, photons which reflect off of teflon experience a combination of variances from each path photons travel and variances in each gain which they undergo. These conditional variances will always be larger than a simple Poisson distribution would indicate [55], so our variance estimate for

$P^{(1)}$ is probably an underestimate.)

Following the random process associated with photons reaching the APD gangs, there is also randomness associated with signal amplification internally in the APDs. These processes are described in detail in [53], but we can summarize that:

1. Optical photons in the liquid xenon arrive at the active layer of the APDs have a wavelength of roughly 178 nm, meaning that each optical photon has an energy of roughly 7.0 eV. The energy required to produce one electron-hole pair in silicon is roughly 3.66 eV, which means that each incident photon produces roughly 1.9 electron-hole pairs; we will define $P_{ia}^{(2)}$ to be the number of electron-hole pairs actually produced from $P_{ia}^{(1)}$ incident photons. The corresponding Fano factor for electrons produced in silicon is roughly 0.1, meaning that in addition to the uncertainty in $P_{ia}^{(1)}$ which we have already characterized, there is an additional uncorrelated variance in $P_{ia}^{(2)}$ equal to $0.1 \langle P_{ia}^{(2)} \rangle$. The correlations in the parameters $P_{ia}^{(2)}$ are therefore:

$$\langle P_{ia}^{(2)} \rangle = 1.9 \cdot P_{ia}^{(1)} \quad (3.7)$$

$$\langle P_{ia}^{(2)} P_{jb}^{(2)} \rangle = (1.9)^2 \langle P_{ia}^{(1)} P_{jb}^{(1)} \rangle + 0.1 \langle P_{ia}^{(2)} \rangle \delta_{ij} \delta_{ab} \quad (3.8)$$

2. Electron-hole pairs are then amplified by an avalanche process inside the APDs. The magnitude of this gain is APD-dependent, generally on the order of 200 – 300, and can be identified by a time-dependent quantity $G_i^P(t)$; we will call the number of output electrons $P_{ia}^{(3)}$. In addition to amplification of existing noise in $P_{ia}^{(2)}$, two additional source of noise are introduced. First,

there is statistical variance in the amplification experienced by each electron due to the randomness of the avalanche process. This variance is dependent on many factors, including the gain, and scales like the square root of the number of electrons; we define the variance on the gain experienced by a single electron as $\sigma_{G_i}^2(t)$. From [53], $\sigma_{G_i}^2(t)$ is approximately equal to $(G_i^P(t))^2$ when $G_i^P(t) = 100$; as a result, we approximate $\sigma_{G_i}^2(t) \approx (G_i^P(t))^2$. Second, there are gain non-uniformities in the diode volume, which contribute variance proportional to the square of the number of initial electron-hole pairs; we will identify the proportionality constant as σ_{NU}^2 , which may depend on time and APD gang. The magnitude of this uncertainty is not well-known, but may be significant. Our current code base treats σ_{NU}^2 as zero; however, future analyses will likely attempt to estimate it, and the derivation which follows will retain it as an aid to that anticipated work. The correlations in the parameters $P_{ia}^{(3)}$ are therefore:

$$\langle P_{ia}^{(3)} \rangle = G_i^P(t_a) P_{ia}^{(2)} \quad (3.9)$$

$$\langle P_{ia}^{(3)} P_{jb}^{(3)} \rangle = G_i^P(t_a) G_j^P(t_b) \langle P_{ia}^{(2)} P_{jb}^{(2)} \rangle + \left[P_{ia}^{(2)} \sigma_{G_i}^2(t_a) + \left(P_{ia}^{(2)} \right)^2 \sigma_{NU}^2 \right] \delta_{ij} \delta_{ab} \quad (3.10)$$

My thinking on non-uniformities has evolved a little bit – clarify in this section why it's such a big deal.

Finally, there is amplification

$$M_{ia} = G_i^E(t) P_{ia}^{(3)} \quad (3.11)$$

associated with the electronics of the APDs which is dependent on time and channel. This includes preamplifier gain, shaper gain, gain associated with the shaping times, and conversion from voltage to ADC counts. We assume that no signal-dependent noise is introduced during this stage, but there is constant electronic noise $N_i[\tau]$ which has been described above. Additionally, the APDs can contribute noise in the form of a dark current which is uncorrelated with signals; this is inseparable from the electronic noise, and so we absorb it into our description of $N_i[\tau]$.

3.2.2 APD Signals and Noise

We have already characterized the overall gain of the APDs using the lightmap described earlier. We know that a typical single-site deposit of 2615 keV at position \vec{x} and time t will produce a signal with expected magnitude $L_i(\vec{x}, t)$ on APD channel i . We must now connect this empirical understanding of our APD signals with the physical understanding in terms of photons and electrons described above.

We have identified the number of photons created by an event a as $P_a^{(0)}$. In reality, an ideal scintillation measurement can only measure this quantity, and not the true energy of the event – there is randomness in the number of photons produced by a known-energy deposit, so $P_a^{(0)}$ will not be perfectly correlated with the energy E_a of the event. However, the number of photons generated is a difficult parameter to understand empirically, and all of our calibrations occur at a known energy even though the corresponding number of deposited photons is unknown.

Simulations using NEST [58] which have been performed within the EXO

group by Liangjian Wen at our bulk electric field indicate that we can expect roughly 82,000 photons from a 2615-keV gamma deposit. We will identify this parameter with c , and express the average relation between energy and photon yield in EXO-200 by

$$P^{(0)} = c \cdot E, \quad (3.12)$$

where E is the energy measured in units of 2615 keV. We will treat this relationship as exact, and attempt to measure E rather than $P^{(0)}$; this means that when we speak of measuring the energy of an event using the APD signals, we really are referring to measuring the “scintillation energy,” or a parameter proportional to the number of emitted photons which on average will equal the true energy of the deposit.

Ignoring variances for a moment, we can combine equations 3.5, 3.7, 3.9, 3.11, and 3.12 to state that on average,

$$M_{ia} = 1.9 \cdot cf_i(\vec{x}_a)G_i^P(t_a)G_i^E(t_a)E_a.$$

We can compare this to our empirical lightmap measurements, which are expressed by $M_{ia} = L_i(\vec{x}_a, t_a)E_a$, and we conclude that:

$$L_i(\vec{x}, t) = 1.9 \cdot cf_i(\vec{x})G_i^P(t)G_i^E(t). \quad (3.13)$$

We recall that it is assumed the lightmap is a separable function, $L_i(\vec{x}, t) = R_i(\vec{x})S_i(t)$, and we can provide the two proportionalities:

$$R_i(\vec{x}) \propto f_i(\vec{x})$$

$$S_i(t) \propto G_i^P(t)G_i^E(t).$$

Although $G_i^E(t)$ is in principle time-dependent, due to environmental effects on the APD electronics and occasional replacement of electronics cards for certain channels, precision data on its value is not readily available. Instead, we use a time-independent estimate

$$G_i^E(t) = 1.1 \cdot 10^{-3} \quad (3.14)$$

ADC counts per electron emitted from the APD; the uncertainty in this figure is dominated by uncertainty in the preamplifier gains, which are controlled by a roughly 5 pF capacitor. Other contributors to the gain include shaper gain of 21.2 (from a combination of amplification and shaping) and a full-scale digitizer range of 2.5 volts for 12-bit digitization.

We do have independent measurements of the APD gains $G_i^P(t)$ available from laser calibration runs. These special runs allow a laser to shine into the detector from a fixed point and with a stable amplitude while the bias voltages on the APDs are varied from an effective unity gain to our standard voltage settings. Using these measurements, we are able to measure $G_i^P(t)$ at weekly intervals from September 2012 to the present time. Before September 2012, some less-reliable laser data is available, but results from that data are not readily available as of this writing.

It would be possible, and should be a goal for future improvements, to make use of this full range of laser data. However, the laser data provides a less uniform history of APD gains over the full data-taking window of EXO-200 from September 2011 to November 2013, and it is much easier to track time-dependent behavior from Thorium source data which have been collected regularly throughout that period.

As a result, a compromise is used to characterize $G_i^P(t)$. One particular laser run, run 4540 (taken on December 13, 2012), is used to fix $G_i^P(t_{4540})$, and the function is extrapolated using Thorium source data with:

$$G_i^P(t) \approx G_i^P(t_{4540}) \cdot S_i(t)/S_i(t_{4540}).$$

This assumption makes use of the approximation that $G_i^E(t)$ is roughly constant in time, which is probably only accurate to one significant figure; therefore when an electronics change is made to a channel, we can expect that the accuracy of $G_i^P(t)$ is no better than one significant figure. These results mean that we can also estimate with the same level of accuracy:

$$f_i(\vec{x}) \approx \frac{S_i(t_{4540})}{G_i^P(t_{4540})} \cdot \frac{R_i(\vec{x})}{1.9 \cdot c G_i^E(t)}.$$

It is then possible to express the full correlations in signal magnitudes in terms of the scintillation energies of deposits as:

$$\langle M_{ia} \rangle = L_i(\vec{x}_a, t_a) E_a \tag{3.15}$$

$$\begin{aligned} \langle M_{ia} M_{jb} \rangle &= L_i(\vec{x}_a, t_a) L_j(\vec{x}_b, t_b) E_a E_b \left[1 + \frac{\sigma_{NU}^2}{(G_i^P(t))^2} \delta_{ij} \delta_{ab} \right] \\ &\quad - L_i(\vec{x}_a, t_a) L_j(\vec{x}_a, t_a) E_a \delta_{ab} / c \\ &\quad + L_i(\vec{x}_a, t_a) G_i^E(t) E_a \delta_{ij} \delta_{ab} \left[(0.1 + 1.9) G_i^P(t) + \frac{\sigma_{G_i}^2(t)}{G_i^P(t)} \right]. \end{aligned} \tag{3.16}$$

We combine this with our knowledge of the noise coefficients:

$$\begin{aligned} \langle \tilde{N}_i[f] \rangle &= 0 \\ \langle \tilde{N}_i[f] \tilde{N}_j[g] \rangle &= \begin{cases} 0 & \text{if } f \neq g \\ \text{known} & \text{if } f = g \end{cases} \end{aligned}$$

and can claim to now have a complete description of the signals and noise observed in the APD channels.

3.3 Derivation

It is now possible to specify the optimization criteria for generating an energy estimate from the APD signals. We wish to identify an energy estimator which is unbiased and has a minimal expected error. For the problem to remain tractable, we will demand that the operator be linear. Furthermore, although the Fourier-transformed waveforms $\tilde{X}_i[f]$ are complex-valued, we will require that the energy estimate be strictly real-valued.

We will therefore take the energy estimator to be of the form:

$$\hat{E}_a = \sum_{if} A_{ifa} \tilde{X}_i^R[f] + B_{ifa} \tilde{X}_i^I[f]. \quad (3.17)$$

The goal of denoising is therefore reduced to identifying the optimal parameters A_{ifa} and B_{ifa} for this estimator.

The error in the energy estimate \hat{E}_a of E_a is defined by:

$$\epsilon_a^2 = \left\langle \left(\hat{E}_a - E_a \right)^2 \right\rangle. \quad (3.18)$$

Our goal is then to minimize ϵ_a^2 under the constraint of no bias, ie. that:

$$\left\langle \hat{E}_a - E_a \right\rangle = 0 \quad (3.19)$$

or, equivalently,

$$\sum_{ifb} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] L_i(\vec{x}_b, t_b) E_b = E_a.$$

However, we will find that it is necessary to specify a slightly stronger constraint. In particular, it will be desirable to ensure that the constraints are as independent of energy as possible to reduce the need to input an estimated energy into our energy estimator; therefore we will freely employ the stronger constraint

$$\sum_{if} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] L_i(\vec{x}_b, t_b) = \delta_{ab} \text{ for all } b, \quad (3.20)$$

which implies the earlier forms and leads to advantageous cancellations of terms.

We now proceed with the optimization. We start by expanding ϵ_a^2 :

$$\begin{aligned} \epsilon_a^2 &= \left\langle \left(\hat{E}_a - E_a \right)^2 \right\rangle \\ &= \left\langle \hat{E}_a^2 \right\rangle - E_a \left\langle \hat{E}_a \right\rangle - \left\langle \hat{E}_a - E_a \right\rangle E_a \end{aligned}$$

We can employ the constraint of equation 3.19 to simplify the second term of this expansion, and eliminate the third altogether; we then proceed:

$$\begin{aligned} \epsilon_a^2 &= \left\langle \hat{E}_a^2 \right\rangle - E_a^2 \\ &= \left\langle \left(\sum_{if} \left[A_{ifa} \tilde{X}_i^R[f] + B_{ifa} \tilde{X}_i^I[f] \right] \right)^2 \right\rangle - E_a^2 \\ &= \left\langle \left(\sum_{if} \left[A_{ifa} \tilde{N}_i^R[f] + B_{ifa} \tilde{N}_i^I[f] \right] \right. \right. \\ &\quad \left. \left. + \sum_{ifb} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] M_{ib} \right)^2 \right\rangle - E_a^2 \end{aligned}$$

The noise \tilde{N} and signal M_{ia} are uncorrelated, so multiplicative cross-terms will have an expectation value of zero:

$$\begin{aligned} \epsilon_a^2 &= \left\langle \left(\sum_{if} \left[A_{ifa} \tilde{N}_i^R[f] + B_{ifa} \tilde{N}_i^I[f] \right] \right)^2 \right\rangle \\ &\quad + \left\langle \left(\sum_{ifb} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] M_{ib} \right)^2 \right\rangle - E_a^2 \end{aligned}$$

