UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Robust Measurement of Mixing Parameters $\sin^2 2\theta_{13}$ and Δm^2_{ee} with Reactor Antineutrinos at Daya Bay

Permalink

https://escholarship.org/uc/item/9ft2b17p

Author

Kramer, Matthew

Publication Date

2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Robust Measurement of Mixing Parameters $\sin^2 2\theta_{13}$ and Δm^2_{ee} with Reactor Antineutrinos at Daya Bay

By

Matthew Kramer

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

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

Fall 2021

Robust Measurement of Mixing Parameters
$\sin^2 2\theta_{13}$ and Δm^2_{ee} with Reactor Antineutrinos
at Daya Bay

Copyright 2021

By

Matthew Kramer

This work is licensed under a
Creative Commons Attribution 4.0 International License.

Abstract

Robust Measurement of Mixing Parameters
$\sin^2 2\theta_{13}$ and Δm^2_{ee} with Reactor Antineutrinos
at Daya Bay

by

Matthew Kramer

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Kam-Biu Luk, Chair

The Daya Bay reactor neutrino experiment measures the oscillation of electron antineutrinos produced by the Daya Bay and Ling Ao-I & II Power Plants (totaling 17.4 $\mathrm{GW_{th}}$) in southern China. The fluxes and spectra of the antineutrinos are observed using eight functionally identical gadolinium-doped liquid scintillator detectors, with a total target mass of 160 tons, divided among two near and one far experimental halls, resulting in baselines of order 500 m (near) and 2 km (far). The oscillation phenomenon appears as a deficit of electron antineutrinos at the far hall relative to the yield at the near halls.

This work describes a spectral measurement of the neutrino mixing parameters $\sin^2 2\theta_{13}$ and Δm^2_{ee} using a six-year sample of over 2 million inverse beta-decay (IBD) events identified via the outgoing neutron's subsequent capture on gadolinium. A defining feature of this analysis is its ability to be efficiently repeated for arbitrary choices of the *IBD cuts*, that is, the criteria used for selecting IBD events. This required the implementation of a flexible and efficient IBD selection, which was coupled to a parallelized upgrade of the official Daya Bay oscillation fitter developed at Lawrence Berkeley National Laboratory. The generalization of the analysis to arbitrary IBD cuts (beyond the two nominal IBD cuts used in official Daya Bay results) involved studies of various efficiencies and backgrounds and their dependence on the cuts, which we describe here.

With this infrastructure, we are able to explore how the best-fit oscillation parameters and their measured uncertainties vary as functions of the cuts, enabling, first, the characterization of the systematic uncertainty associated with the freedom to vary the cuts, and second, the exploration of whether the nominal IBD cuts should be modified in order to gain a reduced uncertainty on the oscillation parameters. We find that, compared to the two nominal cuts, there is no alternative IBD cut that provides a

substantial reduction in uncertainty. The cut-variation study suggests the assessment of an additional systematic uncertainty of 0.0006 on $\sin^2 2\theta_{13}$ and 1.9×10^{-5} eV^2 on Δm^2_{ee}, leading to a 2% increase in the total uncertainty of $\sin^2 2\theta_{13}$ and a 4% increase in that of Δm^2_{ee}. Our final measurement is $\sin^2 2\theta_{13} = 0.0850 \pm 0.0030$ and $\Delta m^2_{ee} = (2.5010 \pm 0.0072) \times 10^{-3}$ eV^2 in the standard three-flavor model.

The methods developed here serve to demonstrate that our oscillation fit is both robust and optimal with respect to variations in the cuts. Given the overall similarity between our oscillation analysis and those used in official Daya Bay results, similar conclusions can be drawn for the latter. Our findings thus validate past Daya Bay results, while future results will be able to benefit from this groundwork, enabling additional validation of the results and broadened assessment of their systematic uncertainties.

To my parents.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, this work owes its existence to the patient guidance and support of my friend and advisor, Professor Kam-Biu Luk, I have been extraordinarily lucky to have him as a mentor, one from whom I have learned so much not only about physics, but also about scientific integrity, service, and being a decent human being. His example truly provides an inspiration to strive toward. Moreover, his detailed and insightful feedback was invaluable over the course of preparing this work.

In a similar vein, I wish to thank the other members of my dissertation committee, Professors Yury Kolomensky and Karl van Bibber, for taking the time to review this work and provide suggestions for improvement. I owe Professor Kolomensky additional thanks for the many 290E seminars that he organized and led over the years. In particular, his lecture slides on statistics remain an extremely useful reference, all these years later. I cannot neglect to also note how much I enjoyed and learned from the many courses I took from him both as an undergraduate and a graduate student.

I have also learned a great deal from the many colleagues I've been privileged to work with on Daya Bay. Naturally, I've worked most closely with those who have formed part of the LBNL analysis group at one time or another. In no particular order, these include Cheng-Ju Lin, Herb Steiner, Dan Dwyer, Pedro Ochoa, Yasu Nakajima, Patrick Tsang, Sam Kohn, Chris Marshall, and Henoch Wong. My extreme gratitude also goes out to my Czech colleague Beda Roskovec; in addition to his collaboration on numerous analysis tasks, I truly appreciate his generosity in volunteering to assist me, and eventually lead the charge, in carrying out the activities of Daya Bay's Data Quality Working Group. My Chinese colleague Wenqiang Gu was also extremely helpful in teaching me about Daya Bay's data quality procedures. I also wish to thank my Russian colleagues, Maxim Gonchar and Kostya Treskov, for numerous delightful conversations on physics, programming, and life, and for their assistance in tracking down some particularly thorny bugs in my IBD selection.

The analysis presented in this work depends on a number of ingredients for which credit is due. Namely, the reactor model, including the predicted flux and spectra and the associated covariance matrix, is due to Christine Lewis, while the LBNL fitter and toy Monte Carlo were built by Yasu Nakajima, Patrick Tsang, Pedro Ochoa, Cheng-Ju Lin, and Henoch Wong. The fast-neutron background prediction comes from Bei-Zhen Hu, and the predictions for the AmC and α–n backgrounds from Lianghong Wei. The

energy nonlinearity model was provided by Yongbo Huang. References to their work can be found in the appropriate sections of the text. Of course, I must also collectively credit the many scientists and engineers on the Collaboration who have contributed to the overall experimental apparatus, the calibration, the reconstruction, and so forth. Daya Bay is a team effort, and I have been fortunate to work with such an incredible team. I've also been fortunate to have access to the computing and storage facilities at NERSC, and I thank the engineers who scramble behind the scenes while incompetent users, such as this author, try as hard as possible to crash the system.

