

ABSTRACT

Title of dissertation: SIGNAL DENOISING FOR THE
EXO-200 DETECTOR AND AN
IMPROVED LIMIT ON $\beta\beta 0\nu$ DECAY

Clayton 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 THE EXO-200 DETECTOR AND AN
IMPROVED LIMIT ON $\beta\beta 0\nu$ DECAY

by

Clayton 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 Betsy Beise (need to ask)

Professor Herbert Breuer (need to ask)

Professor Peter Shawhan (need to ask)

Dean's Representative (?)

Professor Rabindra Mohapatra (would also be a natural choice)

© Copyright by
Clayton Davis
2014

Preface

If needed.

Foreword

If needed.

Dedication

If needed.

Acknowledgments

Thanks all for lots of help...

Table of Contents

List of Figures	vii
List of Abbreviations	ix
1 Theory of $\beta\beta 0\nu$ Decay	1
2 The EXO-200 Detector	8
3 The Lightmap	12
3.1 Four-Dimensional Lightmap	14
3.2 Algorithm Details	16
3.2.1 Event Selection	16
3.2.2 Function Parametrization	19
3.2.3 Convergence and Error Estimation	23
3.3 Visualization	26
3.4 Implications for Signal Reconstruction	29
4 Denoising Theory	33
4.1 Setup	33
4.1.1 The Noise Model	34
4.1.2 Optimization and Constraints	37
4.2 Derivation	38
4.3 Preconditioning	39
4.4 Solver	41
4.5 Implementation	43
5 Low-Field Running	48
A Running Code	49
Bibliography	50

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 occur even if a single β decay would be energetically forbidden.	2
1.2	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.	3
1.3	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 1.2 is the cause.	4
1.4	Since $\beta\beta 0\nu$ requires that the reaction in Figure 1.3 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 $\bar{\nu}$ - ν oscillation as a higher-order process.	5
1.5	The fundamental parameters m_1, m_2, m_3 of the three neutrino mass eigenstates are not directly measured; instead, experiments give different weighted sums of the three that allow these parameters to be constrained. M is the sum of the three mass values; $\langle m_\beta \rangle$ is the effective mass relevant to β decay, and $\langle m_{\beta\beta} \rangle$ is the effective mass probed by $\beta\beta 0\nu$ decay. The plots above indicate the relation, in the case of Majorana neutrinos, among these various parameters based on our current knowledge of Δm^2 and mixing angles (red lines if the smaller Δm^2 belongs to the lighter neutrinos, black lines if it belongs to the heavier neutrinos) and if these parameters were known exactly (hatched areas). Image provided by [1].	7
2.1	Schematic of the inner EXO-200 TPC.	9
3.1	Lightmap position-dependence $R(x, y, z)$ for selected APD gangs. . .	27
3.2	Lightmap position-dependence $R(x, y, z)$ for selected APD gangs. Here extreme anode positions are omitted to permit better contrast for the lightmap in the fiducial volume.	28
3.3	Functions $S(t)$ for selected channels.	30

3.4	Functions $S(t)$ for selected channels.	31
3.5	Functions $S(t)$ for selected channels.	32

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

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

Standard-model $\beta\beta$ decay is the result of the particle interaction $2d \rightarrow 2u + 2e^- + 2\bar{\nu}_e$ mediated by W^- -exchange, as depicted in Figure 1.1. It is effectively the simultaneous occurrence of two β decays from the same nucleus.

$\beta\beta$ decay is the result of a second-order weak interaction, so it has a remarkably long half-life; however, it does not violate any symmetries and is, in this sense, a mundane prediction of the Standard Model. Although many nuclei are expected to decay by $\beta\beta$, the process is thoroughly masked by conventional β decay in most of them. There are, however, isotopes which are forbidden from undergoing β decay by energy conservation, yet permitted to undergo $\beta\beta$ decay. These are in all known instances isotopes with even-even proton-neutron numbers; β decay would take these isotopes to odd-odd, which is energetically disfavored due to the loss of pairing forces between like nuclei, while $\beta\beta$ decay permits these isotopes to remain in a favorable even-even state. Isotopes which follow this pattern are considered $\beta\beta$ candidates.

The detection and study of $\beta\beta$ decay provides an opportunity to test a new class of nuclear matrix computations, and also a chance for experimentalists to push the boundaries of low-background experiments; however, the primary appeal of $\beta\beta$ decay is through the opportunity it provides to probe the nature of neutrinos. It

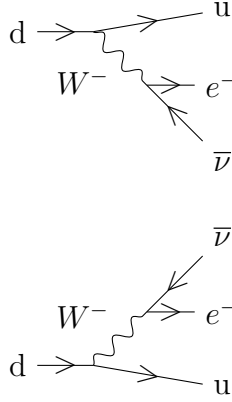


Figure 1.1: Feynman diagram for $\beta\beta 2\nu$ decay. The reaction products are equivalent to two β decays in succession, but this reaction can occur even if a single β decay would be energetically forbidden.

has been suggested as early as 1937 that neutrinos could possess mass through a novel neutrino-antineutrino interaction, provided that the neutrino be its own antiparticle [2]. As evidence mounted through the 1990's that neutrinos must have nonzero mass (see, in particular, the observation of neutrino oscillation in [3]), interest in possible mass-generating mechanisms has increased – including Majorana's mechanism.

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 1.2, in which the two outgoing neutrinos in $\beta\beta$ decay are replaced by one virtual neutrino. This is only possible if neutrinos are their own antiparticles because the timelike arrow of the neutrino cannot be specified. We refer to the standard process as $\beta\beta 2\nu$ decay and the novel process as $\beta\beta 0\nu$ decay.

A clear observation of $\beta\beta 0\nu$ would immediately yield a number of interesting

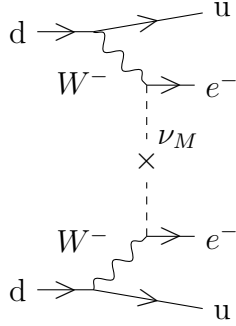


Figure 1.2: Feynman diagram for $\beta\beta 0\nu$ decay. A virtual neutrino mediates the exchange. This is only possible if $\bar{\nu}_R$ can oscillate to ν_L , and the interaction that induces this oscillation also generates neutrino mass.

physical insights. The existence of the process automatically demonstrates non-conservation of total lepton number, since the lepton number increases by 2. Numerous theories have suggested that lepton number may not be conserved [4] [5], but none have yet reported a positive result. In the conventional Standard Model with massless neutrinos, lepton number conservation is an accidental symmetry [6], but since neutrinos are known to have mass, there is no longer *a priori* any reason to expect conservation or non-conservation.

A second conclusion, as discussed above, would be that neutrino mass comes from a Majorana mass term. This term appears in the lagrangian (for each mass eigenstate) as

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

where the superscript-*c* represents charge conjugation. (This is the reason that

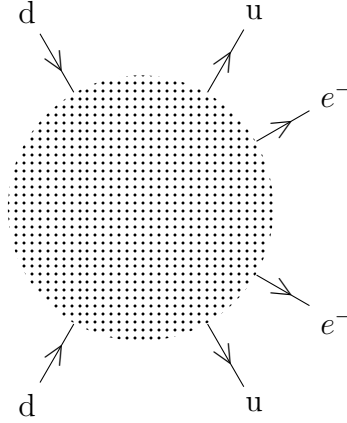


Figure 1.3: A generic $\beta\beta 0\nu$ process, with no assumptions of the underlying mechanism. If $\beta\beta 0\nu$ is observed, it requires that this reaction is possible even though it does not strictly imply that the oscillation in Figure 1.2 is the cause.

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 [1].