Additionally, electronic noise cross-terms between different frequencies will not survive:

$$\begin{aligned}
\epsilon_a^2 &= \left\langle \sum_{ijf} \left[A_{ifa} \tilde{N}_i^R[f] + B_{ifa} \tilde{N}_i^I[f] \right] \left[A_{jfa} \tilde{N}_j^R[f] + B_{jfa} \tilde{N}_j^I[f] \right] \right\rangle \\
&\quad + \left\langle \left(\sum_{ifb} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] M_{ib} \right)^2 \right\rangle - E_a^2 \\
&= \sum_{ijf} \left[A_{ifa} A_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^R[f] \right\rangle + A_{ifa} B_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^I[f] \right\rangle \right. \\
&\quad \left. + B_{ifa} A_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^R[f] \right\rangle + B_{ifa} B_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^I[f] \right\rangle \right] \\
&\quad + \sum_{\substack{ifb \\ jgc}} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] \left[A_{jga} \tilde{Y}_{jc}^R[g] + B_{jga} \tilde{Y}_{jc}^I[g] \right] \langle M_{ib} M_{jc} \rangle - E_a^2
\end{aligned}$$

We will now expand $\langle M_{ib}M_{jc} \rangle$ using equation 3.16 and take advantage of the stronger form of our constraint 3.20 to simplify the expression:

$$\begin{aligned}
\epsilon_a^2 = & \sum_{ijf} \left[A_{ifa}A_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^R[f] \right\rangle + A_{ifa}B_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^I[f] \right\rangle \right. \\
& \left. + B_{ifa}A_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^R[f] \right\rangle + B_{ifa}B_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^I[f] \right\rangle \right] \\
& + \left(\sum_{ifb} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] L_i(\vec{x}_b, t_b) E_b \right)^2 - E_a^2 \\
& - \sum_b \left(\sum_{if} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] L_i(\vec{x}_b, t_b) \right)^2 \frac{E_b}{c} \\
& + \sum_{ifgb} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] \left[A_{iga} \tilde{Y}_{ib}^R[g] + B_{iga} \tilde{Y}_{ib}^I[g] \right] \cdot \\
& L_i(\vec{x}_b, t_b) E_b \left[(0.1 + 1.9) G_i^E(t_b) G_i^P(t_b) \right. \\
& \left. + \frac{G_i^E(t_b)}{G_i^P(t_b)} \sigma_{G_i}^2(t_b) + L_i(\vec{x}_b, t_b) E_b \frac{\sigma_{NU}^2}{(G_i^P(t_b))^2} \right] \\
= & \sum_{ijf} \left[A_{ifa}A_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^R[f] \right\rangle + A_{ifa}B_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^I[f] \right\rangle \right. \\
& \left. + B_{ifa}A_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^R[f] \right\rangle + B_{ifa}B_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^I[f] \right\rangle \right] \\
& + \sum_{ib} \left(\sum_f \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] \right)^2 \cdot \\
& L_i(\vec{x}_b, t_b) E_b \left[(0.1 + 1.9) G_i^E(t_b) G_i^P(t_b) \right. \\
& \left. + \frac{G_i^E(t_b)}{G_i^P(t_b)} \sigma_{G_i}^2(t_b) + L_i(\vec{x}_b, t_b) E_b \frac{\sigma_{NU}^2}{(G_i^P(t_b))^2} \right] \\
& - \frac{E_a}{c}
\end{aligned}$$

We are now in a position to evaluate the partial derivatives of ϵ_a^2 with respect

to the parameters A_{ifa} and B_{ifa} . They are:

$$\begin{aligned}
\frac{\partial \epsilon_a^2}{\partial A_{ifa}} &= 2 \sum_j \left[A_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^R[f] \right\rangle + B_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^I[f] \right\rangle \right] \\
&\quad + 2 \sum_{gb} E_b \tilde{Y}_{ib}^R[f] L_i(\vec{x}_b, t_b) \left[A_{iga} \tilde{Y}_{ib}^R[g] + B_{iga} \tilde{Y}_{ib}^I[g] \right] \cdot \\
&\quad \left[(0.1 + 1.9) G_i^E(t_b) G_i^P(t_b) \right. \\
&\quad \left. + \frac{G_i^E(t_b)}{G_i^P(t_b)} \sigma_{G_i}^2(t_b) + L_i(\vec{x}_b, t_b) E_b \frac{\sigma_{NU}^2}{(G_i^P(t_b))^2} \right] \\
\frac{\partial \epsilon_a^2}{\partial B_{ifa}} &= 2 \sum_j \left[A_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^R[f] \right\rangle + B_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^I[f] \right\rangle \right] \\
&\quad + 2 \sum_{gb} E_b \tilde{Y}_{ib}^I[f] L_i(\vec{x}_b, t_b) \left[A_{iga} \tilde{Y}_{ib}^R[g] + B_{iga} \tilde{Y}_{ib}^I[g] \right] \cdot \\
&\quad \left[(0.1 + 1.9) G_i^E(t_b) G_i^P(t_b) \right. \\
&\quad \left. + \frac{G_i^E(t_b)}{G_i^P(t_b)} \sigma_{G_i}^2(t_b) + L_i(\vec{x}_b, t_b) E_b \frac{\sigma_{NU}^2}{(G_i^P(t_b))^2} \right]
\end{aligned}$$

We will use Lagrange's method to minimize ϵ_a^2 while satisfying our constraints;

we define

$$C_{ab} = \sum_{if} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] L_i(\vec{x}_b, t_b),$$

with ordered indices, and restate the constraints of equation 3.20 as

$$C_{ab} = \delta_{ab}.$$

Then, the partial derivatives of these constrained expressions are:

$$\begin{aligned}
\frac{\partial C_{bc}}{\partial A_{ifa}} &= \tilde{Y}_{ic}^R[f] L_i(\vec{x}_c, t_c) \delta_{ab} \\
\frac{\partial C_{bc}}{\partial B_{ifa}} &= \tilde{Y}_{ic}^I[f] L_i(\vec{x}_c, t_c) \delta_{ab}
\end{aligned}$$

Denoting the set of Lagrange multipliers for ϵ_a^2 with λ_{ab} , where the indices are ordered, and allowing these parameters to absorb constant factors, we can at last

identify the full set of linear equations describing the optimal energy estimator \hat{E}_a :

$$\begin{aligned}
& \sum_j \left[A_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^R[f] \right\rangle + B_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^I[f] \right\rangle \right] \\
& + \sum_{gb} E_b \tilde{Y}_{ib}^R[f] L_i(\vec{x}_b, t_b) \left[A_{iga} \tilde{Y}_{ib}^R[g] + B_{iga} \tilde{Y}_{ib}^I[g] \right] \cdot \\
& \quad \left[(0.1 + 1.9) G_i^E(t_b) G_i^P(t_b) + \frac{G_i^E(t_b)}{G_i^P(t_b)} \sigma_{G_i}^2(t_b) + L_i(\vec{x}_b, t_b) E_b \frac{\sigma_{NU}^2}{(G_i^P(t_b))^2} \right] \\
& + \sum_b \lambda_{ab} \tilde{Y}_{ib}^R[f] L_i(\vec{x}_b, t_b) = 0 \quad \text{for each } i, f \\
& \sum_j \left[A_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^R[f] \right\rangle + B_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^I[f] \right\rangle \right] \\
& + \sum_{gb} E_b \tilde{Y}_{ib}^I[f] L_i(\vec{x}_b, t_b) \left[A_{iga} \tilde{Y}_{ib}^R[g] + B_{iga} \tilde{Y}_{ib}^I[g] \right] \cdot \\
& \quad \left[(0.1 + 1.9) G_i^E(t_b) G_i^P(t_b) + \frac{G_i^E(t_b)}{G_i^P(t_b)} \sigma_{G_i}^2(t_b) + L_i(\vec{x}_b, t_b) E_b \frac{\sigma_{NU}^2}{(G_i^P(t_b))^2} \right] \\
& + \sum_b \lambda_{ab} \tilde{Y}_{ib}^I[f] L_i(\vec{x}_b, t_b) = 0 \quad \text{for each } i, f \\
& \sum_{if} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] L_i(\vec{x}_b, t_b) = \delta_{ab} \quad \text{for each } b
\end{aligned}$$

To simplify the notation, let us define a new function:

$$\begin{aligned}
q(i, b) := & \left((0.1 + 1.9) G_i^E(t_b) G_i^P(t_b) + \frac{G_i^E(t_b)}{G_i^P(t_b)} \sigma_{G_i}^2(t_b) \right. \\
& \left. + L_i(\vec{x}_b, t_b) E_b \frac{\sigma_{NU}^2}{(G_i^P(t_b))^2} \right) L_i(\vec{x}_b, t_b) E_b
\end{aligned} \tag{3.21}$$

which we note should be strictly positive if measured correctly (because all gains and lightmap values should be strictly positive, and all variances are non-negative).

Using this shorthand, we can re-write our system of equations in the more compact

form:

$$\begin{aligned}
& \sum_j \left[A_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^R[f] \right\rangle + B_{jfa} \left\langle \tilde{N}_i^R[f] \tilde{N}_j^I[f] \right\rangle \right] \\
& + \sum_{gb} \tilde{Y}_{ib}^R[f] q(i, b) \left[A_{iga} \tilde{Y}_{ib}^R[g] + B_{iga} \tilde{Y}_{ib}^I[g] \right] \\
& + \sum_b \lambda_{ab} \tilde{Y}_{ib}^R[f] L_i(\vec{x}_b, t_b) = 0 \quad \text{for each } i, f \quad (3.22a)
\end{aligned}$$

$$\begin{aligned}
& \sum_j \left[A_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^R[f] \right\rangle + B_{jfa} \left\langle \tilde{N}_i^I[f] \tilde{N}_j^I[f] \right\rangle \right] \\
& + \sum_{gb} \tilde{Y}_{ib}^I[f] q(i, b) \left[A_{iga} \tilde{Y}_{ib}^R[g] + B_{iga} \tilde{Y}_{ib}^I[g] \right] \\
& + \sum_b \lambda_{ab} \tilde{Y}_{ib}^I[f] L_i(\vec{x}_b, t_b) = 0 \quad \text{for each } i, f \quad (3.22b)
\end{aligned}$$

$$\sum_{if} \left[A_{ifa} \tilde{Y}_{ib}^R[f] + B_{ifa} \tilde{Y}_{ib}^I[f] \right] L_i(\vec{x}_b, t_b) = \delta_{ab} \quad \text{for each } b \quad (3.22c)$$

It is important to note here that energies E_b , which we are intending to measure, do in fact enter into this linear set of equations. At first this would seem to demonstrate that the equations here are impossible to apply, but in fact they remind us that the forms of noise which are correlated with the signal are also dependent on the magnitude of that signal. Statistical fluctuations in the number of photons observed on each APD, for instance, scale like the square root of the number of photons emitted.

We can see, then, that some estimate of energy is necessary to know the relative importance of different forms of noise. But the constraint equation does ensure that our estimate will be unbiased regardless of what energy estimates are fed into the system of equations, so we can be confident that it is only a rough estimate of the energy scale which is needed, and not worry that the optimization will be biased to

estimate energies similar to the estimates we feed in. For this purpose, it is sufficient to use the charge-only energy as an estimate of the scintillation-only energy.

3.4 Matrix Version

Because the system of equations above is linear in A_{ifa} and B_{ifa} , we wish to solve it as a matrix equation. This requires, first, that we choose an ordering of the unknowns; the best ordering will be the one which groups nonzero entries into blocks, since this will allow us to make more efficient use of matrix libraries. We define:

$$\vec{A}_a = (A_{11a}, B_{11a}, A_{21a}, B_{21a}, \dots, B_{i_{max}1a}, A_{12a}, \dots, A_{1f_{max}a}, A_{2f_{max}a}, \dots, A_{i_{max}f_{max}a}) \quad (3.23)$$

so that the entries alternate between A and B , iterating quickly through channels and more slowly through frequencies. Note that only the A terms are included for the maximum frequency; this is because $X_i[t]$ is real-valued, so $\tilde{X}_i[0]$ and $\tilde{X}_i[f_{max}]$ are real-valued as well. We also will find it convenient to define the matrix:

$$\mathbf{A} = \begin{pmatrix} \vdots & & \vdots \\ \vec{A}_1 & \dots & \vec{A}_{a_{max}} \\ \vdots & & \vdots \end{pmatrix} \quad (3.24)$$

which includes the parameters needed for each of the estimators \hat{E}_a .

We can then define the electronic noise blocks as:

$$\mathbf{N}_f = \begin{cases} \text{COV} \left(\tilde{N}_1^R[f], \tilde{N}_1^I[f], \tilde{N}_2^R[f], \dots, \tilde{N}_{i_{max}}^I[f] \right) & \text{for } f < f_{max} \\ \text{COV} \left(\tilde{N}_1^R[f], \tilde{N}_2^R[f], \dots, \tilde{N}_{i_{max}}^R[f] \right) & \text{for } f = f_{max} \end{cases} \quad (3.25)$$

where \mathbf{COV} specifies the covariance matrix of an ordered list of random variables,

$$\mathbf{COV}(x_1, x_2, \dots, x_n) = \begin{pmatrix} \text{cov}(x_1x_1) & \text{cov}(x_1x_2) & \cdots & \text{cov}(x_1x_n) \\ \text{cov}(x_2x_1) & \text{cov}(x_2x_2) & \cdots & \text{cov}(x_2x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(x_nx_1) & \text{cov}(x_nx_2) & \cdots & \text{cov}(x_nx_n) \end{pmatrix}. \quad (3.26)$$

Note that because our noise variables $\tilde{N}_i^R[f]$ and $\tilde{N}_i^I[f]$ have mean zero, we can move smoothly between expectation values of products and covariances:

$$\begin{aligned} \text{cov}(\tilde{N}_i^R[f], \tilde{N}_j^R[f]) &= \langle \tilde{N}_i^R[f] \tilde{N}_j^R[f] \rangle \\ \text{cov}(\tilde{N}_i^R[f], \tilde{N}_j^I[f]) &= \langle \tilde{N}_i^R[f] \tilde{N}_j^I[f] \rangle \\ \text{cov}(\tilde{N}_i^I[f], \tilde{N}_j^R[f]) &= \langle \tilde{N}_i^I[f] \tilde{N}_j^R[f] \rangle \\ \text{cov}(\tilde{N}_i^I[f], \tilde{N}_j^I[f]) &= \langle \tilde{N}_i^I[f] \tilde{N}_j^I[f] \rangle. \end{aligned} \quad (3.27)$$

We pack together the blocks of equation 3.25 into a sparse noise matrix:

$$\mathbf{N} = \begin{pmatrix} \mathbf{N}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{N}_{\mathbf{f}_{\max}} \end{pmatrix}. \quad (3.28)$$

In a similar way, we can define the other noise terms in terms of matrix operations. We will find that it is possible to describe the noise terms correlated with signals as a product of two matrices, $\mathbf{P} = \mathbf{P}_1 \mathbf{P}_2$. First we define the matrix \mathbf{P}_1 , which steps horizontally through the APD channels and signals. For a particular

choice of indices j, b , the corresponding column of \mathbf{P}_1 will be:

$$\mathbf{P}_1(\text{column } j, b) = \begin{pmatrix} \tilde{Y}_{jb}^R[1]\delta_{1j} \\ \tilde{Y}_{jb}^I[1]\delta_{1j} \\ \tilde{Y}_{jb}^R[1]\delta_{2j} \\ \vdots \\ \tilde{Y}_{jb}^I[1]\delta_{i_{max}j} \\ \tilde{Y}_{jb}^R[2]\delta_{1j} \\ \vdots \\ \tilde{Y}_{jb}^R[f_{max}]\delta_{1j} \\ \tilde{Y}_{jb}^R[f_{max}]\delta_{2j} \\ \vdots \\ \tilde{Y}_{jb}^R[f_{max}]\delta_{i_{max}j} \end{pmatrix} \sqrt{q(j, b)}. \quad (3.29)$$

Note that in each column, only a small subset of the entries are nonzero.