This work was prepared entirely using Free Software, which I capitalize to emphasize that such software not only is free-of-charge, but also respects and enables the freedom of the user to inspect, modify, and share the code. I am a strong believer that science should be open and accessible to all, and Free Software is, in my view, essential for achieving this goal. I thus would like to thank the many individuals who have contributed to Linux, Git, LaTeX GCC, ROOT, Python, Julia, Geant4, Numpy, and all of the time-tested scientific libraries responsible for much of the heavy lifting in this field. Of all the software I have used in my work, however, the one that is closest to my heart is GNU Emacs. As far as I am concerned, this 40-year-old relic is still a century ahead of its time, and I cannot imagine working without it.

Finally, I wish to thank my family, my friends, and Cal Sailing Club. I am truly blessed to have all of you in my life.

Preface

In order to provide the reader with a reasonably clear understanding of the complete journey from the reactors and detectors to Daya Bay's raw data and then to our final results, we discuss each step in sequence, aiming to give at least a minimal amount of detail. This author has contributed to a number of these steps, but given the complexity of the experiment and its analysis, our results are necessarily built upon the work of the many others who have contributed to Daya Bay over the years. This thesis focuses primarily on our high-level analysis work, as detailed mainly in Chapters 5, 6, 7 and 9. Discussions of hardware and lower-level analysis are provided for completeness, but kept relatively concise, with additional details provided for some of the areas in which this author contributed personally. Certain elements of the high-level analysis, including the reactor model, the fitting framework, and some of the subdominant background predictions, are adopted from the work of others, to which citations are provided. For the interested reader, we review these subjects in the Appendix.

Working on Daya Bay has given this author the opportunity to contribute to the experiment in a range of areas. In order to provide some context, we list a number of our main contributions, along with references to corresponding sections of the text.[1]

- Onsite involvement in PMT testing and installation, commissioning of EH2-AD2 and EH3-AD4, and data-taking (Chapter 2).
- Implementation of the algorithm for calculating the calibrated time of each PMT hit, and preparation of the calibration constants (Section 3.2).
- Involvement in development and implementation of the algorithm for calculating the amount of calibrated charge for each PMT hit, with emphasis on the response of the electronics to multiple hits spaced closely in time (Section 3.3.1).
- Involvement in the generation of calibration constants for the energy reconstruction (Section 4.2.2) and the automation of this system.
- Implementation and validation (but not the original development) of the time-dependent nonuniformity correction for the energy reconstruction (Section 4.2.3).

[1] The order of this list roughly follows that of the text.

- Contributions to and maintenance of various components of `NuWa`, Daya Bay's offline analysis framework (Section 5.1), particularly the components related to time, charge, and energy calibration and reconstruction.

- Involvement in the migration of Daya Bay's data production process from NERSC's legacy PDSF computing cluster to the Cori high-performance computing system, requiring a bottom-up redesign of the production system. The ∼500 TB of data in the P17B dataset was successfully processed on Cori, with an order-of-magnitude improvement in production time.

- A modular and extensible C++ framework for high-performance processing of event streams in ROOT format, allowing for filtering events, transforming them, collecting statistics, and searching for time-correlated event clusters, such as IBD double coincidences [1]. Arbitrarily complex event selections can be rapidly implemented on top of this framework. A suite of scripts and tools enable deployment on Slurm-compatible high-performance computing systems for processing petabyte-scale datasets.

- An IBD (and singles) selection [2] built on the aforementioned framework (Chapter 5). The performance and flexibility of this implementation makes possible the rapid and systematic exploration of changes to the IBD selection criteria (Chapter 9).

- Cross-checks of the consistency of IBD candidate samples between different analysis groups, prior to publication of official oscillation results.

- Co-convenorship of the Daya Bay Data Quality Working Group, sharing responsibility for data quality in Daya Bay publications since 2017 (Section 5.4). Development of various data quality procedures, tools, and documentation. Maintenance and repair of data quality database. Investigation of anomalous data.

- Development of a website [3] for data quality activities, enabling exploration of data quality via various metrics, interactive (un)tagging of bad data, and reviews of the tagging decisions made by data quality shifters (Figure 5.4). This website made it much easier to carry out a data quality shift, increasing participation in this important role by members of the Collaboration.

- Original development of the LBNL ^{9}Li analysis (Section 6.3), which was further extended by Marshall [4] to support spectrum extraction and improved calculations of efficiencies and uncertainties. This work forms, together with other independent evaluations, the basis of the ^{9}Li/^{8}He rates used in official Daya Bay results.

- A novel method for analytically calculating the singles rate (and, by extension, the accidentals rate and multiplicity cut efficiency) using the Lambert W function (Section 7.5).
- A Monte Carlo method for computing the statistical uncertainty of the accidentals rate (Section 7.7.2).
- Various contributions to (but not the original development of) the LBNL toy Monte Carlo and oscillation fitter (Chapters 8 and C), including reorganization of the code, documentation of internals, updates for new datasets, integration with our IBD selection, and, most notably, significant performance improvements. Using aggressive multi-threading and process-based concurrency, runtime was reduced by an order of magnitude (Section 9.1.1), making possible our cut-variation study.
- Monte Carlo methods for predicting the muon veto efficiency from the rates of different classes of muons (Section 9.2.1).
- Techniques to correct background rates for changes in IBD selection criteria, without rerunning the original background analyses (Chapter 9).
- Development of methods for measuring the neutron-capture spectra and calculating the efficiency of the delayed-energy cut (Section 9.3.2).
- An implementation of vertex cuts for the nGd oscillation analysis (Section 9.5), an analysis in which such cuts have previously been minimally investigated (in contrast to the nH analysis.)
- Ongoing use of the infrastructure described in this thesis, particularly the ability to study cut variations (Chapter 9), in the preparation of upcoming Daya Bay oscillation-analysis results. These methods enable demonstration of the minimal cut-sensitivity of the results, and assignment of a systematic to account for any residual sensitivity.

A note on terminology: Throughout this work, the term "spectrum" will typically be used to refer to both the *rate* and the *spectral shape* (e.g., in terms of reconstructed energy) of some process. More concretely, a spectrum can be thought of as an energy histogram. When it is clear from the context, "spectrum" may occasionally refer to only the spectral shape; when we need to be more explicit, "shape" will be used in situations where the rate is irrelevant.

Chapter 1

Physics of neutrinos

1.1 Neutrinos and the Standard Model

The Standard Model (SM) of particle physics [5] has proven to be an enormous success. From a handful of ingredients—three gauge groups, three generations of quarks and leptons, the Higgs field, and 18 free parameters—the SM provides a succinct and precise description of nature that agrees incredibly well with the bulk of experimental observations.