It is worth noting that the interaction term above assumes no mediating particles in the mechanism. However, if $\beta\beta 0\nu$ decay is observed, it leads very generally to a conclusion that neutrinos have an *effective* Majorana mass interaction [7]. We can see this by examining the generic diagram for $\beta\beta 0\nu$ decay in Figure 1.3. No matter what the mechanism of the decay is, we can insert this diagram into the larger one in Figure 1.4 and see that an effective neutrino-antineutrino mass vertex is implied by the existence of $\beta\beta 0\nu$ decay.

If the mechanism of $\beta\beta 0\nu$ decay really is, at tree level, a neutrino mass term

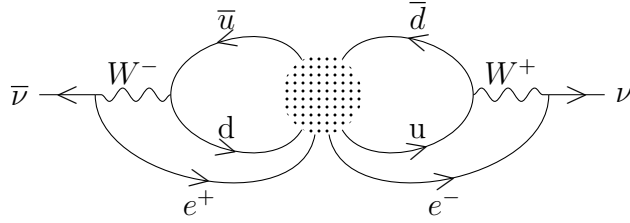


Figure 1.4: Since $\beta\beta 0\nu$ requires that the reaction in Figure 1.3 occurs, we can insert that diagram into this larger one and assert that regardless of the underlying mechanism, $\beta\beta 0\nu$ implies an effective neutrino mass generated by $\bar{\nu}$ - ν oscillation as a higher-order process.

as shown in Figure 1.2, then it is possible to infer the mass scale of the neutrino from the observed rate of decay. To lowest order we can assert that

$$[T_{1/2}^{0\nu}]^{-1} \propto \langle m_{\beta\beta} \rangle^2 = \left| \sum_k m_k U_{ek}^2 \right|^2,$$

where U_{ek} is the neutrino mixing matrix and the m_k identify the three neutrino mass eigenstates. The constant of proportionality is isotope-dependent; state-of-the-art nuclear matrix computations can provide it within a factor of 2 – 3 [8]. We will not consider the complex challenges of these computations here.

There are bounds on the values that may be seen, based on previous results in $\beta\beta 0\nu$ decay and other measurements. Figure 1.5 shows how the various mass parameters can be related to each other. M , the combined masses of all three neutrino flavors, can be limited by cosmological observations; recent results from Planck combined with WMAP and baryon acoustic oscillations have restricted $M < 0.23$ eV [9]. Measurements of the β decay spectrum of Tritium have yielded bounds on $\langle m_{\beta} \rangle$, the weighted average mass of the electron-neutrino; this parameter is

currently known to be less than 2.5 eV [10], and the KATRIN experiment is expected to begin data-taking in 2012 with a sensitivity down to 0.2 eV [11].

The strongest limits on $\langle m_{\beta\beta} \rangle$ have been published by the Heidelberg-Moscow experiment studying ^{76}Ge . The results of this experiment have been interpreted in conflicting ways by its collaboration: the faction that asserts a positive detection of $\beta\beta 0\nu$ has most recently published a value of $\langle m_{\beta\beta} \rangle = 0.32 \pm 0.03$ eV [12], while the full collaboration's strongest limit was $\langle m_{\beta\beta} \rangle < 0.35$ eV (both neglecting uncertainty in the nuclear matrix elements) [13]. In addition to the disagreement within the Heidelberg-Moscow collaboration, though, there is also widespread agreement that it is unreliable to draw conclusions on $\langle m_{\beta\beta} \rangle$ from only one isotope due to the poorly-understood uncertainties in the nuclear matrix elements [8]. All of this means that there is significant interest by many collaborations, working with many different isotopes in parallel, to probe below the threshold set by the Heidelberg-Moscow experiment to confirm or refute the claimed discovery.

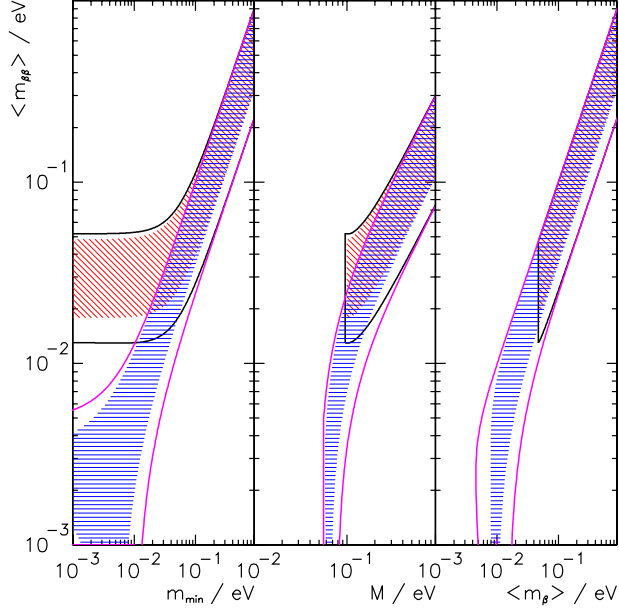


Figure 1.5: The fundamental parameters m_1 , m_2 , m_3 of the three neutrino mass eigenstates are not directly measured; instead, experiments give different weighted sums of the three that allow these parameters to be constrained. M is the sum of the three mass values; $\langle m_{\beta} \rangle$ is the effective mass relevant to β decay, and $\langle m_{\beta\beta} \rangle$ is the effective mass probed by $\beta\beta 0\nu$ decay. The plots above indicate the relation, in the case of Majorana neutrinos, among these various parameters based on our current knowledge of Δm^2 and mixing angles (red lines if the smaller Δm^2 belongs to the lighter neutrinos, black lines if it belongs to the heavier neutrinos) and if these parameters were known exactly (hatched areas). Image provided by [1].

Chapter 2: The EXO-200 Detector

Among the experiments with potential to match the sensitivity of Heidelberg-Moscow is the Enriched Xenon Observatory (EXO). In particular, the EXO-200 detector is designed to hold 200 kg of Xe enriched to 80% in ^{136}Xe , and anticipates probing down to $\langle m_{\beta\beta} \rangle \sim 0.1$ eV [14]. Xenon is a convenient material to study because it is a noble element, making it easy to purify chemically. It has no long-lived radioactive isotopes, so there is no concern of cosmogenic activation of the Xenon while aboveground. It has a fairly high Q-value of 2.458 MeV, giving it a larger phase-space for its decay products and putting its peak out of reach of many common backgrounds. Xenon also provides two means to measure energy from a decay: under the influence of an electric field, ionization charge in pure Xenon will drift; and regardless of electric field, excited Xenon scintillates. The ratio of scintillation to ionization depends on electric field and the density of the energy deposit; thus, at high enough electric field discrimination is possible between high-energy-density α -decay and lower-energy-density β - and γ -decay. Another benefit of Xenon is that, since it can act as an energy detector in its pure gaseous or liquid state, it is easy to make a large monolithic Xenon detector. We will here review how the EXO-200 experiment has been constructed to exploit these properties and

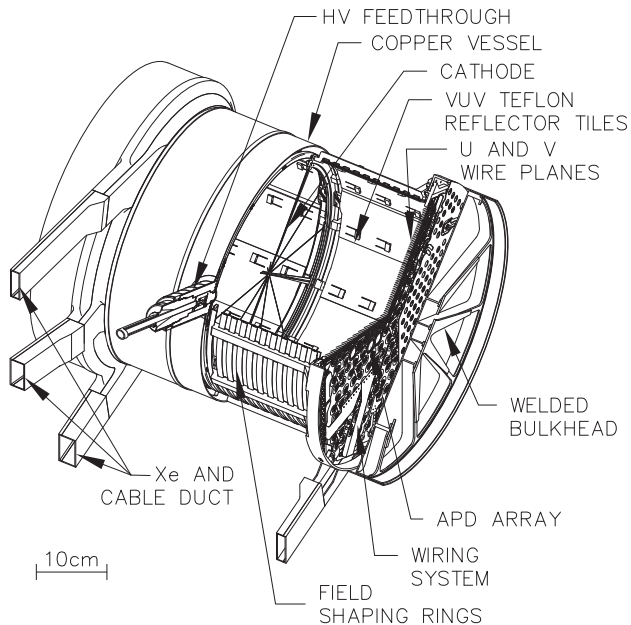


Figure 2.1: Schematic of the inner EXO-200 TPC.

explore down to interesting neutrino masses.