This means that \mathbf{P}_2 should be a matrix with rows that step through the APD channels and signals as well. For a particular choice of indices j, b , the corresponding row of \mathbf{P}_2 will be:

$$\mathbf{P}_2(\text{row } j, b) = \begin{pmatrix} \tilde{Y}_{jb}^R[1]\delta_{1j} & \tilde{Y}_{jb}^I[1]\delta_{1j} & \tilde{Y}_{jb}^R[1]\delta_{2j} & \cdots & \tilde{Y}_{jb}^I[1]\delta_{i_{max}j} & \tilde{Y}_{jb}^R[2]\delta_{1j} & \cdots \\ \cdots & \tilde{Y}_{jb}^R[f_{max}]\delta_{1j} & \tilde{Y}_{jb}^R[f_{max}]\delta_{2j} & \cdots & \tilde{Y}_{jb}^R[f_{max}]\delta_{i_{max}j} \end{pmatrix} \sqrt{q(j, b)}. \quad (3.30)$$

Since we will be using the product $\mathbf{P} = \mathbf{P}_1\mathbf{P}_2$, it clearly must be the case that the ordering of columns in \mathbf{P}_1 is the same as the ordering of rows in \mathbf{P}_2 . Because in practice there will always be more APD channels than APD signals, it will generally

be advantageous to iterate through j fastest, and step through b more slowly. We can also see that $\mathbf{P}_1 = \mathbf{P}_2^\top$, so

$$\mathbf{P} = \mathbf{P}_2^\top \mathbf{P}_2. \quad (3.31)$$

The constraint equations are represented by a matrix \mathbf{C} with rows that step through the APD signals; for a particular choice of b the corresponding row of \mathbf{C} will be:

$$\mathbf{C}(\text{row } b) = \begin{pmatrix} \tilde{Y}_{1b}^R[1]L_1(\vec{x}_b, t_b) & \tilde{Y}_{1b}^I[1]L_1(\vec{x}_b, t_b) & \tilde{Y}_{2b}^R[1]L_2(\vec{x}_b, t_b) \\ \dots & \tilde{Y}_{i_{max}b}^I[1]L_{i_{max}}(\vec{x}_b, t_b) & \tilde{Y}_{1b}^R[2]L_1(\vec{x}_b, t_b) \\ \dots & \tilde{Y}_{1b}^R[f_{max}]L_1(\vec{x}_b, t_b) & \tilde{Y}_{2b}^R[f_{max}]L_2(\vec{x}_b, t_b) \\ & \dots & \tilde{Y}_{i_{max}b}^R[f_{max}]L_{i_{max}}(\vec{x}_b, t_b) \end{pmatrix}. \quad (3.32)$$

Finally it is possible to restate the full linear equation specified above in matrix form, where we can simultaneously include the equations for all of the estimators \hat{E}_a :

$$\begin{pmatrix} \mathbf{N} + \mathbf{P} & \mathbf{C}^\top \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \mathbf{A} = \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix}. \quad (3.33)$$

The matrix \mathbf{N} is the same as the covariance matrix of the electronic noise, so it must be symmetric and positive-semidefinite. Similarly, we have seen that $\mathbf{P} = \mathbf{P}_2^\top \mathbf{P}_2$, implying that \mathbf{P} must be symmetric and positive-semidefinite. This means that there is a Cholesky decomposition $\mathbf{N} + \mathbf{P} = \mathbf{L}\mathbf{L}^\top$, where \mathbf{L} is a lower-triangular matrix, and that $\mathbf{N} + \mathbf{P}$ is symmetric and positive semi-definite. This implies that the full matrix equation is symmetric; however, it is insufficient for

showing that the full system is positive semi-definite, and in fact we will see that this is not the case.

We also note the important observation that this matrix is sparse. Given roughly 1000 frequency components in our signals, only 0.1% of the entries of \mathbf{N} are non-zero. And with 70 APD channels and only a small number of signals we wish to denoise, the combined number of non-zero entries in \mathbf{P}_1 and \mathbf{P}_2 is easily an order of magnitude less than the number of non-zero entries in \mathbf{N} . As a result, we can conclude that this matrix is extremely sparse, and should ensure that whatever method we use to solve these equations takes advantage of this sparsity.

3.5 Preconditioning

We have specified a matrix equation which we now wish to solve. However, it is generally recommended that large matrix equations should not be solved directly. Instead, it is recommended that the matrix should be preconditioned. This means that if a matrix equation $\mathbf{A}\vec{x} = \vec{b}$ should be solved, one should first find a matrix \mathbf{A}' which is approximately equal to \mathbf{A} and which is easily invertible. Finding the right balance between the quality of the approximation $\mathbf{A}' \approx \mathbf{A}$ and the difficulty of computing the inverse $(\mathbf{A}')^{-1}$ is more art than science, but we will present here a preconditioning matrix which seems to strike a good balance for this particular system.

The matrix equation we wish to solve takes the form:

$$\begin{pmatrix} \mathbf{N} + \mathbf{P} & \mathbf{C}^\top \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \mathbf{A} = \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix}$$

Let us define $\mathbf{D} = \text{diag}(\mathbf{N})$ and approximate

$$\begin{pmatrix} \mathbf{N} + \mathbf{P} & \mathbf{C}^\top \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \approx \begin{pmatrix} \mathbf{D} & \mathbf{C}^\top \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \quad (3.34)$$

This approximate form is easy to invert (assuming every frequency component of every channel has non-zero noise and every constraint is satisfiable, which is true in practice), and can be used to precondition our problem for better numerical behavior.

We can further factor this approximate form of the matrix:

$$\begin{aligned} \begin{pmatrix} \mathbf{D} & \mathbf{C}^\top \\ \mathbf{C} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{0} \\ \mathbf{C}\mathbf{D}^{-1/2} & \mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{D}^{-1/2}\mathbf{C}^\top \\ \mathbf{0} & -\mathbf{H}^\top \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}\mathbf{D}^{-1/2} & \mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{D}^{-1/2}\mathbf{C}^\top \\ \mathbf{0} & -\mathbf{H}^\top \end{pmatrix} \begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \end{aligned} \quad (3.35)$$

where \mathbf{H} is a lower-diagonal square matrix defined uniquely as the Cholesky decomposition $\mathbf{H}\mathbf{H}^\top = \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^\top$. (The solution is guaranteed to be real because \mathbf{D} is positive-semidefinite in all cases, and positive-definite in practice – it consists of noise variance terms, which are strictly positive in practice.) Note that $\mathbf{C}\mathbf{D}^{-1}\mathbf{C}^\top$ has dimensions equal to the number of signals in an event, which in practice will be small enough that we can compute its Cholesky decomposition $\mathbf{H}\mathbf{H}^\top$ directly.

The two diagonal factors are easily inverted; we further define

$$\mathbf{K}_1 = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{CD}^{-1/2} & \mathbf{H} \end{pmatrix} \quad \mathbf{K}_1^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{H}^{-1}\mathbf{CD}^{-1/2} & \mathbf{H}^{-1} \end{pmatrix} \quad (3.36)$$

$$\mathbf{K}_2 = \begin{pmatrix} \mathbf{I} & \mathbf{D}^{-1/2}\mathbf{C}^\top \\ \mathbf{0} & -\mathbf{H}^\top \end{pmatrix} \quad \mathbf{K}_2^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{D}^{-1/2}\mathbf{C}^\top\mathbf{H}^{-\top} \\ \mathbf{0} & -\mathbf{H}^{-\top} \end{pmatrix}, \quad (3.37)$$

where $\mathbf{M}^{-\top}$ represents the inverse of the transpose of \mathbf{M} . We then follow the standard prescription for preconditioning a linear system by replacing our original matrix equation with the new form:

$$\mathbf{K}_1^{-1} \begin{pmatrix} \mathbf{D}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{N} + \mathbf{P} & \mathbf{C}^\top \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{D}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{K}_2^{-1} \tilde{\mathbf{A}} = \mathbf{K}_1^{-1} \begin{pmatrix} \mathbf{D}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix}$$

$$\tilde{\mathbf{A}} = \mathbf{K}_2 \begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{A}$$

which reduces to

$$\mathbf{K}_1^{-1} \left[\begin{pmatrix} \mathbf{D}^{-1/2}\mathbf{ND}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{D}^{-1/2}\mathbf{PD}^{-1/2} & \mathbf{D}^{-1/2}\mathbf{C}^\top \\ \mathbf{CD}^{-1/2} & \mathbf{0} \end{pmatrix} \right] \mathbf{K}_2^{-1} \tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}^{-1} \end{pmatrix} \quad (3.38)$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{D}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{K}_2^{-1} \tilde{\mathbf{A}} \quad (3.39)$$

This preconditioned set of equations is numerically quite stable. It is also possible to provide an excellent initial guess $\tilde{\mathbf{A}}_0$ by taking the approximate form of the matrix to be exact, yielding

$$\tilde{\mathbf{A}}_0 = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}^{-1} \end{pmatrix}$$

Based on these various advantages, it is this system which we attempt to solve rather than the original form. Below we define matrices \mathbf{M} and \mathbf{B} for convenience, and summarize the system to be solved:

$$\mathbf{M} = \mathbf{K}_1^{-1} \left[\begin{pmatrix} \mathbf{D}^{-1/2} \mathbf{N} \mathbf{D}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{D}^{-1/2} \mathbf{P} \mathbf{D}^{-1/2} & \mathbf{D}^{-1/2} \mathbf{C}^\top \\ \mathbf{C} \mathbf{D}^{-1/2} & \mathbf{0} \end{pmatrix} \right] \mathbf{K}_2^{-1} \quad (3.40)$$

$$\mathbf{B} = \tilde{\mathbf{A}}_0 = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}^{-1} \end{pmatrix} \quad (3.41)$$

$$\mathbf{M} \tilde{\mathbf{A}} = \mathbf{B} \quad (3.42)$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{D}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{K}_2^{-1} \tilde{\mathbf{A}} \quad (3.43)$$

3.6 Solver

The matrix $\tilde{\mathbf{A}}$ for which we attempt to solve will have as many columns as there are signals we wish to denoise within an event; often, this will only be one, and $\tilde{\mathbf{A}}$ will simply be a column vector. However, there certainly are many events with multiple scintillation signals; we will also later discuss extended applications of denoising where we may treat events as having additional signals. In these cases the number of columns may be larger. In such cases it would be possible to solve for each column of $\tilde{\mathbf{A}}$ independently, but more efficient to solve the entire system simultaneously. In this way, information obtained from multiplying the matrix by one column may be exploited to solve the other columns as well, effectively multiplying the benefit from each matrix-multiplication by the number of columns of $\tilde{\mathbf{A}}$.

Matrix-solving algorithms which handle multiple vectors simultaneously are commonly referred to as block algorithms. We will use the Block-BiCGSTAB algorithm [41], an iterative method based on the popular stabilized biconjugate-gradient (BiCGSTAB) method. BiCGSTAB and Block-BiCGSTAB are notable for being able to deal with arbitrary matrices rather than requiring that the matrix be positive-definite (which our matrix is not). The algorithm is reproduced in the form used by our code as Algorithm 1. This algorithm only requires that we are able to left-multiply vectors and matrices by \mathbf{M} quickly, which makes it possible to exploit the sparsity of \mathbf{M} .

Since each column of $\tilde{\mathbf{A}}$ corresponds to one signal we are refitting, the norm of the corresponding column of \mathbf{R} tells us how well we have identified the optimal magnitude estimator for that signal; thus, to test for termination we should evaluate the norm of each column of \mathbf{R} , and while any of those norms are above some threshold, the solver must be permitted to continue. Because our resolution is only on the order of 1%, solutions generally need not be very accurate; in practice, we establish reasonable thresholds by plotting resolution against threshold for a small subset of data, and observing when the resolution stops improving.