However, even if we ignore the glaring absence of gravity in the theory (a difficulty inherent to quantum field theory itself), there are clear and exciting signs that new physics must lie beyond the Standard Model. For instance, the SM fails to explain dark matter, dark energy, cosmic inflation, the lightness of the electroweak scale, and the matter/antimatter asymmetry of the Universe, among other puzzles.

These are difficult problems which may take decades to resolve, and in most cases there is little clarity on how the SM will need to be extended along the way. On the other hand, the last half-century has produced overwhelming evidence of another Beyond the Standard Model (BSM) effect, one that admits a relatively successful quantitative description: neutrino mixing. Before discussing this phenomenon, it is worthwhile to review the story of the neutrino itself.

More than a century ago, measurements of nuclear beta decay gave the surprising result that the energy of the outgoing electron was continuously distributed [6], in stark contrast to the discrete lines observed in alpha and gamma decay. If, like alpha and gamma decay, beta decay were a two-body process, then a continuous spectrum would seem to imply the violation of energy and momentum conservation. Furthermore, it had been observed that nuclear spin is either integral (for even mass numbers) or half-integral (for odd mass numbers), implying that the nuclear spin can only change by an integer during beta decays, which conserve the mass number. And yet, the electron has a spin of 1/2, so a two-body process would also imply the non-conservation of angular momentum. Did all of this mean that, alas, it was

necessary to discard the most sacred conservation laws of physics?

An alternative resolution, one that would avoid violations of the conservation laws, was proposed in 1930 by Pauli, who postulated the existence of a yet-unobserved, light, neutral particle, or "neutron", contained in the nucleus and emitted along with the electron in beta decay [7]. Chadwick's 1932 discovery of the actual neutron led Fermi and others to re-dub Pauli's hypothetical particle as the *neutrino*. In 1933, Fermi's theory of beta decay [8], which incorporated the neutrino, was successful in reproducing the measured electron spectra, but the neutrino itself would not be directly observed for another two decades.

In 1956, antineutrinos from a nuclear reactor were detected by the Cowan-Reines experiment, marking the first direct confirmation of the neutrino's existence [9]. Over the decades that followed, it was found that neutrinos come in three distinct "flavors"—electron, muon, and tau, corresponding to their charged lepton partners—and that neutrinos lack any discernible mass. These qualities would eventually be incorporated into the Standard Model, which crystallized in the 1970s after a period of remarkable theoretical and experimental progress in particle physics.

Since neutrinos have only ever been observed via their weak interactions, which solely involve the left-handed neutrino state, there is no direct experimental evidence for a right-handed neutrino. Additionally, the observed kinematics of beta decay are, within experimental limits, consistent with the neutrino being massless. Accordingly, in the SM, the three neutrinos are massless left-handed Weyl spinors that interact via the W and Z bosons. From a theoretical standpoint, masslessness is appealing, as it avoids the need to imbue the theory with right-handed neutrinos (which have never been observed) or Majorana mass terms (which are not present for any other SM particle). Although some puzzling neutrino observations, discussed in Section E.1, had been known since the late 1960s, massive neutrinos were seldom given serious consideration as the explanation. But as experimental evidence continued to mount, this wall would eventually have to crumble, leading to the revolution that has transpired over the last few decades. Before detailing this history, we provide an overview of the underlying physics, the parameters of which will be referred to repeatedly in our historical discussion.

1.2 Neutrino oscillation physics

Neutrino oscillations are the consequence of two facts: First, that the flavor eigenstates are not the same as the mass eigenstates (*mixing*), and second, that the mass eigenvalues aren't fully degenerate (implying that at least one is nonzero). As a result, a flavor eigenstate is a superposition of mass eigenstates which each undergo phase rotation at their own rates; the mass components thus interfere to produce different flavor compositions over time, leading to the observation of oscillations.

The neutrino fields in the flavor and mass bases are related by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) mixing matrix, U_{PMNS} (or simply U):

$$\nu_\alpha = U_{\alpha i}\nu_i, \tag{1.1}$$

that is, for three generations,

$$\begin{pmatrix}\nu_e\\ \nu_\mu\\ \nu_\tau\end{pmatrix} = \underbrace{\begin{pmatrix}U_{e1} & U_{e2} & U_{e3}\\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3}\\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3}\end{pmatrix}}_{U_{\text{PMNS}}} \begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix}. \tag{1.2}$$

To determine the relationships among the neutrino *states,* as opposed to the fields, we first note that the field operator ν has the effect of annihilating a neutrino state. Conversely, its adjoint $\nu^\dagger$ can be used to create a neutrino state $|\nu\rangle$ out of the vacuum $|0\rangle$. Considering the case of creating a flavor eigenstate $|\nu_\alpha\rangle$, we have:

$$\begin{aligned}|\nu_\alpha\rangle &= \nu_\alpha^\dagger |0\rangle\\ &= (U_{\alpha i}\nu_i)^\dagger |0\rangle\\ &= U_{\alpha i}^* \nu_i^\dagger |0\rangle\\ &= U_{\alpha i}^* |\nu_i\rangle .\end{aligned} \tag{1.3}$$

That is, the *states* (as opposed to the fields) are related by the complex conjugate of U. For antineutrinos (annihilated by the $\nu^\dagger$ operator), the situation is reversed:

$$\begin{aligned}\nu_\alpha^\dagger &= U_{\alpha i}^* \nu_i^\dagger\\ |\bar{\nu}_\alpha\rangle &= U_{\alpha i} |\bar{\nu}_i\rangle .\end{aligned} \tag{1.4}$$

U_{PMNS} can be parameterized in terms of the three mixing angles, θ_{12}, θ_{23}, and θ_{13}, along with a CP-violating complex phase δ_{CP}:[1]

$$\begin{aligned}U &= \begin{pmatrix}1 & 0 & 0\\ 0 & c_{23} & s_{23}\\ 0 & -s_{23} & c_{23}\end{pmatrix}\begin{pmatrix}c_{13} & 0 & s_{13}e^{-i\delta_{\text{CP}}}\\ 0 & 1 & 0\\ -s_{13}e^{i\delta_{\text{CP}}} & 0 & c_{13}\end{pmatrix}\begin{pmatrix}c_{12} & s_{12} & 0\\ -s_{12} & c_{12} & 0\\ 0 & 0 & 1\end{pmatrix}\\ &= \begin{pmatrix}c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{\text{CP}}}\\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{\text{CP}}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{\text{CP}}} & s_{23}c_{13}\\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{\text{CP}}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{\text{CP}}} & c_{23}c_{13}\end{pmatrix}.\end{aligned} \tag{1.5}$$

[1] Although a general 3×3 unitary matrix includes six complex phases, most of the phases in a 3×3 *mixing* matrix are physically meaningless. If the neutrinos are Dirac particles, with distinct particle and antiparticle fields, then five of the phases can be absorbed into the definitions of the fields, leaving only one physical phase. It could be inserted anywhere in the factorization of U as long as unitary is preserved, but by convention the definition in Equation 1.5 is what is used. Conversely, if the neutrinos are Majorana particles, where $\nu = \bar{\nu}$ (as discussed in Section E.3.2), only three phases can be absorbed, leaving two physical *Majorana phases* in addition to δ_{CP}.

where $c_{12} \equiv \cos\theta_{12}$ and $s_{12} \equiv \sin\theta_{12}$, etc. It is these mixing angles (or trigonometric functions thereof), and not the matrix elements themselves, that are directly measured by experiments.