The core of the EXO-200 detector is a cylindrical time-projection chamber (TPC), depicted in Figure 2.1, under a potential of roughly 400 V/cm. The electric field is shaped to be parallel to the axis of the TPC throughout its fiducial volume. The cathode is located in the middle of the TPC, and charge is collected on grids of parallel wires on either end of the TPC, giving a high-quality probe of the total ionization charge and location along one axis. Also on both ends of the TPC, in front of these collection wires is another grid of parallel wires, orthogonal to the collection wires, which identifies an induced signal from passing ionization charge to give another position coordinate. Finally, on both endcaps of the TPC are many large-area avalanche photodiodes (LAAPDs) which can detect the scintillation light from the primary decay [15]. The time difference between scintillation and charge

collection gives a third position coordinate from a decay, and the scintillation also can give a secondary measurement of the energy of the deposit. As noted above, comparing the scintillation and collected ionization allows α -decays to be identified, removing one broad source of backgrounds.

The Xenon must be kept extremely pure, both to reduce backgrounds and to permit the ionization charges to drift farther with less attenuation. To accomplish this, the detector has two pipes feeding into it so that Xenon may be circulated out of the detector continuously for repurification. This is done using SAES Zirconium getters, which filter out all chemically reactive elements and return only noble elements. Radioactive isotopes which are not filtered out are ^{85}Kr (which has a Q-value much lower than that of ^{136}Xe , making it fairly harmless) and all forms of Radon (some of whose daughter products can deposit energy around the Q-value of ^{136}Xe).

The TPC itself is made of oxygen-free high-thermal-conductivity (OFHC) copper that was kept under concrete shielding during its time aboveground to protect it from neutron-induced activation. Nevertheless, the copper may be a significant source of backgrounds in the energy range of interest, primarily due to natural contamination by ^{40}K , ^{238}U and ^{232}Th [16]. To mitigate this, significant effort has been expended in making the copper walls as thin as possible, averaging 1.5 mm thick. The ductility of copper makes such thin walls technically challenging, and a significant infrastructure has been developed to maintain the pressure of ultrapure HFE coolant, in which the TPC is immersed, so that the pressure difference can be controlled to ~ 10 torr.

Other significant sources of background may be the charge collection wires that

pass through the TPC and the LAAPDs on its ends. Both backgrounds have been minimized by limiting the total mass and carefully selecting materials to use. The charge collection wires are photo-etched from phosphor bronze; the etching process was found to contribute to surface contamination, which was minimized by the use of clean etching agents and the use of a more aggressive cleaning procedure after etching [16]. The LAAPDs are designed to minimize the radioactive contamination they introduce. They were designed without standard ceramic encapsulation; instead the LXe is itself relied upon to insulate them. Additionally, Aluminum used in the LAAPDs was found to be high-background, and was replaced with an acceptable substitute. These efforts have reduced the activity introduced by the LAAPDs [16] [15].

These are a few of the critical subsystems that were screened prior to use. More details will be presented in a forthcoming detector paper to be written by the EXO collaboration.

Chapter 3: The Lightmap

The EXO detector was designed with roughly 450 APDs ganged into 74 channels of six to seven APDs each. These APDs were set into the two endcaps of the cylindrical EXO detector. 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. Additional stepwise changes in APD behavior come from changes to the electronics which have been implemented multiple times during the course of the experiment.

The different positions of the APDs means that given the same amount of energy deposited into the detector, a channel may show a larger or smaller signal depending on the location of the energy deposit. Additionally, there will be time-dependent changes: gain will drift in each APD at independent rates, and stepwise changes to the APD signals comes from changes to electronics and disabling of channels.

As a result, when we attempt to characterize the APD response of our detector, it is critical to map out the time and space dependence of that response. Early attempts to do this 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 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 70-71 active APD channels could be treated 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 can 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.

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. 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.

3.1 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 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 $L(x, y, z, t)$.

This may at first seem infeasible. After all, by adding an independent time argument it appears that rather than measuring a lightmap independently from each calibration campaign, we must measure one independently for each time bin. 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 (x, y, z) 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

$$L(x, y, z, t) = R(x, y, z)S(t) \quad (3.1)$$

Is this simplification fully motivated by the detector? It must be admitted that it is not. The first point is fairly accurate: we must trust that the electric field and the reflectivity of detector surfaces are constant in time, but basically 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 limits scintillation resolution. Currently it is assumed to be a small effect and ignored.

So, treating equation 3.1 as valid, one simple general scheme to find R and S iteratively is:

1. Compile, as well as possible, a list of single-site 2615-keV events from Thorium-228 source data. Some Compton-scattered events will inevitably leak into this dataset, but the selection criteria should be designed to minimize this leakage while maximizing the fraction of true 2615-keV events accepted.
2. Initially, assume a constant function $S(t) = 1$ for all APD channels.

3. For each APD channel, estimate $R(x, y, z) = L(x, y, z, t)/S(t)$ from the set of events tabulated in step 1.
4. For each APD channel, estimate $S(t) = L(x, y, z, t)/R(x, y, z)$ from the set of events tabulated in step 1.
5. If convergence has not yet been reached, return to step 3.

Although this algorithm is not necessarily guaranteed to achieve convergence at all, in practice $S(t)$ is close to constant, resulting in fairly rapid convergence.

3.2 Algorithm Details

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

3.2.1 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. [17]

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%) has not demonstrated any significant improvement in the quality of the lightmap; 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 we cannot afford to do. (The statistics from strong Thorium source runs are too valuable, particularly because they are the primary source of statistics in some regions of the detector.)

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

3.2.2 Function Parametrization

Because we are attempting to empirically measure the functions $R(x, y, z)$ and $S(t)$ from a finite dataset, we need to specify some more limited form for them to take. All current approaches to describing $R(x, y, z)$ first bin the detector volume into three-dimensional voxels, and then define $R(x, y, z)$ 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 (many of which lie outside 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 these regions may have quite significant statistical error.

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 1cm (WHAT IS IT?) 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 [17] 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 is coarser than with the current cubic voxels being used; this is made possible by the larger quantity of available statistics from running much longer, 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 [18] 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 [18] for a detailed explanation of the related issues in kernel estimation. (SHOULD HAVE A CHAPTER ON REGRESSION AS WELL – BUT UNFORTUNATELY LEFT IT IN THE OFFICE, AND CAN’T ACCESS ONLINE. RETURN TO THIS.) Attempting to use a locally adaptive kernel regression for $R(x, y, z)$ 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, it has generally been

the practice of the collaboration to take 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 $R(x, y, z)$, 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.

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 $R(x, y, z)$ 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.

3.2.3 Convergence and Error Estimation

The topics of convergence and error estimation for the lightmap are perhaps the most rudimentary; a full analysis of the lightmap cannot be complete without them. However, for the current analysis they are approached essentially by a series of guesses; the lightmap is then used in further analysis, and the validity of the lightmap is established by its successful use in that downstream work.

Error estimation is not attempted on the current lightmap. We do study the number of hits observed in each position voxel of the detector (see section 3.3), and presume that the most significant source of error comes from low-statistics regions of the detector. However, no explicit estimation of the lightmap uncertainty has been attempted.