3.7 Implementation

Although the previous section described the algorithm used for solving these matrix equations, in practice the processing time required is still significant; combined with the very large number of events which we hope to denoise (hundreds of millions), we

Algorithm 1 Block-BiCGSTAB Algorithm

```
1:  $\tilde{\mathbf{A}} \leftarrow \tilde{\mathbf{A}}_0$ 

2:  $\mathbf{R} \leftarrow \mathbf{B} - \mathbf{M}\tilde{\mathbf{A}}$ 

3:  $\mathbf{S} \leftarrow \mathbf{R}$ 

4:  $\tilde{\mathbf{R}}_0 \leftarrow \mathbf{R}$ 

5: loop

6:    $\mathbf{V} \leftarrow \mathbf{M}\mathbf{S}$ 

7:   Solve  $(\tilde{\mathbf{R}}_0^\top \mathbf{V})\alpha = (\tilde{\mathbf{R}}_0^\top \mathbf{R})$  for  $\alpha$ .

8:    $\mathbf{R} \leftarrow \mathbf{R} - \mathbf{V}\alpha$ 

9:    $\mathbf{T} \leftarrow \mathbf{M}\mathbf{R}$ 

10:   $\omega \leftarrow \langle \mathbf{T}, \mathbf{S} \rangle_F / \langle \mathbf{T}, \mathbf{T} \rangle_F$ , where  $\langle \cdot, \cdot \rangle_F$  is the Frobenius matrix norm.

11:   $\tilde{\mathbf{A}} \leftarrow \tilde{\mathbf{A}} + \mathbf{S}\alpha + \omega \mathbf{R}$ 

12:   $\mathbf{R} \leftarrow \mathbf{R} - \omega \mathbf{T}$ 

13:   $\mathbf{R}$  currently equals  $\mathbf{B} - \mathbf{M}\tilde{\mathbf{A}}$ 

14:  if all columns of  $\mathbf{R}$  have sufficiently small magnitudes then

15:    return  $\tilde{\mathbf{A}}$ 

16:  end if

17:  Solve  $(\tilde{\mathbf{R}}_0^\top \mathbf{V})\beta = -(\tilde{\mathbf{R}}_0^\top \mathbf{T})$  for  $\beta$ .

18:   $\mathbf{S} \leftarrow \mathbf{R} + (\mathbf{S} - \omega \mathbf{V})\beta$ 

19: end loop
```

still must seek out any possible speed-ups. There are certain implementation tricks which can be applied with significant impact. These are described here.

The majority of non-zero matrix entries comes from the electronic noise matrix \mathbf{N} , making it the computational bottleneck of the algorithm. \mathbf{N} is block-diagonal, as seen in equation 3.28; if we assume roughly 70 APD channels, then the last block of \mathbf{N} will have 70^2 non-zero entries and all other blocks will have $4 \cdot 70^2$ non-zero entries. With waveforms having 2048 samples, and ignoring the baseline component, the problem of multiplying a vector by the electronic noise terms thus consists of 1024 matrix-matrix multiplications, with each left matrix containing roughly 20,000 non-zero terms and each right matrix consisting of $O(1)$ columns and roughly 150 rows. (The number of columns in the right matrix is equal to the number of signals, which may be sometimes 3 – 5 or more; however, it is always much smaller than 150, and describing it as $O(1)$ captures this fact.)

The key points here which create an opportunity for optimization are that:

- The electronic noise terms generally do not change event-by-event. They are treated as constant across many runs.
- It is possible to multiply two $N \times N$ matrices together much faster than it is possible to perform N multiplications of an $N \times N$ matrix with N $N \times 1$ matrices. This can be achieved with a combination of algorithms faster than naive $O(N^3)$ speed and exploitation of low-level computer hardware features such as the CPU cache. [22] [38]

What we wish to do, then, is reorganize the solver algorithm so that whenever

multiplication by the electronic noise matrix is required, rather than performing that multiplication on a “skinny” $150 \times O(1)$ matrix, we pack together many such “skinny” matrices into a single matrix with many columns. Multiplication can then be performed in bulk; and individual columns from the result can be extracted and used as before.

Additionally, it is important that matrix multiplication be made as fast as possible. It has long been known that matrix multiplication provides significant opportunities for low-level optimizations. For example, most of the time taken by naive matrix multiplication is spent fetching and writing data to and from RAM. Significant speedups can be achieved by minimizing the number of CPU cache misses, which can be accomplished by operating on matrices in blocks with size chosen so they fit entirely in the cache. Multiplication instructions can also often be packed into vector instructions such as SSE or AVX. For extremely large matrices, there are even algorithms which require fewer than $O(N^3)$ multiplications, and these can sometimes be beneficial. [38]

Optimization of matrix-matrix multiplication is a large field, but fortunately there are a number of well-tuned software libraries implementing matrix multiplication well-tuned to specific machines. These libraries generally provide the standard Basic Linear Algebra Subroutines (BLAS) interface, making them interchangeable with ease. In this instance, Intel’s MKL library has been used; benefits to this implementation are its availability on scientific computing platforms and its ability to adapt to heterogeneous compute clusters by detecting the architecture of the machine on which it is run. [46]

Finally, implementation at NERSC has the interesting feature that NERSC’s computing systems are designed for multi-core processing. The Hopper and Edison computing systems at NERSC allocate whole nodes, each of which contains 24 cores. Although it is possible to simply run 24 independent processes on each node, memory on NERSC machines is highly constrained; memory use can be reduced by sharing certain large data structures across multiple cores on a node. Additionally, as more events are packed together for electronic-noise multiplication, greater savings can be realized; so it is better to handle many events in coordination. To this purpose, a multi-threaded version of the program has been developed to exploit portions of the code which are conveniently distributed. Because NERSC nodes have a Non-Uniform Memory Access (NUMA) architecture, processes are constrained to only run on cores with a similar memory access pattern; on Hopper this results in four 6-threaded processes per node, while on Edison this results in two 12-threaded processes per node.

As a result, and because the bulk of our computational time is spent multiplying by \mathbf{N} , we pursue the following strategies to improve code performance:

1. One denoising process will be run on each NUMA node of a NERSC machine. Processes should be run in “strict memory mode,” where a process only has access to memory resources on its own NUMA node.
2. Within a process, algorithm 1 should be performed in parallel on all available cores to maximize CPU utilization.
3. When a thread encounters the need for a matrix multiplication in lines 2, 6,

or 9, the thread should not multiply by \mathbf{N} . Instead, it should push the left hand side into new columns of a matrix \mathbf{RHS} of “thin” matrices needing to be multiplied by \mathbf{N} .

4. When enough columns have accumulated in \mathbf{RHS} , collect the running threads; these threads should each perform some fraction of the matrix multiplications between blocks of \mathbf{N} and the corresponding rows of \mathbf{RHS} . When this is complete, the threads should resume their processing of algorithm 1 in parallel on the available pool of events.
5. Every matrix multiplication should be performed by a direct call to the Intel MKL library. To facilitate this process, \mathbf{N} should be stored in a format that ensures its blocks are accessed in a contiguous segment of memory. Furthermore, because there is ample opportunity for higher-level parallelization, at no time should a multithreaded BLAS operation be invoked; this reduces computational overhead of parallelization.

These represent the most significant set of computational optimizations. However, there are also smaller optimizations which can still produce non-negligible speedups. The naive Block-BiCGSTAB algorithm (in the case of only one signal in an event) requires two matrix-vector multiplications, but it also requires a number of vector-vector operations, and the cumulative cost of these vector-vector operations can become non-negligible. One obvious shortcut is to reuse our computation of $\tilde{\mathbf{R}}_0^\top \mathbf{V}$ from line 7 of Algorithm 1 to save work in line 17.

Another speedup for our vector-vector operations consists of combining mul-

tiple operations into one step. We see on line 7 that both $\tilde{\mathbf{R}}_0^\top \mathbf{V}$ and $\tilde{\mathbf{R}}_0^\top \mathbf{R}$ are computed. As described earlier in the context of multiplication of vectors by electronic noise, it is possible to take advantage of caching and speed up vector operations by packing multiple vectors together into a more square-shaped matrix. In this case, rather than computing these two products separately, we can pack \mathbf{V} and \mathbf{R} together and compute

$$\tilde{\mathbf{R}}_0^\top \begin{pmatrix} V & R \end{pmatrix}$$

as one step. The rest of the algorithm will not be affected by the fact that these two matrices are stored contiguously in memory.

3.8 Denoising in Practice

The preceding description of denoising has been intended to convey our fullest current understanding of denoising as it should be performed. However, due to the computational intensity of this algorithm it is not feasible to conduct routine data processing of the full EXO-200 dataset whenever updates to the code base are implemented. As a result, the most recent denoised data available from EXO-200 demonstrates certain deviations from the description above. It is known, and has been demonstrated, that correcting some of these flaws can lead to further improvements in resolution, but no physics analysis will be performed with those improvements until the computational cost of denoising is again deemed worthwhile.

In this section we outline the places where the denoising used for the current physics analysis differs from the description above. Differences in implementation

also exist, but will not be described here. The differences of denoising are:

- The signal variance coming from sources internal to the APDs was not fully understood by this author when denoising was first implemented, and as a result the signal variance is significantly underestimated. Specifically, in place of equations 3.8 and 3.10 we take

$$\left\langle P_{ia}^{(3)} P_{jb}^{(3)} \right\rangle = (1.9)^2 \cdot G_i^P(t_a) G_j^P(t_b) \left\langle P_{ia}^{(1)} P_{jb}^{(1)} \right\rangle.$$

- We have described with equation 3.14 that the electronic gain is treated as having a constant value of $1.1 \cdot 10^{-3}$ ADC counts per electron emitted from the APDs. The wires use a similar set of electronics; however, their gain is nominally set to a different value of roughly $6.3 \cdot 10^{-4}$ ADC counts per electron collected on the wires, and this author confused the two numbers. As a result, $G_i^E(t)$ was underestimated by almost a factor of two; the effect of this is to overestimate by almost a factor of two the number of photons being deposited on APDs, which again causes us to underestimate the signal variance introduced within the APDs.

The effect of both of these issues is, loosely speaking, to fool denoising into believing its photon statistics are greater than they truly are; this, in turn, leads it to prefer a higher concentration of signal (more weight on fewer channels) than is ideal. Thus, the resolution is negatively impacted by these mistakes; the impact has been measured in subsets of the the EXO dataset and this negative impact is apparent.

However, neither of these mistakes have any impact on the correctness of the constraints. As a result, the denoised energy measurements should still be unbiased, even if they have worse variance than necessary. Results from the use of this data will be described in chapter 6.

3.9 Future Plans

As outlined here, denoising is capable of producing optimal estimates of the total energy emitted in scintillation by a particular event; however, it should be noted that this algorithm can be applied to a more general range of problems in the EXO-200 detector with only minor modifications.

3.9.1 Anticorrelated Cluster Energies

One important generalization is in our definition of “signals.” In the body of this chapter, we have worked on the assumption that a signal consists of a simultaneous set of energy deposits in the bulk of the detector, possibly at multiple locations. We do this because it must be possible for denoising to break degeneracies between signals – by specifying that two different signals must have occurred at different times, we give the algorithm a means to discriminate between light coming from the two different signals and assign energy properly to each of them.

Historically, time was the only useful variable for this sort of discrimination; because multiple APD channels were added together and all APD channels were treated as identical in this sum, a significant amount of information was discarded

and there were no remaining parameters to use for signal separation except for the signal time.

However, this denoising algorithm makes use of additional information: the relative magnitudes of each APD gang. By depending on a highly detailed lightmap, we have incorporated information about which channels should see light most from energy deposits at different positions. As a result, if we permit two “signals” which are simultaneous in time, it should be possible to separate their energy content by comparing their positions to the set of channels which see more light. This would allow us to produce anti-correlated energy measurements of individual clusters, which should give a significant improvement in the energy resolution achievable for clusters within an event.

In principle, this is entirely feasible – in the extreme case, when two clusters are at opposite ends of the detector we can expect them to produce light on an entirely different set of channels, and it should be quite easy to separate light produced by each. However, this extension of our algorithm does present some practical challenges.

In EXO’s reconstruction, light signals which are found close together in time are aggressively merged together, ensuring that denoising will never encounter light signals which are very close in time; this assures us that two signals separated in time will be well-separated, and there will not be a near-degeneracy which could render the solver algorithm unstable.

However, charge clusters are merged much less aggressively, and some bugs exist which have minimal effect on the main analysis but permit the existence of

clusters very near to each other. For denoising to separate simultaneous light signals solely based on their relative position, they must be separated far enough that their respective light yields on nearby APD channels are significantly different, which is not at all guaranteed by reconstruction. As a result, denoising can only fit simultaneous signals separately if it performs a re-clustering process in which clusters occurring near each other are artificially merged together for the purpose of denoising. The threshold for merging clusters together would require further study, and might be different in each of the three coordinate directions.

Such an analysis would not have a significant impact on single-site data, so it would have minimal impact on our primary goal of searching for double-beta decay. However, it could have a significant impact on a search for excited-state two-neutrino double-beta decay by improving the resolution of gamma lines associated with Barium de-excitation. Barabash states in [16] that the dominant mode of excited-state $2\nu\beta\beta$ decay will be to the 0_1^+ state, which in ^{136}Ba would lead to the emission of a 760 keV and 819 keV gamma in coincidence; if denoising could be accomplished on individual clusters, the sensitivity to these gamma lines could be significantly enhanced.

There are also potential benefits to our understanding of the detector itself. Currently the only mono-energetic beta calibration available for EXO is the double-escape peak, formed when a gamma particle enters the detector, pair-produces an electron and positron, and the positron interacts with an electron and emits gammas which travel sufficiently far from the initial interaction site so that the initial electron interactions and gamma interactions can be clearly separated. Currently we have

the ability to separate charge yield from gammas and the initial electron in some of these events, allowing us to understand the different charge yields from gammas and betas. However, we have no similar capabilities for light yield because the gamma and beta interactions are treated as just one “signal.” If we could separate the light yield from these spatially separated clusters, we could use these events to understand both the light and charge yield for different types of interactions.

Another type of detector study which might benefit from anticorrelated cluster energies is the Compton telescope. This is a technique that combines our knowledge of cluster energies and scattering angles, along with the Compton scattering formula [26], which allows us to produce estimates of the origin of the incident gamma. Such a technique permits us to study the sources of our backgrounds, informing future attempts to reduce them by indicating which materials had the most significant negative effect. The accuracy by which we place the origin of the incident gammas is strongly related to the accuracy of our measurements of the energies of the individual clusters, so it is expected that improving the accuracy of our cluster energies would have a positive impact on our ability to locate the backgrounds of our experiment.