Physically speaking, neutrinos are produced in local processes that conserve energy and momentum. A fully microscopic treatment of neutrino oscillation would therefore require accounting for the neutrino's energy-momentum spread and its entanglement with the other interaction products (as required for energy-momentum conservation). However, for a bulk flux of neutrinos from a known source, the oscillation probability can be derived by simply considering a spacetime-filling plane wave of well-defined energy *or* momentum (but not both). For the physical systems of interest in neutrino oscillation experiments, this approach gives results that agree (for all practical purposes) with those of a fully self-consistent theoretical treatment [10]. Indeed, when Daya Bay's data is fit to models that treat the neutrino as a wave packet, the results are completely consistent with the plane-wave approach [11]. As such, for the purposes of this illustrative derivation, we proceed with the plane-wave model.

To calculate the oscillation probability over baseline L for a neutrino of energy E and initial flavor α, we roughly follow[2] [12] and consider a stationary neutrino state $|\nu(x)\rangle$ defined such that[3]

$$|\nu(0)\rangle = |\nu_\alpha\rangle \,. \tag{1.6}$$

That is, at the origin, $|\nu\rangle$ is a flavor eigenstate, consisting of a mixture of mass eigenstates:

$$|\nu(0)\rangle = U^*_{\alpha i} \, |\nu_i\rangle \,. \tag{1.7}$$

The spatial dependence of $|\nu(x)\rangle$ is determined by the fact that the state has a fixed energy E. Each mass-eigenstate component i of $|\nu(x)\rangle$ is then a plane wave with momentum

$$p_i = \sqrt{E^2 - m_i^2} = E - \frac{m_i^2}{2E} - \mathcal{O}\left(\frac{m_i^4}{E^3}\right). \tag{1.8}$$

Here, we assume that the masses are very small, below an eV^2, which at the $\mathcal{O}(\text{MeV})$ energy scales we consider, implies that the higher-order terms in Equation 1.8 are

[2] The derivation in [12] begins with a neutrino of well-defined *momentum* rather than energy. To leading order, the results are the same whether we fix the energy, momentum, or even the velocity. Throughout this work, energy is the variable of interest, so for clarity we choose it to be the fixed quantity here.

[3] In what follows, flavor indices are represented by Greek letters, while mass indices use Roman letters. Also, without loss of generality, we consider only one spatial dimension in this treatment. Time-dependence is ignored, since we are working with an energy eigenstate.

negligible.[4] We then have

$$\begin{aligned} |\nu(x)\rangle &= e^{ip_i x} U^*_{\alpha i} |\nu_i\rangle \\ &\approx \exp\left(iEx - i\frac{m_i^2}{2E}x\right) U^*_{\alpha i} |\nu_i\rangle . \end{aligned} \tag{1.9}$$

The probability of measuring $|\nu(L)\rangle$ to be of flavor β is then

$$\begin{aligned} P(\alpha \to \beta) = |\langle \nu_\beta \,|\, \nu(L)\rangle|^2 &= \left|\langle \nu_j| \, U_{\beta j} \exp\left(-i\frac{m_i^2 L}{2E}\right) U^*_{\alpha i} |\nu_i\rangle\right|^2 \\ &= \left|U^*_{\alpha i} U_{\beta i} \exp\left(-i\frac{m_i^2 L}{2E}\right)\right|^2 \\ &= U^*_{\alpha j} U_{\beta j} U_{\alpha i} U^*_{\beta i} \exp\left(-i\frac{\Delta m_{ji}^2 L}{2E}\right), \end{aligned}$$

where

$$\Delta m_{ji}^2 \equiv m_j^2 - m_i^2. \tag{1.10}$$

This can be rewritten by separating the terms for $i = j$ and $i \neq j$, giving

$$P(\alpha \to \beta) = \sum_j |U_{\alpha j}|^2 |U_{\beta j}|^2 + 2\,\mathfrak{Re} \sum_{j>i} U^*_{\alpha j} U_{\beta j} U_{\alpha i} U^*_{\beta i} \exp\left(-i\frac{\Delta m_{ji}^2 L}{2E}\right), \tag{1.11}$$

where, for clarity, we are explicitly indicating the summations. Going further, we can employ the unitarity relation

$$U_{\alpha j} U^*_{\beta j} = \delta_{\alpha\beta} \tag{1.12}$$

which, upon squaring, gives

$$\sum_j |U_{\alpha j}|^2 |U_{\beta j}|^2 = \delta_{\alpha\beta} - 2 \sum_{j>i} \mathfrak{Re}(U^*_{\alpha j} U_{\beta j} U_{\alpha i} U^*_{\beta i}). \tag{1.13}$$

Substituting this into Equation 1.11, and using Euler's identity, along with the trigonometric identity $1 - \cos 2\varphi = 2\sin^2\varphi$, we finally get

$$\begin{aligned} P(\alpha \to \beta) = \delta_{\alpha\beta} &- 4 \sum_{j>i} \mathfrak{Re}(U^*_{\alpha j} U_{\beta j} U_{\alpha i} U^*_{\beta i}) \sin^2\left(\frac{\Delta m_{ji}^2 L}{4E}\right) \\ &+ 2 \sum_{j>i} \mathfrak{Im}(U^*_{\alpha j} U_{\beta j} U_{\alpha i} U^*_{\beta i}) \sin\left(\frac{\Delta m_{ji}^2 L}{2E}\right). \end{aligned} \tag{1.14}$$

[4] In the case of the electron neutrino, this smallness was obvious from the earliest observations of beta decay kinematics. For the muon neutrino, direct measurements of pion decay [13] provided a less-stringent upper limit on the mass of $\mathcal{O}(1\text{ MeV})$, but later oscillation measurements of the mass-squared splittings showed that all three eigenstates indeed possess similarly tiny masses.

Note that this result applies to *neutrinos*, not antineutrinos. For the latter, Equation 1.4 implies that the initial state $|\bar{\nu}_\alpha(0)\rangle$ can be written as

$$|\bar{\nu}_\alpha(0)\rangle = U_{\alpha i}\,|\bar{\nu}_i\rangle \tag{1.15}$$

(compare to Equation 1.7, and note the lack of complex conjugation on the matrix element). The preceding derivation then produces Equation 1.14 with the complex conjugations swapped.