One difficulty with estimating the errors in the lightmap function $L(x, y, z, t)$ comes from the correlation between errors in $R(x, y, z)$ 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 $R(x, y, z)$ (if $S(t)$ is assumed to be perfectly known) or $S(t)$ (if $R(x, y, z)$ is assumed to be perfectly known). But in reality, all of the measurements of $R(x, y, z)$ are correlated with all of the measurements of $S(t)$, meaning that even the errors of different voxels of $R(x, y, z)$ 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 $R(x, y, z)$ 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 $R(x, y, z)$ and $S(t)$ together. In the current scheme, $R(x, y, z)$ 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, this is still the sort of problem which is routinely managed in large-scale applications, and it would only need to be solved occasionally when it is necessary to derive a new lightmap, so it is not infeasible to imagine attempting it. 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 correlations between errors of the different voxels of $R(x, y, z)$ or runs in $S(t)$, but it would come closer than the previous naive method.

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 the paradoxical tradeoff that 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 $R(x, y, z)$, 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 $R(x, y, z)$; 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, we currently perform three to four iterations and claim that convergence is generally roughly 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.

3.3 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 $R(x, y, z)$ for each APD gang as a set of ROOT TH3D objects, and fortunately ROOT provides a number of excellent plotting features suitable for a three-dimensional dataset. In figure 3.1, it is possible to view the values of $R(x, y, z)$ 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

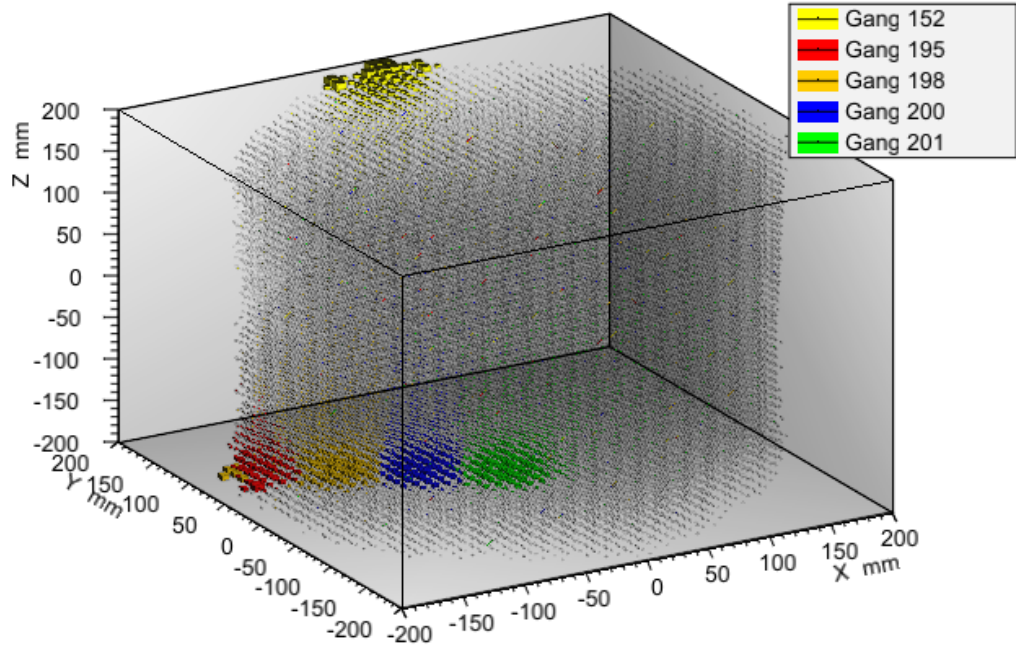


Figure 3.1: Lightmap position-dependence $R(x, y, z)$ for selected APD gangs.

concentrated on a single APD gang.

Figure 3.1 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 3.2 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

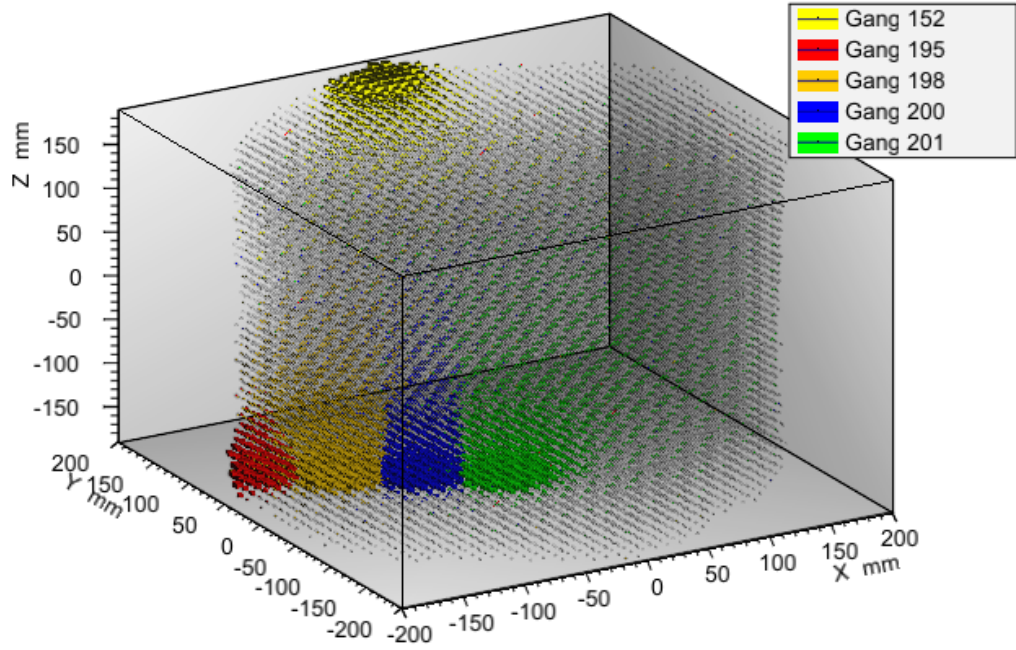


Figure 3.2: Lightmap position-dependence $R(x, y, z)$ for selected APD gangs. Here extreme anode positions are omitted to permit better contrast for the lightmap in the fiducial volume.

aligned with the event.

- We can also see, from gang 201 (green), that events can produce significant signal on APD gangs which 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.

ALSO TALK ABOUT GAIN VS. TIME, HITS PER VOXEL, OTHERS?

3.4 Implications for Signal Reconstruction

Prior to this work, it was standard to sum all of the scintillation waveforms from gangs on a single anode plane. The lightmap we have produced provides a trove of information about how light propagates through the detector, and as such it can be used to identify qualitatively how scintillation signal reconstruction might be improved.