The last benefit to this approach is more technical: evaluating the lightmap for a multi-site event requires an energy-weighted combination of the values of the lightmap at individual cluster locations. This process is complicated and error-prone, and can be avoided entirely when we treat each signal instead as having a single well-defined position.

3.9.2 Wire Denoising

The current EXO-200 resolution is limited by the scintillation resolution of the APDs. As a result, little benefit is anticipated from denoising the wires. However, it is possible that a future detector could reduce the scintillation noise to a point where resolution of the wires is non-negligible; alternatively, a future detector could have noise which is correlated between wires and APDs, making it possible to use noise information from the wires to improve our scintillation resolution even though no direct benefit is expected to the wire resolution itself.

In any case, it is possible to apply equations 3.22 to a system with a combination of wires and APDs without significant adjustment. The template functions $Y[t]$ will of course be significantly different, and must include a time offset to account for the different time of arrival of the electrons and photons. The “lightmap” of charge will be far less diffuse than for APDs because charge from any location generally drifts directly to only a small number of wires along externally imposed field lines. And finally, because electron statistics should be negligible and signal amplification should not be a noisy process, for wire channels we would modify equation 3.21 to state that $q(i_{wire}, b) = 0$. In this sense, denoising of wires consists of minimizing the variance in the energy estimator due to electronic noise, and using the wires to improve our denoising of APDs consists of using information about the wires to reduce variance due to electronic noise without expecting any correlations between APD and wire signals, all while ensuring that the energy estimates are unbiased by retaining the same constraints.

Chapter 4: Noise Measurements

Chapter 5: The Lightmap

The EXO detector was designed with roughly 450 APDs ganged into 74 data channels of six to seven APDs each. Three of the APD ganged channels were disabled due to noisy components before physics data collection began; a fourth channel was disabled in February 2012 due to increasing noise in its components. All of the APDs were set into the two endcaps of the cylindrical EXO detector. To improve light collection, teflon sheets cover the sides of the detector and reflect light back into the liquid xenon rather than allowing it to be absorbed by the vessel walls.

The different positions of the APDs means that given the same amount of energy deposited into the detector, channels whose APDs are nearest to the deposited energy may show significantly larger signals than channels with APDs far from the deposit. As a result, it is critical to map out the response of the APDs as a function of deposit position \vec{x} .

Furthermore, the APDs and APD electronics will show time-dependent changes: gain will drift in each APD at independent rates, and stepwise changes to the APD signals comes from disabling of channels and changes to the electronics which have been implemented multiple times during the course of the experiment. This means

that in addition to mapping the response of the APDs as a function of deposit position \vec{x} , we must also map it as a function of deposit time t .

Early attempts to generate this mapping focused on periodic campaigns to collect large amounts of known-energy data and characterize the APD response at a moment in time. These campaigns would generally consist of days of source data. The strong Thorium-228 source would be used, and an expert would be on-site to position the source in a wide range of locations, to ensure that a single campaign was independently sufficient to characterize the APD response. The 2615-keV single-site gamma line of the Thallium-208 daughter product of Thorium-228 was isolated in offline analysis, and signals from these events were used to measure the signal magnitude from a known-energy single-position deposit in the detector.

Even with such a significant quantity of source data, statistics were found to be insufficient. Some regions of the detector were difficult to illuminate, and signals on individual channels were small. To simplify the problem, signals on each endcap were summed together (without gain corrections) so that the 70 or 71 active APD channels could be treated offline as two large-magnitude APD channels instead. This increased the size of the signals; it also made the spatial dependence of the response smoother, so that a sparser distribution of data could still be sufficient to characterize the response. This permitted the creation of an APD-plane lightmap which allowed EXO to produce its first position-dependent corrections to scintillation energy. [44]

Although the APD-plane lightmap produced significant improvements in the energy resolution achieved by the EXO detector, inevitably it was only an incomplete

characterization of the APD response. Indeed, because for $\beta\beta 0\nu$ studies we place our fiducial cut as close to the edges of the detector as possible, in certain regions of the detector our scintillation signal might be highly concentrated on one or a small number of channels. Summing together multiple channels is, in this sense, a lossy form of compression of the data, and it is tempting to see whether we can extract better physics if it is avoided. The denoising algorithm described in chapter 3 describes one such analysis which can exploit knowledge of the behavior of individual APD channels; others might include reducing our scintillation thresholds by allowing us to search for signals in subsets of the APDs rather than only in sums across entire endcaps. Thus, characterizing the APD yield on individual APD-gang channels is not in itself an important component of our analysis, but may be expected to provide a critical tool for more advanced scintillation analysis.

5.1 Four-Dimensional Lightmap

As described above, in constructing an individual-APD lightmap we face two conflicting needs: we must use as much data as possible to handle the faster spatial dependence and smaller signals expected, but if too much data is included then we run the risk of combining data taken when an APD or the electronics had a different gain. If we truly wish to use all available data, then we will need to simultaneously understand the full time-dependence of the gain. In other words, rather than forming a small number of independently-measured three-dimensional lightmaps, we will need to measure a four-dimensional lightmap $L(\vec{x}, t)$. Since we will use Thorium-228

source data to generate the lightmap, we require that our lightmap should predict the magnitude of the signals on each APD gang induced by a single-site deposit of energy 2615 keV; figure 3.2 shows our definition of a unit-magnitude signal, which is the same as was used in chapter 3.

This may at first seem infeasible. After all, adding an extra time dimension to the arguments of the lightmap is equivalent to generating a new lightmap for every time bin, which is exactly the situation we wished to avoid. But we can make a key simplifying assumption that the lightmap is separable. In a physical sense, we can assume that:

1. From a given position \vec{x} photons deposit on each APD channel at a constant rate.
2. Each APD channel in turn magnifies and shapes its signal with a gain which may vary in time, but does not depend on the point of origin of the photons.

So, we demand that the lightmap have the much simpler form

$$L(\vec{x}, t) = R(\vec{x})S(t) \tag{5.1}$$

Is this simplification fully motivated by the detector? Although it is a reasonable approximation, there are details which it omits. The first point is fairly accurate: we must trust that the electric field and the reflectivity of detector surfaces are constant in time, but these conditions are assumed to hold within EXO-200.

However, each channel can draw its signal from multiple APDs, each of which has an independent time-varying gain. Photons from a deposit may preferentially

sample the gain from the closest APD within a channel, so deposits in different locations may track more closely the gain of the closest APD within a channel, and the second point may not be accurate. Studying the impact of this effect should be a topic for future study, particularly to understand to what degree it limits the effectiveness of a lightmap; it may be relevant to nEXO planning to understand how ganging together APDs reduces the quality of the approximation of equation 5.1. The current analysis will treat it as a small effect.

So, treating equation 5.1 as valid, one simple general scheme to find R and S iteratively is described in algorithm 2. Step 1 of that algorithm is to select 2615-keV single-site events; it is understood that compton-scattered events will inevitably leak into our selection window, but we use the best available energy measurements to reduce that likelihood and trust that as analysis improvements continue to sharpen our resolution, the number of compton-scattered events leaking through becomes smaller and the quality of the lightmap will continue to improve. The rest of the algorithm then describes how we can efficiently converge on estimates of $R(\vec{x})$ and $S(t)$ through an iterative approach. We do not prove that the algorithm is guaranteed to converge; however, in practice $S(t)$ is close to constant, resulting in fairly rapid convergence.

5.2 Algorithm Details

The previous section outlines a general algorithm for computing a four-dimensional lightmap from source data. In this section we will specify the implementation choices

Algorithm 2 Generating a Lightmap

- 1: Tabulate, using the best energy measurement available, a list of single-site 2615-keV events from Thorium-228 source data.
 - 2: Set $S(t) = 1$ for all APD channels.
 - 3: **repeat**
 - 4: For each APD channel, estimate $R(\vec{x}) = L(\vec{x}, t)/S(t)$ from the set of events tabulated in line 1.
 - 5: For each APD channel, estimate $S(t) = L(\vec{x}, t)/R(\vec{x})$ from the set of events tabulated in line 1.
 - 6: **until** convergence is reached.
-

which are made in the code currently being used. In attempt to encourage future experimentation, alternative options will also be listed in some detail, along with some motivations for these alternatives. Time has not permitted testing of all of these options, but it is hoped that they will be explored in the future.

5.2.1 Event Selection

The single-site Thorium-228 spectrum is well-peaked at 2615 keV, making it an excellent monoenergetic calibration line. Our challenge is to select truly 2615-keV events as efficiently as possible, while simultaneously avoiding near-2615-keV events which may leak into the dataset.

This is an inherently iterative process. The first lightmap had to be constructed from an ionization-only spectrum because no useful position-corrected scintillation measurement existed; the resolution of the ionization-only spectrum was

quite poor, roughly 3.5% (sigma) of the energy, as described in Steve Herrin’s thesis, and severe compromises had to be made to keep the Compton-shoulder leakage down to acceptable levels. That work required events to have ionization between 0.33σ and 3σ to the right of the ionization peak. [44]

Such a strongly asymmetric cut was chosen to avoid leakage from Compton shoulder events, which it did successfully. However, because of anticorrelation between scintillation and ionization, choosing events whose ionization had oscillated high means this cut was also preferring events whose scintillation had oscillated low, introducing a bias in the lightmap magnitude. Additionally, the cut permits less than 37% of desired events to be accepted, which is a substantial loss of efficiency that has a significant impact in certain low-statistics regions of the detector.

The current work has had the benefit of an existing position-corrected scintillation measurement, leading to a rotated energy spectrum with resolution of roughly 1.8%. As a result, it is possible to accept a wider cut (in sigmas) while still keeping Compton shoulder leakage small. We currently accept all events within 2σ of the peak, with better than 95% acceptance and only small leakage. An additional benefit is that this improved resolution allows us to use a symmetric acceptance region, avoiding the implicit bias introduced from the earlier asymmetric cut window.

Preliminary investigations have been conducted to see what impact is observed from using the denoised scintillation measurements which are the subject of this work. Presently the improvement in resolution (to roughly 1.4%) UPDATE WITH REAL NUMBER when the paper is finalized has not demonstrated any significant improvement in the quality of the lightmap beyond what we could achieve with

1.8% resolution; however, this is a topic of continuing study which will certainly be revisited.

Beyond the cuts described above (single-site within an appropriate energy window), other basic cuts are applied to ensure only well-behaved scintillation signals are selected:

- Charge and light must individually lie within a reasonable range.
- The charge-to-light ratio must not identify an event as an alpha decay.
- The scintillation time must be well-separated from any other scintillation signals in same event frame.
- All charge signals which are assigned to this scintillation signal must be assigned unambiguously.
- All three position coordinates must be known. (No fiducial cut is placed, since that would restrict the volume where the lightmap is specified.)

Many of these cuts are probably no longer necessary – as the resolution has improved, the chances of contamination have decreased and we can be a bit less cautious about accepting events. It is worth remembering, though, that our main analysis only accepts events with a single scintillation signal, something which for us would carry too high a cost in lost statistics. (Strong Thorium runs produce a significant rate of events with multiple scintillation signals. These statistics are quite valuable, particularly because they are the primary source of statistics in some hard-to-reach

regions of the detector.) As a result, it is important to maintain some caution with cuts designed to handle those events more robustly.

5.2.2 Function Parametrization

Because we are attempting to empirically measure the functions $R(\vec{x})$ and $S(t)$ from a finite dataset, we need to specify some more limited form for them to take. All current approaches to describing $R(\vec{x})$ first bin the detector volume into three-dimensional voxels, and then define $R(\vec{x})$ to interpolate linearly between known values at the center of the voxels. The size of these voxels must be chosen with some care; if they are too small, then low per-voxel statistics will cause the statistical errors on the signal magnitude to dominate, whereas if the voxels are too large then the spatial variation of the lightmap will not be fully captured.

In the current lightmap, the detector is binned into 1cm^3 voxels; the detector is easily contained within a box with sides 40cm long, leading to 64,000 voxels (of which roughly 20% lie outside of the detector and will be empty). The choice of voxel size was made based on the size of an individual APD. The APDs are roughly 2.5cm in diameter, so very near the anode we would like for the size of a voxel to be much smaller than 2.5cm . When this is done, much of the detector has entirely sufficient statistics per voxel; however, there are some regions of the detector with fewer than ten hits per voxel, indicating that $R(\vec{x})$ may have quite significant statistical error in these regions.

It is possible to justify a choice of larger voxels. The APDs lie at $\pm 204\text{mm}$

from the cathode, which means that there is more than 2cm between our fiducial volume and the APDs. At this distance, the dependence of the lightmap on x and y is expected to be much slower than at the APD plane, indicating that perhaps 2cm binning in x and y may be sufficient. Additionally, z -dependence of the lightmap is expected to be fairly smooth throughout the detector; since we interpolate linearly between voxels, it may be possible to use a z -binning much coarser than 1cm . This may be a topic for future investigation.

It is also worth mentioning that less intuitive voxel geometries have been tried in the past. The older APD-plane lightmap [44] used a cylindrical-coordinate binning; binning in r was selected to make the bin volumes roughly constant along the r axis, binning in the angular coordinate was uniform, and binning in z was chosen to be coarser near the anodes and finer near the cathode to reflect faster variation there. In all cases the binning was coarser than with the current cubic voxels being used; our finer binning is made possible by the larger quantity of available statistics from using the full dataset, and is justified by the potential for the yield on a single APD gang to vary more rapidly than the yield averaged across an entire APD plane.

It is well-known [57] that when it is necessary to estimate a multivariate function from limited statistics, a choice of binning can have a significant impact on the result. It is preferable to use an unbinned technique such as kernel regression. In particular, our data density is highly non-uniform, and it should be possible to measure the lightmap with high fidelity in regions of high statistics, while smoothly transitioning to a coarser interpolation in regions of low statistics to minimize the impact of uncertainty from individual hits. State-of-the-art solutions to this problem

would rely on locally adaptive kernel regression; see [57] for a detailed explanation of the related issues in non-parametric multivariate regression. Attempting to use a locally adaptive kernel regression for $R(\vec{x})$ should be considered a highly appealing extension to the algorithm for generating a lightmap.