Now that we have defined the mixing angles and mass-squared splittings, it is worthwhile to enumerate the global best-fit values of these parameters, since the values will be alluded to in subsequent discussions. Table 1.1 lists the values as compiled by the Particle Data Group [14]. Currently, it is unknown whether the neutrino mass spectrum consists of two "light" states and one "heavy" state, or vice versa. This issue is discussed further in Section E.3.2. The two possible *orderings* (or *hierarchies*) are labeled "normal" and "inverted", respectively, as illustrated in Figure 1.1. Since the ordering affects the existing measurements of θ_{23} and Δm^2_{23}, Table 1.1 gives values for both cases.

Parameter	Value	Comment
$\sin^2\theta_{12}$	0.307 ± 0.013	
$\sin^2\theta_{23}$	0.546 ± 0.021	Normal order
$\sin^2\theta_{23}$	0.539 ± 0.022	Inverted order
$\sin^2\theta_{13}$	$(2.20 \pm 0.07) \times 10^{-2}$	
Δm^2_{21}	$(7.53 \pm 0.18) \times 10^{-5}$ eV^2	
Δm^2_{32}	$(2.453 \pm 0.033) \times 10^{-3}$ eV^2	Normal order
Δm^2_{32}	$(-2.524 \pm 0.034) \times 10^{-3}$ eV^2	Inverted order

Table 1.1: The global best-fit oscillation parameters, from [14]. Although Δm^2_{31} is not listed, it can trivially be calculated as $\Delta m^2_{32} + \Delta m^2_{21}$ (for either ordering, given the sign convention used here).

In an experiment where we are looking for the *disappearance* of (anti)neutrinos of a given flavor, our interest is in $P(\alpha \to \alpha)$. In this case, the behavior of neutrinos and antineutrinos is the same, and Equation 1.14 simplifies to

$$P(\alpha \to \alpha) = 1 - 4\sum_{j>i} |U_{\alpha j}|^2 |U_{\alpha i}|^2 \sin^2\left(\frac{\Delta m^2_{ji} L}{4E}\right). \tag{1.16}$$

The matrix elements from Equation 1.5 can then be inserted in order to derive the disappearance probability for a particular flavor. For electron antineutrinos, as at

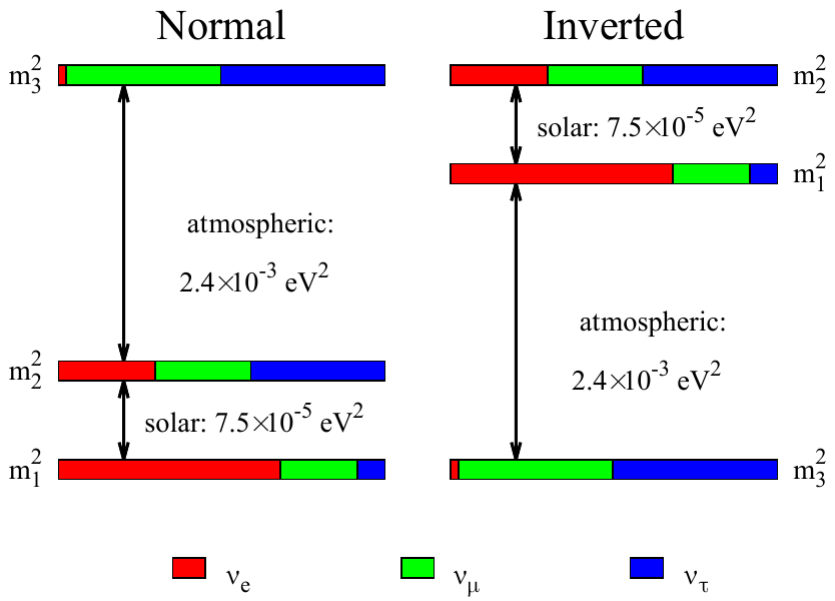

Figure 1.1: The two possible *mass hierarchies*, i.e., orderings of neutrino masses, as allowed by current data. From [15].

Daya Bay, the survival probability is

$$\begin{aligned} P(\bar{\nu}_e \to \bar{\nu}_e) = 1 &- \cos^4\theta_{13}\sin^2 2\theta_{12}\sin^2\Delta_{21} \\ &- \sin^2 2\theta_{13}\left(\cos^2\theta_{12}\sin^2\Delta_{31} + \sin^2\theta_{12}\sin^2\Delta_{32}\right), \end{aligned} \tag{1.17}$$

where we've introduced the notation

$$\Delta_{ji} \equiv \frac{\Delta m^2_{ji} L}{4E} \approx \frac{1.267\,\Delta m^2_{ji}\,[\mathrm{eV}^2]\;L\,[\mathrm{m}]}{E\,[\mathrm{MeV}]}.$$

Based on the value of Δm^2_{21} given in Table 1.1, at the baselines of Daya Bay $\sin^2\Delta_{21}$ is so small that the experiment has no ability to constrain θ_{12} (nor Δm^2_{21} itself). In this analysis, then, θ_{12} and Δm^2_{21} are fixed to the global values. This leaves θ_{13}, $|\Delta m^2_{31}|$, and $|\Delta m^2_{32}|$. However, given that the difference between $|\Delta m^2_{31}|$ and $|\Delta m^2_{32}|$ is less than one part in thirty, distinguishing between their corresponding phases would require some combination of a high-resolution detector and a baseline long enough to stretch out the phase difference. At Daya Bay, the baseline of ∼1 km is approximately one oscillation length, where the detector resolution is insufficient to resolve the difference. On the other hand, since Δm^2_{21} is well constrained and known to be positive, we can tightly relate $|\Delta m^2_{31}|$ and $|\Delta m^2_{32}|$:

$$|\Delta m^2_{31}| = \begin{cases} |\Delta m^2_{32}| + \Delta m^2_{21}, & \text{normal hierarchy } (\Delta m^2_{32} > 0), \\ |\Delta m^2_{32}| - \Delta m^2_{21}, & \text{inverted hierarchy } (\Delta m^2_{32} < 0), \end{cases} \tag{1.18}$$

Thus, by using this relation to eliminate one parameter, it is possible to perform a two-parameter fit of Equation 1.17 directly, *provided that the mass hierarchy is specified*. Since the mass hierarchy is currently unknown, two sets of results must be reported; furthermore, they are subject to change if an improved determination of Δm^2_{21} is ever published.