Most significantly, figure 3.2 demonstrates that

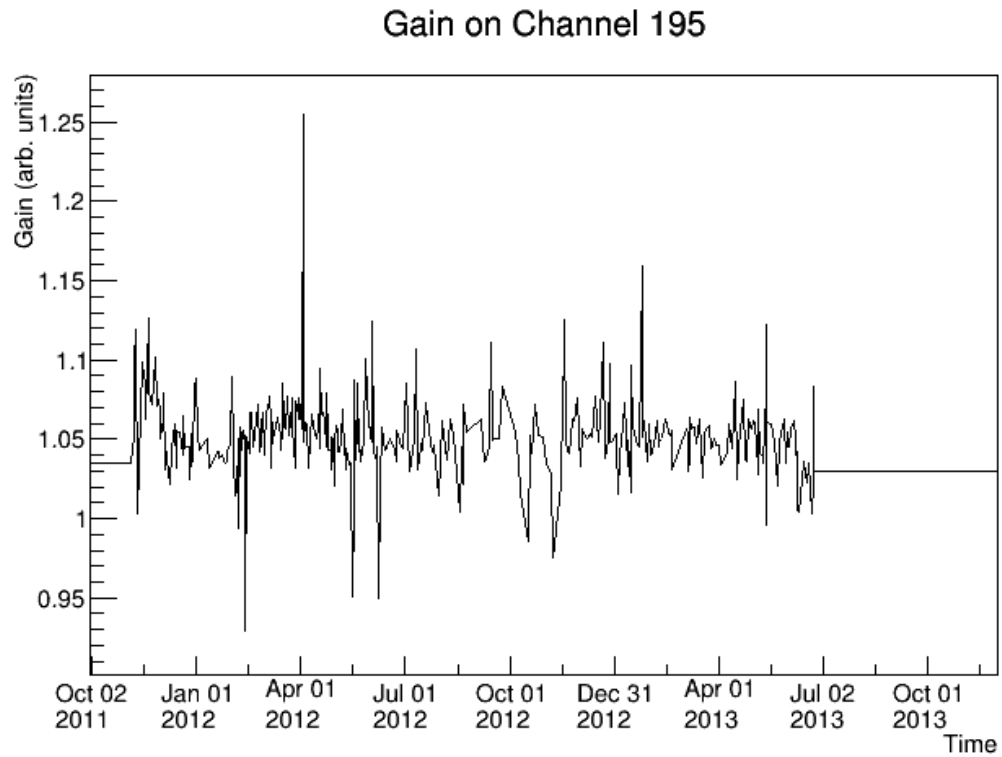
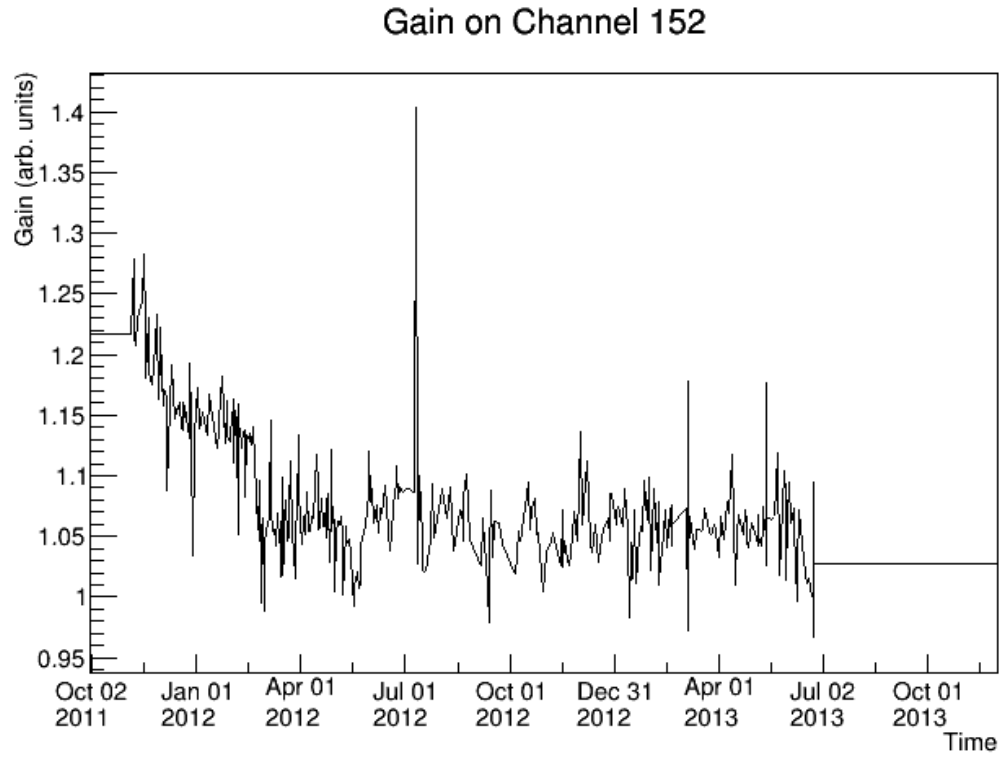


Figure 3.3: Functions $S(t)$ for selected channels.

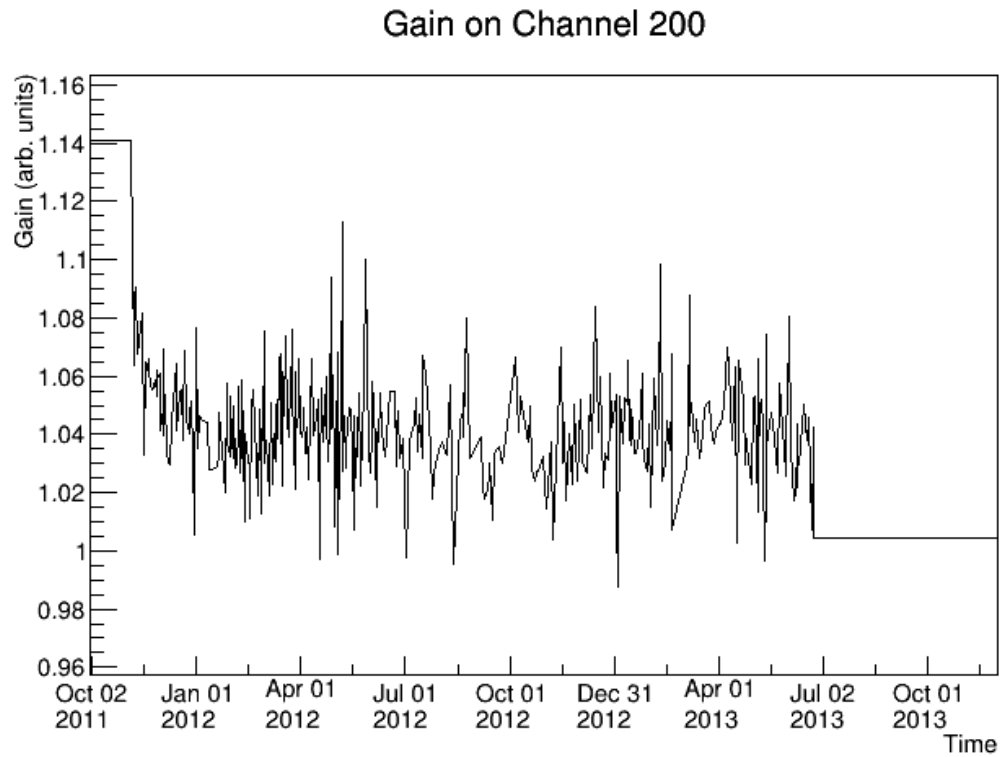
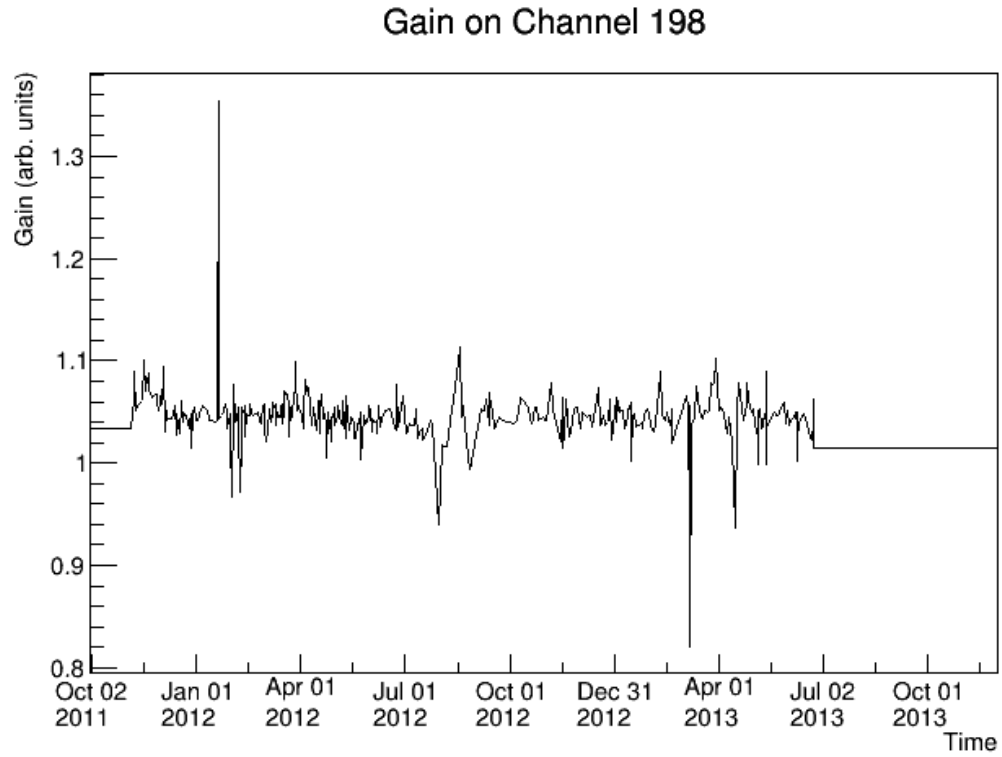