The parametrization of $S(t)$ presents a very different set of choices. Thorium data is taken in bursts, with the time between mostly filled by low-background runtime. When we choose to treat $S(t)$ as a smoothly varying function, it becomes critical to interpolate properly – after all, the low-background runtime is the critical part of the experiment. (If we only produce a lightmap accurate during source runtime, we will measure an energy resolution from source data which is not borne out in the low-background data, so we should in fact be able to give some guarantee that $S(t)$ is almost as accurate during the low-background runtime as during the source runtime.) Fortunately, it is generally true that the time variation of the APD response is quite slow; exceptions are generally due to changes in the electronics which should occur at well-specified times.

Currently each source run is treated as a burst of data taken at the midpoint of the run, and $S(t)$ is measured at that point only from the data of that run; then $S(t)$ is linearly interpolated between those points. In principle it is possible that between two source runs an electronics change may have been performed, which would mean that a better interpolation would be to assume $S(t)$ is flat except for a step at the time of the electronics change; in practice, though, EXO has generally taken a source run immediately before and after an electronics change, so no high-value data is taken during that interval and this detail can be ignored.

Another concern with this method is what should be done with short source runs. If a run is too short, then the measurement of $S(t)$ coming from that run may have significant errors. We currently mitigate this issue by entirely throwing out all data from runs with fewer than 100 usable events. We justify this approach by claiming that even though those events might in principle have contributed some useful information on $R(\vec{x})$, without a good measurement of the relevant $S(t)$ it is impossible to use that data.

It would be better if we made use of our knowledge that $S(t)$ changes slowly or at well-defined moments. In the future, it would be useful if instead we performed smoothed interpolations between electronics changes. This could be done in the same style as EXO’s lifetime corrections, using polynomial fits, where the degree of the polynomial could be determined by eye. A candidate set of time intervals when $S(t)$ might be treated as constant could be the same as the set of constant noise windows described in chapter 4. It would be possible to check whether this time binning is fine enough by looking at the Thorium scintillation peak position versus time in denoised data – if it appears to drift in time within a window where $S(t)$ is treated as constant, then it is likely that $S(t)$ needs to be binned more finely.

Alternatively, often there will be a long string of consecutive source runs which should be combined into one higher-statistics dataset. This process would need to be done essentially by hand, and has not been performed for the current analysis, but it could benefit the accuracy of the $S(t)$ function and also recover some statistics in cases where the individual runs might be too short for inclusion in the lightmap dataset.

The choice of binning or parametrization for $R(\vec{x})$ and $S(t)$ can have a profound impact on the accuracy of the lightmap, and the current analysis has only skimmed the surface in trying options. It is hoped that further work on the lightmap will include significant investigation in these topics.

5.2.3 Convergence and Error Estimation

Our treatment of the convergence of algorithm 2 and the resulting statistical errors in $L(\vec{x}, t)$ is extremely rudimentary. Broadly convergence is checked not by analysis of the lightmap itself, but by verifying that the energy resolution from denoising is not improved by further iterations through algorithm 2. Lightmap errors do not directly enter into the calculations of denoising (which treats the lightmap as perfectly known), so they are not estimated at all in the current analysis.

As a proxy for estimation of the statistical errors in the lightmap, we instead study the number of hits observed in each position voxel of the detector (see section 5.4), and presume that the most significant source of error comes from low-statistics regions of the detector. This information can motivate future data-taking campaigns to probe the light response in those regions of the detector and reduce these uncertainties. However, no explicit estimation of the lightmap uncertainty has been attempted.

One difficulty with estimating the errors in the lightmap function $L(\vec{x}, t)$ comes from the correlation between errors in $R(\vec{x})$ and $S(t)$. If one of those two components were known perfectly, then we could treat the fit uncertainties of each hit as

independent errors, and propagate those errors into an uncertainty for each voxel of $R(\vec{x})$ (if $S(t)$ is assumed to be perfectly known) or $S(t)$ (if $R(\vec{x})$ is assumed to be perfectly known). But in reality, all of the measurements of $R(\vec{x})$ are correlated with all of the measurements of $S(t)$, meaning that even the errors of different voxels of $R(\vec{x})$ or different times in $S(t)$ are correlated with each other. Fully characterizing these errors could evolve into a significant project in itself, and it is not clear that the payoff could justify such an investment of effort or computation.

Although a full estimate of lightmap errors with correlations would probably be computationally quite intensive to produce, it might be possible to estimate the independent errors of each voxel of the position lightmap $R(\vec{x})$ or each independent run making up $S(t)$ by treating the other function as perfectly known, as described above. This would generally produce an underestimate of the uncertainty in each, but the estimate might still give some benefit.

To produce a more accurate estimate of errors, it would probably be necessary to do a simultaneous fit by varying both $R(\vec{x})$ and $S(t)$ together. In the current scheme, $R(\vec{x})$ contains far more complexity than $S(t)$ with roughly 50,000 non-empty voxels, so for each APD gang we would need to simultaneously vary roughly 50,000 parameters to obtain the optimal lightmap. It is exactly this time-intensive process which was avoided by choosing the iterative approach for measuring the lightmap; however, it would only need to be performed occasionally when deriving a new lightmap, so it is not infeasible to imagine attempting this project. The prospect of feeding in a high-quality guess obtained from the iterative method presents an additional significant time-saver. This method would not fully account for the cor-

relations between errors of the different voxels of $R(\vec{x})$ or runs in $S(t)$, but it would come closer than the naive method described earlier.

However, it is also possible that identifying the error of each voxel or run is heading down the wrong path. As described earlier, it is likely that the optimal (lowest-error) method for estimating the lightmap will be an unbinned method such as a locally-adaptive kernel regression method. Error estimation in kernel regression presents significant additional challenges compared to errors from binned parametrizations. At present, I am not aware of any simple scheme to manage this difficulty, meaning that there may be a paradoxical tradeoff that using the best method for minimizing lightmap errors simultaneously makes those errors infeasible to estimate.

Note that under the iterative method, the most naive method for estimating errors is not valid. It might appear that when $S(t)$ is initialized to a constant value of 1, we could also assign to it some constant error. Then when we compute $R(\vec{x})$, we could propagate independent errors from the fit uncertainties of signals and from $S(t)$; and when we compute $S(t)$ we could propagate independent errors from the fit uncertainties and from $R(\vec{x})$; and so forth. However, such a scheme provides no compensating feedback mechanism to force the iterated errors to a reasonable or accurate value, so there is no reason to believe the errors from such a system (if, that is, they converge at all). This difficulty may underscore the fundamental challenge associated with measuring the lightmap errors within our scheme – iterative methods are well-suited to solving a system, but when correlated errors are mistreated as uncorrelated an iterative method can easily magnify the

impact of that mistreatment.

On the topic of convergence, it has been mentioned already that by starting with a generally accurate initial guess for $S(t)$ as constant, the iterative solution method tends to converge rapidly. As a result, and because iterations take only a few hours to perform on a single machine, we currently perform three iterations and claim that convergence is approximately reached. Ideally, we would require that none of the position voxels or runs change value within an iteration by more than some fraction of their estimated errors; but in the absence of estimated errors, this is of course impossible.

It would be possible, in a conservative approach, to require each value to converge to some small fraction of an ADC unit, ensuring that the convergence is better than our ability to measure signals. Such a requirement might force us to compute more iterations than are truly warranted by our lightmap errors, but given the reasonable speed of each iteration, such a method still might not be unreasonable.

5.3 Implementation for this Analysis

The physics analysis which will be described in this paper includes data from EXO runs 2464 to 5597, which were taken between October 5, 2011 and September 1, 2013. However, at the time when the initial denoising processing was begun, the tentative range of runs to be used only extended up to 5367 on June 24, 2013. As a result, calibrated data at that point was only available up to June 24, 2013, and the lightmap had only been generated using the same set of Thorium source data.

When the run range for the present analysis was extended, a significant portion of the dataset had already been denoised with the lightmap based only on data extending to June 24, 2013. Although the possibility of creating a new lightmap was considered, this would have required a re-analysis and re-verification of the existing denoised dataset. As a result, the existing lightmap continued to be used. The functions $S(t)$, which had only been directly measured up to June 24, 2013, were assumed to remain constant between June 24 and September 1, 2013; no known changes to the APDs occurred during this time, nor did any known environmental factors change, so this assumption was considered acceptable. Future lightmaps of course will make use of a more complete dataset rather than relying on extrapolation.

Furthermore, one bug was discovered in the lightmap which was used for denoising. A set of Radium-226 source data was taken before the corresponding identifier could be set in data files, and as a result they were provisionally labeled as Thorium runs. These runs were mistakenly incorporated into the generated lightmap, and events from the 2204-keV gamma line of the Bismuth-214 daughter product of Radium-226 were selected and handled as though they were legitimate 2615-keV events.

It was possible to partially remedy this after the fact: since currently there is a one-to-one relationship between individual source runs and their corresponding points in the functions $S(t)$, it was possible to artificially erase the data points of $S(t)$ originating from Radium runs. It is not so easy to remove the impact they may have had on $R(\vec{x})$. However, the Radium runs were taken at a location which is extremely well-populated by Thorium data throughout the life of the detector; since

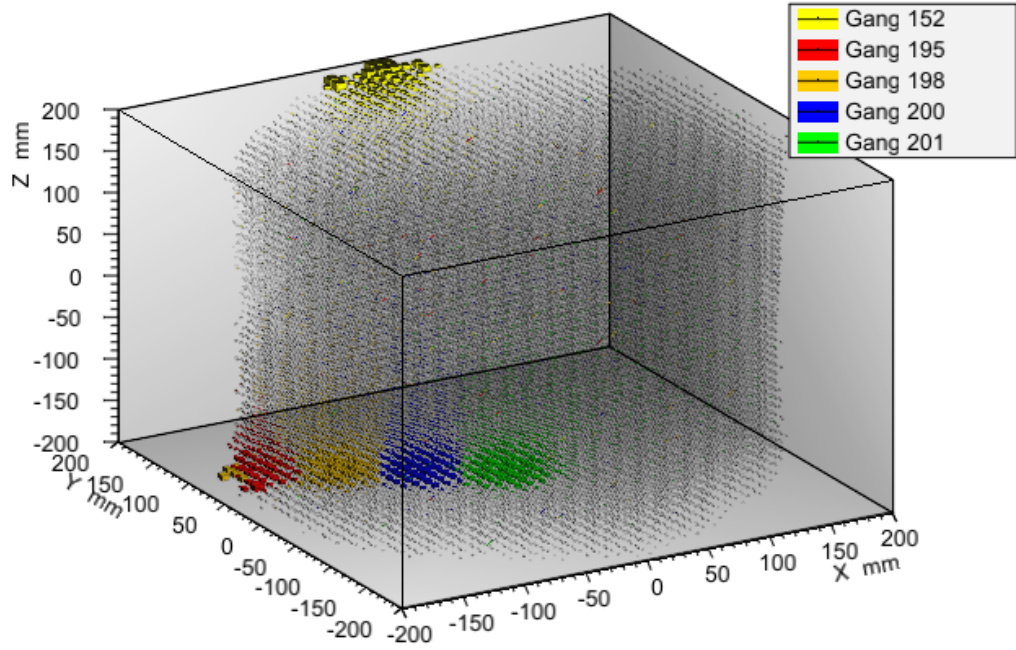


Figure 5.1: Lightmap position-dependence $R(\vec{x})$ for selected APD gangs.

the quantity of Radium data is small by comparison, the effect on $R(\vec{x})$ is presumed to be quite small. Subsequent studies using a correctly-generated lightmap on small subsets of data indicate that the effect is indeed negligible, as expected.

5.4 Visualization

Although it is not strictly necessary to be able to visually inspect the lightmap for it to be useful, nevertheless it is reassuring to see that the lightmap is qualitatively similar to intuitive expectations of how light should propagate through the detector.

We have chosen to store the position-dependent $R(\vec{x})$ for each APD gang as a set

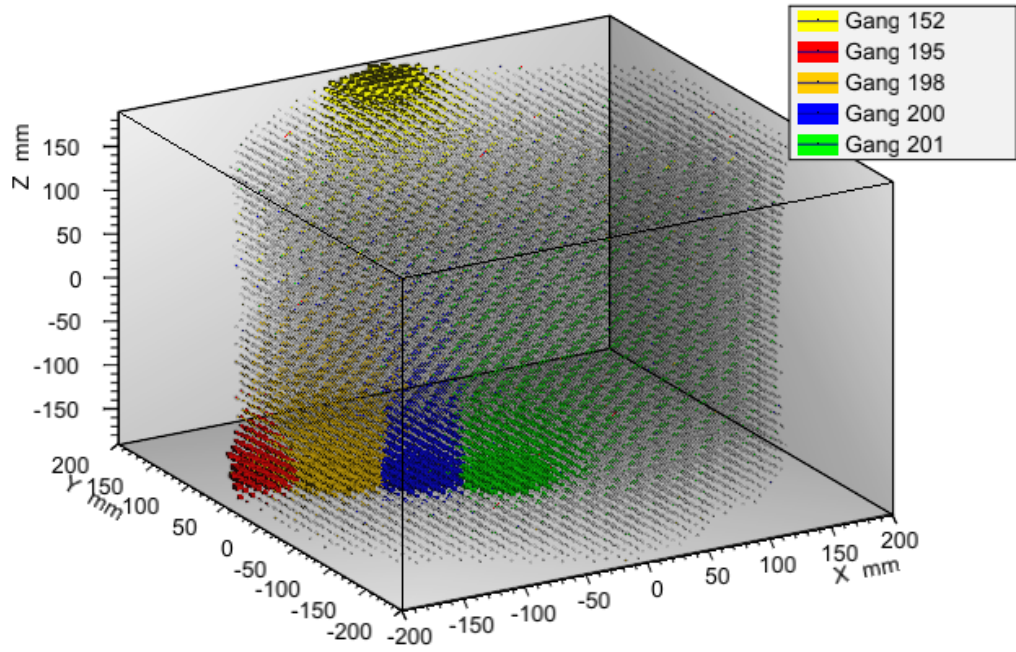


Figure 5.2: Lightmap position-dependence $R(\vec{x})$ for selected APD gangs. Here extreme anode positions are omitted to permit better contrast for the lightmap in the fiducial volume.