Alternatively, we can recast Equation 1.17 in a form that refers only to an empirical *effective* mass splitting Δm^2_{ee}:

$$P(\bar{\nu}_e \to \bar{\nu}_e) \approx 1 - \cos^4\theta_{13}\sin^2 2\theta_{12}\sin^2\Delta_{21} - \sin^2 2\theta_{13}\sin^2\Delta_{ee}, \tag{1.19}$$

where

$$\Delta_{ee} \equiv \frac{\Delta m^2_{ee} L}{4E}. \tag{1.20}$$

It must be noted that Equation 1.19 is not an exact re-parameterization of Equation 1.17, since the former contains only two frequencies rather than three. Again, however, Daya Bay cannot resolve between Δ_{31} and Δ_{32}, so there is no practical loss in sensitivity with this approach. The advantage of it is that it produces a single value that is independent of the mass hierarchy and immune to changes wrought by updates to Δm^2_{21}.

Δm^2_{ee}, as used here, has an *operational,* rather than a physical, definition: It is simply the value that, when inserted into Equation 1.19, gives the best fit to the data. In Daya Bay's case, however, Δm^2_{ee} can be related to physical quantities to a very good degree of approximation. Returning to Equation 1.17, it can be shown [16] that, in Daya Bay's range of L/E,

$$\cos^2\theta_{12}\sin^2\Delta_{31} + \sin^2\theta_{12}\sin^2\Delta_{32} \approx \sin^2(\Delta_{32} \pm \phi), \tag{1.21}$$

where

$$\phi \equiv \arctan\left(\frac{\sin 2\Delta_{21}}{\cos 2\Delta_{21} + \tan^2\theta_{12}}\right), \tag{1.22}$$

and "+" ("–") corresponds to the normal (inverted) hierarchy. Inserting Equation 1.21 into Equation 1.17 and comparing to Equation 1.19 gives the approximate relation

$$\Delta m^2_{ee} \approx |\Delta m^2_{32}| \pm \Delta m^2_{\phi}/2, \tag{1.23}$$

where

$$\Delta m^2_{\phi} = \frac{4\phi E}{L}. \tag{1.24}$$

Although Δm^2_{ϕ} depends on L/E, for Daya Bay it is essentially constant and equal to $\cos^2\theta_{12}\Delta m^2_{21}$. Inserting that into Equation 1.23, we find that, for both mass hierarchies,

$$\Delta m^2_{ee} \approx |\Delta m^2_{31}\cos^2\theta_{12}| + |\Delta m^2_{32}\sin^2\theta_{12}|. \tag{1.25}$$

For all intents and purposes, this approximation is exact at Daya Bay. Hence Equations 1.25 and 1.18 can be used to relate Daya Bay's measured Δm^2_{ee}, as operationally defined by Equation 1.19, to the physical mass splittings for the normal and inverted hierarchies (NH and IH, respectively):

$$|\Delta m^2_{31}| = \Delta m^2_{ee} + \sin^2\theta_{12}\Delta m^2_{21}, \qquad |\Delta m^2_{32}| = \Delta m^2_{ee} - \cos^2\theta_{12}\Delta m^2_{21} \qquad \text{[NH]} \tag{1.26}$$

$$|\Delta m^2_{31}| = \Delta m^2_{ee} - \sin^2\theta_{12}\Delta m^2_{21}, \qquad |\Delta m^2_{32}| = \Delta m^2_{ee} + \cos^2\theta_{12}\Delta m^2_{21} \qquad \text{[IH]} \tag{1.27}$$

Of course, in doing so, one must still specify the mass hierarchy and Δm^2_{21}. As a sanity check, Daya Bay has compared the results of fitting Equation 1.17 and Equation 1.19, finding that, in the end, both techniques give the same values of $|\Delta m^2_{32}|$ and $|\Delta m^2_{31}|$. For the sake of simplicity, this analysis will use Equation 1.19 and fit Δm^2_{ee}.

1.3 Further reading

The history of neutrino oscillations is a fascinating one. For the interested reader, we give a general overview in Section E.1. In Section E.2, we provide additional historical discussion of reactor neutrino experiments.

Chapter 2

Daya Bay experiment

Introduction

The Daya Bay experiment was designed to measure θ_{13} by observing the antineutrinos produced by the six 2.9 GW_{th} nuclear reactors of the Daya Bay and Ling Ao power plants, located in Shenzhen in southern China. A total of eight functionally identical antineutrino detectors (ADs) were deployed, each containing a target of 20 tons of gadolinium-doped liquid scintillator (GdLS). Four of the ADs were evenly divided among two near halls (∼350-600 m baselines from the cores), and the remaining four were placed in a single far hall (∼1500-1950 m baselines). Shielding from cosmic rays was provided by ∼100 m and ∼300 m, respectively, of mountainous overburden at the near and far halls. The ADs in each hall were immersed in instrumented water pools which provided shielding from ambient radioactivity and detection of Cherenkov radiation from atmospheric muons. Redundant detection of muons, as well as directional information, were made available by resistive plate chambers (RPCs) laid on top of the water pools. In this chapter we discuss further details of the layout, the detectors, and the shielding and vetoing system of the experiment.

2.1 Site layout

As shown in Figure 2.1, the power reactors are divided into three nuclear power plants (NPPs), each containing two cores (illustrated in Figure 2.2). One of the two clusters contains the Daya Bay NPP (cores D1 and D2), while the other cluster consists of the Ling Ao (L1 and L2) and Ling Ao-II (L3 and L4) NPPs. EH1 is located around 350 m from the Daya Bay NPP, while EH2 is roughly 500 m from the two Ling Ao NPPs. The far hall, EH3, in turn is located about 1900 m from the Daya Bay NPP and 1500 m from the Ling Ao NPPs. The measured baselines, as determined from a combined GPS and total station theodolite survey, are given in Table 2.1. The uncertainty of ∼2 cm in these measurements (less than 0.01% in the worst case)

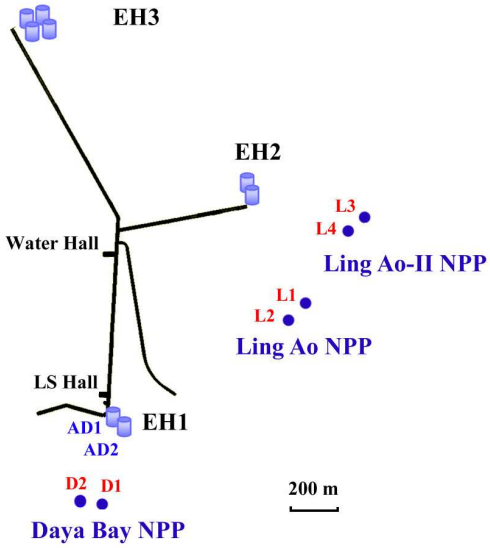

Figure 2.1: The layout of the Daya Bay experiment. The black lines represent horizontal tunnels into the mountain. From [17].

is negligible relative to the other uncertainties in the oscillation analysis. Likewise, simulations of reactor operation have determined that the centroid of $\bar{\nu}_e$ emission lies within ∼2 cm of each core's center [19]. Although $\bar{\nu}_e$ emission is distributed across the volume of each core (3.7 m high and 3 m in diameter), this spread has negligible effects at Daya Bay's baselines [19][1]. Altogether, then, and in keeping with the official practices of the Collaboration, we treat each reactor as a point source, and we asume that the baselines are known exactly.