Figure 3.4: Functions $S(t)$ for selected channels.

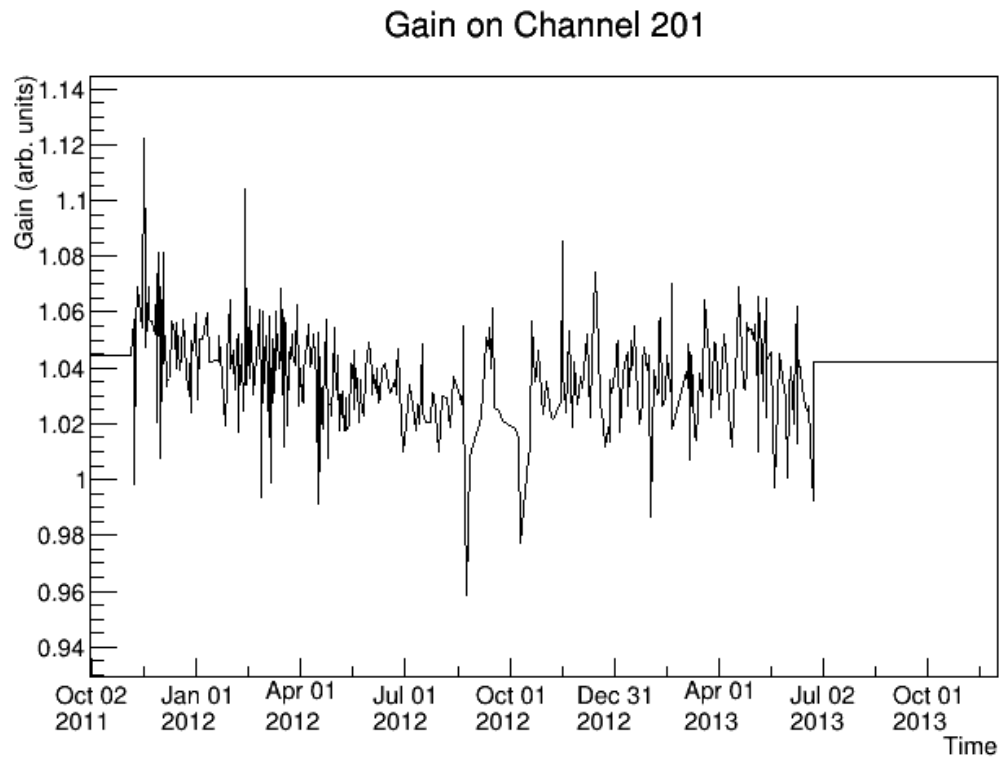


Figure 3.5: Functions $S(t)$ for selected channels.

Chapter 4: Denoising Theory

4.1 Setup

We first establish a number of notational conventions:

- i, j will represent indices over all channels.
- k, ℓ will represent only APD channel indices.
- m, n will represent only wire channel indices.
- t will represent the time indices of a discrete waveform.
- f, g will represent the frequency indices of Fourier-transformed waveforms.
- r, s will represent indices of signals in an event.
- For a waveform $*[t]$, 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.

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

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

$$\tilde{X}_i[f] = \sum_s C_s \tilde{Y}_{is}[f] + \tilde{N}_i[f].$$

4.1.1 The Noise Model

We will assume that there are two sources of noise in this model. First, the electronic noise $N_i[t]$ is assumed to be randomly distributed with mean $\langle \tilde{N}_i[f] \rangle = 0$ for all i, f . (This can also be interpreted as meaning that the phase of each frequency component is random.) We additionally demand that electronic noise components with different frequencies are uncorrelated, $\langle \tilde{N}_i[f] \tilde{N}_j[g] \rangle = 0$ when $f \neq g$. Importantly, though, we do permit correlations in noise across different channels, so $\langle \tilde{N}_i[f] \tilde{N}_j[f] \rangle$ may be non-zero for any i, j, f . These electronic-noise correlations will be measured in external data [DESCRIBE ELSEWHERE]; for the present purpose, we simply assume they are known.

The second random variable will be the magnitude of the signal itself, C_s . When an energetic decay occurs in bulk of the detector, it will release some number of photons and electrons which reflects the energy of the decay; however, the number

of photons and electrons actually collected by sensors is randomly distributed. We now consider the sources of that randomness.

On the wire channels, all charge is directed by the electric field to deposit in a single location; so the dominant source of randomness comes from impurities. Electrons are absorbed by impurities through their trajectory across the bulk of the Xenon; the expected quantity of charge decays exponentially in time with some half-life T . We can describe the number of lost electrons by a Poisson process. However, a typical deposit in our Q-value will produce tens of thousands of electrons, from which at most about 10% may be lost to purity, making Poisson noise negligible in this situation. Moreover, there is no available information in the charge channel which can reduce this source of noise; so it is proper to associate this source of uncertainty as a part of the irreducible detector resolution in the charge channel, and we will not treat it further.

On the APD channels, we detect photons which travel from the source of the energy deposit to the face of the APD. The paths of optical photons, and their reflection off of detector surfaces, is highly variable, so it is reasonable to model the number of photons depositing on any particular APD as Poisson-distributed. A typical deposit in our Q-value may deposit 10 – 100 photons on each APD, depending strongly on the location of the deposit; we find that on a single APD channel Poisson noise is very important. Suppose that for a deposit of energy E , we expect an average of Q_{tot} photons to be produced and Q_k photons to deposit on APD channel k . Then the number of photons P_k deposited on APD channel k will follow a multinomial distribution with the properties:

- $\langle P_k \rangle = Q_k$
- $\langle P_k P_\ell \rangle = Q_k Q_\ell \left(1 + \delta_{k\ell} Q_k^{-1/2} Q_\ell^{-1/2} - Q_{tot}^{-1} \right)$

Note that this implies that to fully understand the noise on our APDs, we must have some a priori estimate of Q_{tot} ; this requires us to have an a priori estimate of the deposit energy. It will be important to ensure our estimator does not bias us toward whatever estimate is provided; we will return to this point.

Our noise model, as we have said, is

$$\tilde{X}_i[f] = \sum_s C_s \tilde{Y}_{is}[f] + \tilde{N}_i[f].$$

But it will be advantageous to fix the normalization of \tilde{Y} to something consistent. In the case of wires, we use the usual normalization encountered in reconstruction: a deposit waveform should be normalized so that its peak lies 1 ADC above the baseline. Furthermore, in our model we must characterize the expected signal on every waveform from a given deposit; to generate the expected induced signals on neighboring wires, we use the same digitizer used in Monte Carlo, and scale by the same factor used on the deposit channel.