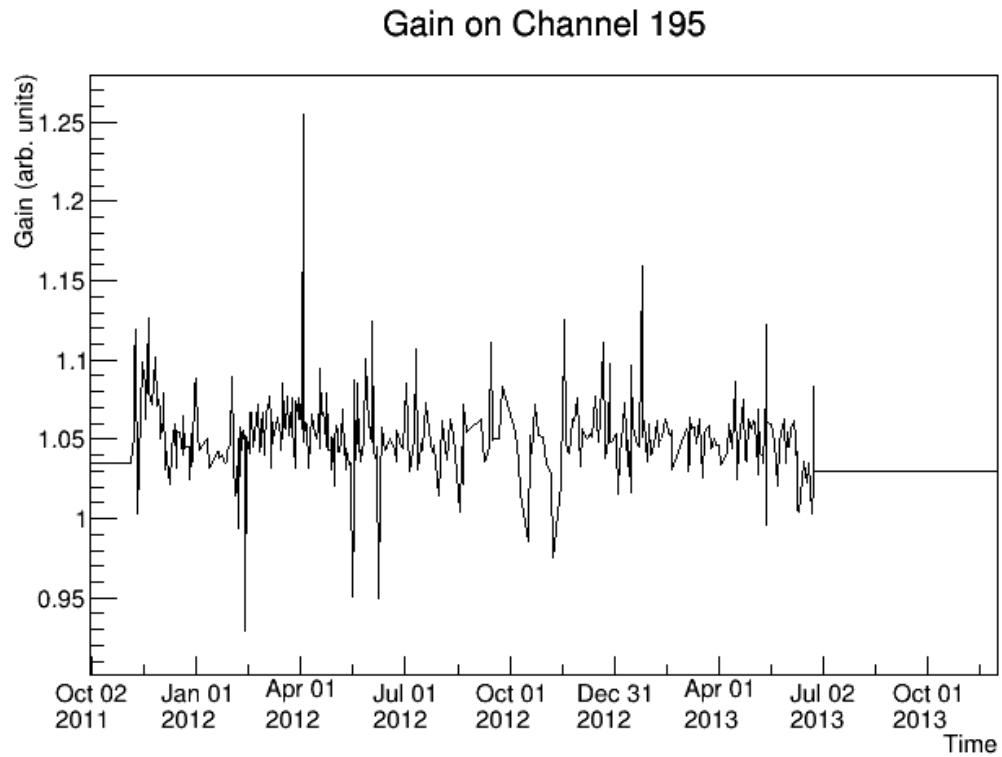
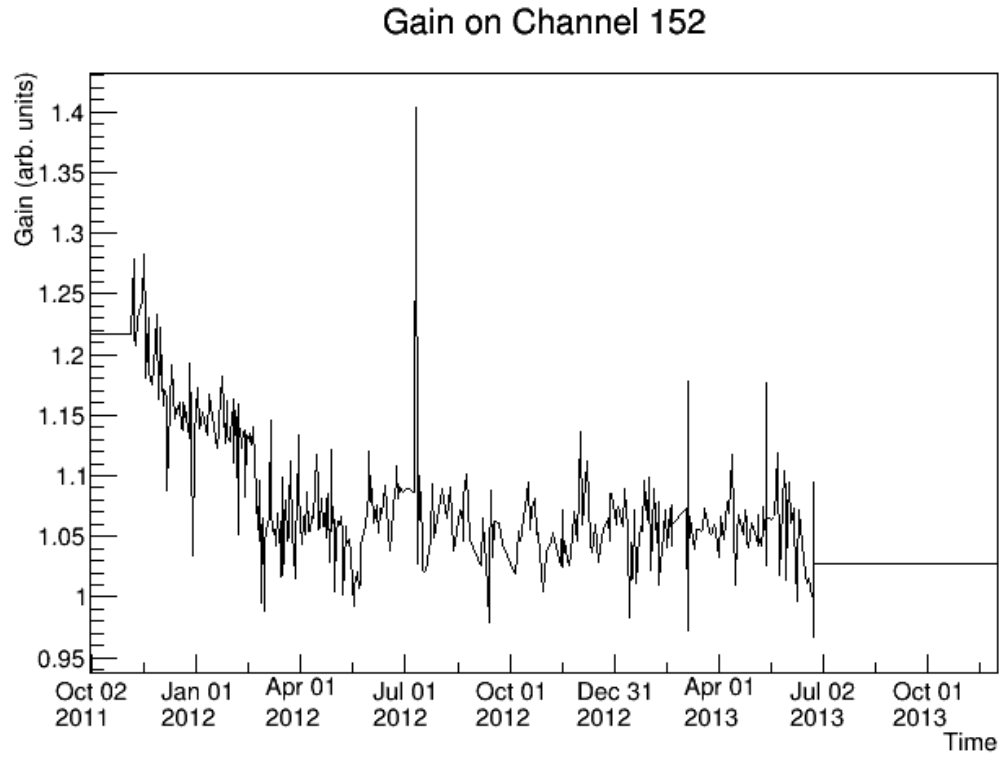


Figure 5.3: Functions $S(t)$ for selected channels.

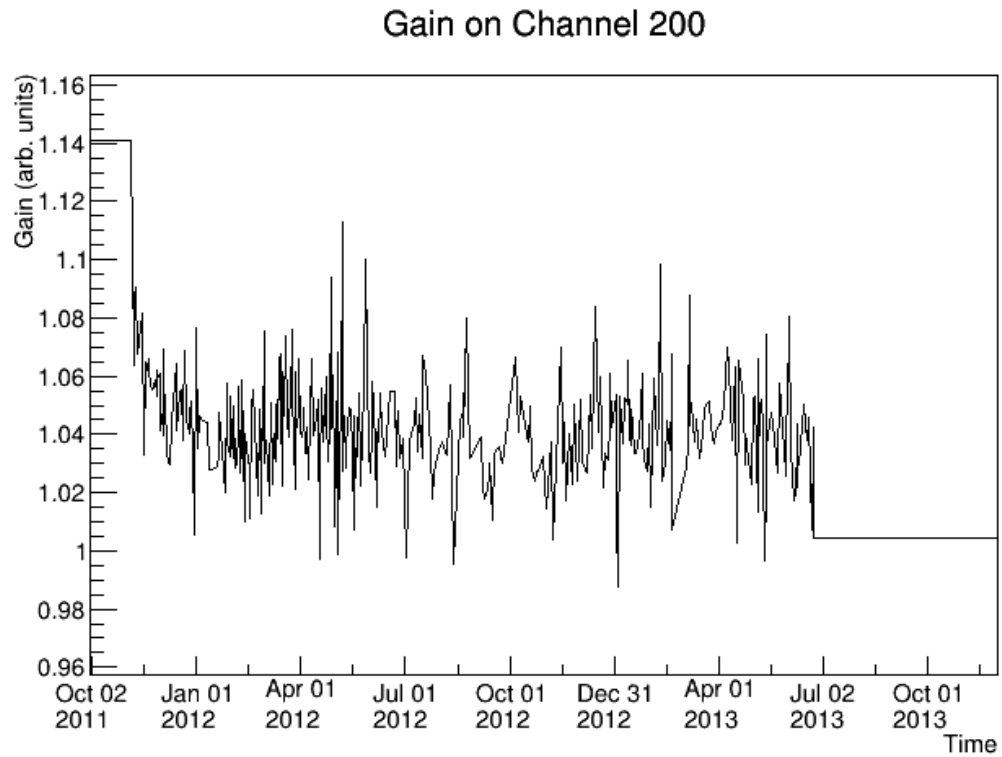
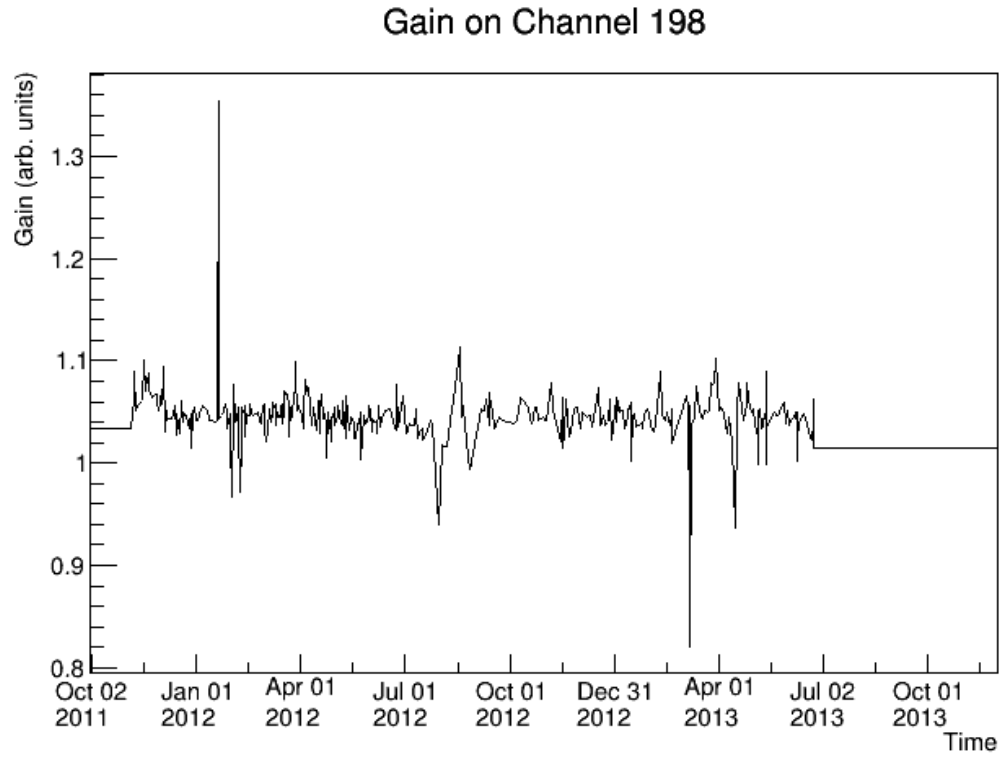


Figure 5.4: Functions $S(t)$ for selected channels.

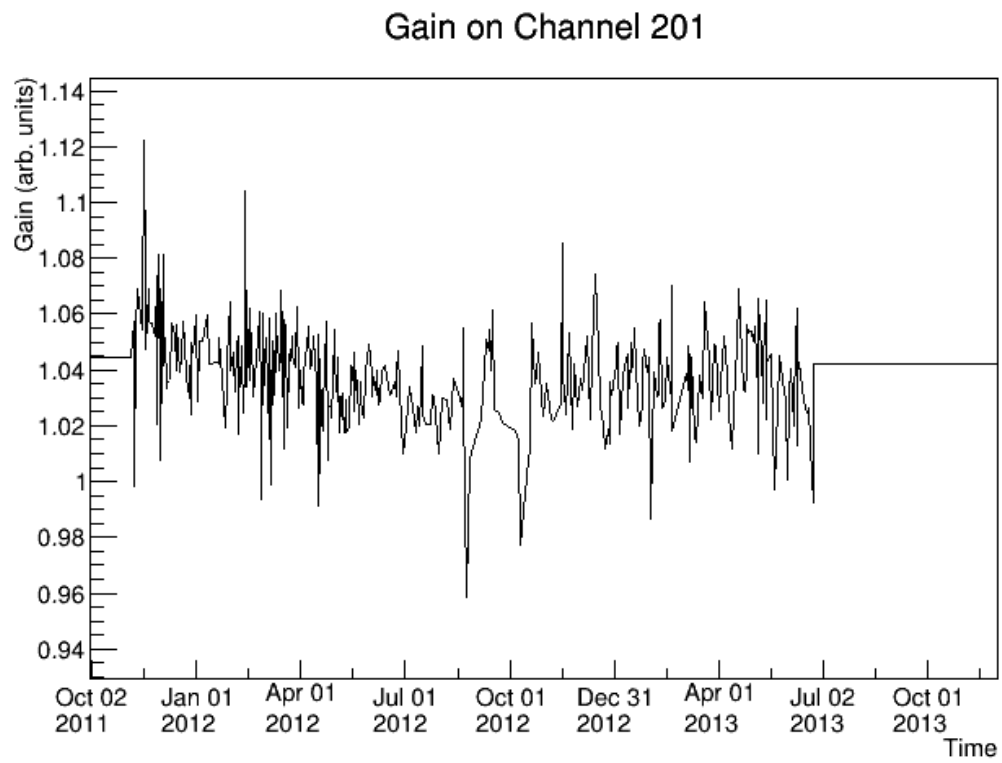


Figure 5.5: Functions $S(t)$ for selected channels.

of ROOT TH3D objects, and fortunately ROOT provides a number of excellent plotting features suitable for a three-dimensional dataset. In figure 5.1, it is possible to view the values of $R(\vec{x})$ for a sample of gangs on top of each other. Larger boxes and denser color indicates a higher yield on the APD gang in question, and it is immediately apparent that events near the anodes produce light which is highly concentrated on a single APD gang.

Figure 5.1 shows the highest yield on the very boundary of the detector, well outside of our fiducial volume; to permit us to more easily view contrast inside of the detector, figure 5.2 shows the same map while omitting the most extreme bin near either anode. Viewing this map, we can see more interesting characteristics of the lightmap:

- Events near the anodes show a high signal concentration on the one gang nearest to their position; however, even deep into the detector near the cathode it is still possible to see the higher concentration of signal on the gang directly aligned with the event.
- We can also see, from gang 201 (green) in this visualization, that events can produce significant signal on APD gangs which they are not directly aligned (in the Z direction); yield on gang 201 can be seen to decrease smoothly in all directions.
- APD gangs in the corners of the detector, such as gang 195 (red), are not effective at measuring light from events which are far away; even directly above gang 195, it is clear that gang 198 is more effective at collecting light

farther away than about five to ten centimeters. This can be attributed to the reflection of photons by teflon, which may enhance the light yield on gangs which are not hidden in corners.