2.2 Antineutrino detectors

The design of the Daya Bay ADs is shown in Figure 2.3. Each AD is made of a cylindrical stainless steel vessel (SSV), 5 m in height and diameter, containing two nested cylinders of UV-transparent acrylic. The inner acrylic vessel (IAV), 3 m in height and diameter, 10 mm thick, contains the target mass of 20 t of gadolinium-doped liquid scintillator (GdLS)[2], containing 0.1% Gd by mass. Surrounding it is the outer acrylic vessel (OAV), 4 m in height and diameter, 18 mm thick, which contains

[1]The same applies to the nonzero volume of the ADs

[2]Daya Bay's liquid scintillator consists of linear alkyl benzene (LAB) as the solvent, 3 g/L of 2,5-diphenyloxazole (PPO) as the fluor, and 15 mg/L of p-bis-(o-methylstyril)-benzene (bis-MSB) as the wavelength shifter. For the GdLS, ^{nat}Gd was added in the form of a complex with 3,5,5-trimethylhexanoic acid (TMHA). Further details can be found in [20].

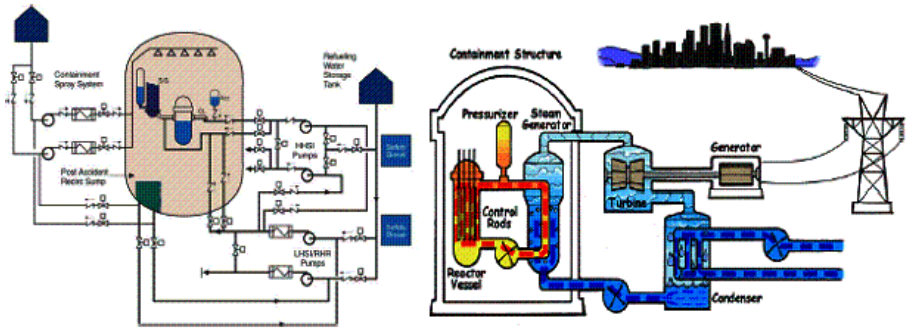

Figure 2.2: The design of the CPR-1000 reactors used at the Ling Ao II power plant. The CPR-1000 is an upgraded variant of the similar M310 reactors used at the Daya Bay and Ling Ao I plants. From [18].

		Reactor baseline [m]					
Hall	Detector	D1	D2	L1	L2	L3	L4
EH1	AD1	362.38	371.76	903.47	817.16	1353.62	1265.32
	AD2	357.94	368.41	903.35	816.90	1354.23	1265.89
EH2	AD3	1332.48	1358.15	467.57	489.58	557.58	499.21
	AD8	1337.43	1362.88	472.97	495.35	558.71	501.07
EH3	AD4	1919.63	1894.34	1533.18	1533.63	1551.38	1524.94
	AD5	1917.52	1891.98	1534.92	1535.03	1554.77	1528.05
	AD6	1925.26	1899.86	1538.93	1539.47	1556.34	1530.08
	AD7	1923.15	1897.51	1540.67	1540.87	1559.72	1533.18

Table 2.1: Baselines between geometric centers of the ADs and of the reactor cores. From [19].

21 t of *Gd-free* liquid scintillator (LS). This "gamma catcher" volume ensures the $\geq$90% containment and measurement of gamma rays produced near the edge of the GdLS, while also providing additional target mass for studies that make use of neutron capture on hydrogen instead of on gadolinium. Between the OAV and the inner wall of the SSV, a 37 t volume of transparent mineral oil (MO) provides shielding from radioactivity in the detector materials, in addition to its role in balancing the stress on the OAV wall. All liquids were formulated to have densities within 1% of each other (Table 2.2), in order to minimize stresses on the vessel walls [21].

Although the ADs were filled at different times (especially in the cases of EH2-AD2 and EH3-AD4, which were commissioned after the experiment had already begun

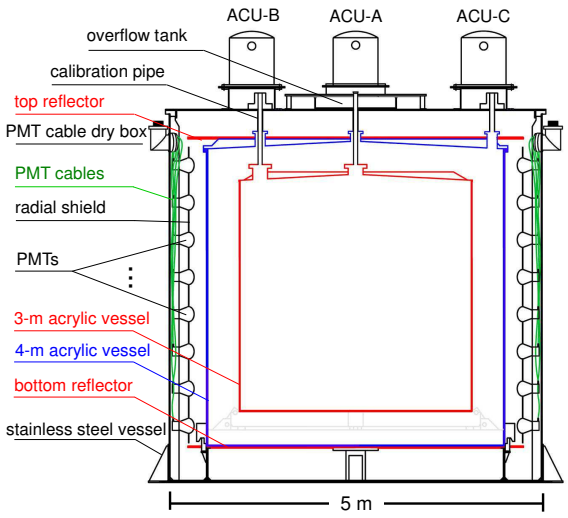

Figure 2.3: Structure of a Daya Bay antineutrino detector. From [19].

Liquid	Density (g/ml)
Gd-LS	0.860
LS	0.859
MO	0.851

Table 2.2: Densities of the liquids in the ADs [17].

taking data), each liquid (GdLS, LS, and MO) was prepared as one large batch in 2011 and kept in storage tanks that were subsequently used, in equal proportions, for all 8 ADs. This procedure minimized any variations in liquid properties between the ADs. Given that the antineutrino detection efficiency is directly proportional to the number of target protons (i.e., the mass) in the GdLS volume, it was crucial to accurately measure the liquid mass in each AD, as any errors in this step would translate into a bias on θ_{13}. Toward this end, the filling process employed precision weigh-bridge load cells to measure the mass of an intermediate holding tank used in the filling process. After the holding tank was filled from the storage tanks, its mass was recorded. Its mass was measured again after the liquid was transferred to the detector; the difference then gave the total mass transferred. The filling process was periodically stopped to add and remove calibration masses to the holding tank in order to track any calibration drifts. Filling proceeded until the overflow tank above each volume was filled to about 1/3 capacity.