On APDs, there is no advantaged channel which should be used to set a scale, as we did for wires. However, there is a preferred energy scale: the lightmap we use is generated from photopeak events at the Thallium-208 gamma line at 2615 keV, and this normalization is preserved in the lightmap file format. So, the lightmap explicitly records the signal size (in ADC counts from baseline to peak) expected from a 2615 keV gamma-like deposit at all positions and times, and we can preserve that normalization here. APD-channel \tilde{Y}_{is} are therefore normalized to the number

of ADC counts expected from a 2615 keV deposit, and C_{is} indicates the measured energy of the deposit in units of 2615 keV.

We can specify these factors explicitly. Let $M_k(x, t)$ be the yield in ADC counts expected on APD channel k from a single-site 2615 keV deposit at position x and time t . Let $\tilde{Y}_\tau[f]$ represent a 1-ADC APD signal occurring at time τ within a waveform. (We assume for now, as we do in the rest of the analysis, that the shaping times on all APDs are identical.)

Let E_{exp} be the a priori expected energy of the deposit, in units of 2615 keV; if we expect c photons per 2615 keV deposited, we then can use $Q_{tot} = c \cdot E_{exp}$. We can also write Q_{ks} in terms of C_s if we have some conversion factor between ADC counts and incident photons on an APD channel k . Let $g_k(t)$ be this gain factor, as a function of time; then

$$\frac{Q_{ks} \cdot g_k(t)}{M_k(x, t)} = E_{exp} = \langle C_s \rangle.$$

4.1.2 Optimization and Constraints

For each signal, we wish to identify a denoising operator which acts on the Fourier components of all of the waveforms and yields an estimate of the magnitude of that signal. To make the problem tractable, we stipulate that this operator must be linear; so our estimate of the unknown parameter C_s will be

$$\hat{C}_s = \sum_{if} A_{ifs} \tilde{X}_i^R[f] + B_{ifs} \tilde{X}_i^I[f]$$

For each event and each signal, the goal of denoising is to find a set of coeffi-

cients A_{ifs} , B_{ifs} which minimizes the expected square-error in this quantity,

$$\epsilon = \left\langle \left(C_s - \hat{C}_s \right)^2 \right\rangle$$

It will turn out to be necessary, to make the problem tractable, to make the additional assumption that this denoising operator produce an unbiased estimate. This is expressed by the constraint:

$$\left\langle C_s - \hat{C}_s \right\rangle = 0$$

Although this constraint is necessary for mathematical tractability, it is worth noting that it is not required by physics. Releasing this constraint would permit biased estimators, but might permit the use of a narrower region of interest with better signal-to-background ratio. For instance, it might permit a mono-energetic peak to appear with asymmetric tails. A future project could involve releasing this constraint and checking whether the resulting resolution is improved; however, the increase in required computation makes this infeasible for us.

4.2 Derivation

We are now in a position to derive the optimal denoising estimator. We start by expanding the square-error value:

$$\begin{aligned} \epsilon &= \left\langle \left(C_s - \hat{C}_s \right)^2 \right\rangle \\ &= \left\langle \left(C_s - \sum_{if} \left[A_{ifs} \tilde{X}_i^R[f] + B_{ifs} \tilde{X}_i^I[f] \right] \right)^2 \right\rangle \\ &= \left\langle \left(C_s - \sum_{if} \left[A_{ifs} \tilde{N}_i^R[f] + B_{ifs} \tilde{N}_i^I[f] \right] - \sum_{ifr} C_r \left[A_{ifs} \tilde{Y}_{ir}^R[f] + B_{ifs} \tilde{Y}_{ir}^I[f] \right] \right)^2 \right\rangle \end{aligned}$$

We assume that the electronic and Poisson noise terms are uncorrelated:

4.3 Preconditioning

The matrix equation we wish to solve takes the form:

$$\begin{pmatrix} E + P & L \\ L^\top & 0 \end{pmatrix} X = \begin{pmatrix} 0 \\ I \end{pmatrix}$$

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

$$\begin{pmatrix} E + P & L \\ L^\top & 0 \end{pmatrix} \approx \begin{pmatrix} D & L \\ L^\top & 0 \end{pmatrix}$$

This approximate form is easy to invert, and can be used to precondition our problem for better numerical behavior.

We can further factor this approximate form of the matrix:

$$\begin{aligned} \begin{pmatrix} D & L \\ L^\top & 0 \end{pmatrix} &= \begin{pmatrix} D^{1/2} & 0 \\ L^\top D^{-1/2} & -H^\top \end{pmatrix} \begin{pmatrix} D^{1/2} & D^{-1/2}L \\ 0 & H \end{pmatrix} \\ &= \begin{pmatrix} D^{1/2} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ L^\top D^{-1/2} & -H^\top \end{pmatrix} \begin{pmatrix} I & D^{-1/2}L \\ 0 & H \end{pmatrix} \begin{pmatrix} D^{1/2} & 0 \\ 0 & I \end{pmatrix} \end{aligned}$$

where H is defined uniquely by $H^\top H = L^\top D^{-1}L$. (The solution is guaranteed to be real because D is positive-semidefinite – it consists of noise variance terms, which are non-negative.)

The two diagonal factors are easily inverted; we further define

$$\begin{aligned} K_1 &= \begin{pmatrix} I & 0 \\ L^\top D^{-1/2} & -H^\top \end{pmatrix} & K_1^{-1} &= \begin{pmatrix} I & 0 \\ H^{\top-1} L^\top D^{-1/2} & -H^{\top-1} \end{pmatrix} \\ K_2 &= \begin{pmatrix} I & D^{-1/2}L \\ 0 & H \end{pmatrix} & K_2^{-1} &= \begin{pmatrix} I & -D^{-1/2}LH^{-1} \\ 0 & H^{-1} \end{pmatrix} \end{aligned}$$

and then follow the standard proscription for preconditioning a linear system:

$$K_1^{-1} \begin{pmatrix} D^{-1/2} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} E + P & L \\ L^\top & 0 \end{pmatrix} \begin{pmatrix} D^{-1/2} & 0 \\ 0 & I \end{pmatrix} K_2^{-1} Z = K_1^{-1} \begin{pmatrix} D^{-1/2} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 \\ I \end{pmatrix}$$

$$Z = K_2 \begin{pmatrix} D^{1/2} & 0 \\ 0 & I \end{pmatrix} X$$

which reduces to

$$K_1^{-1} \left[\begin{pmatrix} D^{-1/2} E D^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} D^{-1/2} P D^{-1/2} & D^{-1/2} L \\ L^\top D^{-1/2} & 0 \end{pmatrix} \right] K_2^{-1} Z = \begin{pmatrix} 0 \\ -H^\top{}^{-1} \end{pmatrix}$$

$$X = \begin{pmatrix} D^{-1/2} & 0 \\ 0 & I \end{pmatrix} K_2^{-1} Z$$

This preconditioned set of equations is numerically quite stable. It is also possible to provide an excellent initial guess Z_0 by taking the approximate form of the matrix to be exact, yielding

$$Z_0 = \begin{pmatrix} 0 \\ -H^\top{}^{-1} \end{pmatrix}$$

Based on these various advantages, it is this system which we attempt to solve rather than the original form. Below we define matrices A and B for convenience,

and summarize the system to be solved:

$$\begin{aligned}
A &= K_1^{-1} \left[\begin{pmatrix} D^{-1/2} E D^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} D^{-1/2} P D^{-1/2} & D^{-1/2} L \\ L^\top D^{-1/2} & 0 \end{pmatrix} \right] K_2^{-1} \\
B &= Z_0 = \begin{pmatrix} 0 \\ -H^\top{}^{-1} \end{pmatrix} \\
AZ &= B \\
X &= \begin{pmatrix} D^{-1/2} & 0 \\ 0 & I \end{pmatrix} K_2^{-1} Z
\end{aligned}$$