Figures 5.3, 5.4, and 5.5 show the functions $S(t)$ for the same sample of gangs. The vertical scale can be treated as having arbitrary units; we note that all of these functions lie roughly around 1, which is attributed to the initial placement of $S(t) = 1$ in algorithm 2. We draw the following primary observations from these plots:

- Many of the gangs show their values of $S(t)$ decreasing rapidly up to around February 2012; some of the gangs also show a sharp decrease in value at that point. The decrease in gain corresponds with observations which were made at the time, leading to a decision to replace electronics on some APD channels on February 23, 2012. Thus, we do see “real” features from these plots.
- The functions $S(t)$ are otherwise dominated by jitter between points, indicating that we do not collect enough statistics from each run to sufficiently constrain $S(t)$. This is taken as the strongest evidence that we should use a coarser time binning for $S(t)$, as described in section 5.2.2. Preliminary work has been performed to do this with more recent lightmaps, but no studies have evaluated the impact on energy resolution.
- Individual points on these functions may spike by as much as 40%. These points have been investigated, and their cause is not understood. These jumps,

along with the overall jitter, will certainly be reduced by the use of a coarser time binning.

Chapter 6: Denoising Results

Chapter 7: Conclusions and Future Plans

Future plans to mention:

- Run1 contains up to 90 livetime-days, which is not negligible; denoising should have a significant improvement in resolution to offer even though APD gains were lower then. (Personal view – our next paper should include a combined analysis of Run1+Run2abcd, and whatever of Run3 is salvagable.)
- Further improvements in denoising are known to be achievable.
- Electronics upgrade could be a future hardware improvement.
- Restate some of the things listed in denoising, like anticorrelated cluster energies.

Bibliography

- [1] K. Abe, et al. Search for differences in oscillation parameters for atmospheric neutrinos and antineutrinos at Super-Kamiokande. *Phys. Rev. Lett.*, 107:241801, Dec 2011. doi:10.1103/PhysRevLett.107.241801.
- [2] P. Adamson, et al. Measurement of the neutrino mass splitting and flavor mixing by MINOS. *Phys. Rev. Lett.*, 106:181801, May 2011. doi:10.1103/PhysRevLett.106.181801.
- [3] M. Agostini, et al. Results on neutrinoless double- β decay of ^{76}Ge from phase I of the GERDA experiment. *Phys. Rev. Lett.*, 111:122503, Sep 2013. doi:10.1103/PhysRevLett.111.122503.
- [4] J. B. Albert, et al. Improved measurement of the $2\nu\beta\beta$ half-life of ^{136}Xe with the EXO-200 detector. *Phys. Rev. C*, 89:015502, Jan 2014. doi:10.1103/PhysRevC.89.015502.
- [5] Josh Albert, et al. Energy correction technote for $0\nu\beta\beta$ analysis (spring 2013), 2013. Internal EXO Document 007.
- [6] Josh Albert, et al. Neutron capture gamma tech note for $0\nu\beta\beta$ analysis (summer 2013), 2013. Internal EXO Document 021.
- [7] F. P. An, et al. Improved measurement of electron antineutrino disappearance at Daya Bay. *Chinese Physics C*, 37(1):011001, 2013.
- [8] J. Angrik et al. KATRIN design report 2004. Technical report, KATRIN Collaboration, 2005.
- [9] J. Argyriades, et al. Measurement of the double- β decay half-life of ^{150}Nd and search for neutrinoless decay modes with the NEMO-3 detector. *Phys. Rev. C*, 80:032501, Sep 2009. doi:10.1103/PhysRevC.80.032501.
- [10] J. Argyriades, et al. Measurement of the two neutrino double beta decay half-life of Zr-96 with the NEMO-3 detector. *Nuclear Physics A*, 847(34):168, 2010. ISSN 0375-9474. doi:http://dx.doi.org/10.1016/j.nuclphysa.2010.07.009.
- [11] C. Arnaboldi, et al. Results from a search for the $0\nu\beta\beta$ -decay of ^{130}Te . *Phys. Rev. C*, 78:035502, Sep 2008. doi:10.1103/PhysRevC.78.035502.

- [12] V. N. Aseev, et al. Upper limit on the electron antineutrino mass from the Troitsk experiment. *Phys. Rev. D*, 84:112003, Dec 2011. doi:10.1103/PhysRevD.84.112003.
- [13] G. Audi, et al. The AME2012 atomic mass evaluation. *Chinese Physics C*, 36(12):1287, 2012.
- [14] M Auger, et al. The EXO-200 detector, part I: detector design and construction. *Journal of Instrumentation*, 7(05):P05010, 2012.
- [15] F. T. Avignone, III, S. R. Elliott, and J. Engel. Double beta decay, majorana neutrinos, and neutrino mass. *Reviews in Modern Physics*, 80:481, 2008.
- [16] A. S. Barabash. Double beta decay: present status. *Phys.Atom.Nucl.*, 73:162, 2010.
- [17] A.S. Barabash and V.B. Brudanin. Investigation of double-beta decay with the NEMO-3 detector. *Physics of Atomic Nuclei*, 74(2):312, 2011. ISSN 1063-7788. doi:10.1134/S1063778811020062.
- [18] J. Barea, J. Kotila, and F. Iachello. Limits on neutrino masses from neutrinoless double- β decay. *Phys. Rev. Lett.*, 109:042501, Jul 2012. doi:10.1103/PhysRevLett.109.042501.
- [19] M.J. Berger, et al. XCOM: Photon cross section database (version 1.5). [Online] Available: <http://physics.nist.gov/xcom> [Sunday, 23-Feb-2014 03:19:51 EST], 2010. National Institute of Standards and Technology, Gaithersburg, MD.
- [20] J Beringer, et al. Review of particle physics. *PHYSICAL REVIEW D*, 86(1):1504, 2012. ISSN 1550-7998.
- [21] M.R. Bhat. Evaluated nuclear structure data file (ENSDF). In SyedM. Qaim, editor, *Nuclear Data for Science and Technology*, Research Reports in Physics, pages 817–821. Springer Berlin Heidelberg, 1992. ISBN 978-3-642-63473-4. doi:10.1007/978-3-642-58113-7_227. Data extracted using the NNDC On-Line Data Service from the ENSDF database, file revised as of February 24, 2014.
- [22] L. S. Blackford, et al. An updated set of basic linear algebra subprograms (BLAS). *ACM Trans. Math. Soft.*, 28(2):135, 2002.
- [23] Planck Collaboration. Planck 2013 results. XVI. cosmological parameters. *Astronomy and Astrophysics*, under review.
- [24] SINDRUM II Collaboration. Improved limit on the branching ratio of $\mu \rightarrow e^+$ conversion on titanium. *Physics Letters B*, 422:334, 1998.
- [25] Arthur H. Compton. A quantum theory of the scattering of x-rays by light elements. *Phys. Rev.*, 21:483, May 1923. doi:10.1103/PhysRev.21.483.

- [26] Arthur H. Compton. A quantum theory of the scattering of x-rays by light elements. *Physical Review*, 21:483, 1923.
- [27] E. Conti, et al. Correlated fluctuations between luminescence and ionization in liquid xenon. *Phys. Rev. B*, 68:054201, Aug 2003. doi:10.1103/PhysRevB.68.054201.
- [28] C. E. Dahl. *The physics of background discrimination in liquid xenon, and first results from Xenon10 in the hunt for WIMP dark matter*. Ph.D. thesis, Princeton University, 2009.
- [29] E.-I. Esch, et al. The cosmic ray muon flux at WIPP. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 538(13):516 , 2005. ISSN 0168-9002. doi:http://dx.doi.org/10.1016/j.nima.2004.09.005.
- [30] R. Arnold et al. Search for neutrinoless double-beta decay of ^{100}Mo with the NEMO-3 detector. Preprint available at <http://arxiv.org/abs/1311.5695>.
- [31] EXO. **Figuring this stuff out**. *Nature*, 2014.
- [32] G. Feinberg, P. Kabir, and S. Weinberg. Transformation of muons into electrons. *Phys. Rev. Lett.*, 3:527, 1959.
- [33] Joseph A. Formaggio. Year one: Direct neutrino masses post Planck. In *TAUP2013*. To be published by Elsevier.
- [34] Y. Fukuda, et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562, Aug 1998. doi:10.1103/PhysRevLett.81.1562.
- [35] A. Gando, et al. Constraints on θ_{13} from a three-flavor oscillation analysis of reactor antineutrinos at KamLAND. *Phys. Rev. D*, 83:052002, Mar 2011. doi:10.1103/PhysRevD.83.052002.
- [36] A. Gando, et al. Limit on neutrinoless $\beta\beta$ decay of ^{136}Xe from the first phase of KamLAND-Zen and comparison with the positive claim in ^{76}Ge . *Phys. Rev. Lett.*, 110:062502, Feb 2013. doi:10.1103/PhysRevLett.110.062502.
- [37] M. C. Gonzalez-Garcia and Yosef Nir. Neutrino masses and mixing: evidence and implications. *Rev. Mod. Phys.*, 75:345, Mar 2003. doi:10.1103/RevModPhys.75.345.
- [38] Kazushige Goto and Robert A. van de Geijn. Anatomy of high-performance matrix multiplication. *ACM Trans. Math. Softw.*, 34(3):12:1, May 2008. ISSN 0098-3500. doi:10.1145/1356052.1356053.
- [39] EXO Monte Carlo Group. EXO 200 montecarlo documentation, 2013. Internal EXO Document 010.

- [40] EXO Reconstruction Analysis Group. Event reconstruction in EXO-200, 2013. Internal EXO Document 008.
- [41] A. El Guennouni, K. Jbilou, and H. Sadok. A block version of BiCGSTAB for linear systems with multiple right-hand sides. *Electronic Transactions on Numerical Analysis*, 16:129, 2003.
- [42] Carter Hall and John Wilkerson. To appear in Annual Reviews of Nuclear and Particle Physics.
- [43] Ryan Herbst. EXO electronics functional specification, 2010. Internal EXO documentation.
- [44] S. Herrin. *Double Beta Decay in Xenon-136: Measuring the Neutrino-Emitting Mode and Searching for Majoron-Emitting Modes*. Ph.D. thesis, Stanford University, 2013.
- [45] P.E. Hodgson, E. Gadioli, and E. Gadioli Erba. *Introductory Nuclear Physics*. Oxford Science Publications, 2003.
- [46] Intel. *Reference Manual for Intel Math Kernel Library*, 11.1 edition.
- [47] J. Kotila and F. Iachello. Phase-space factors for double- β decay. *Phys. Rev. C*, 85:034316, Mar 2012. doi:10.1103/PhysRevC.85.034316.
- [48] D.S. Leonard, et al. Systematic study of trace radioactive impurities in candidate construction materials for EXO-200. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 591(3):490 , 2008. ISSN 0168-9002. doi:http://dx.doi.org/10.1016/j.nima.2008.03.001.
- [49] E. Majorana. Teoria simmetrica dellelettrone e del positrone. *Nuovo Cimento*, 14:171, 1937.
- [50] Rabindra N. Mohapatra and Goran Senjanović. Neutrino mass and spontaneous parity nonconservation. *Phys. Rev. Lett.*, 44:912, Apr 1980. doi:10.1103/PhysRevLett.44.912.
- [51] M Moszynski, M Szawlowski, M Kapusta, and M Balcerzyk. Large area avalanche photodiodes in scintillation and x-rays detection. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 485(3):504 , 2002. ISSN 0168-9002. doi:http://dx.doi.org/10.1016/S0168-9002(01)02117-9.
- [52] R. Neilson. APD installation: Selection and ganging, 2009. Presented to the EXO collaboration at the February 2009 EXO week meeting.

- [53] R. Neilson, et al. Characterization of large area APDs for the EXO-200 detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 608(1):68 , 2009. ISSN 0168-9002. doi:<http://dx.doi.org/10.1016/j.nima.2009.06.029>.
- [54] H. Nishino, et al. Search for proton decay via $p \rightarrow e^+\pi^0$ and $p \rightarrow \mu^+\pi^0$ in a large water cherenkov detector. *Phys. Rev. Lett.*, 102:141801, Apr 2009. doi:10.1103/PhysRevLett.102.141801.
- [55] Sheldon Ross. *A First Course in Probability*. Pearson Prentice Hall, 2006.
- [56] J. Schechter and J.W.F. Valle. Neutrinoless double- β decay in $su(2)\times u(1)$ theories. *Phys. Rev. D*, 25:2951, 1982.
- [57] David W. Scott. *Multivariate density estimation : theory, practice, and visualization*. John Wiley and Sons, Inc., 1992.
- [58] M Szydagis, et al. NEST: a comprehensive model for scintillation yield in liquid xenon. *Journal of Instrumentation*, 6(10):P10002, 2011.
- [59] S. Umehara, et al. Neutrino-less double- β decay of ^{48}Ca studied by $\text{CaF}_2(\text{Eu})$ scintillators. *Phys. Rev. C*, 78:058501, Nov 2008. doi:10.1103/PhysRevC.78.058501.
- [60] Joachim Wolf. The {KATRIN} neutrino mass experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 623(1):442 , 2010. ISSN 0168-9002. doi:<http://dx.doi.org/10.1016/j.nima.2010.03.030>. 1st International Conference on Technology and Instrumentation in Particle Physics.
- [61] L. Yang. EXO-200 electronics upgrade study report, 2013. Internal EXO Document 012.
- [62] Kai Zuber. Summary of the workshop on: Nuclear matrix elements for neutrinoless double beta decay. Available at <http://arxiv.org/abs/nucl-ex/0511009>.