4.4 Solver

The matrix Z for which we attempt to solve will have as many columns as there are signals we wish to simultaneously refit within an event; in the case of light-only denoising and with the restriction that we only examine events with one scintillation signal, Z will always have exactly one column. In general, though, it may be desirable to denoise events with multiple scintillation signals; or it may be beneficial to denoise certain wire channels (which may include wire signals) simultaneously with the APD channels. In these cases the number of columns may be larger – five or six columns may not be unusual. In such cases it would be possible to solve for each column of Z 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 Z .¹

The algorithm which we have selected for solving this system is the block variant of the stabilized biconjugate-gradient method (BI-BiCGSTAB) [CITE THIS]. It is reproduced here in the form used by our code.

$$R \leftarrow B - AZ_0$$

$$P \leftarrow R$$

$$\widetilde{R}_0 \leftarrow R$$

while any column of R has a magnitude greater than desired **do**

$$V \leftarrow AP$$

$$\text{Solve } (\widetilde{R}_0^\top V)\alpha = (\widetilde{R}_0^\top R) \text{ for } \alpha.$$

$$R \leftarrow R - V\alpha$$

$$T \leftarrow AR$$

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

$$Z \leftarrow Z + P\alpha + \omega R$$

$$R \leftarrow R - \omega T$$

R currently equals $B - AZ$; if it is small enough, terminate.

$$\text{Solve } (\widetilde{R}_0^\top V)\beta = -(\widetilde{R}_0^\top T) \text{ for } \beta.$$

$$P \leftarrow R + (P - \omega V)\beta$$

¹In practice the improvement is not quite so impressive, for two reasons: the vectors which are most useful toward solving one column may not be the most useful toward solving another; and some columns may achieve a solution earlier than others, yet are difficult to terminate without restarting the solver entirely. These challenges have not appeared to be significant for this particular system.

end while

Since each column of Z corresponds to one signal we are refitting, the norm of the corresponding column of 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 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.

4.5 Implementation

Although the previous section described the algorithm used for solving these matrix equations, in practice the processing time required is significant. There are certain implementation tricks which can be applied with significant impact. These are described here.

FILL-IN: Poisson multiplication using common factors.

After reducing the complexity of handling Poisson terms in this way, the handling of electronic noise terms in matrix multiplication becomes the computational bottleneck. The structure of the problem is that the matrix containing electronic noise terms is block-diagonal; that is, it consists of one dense block per frequency component, with each block lying on the diagonal and where the number of rows and columns is twice the number of channels being denoised (except for the last

frequency component, where imaginary terms are identically equal to zero). The problem of handling electronic noise terms thus consists of 1024 matrix-matrix multiplications, with the left matrix of approximate size 300×300 and the right matrix of approximate size $300 \times O(1)$.²

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. [Provide a citation!]

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” $300 \times O(1)$ 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

²The 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 300, and describing it as $O(1)$ captures this fact.

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.

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 (MKL is installed on both the SLAC and NERSC computing systems.) and its adaptability (MKL will detect a machine's architecture at runtime, and select a well-tuned codepath then; this is in contrast to some other BLAS implementations which must be compiled on the particular machine where they will be used.).

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.

Outline new algorithm demonstrating multi-threading and batch handling.

E is a list of events.

while there are events to read **do**

Push the next event E_i onto E .

Compute A_i and Z_0^i , the matrix and initial guess for event E_i .

$$R_i \leftarrow B - A_i Z_0^i$$

$$P_i \leftarrow R_i$$

end while

$$R \leftarrow B - AZ_0$$

$$P \leftarrow R$$

$$\widetilde{R}_0 \leftarrow R$$

while any column of R has a magnitude greater than desired **do**

$$V \leftarrow AP$$

$$\text{Solve } (\widetilde{R_0}^\top V)\alpha = (\widetilde{R_0}^\top R) \text{ for } \alpha.$$

$$R \leftarrow R - V\alpha$$

$$T \leftarrow AR$$

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

$$Z \leftarrow Z + P\alpha + \omega R$$

$$R \leftarrow R - \omega T$$

R currently equals $B - AZ$; if it is small enough, terminate.

$$\text{Solve } (\widetilde{R_0}^\top V)\beta = -(\widetilde{R_0}^\top T) \text{ for } \beta.$$

$$P \leftarrow R + (P - \omega V)\beta$$

end while

Todo: Figure out better notation for transpose inverse than $X^{\top^{-1}}$.

Chapter 5: Low-Field Running

I should write up at least a brief analysis of the low-field data, with denoising. In particular, demonstrate resolution vs field. This should be doable with the standard denoising lightmap, and I'll just expect the scintillation to peak somewhere other than 2615 keV.

Chapter A: Running Code

I will describe exactly where to access my code, provide an overview of how it works on an implementation level, and give steps for how to run it.

Bibliography

- [1] K. Nakamura et al. (Particle Data Group). Something from pdg. *J. Phys. G*, 37:075021, 2010.
- [2] E. Majorana. Teoria simmetrica dellelettrone e del positrone. *Nuovo Cimento*, 14:171, 1937.
- [3] Y. Fukuda et al. (Super-Kamiokande Collaboration). Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:15621567, 1998.
- [4] H. Nishino et al. (Super-Kamiokande Collaboration). 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, 2009.
- [5] SINDRUM II Collaboration. Improved limit on the branching ratio of $\mu \rightarrow e^+$ conversion on titanium. *Physics Letters B*, 422:334–338, 1998.
- [6] G. Feinberg, P. Kabir, and S. Weinberg. Transformation of muons into electrons. *Phys. Rev. Lett.*, 3:527, 1959.
- [7] J. Schechter and J.W.F. Valle. Neutrinoless double- β decay in $su(2) \times u(1)$ theories. *Phys. Rev. D*, 25:2951, 1982.
- [8] S. R. Elliott F. T. Avignone III and J. Engel. Double beta decay, majorana neutrinos, and neutrino mass. *Reviews in Modern Physics*, 80:481, 2008.
- [9] Planck Collaboration. Planck 2013 results. xvi. cosmological parameters. *Astronomy and Astrophysics*, under review.
- [10] V. M. Lobashev et al. Direct search for mass of neutrino and anomaly in the tritium beta-spectrum. *Physics Letters B*, 460:227–235, 1999.
- [11] J. Wolf (KATRIN Collaboration). The katrin neutrino mass experiment. *Nuclear Instruments and Methods in Physics Research A*, 623:442, 2010.

- [12] H. V. Klapdor-Kleingrothaus and I. V. Krivosheina. Evidence for neutrinoless double beta decay. *Mod. Phys. Lett. A*, 21:20:1547, 2006.
- [13] H. V. Klapdor-Kleingrothaus et al. Latest results from the heidelberg-moscow double beta decay experiment. *Eur. Phys. J. A*, 12:147, 2001.
- [14] C. Hall. Status of the exo double beta decay search. *ICHEP 2010 Conference*, 2010.
- [15] R. Neilson et al. Characterization of large area apds for the exo-200 detector. *Nucl. Instrum. Meth. A*, 608:68, 2009.
- [16] D.S. Leonard et al. Systematic study of trace radioactive impurities in candidate construction materials for exo-200. *Nucl. Instrum. Meth. A*, 591:490, 2008.
- [17] S. Herrin. Double beta decay in xenon-136: Measuring the neutrino-emitting mode and searching for majoron-emitting modes. *Thesis (Stanford University)*, 2013.
- [18] David W. Scott. *Multivariate density estimation : theory, practice, and visualization*. 1992.