The final mass of each liquid was determined by subtracting the load cell readings before and after filling. An additional subtraction was performed to account for the mass in each overflow tank. Since the load cell had been programmed by the

manufacturer to assume a value of the gravitational acceleration g which was 0.18% higher[3] than that at Daya Bay, a correction of this size was applied to the absolute mass scale. A further 0.13% correction accounted for the nitrogen gas that filled the holding tank after pumping. The calibration measurements ultimately revealed a maximum drift of ±2 kg (0.01% of 20 t). Combined with the uncertainty from the additional corrections, this gave a total error on the GdLS mass of ±3 kg (0.015%), far below the goal of ±0.2% [22].

In order to convert the mass of the GdLS into the number of target protons, the chemical composition of the GdLS must be accurately determined. Samples of liquids were taken from each AD in order to be characterized. The proportion of Gd was measured using X-ray fluorescent spectroscopy, with a precision of ∼2% [21]. Meanwhile, the fractions of carbon, hydrogen, and nitrogen were obtained from combustion analysis. The liquids from all ADs were consistent within the measurement error of 0.3%. Based on this analysis, it was determined that the GdLS contains 7.169×10^{25} protons/kg [22]. For the oscillation analysis, which is only sensitive to AD-to-AD relative variations in detection uncertainty, the 0.3% composition uncertainty is not included in the target mass uncertainty, since all measurements indicated that the ADs contain identical liquids. The GdLS masses and number of target protons are tabulated in Table 2.3 for each AD.

AD	Target mass (kg)	Target protons ($\times 10^{26}$)
EH1-AD1	19941 ± 3	14296 ± 2
EH1-AD2	19966 ± 3	14314 ± 2
EH2-AD1	19891 ± 3	14260 ± 2
EH2-AD2	19945 ± 3	14299 ± 2
EH3-AD1	19913 ± 3	14276 ± 2
EH3-AD2	19991 ± 3	14332 ± 2
EH3-AD3	19892 ± 3	14261 ± 2
EH3-AD4	19931 ± 3	14289 ± 2

Table 2.3: Target masses and number of target protons in each AD. From [19].

Within the MO volume, the inner sidewall of the SSV supports 192 8-inch Hamamatsu R5912 photomultipler tubes (PMTs) to detect the light from scintillation in the scintillator. The PMTs are arranged in eight rings of 24 tubes whose photocathodes protrude from matte-black radial shields that fully cover the sidewalls, preventing light from reflecting off the walls. This simplifies the optical characteristics of the ADs, reducing the complexity of vertex reconstruction. Conversely, however, reflective discs

[3] Amusingly, if this g were really the one at the load-cell factory, then, assuming a uniform spherical Earth, we would conclude that either Daya Bay sits at an elevation above ∼5,000 m, or that it is buried under the sea. Both possibilities are known to be false.

are installed at the top and bottom of the OAV, improving both the energy resolution and the uniformity of light collection. To further reduce nonuniformity effects, each PMT is outfitted with a FINEMET truncated conical magnetic shield to minimize azimuthal variations in PMT response caused by the Earth's magnetic field [23].

2.2.1 PMT electronics

Each PMT is positively biased via a single coaxial cable in order to achieve a gain within 5% of 10^7 (corresponding to ∼20 ADC counts per PE). Due to intrinsic differences between PMTs, the necessary high voltage varies from 1300 to 1700 kV. Collected charge is passed through a passive decoupling circuit which removes the HV offset; this brief (∼20 ns), unbiased pulse is then passed to the front-end electronics (FEE), where it is split and sent to two separate circuits. One of the circuits contains a discriminator with a threshold set to ∼0.25 photoelectrons (PE), which initiates a TDC counter (of 1.6 ns resolution) to record the presence and time of the "hit". The other circuit is a CR-$(RC)^4$ shaper which stretches each pulse to a length of ∼200 ns; the shaped pulse is then split and sent to both a ×10 "high-gain" amplifier and a ×0.5 "low-gain" attenuator (for neutrino-like and muon-like events, respectively) and, finally, the two shaped and rescaled pulses are sampled by a 40 MHz 12-bit ADC. The output of a hardware-based peak-finding algorithm is then recorded as the raw amplitude of the pulse, for both the high-gain and low-gain circuits. Meanwhile, the average of the four ADC samples immediately preceding the over-threshold condition is recorded as the *pre-ADC* or *pedestal* value of the hit, from which the peak ADC value is subtracted in determining the charge, as discussed in Section 3.3[4].

Although every hit is initially observed in this manner by the hardware, it is only recorded in the DAQ's output stream if a trigger is issued for the AD as a whole. An AD can be triggered when the total observed charge is above a software-specified threshold, or when the number of "hit" channels is above threshold[5]. Each FEE board, which reads up to 16 channels, sends two signals to the AD's *local trigger board* (LTB): a digital count of recently-hit channels (NHIT) and an analog sum of the charge across all channels (ESUM)[6]. The LTB combines the inputs from all of the FEEs

[4]The pre-ADC, in general, is *not* taken from the four samples immediately preceding the *peak* sample, because the peak of the shaped curve occurs some 100 ns after the over-threshold condition, as dictated by the time constant of the shaping circuit. Given the sample period of 25 ns, this implies a 4-5 sample lag between the over-threshold condition and the peak. Thus, the pre-ADC is typically taken from the average of the 5th through 9th samples preceding the peak, give or take. This, however, changes when there are multiple closely-spaced hits, as illustrated later in Figure 3.5.

[5]Additional trigger types include *prescaled* triggers, issued at ∼10 Hz for monitoring of low-level activity; *calibration* triggers, issued in sync with LED pulses during weekly calibrations; and *cross* triggers, issued under certain conditions when another detector is triggered.

[6]To be precise, the ESUM signal is generated as the analog sum of all PMT signals, each integrated with a 50 ns shaping time. The analog sum is then passed to a discriminator whose threshold is set by a programmable DAC [24].

and issues a trigger when either NHIT or ESUM are above threshold; for ordinary physics data-taking, the thresholds are NHIT $\geq$ 45 or ESUM $\geq$ 65 photoelectrons ($\sim$0.4 MeV). When a trigger is issued, a GPS-synchronized clock (25 ns resolution) records the overall event timestamp, each hit's TDC is stopped (indicating the offset of each hit relative to the event timestamp), and the readout (including the TDC, high-gain ADC, and low-gain ADC for each hit within 1.2 μs of the trigger) is sent to the DAQ system. Typical trigger rates for the eight ADs are given in Table 2.4.

Detector	Trigger rate (Hz)
EH1-AD1	273
EH1-AD2	268
EH2-AD1	215
EH2-AD2	211
EH3-AD1	131
EH3-AD2	124
EH3-AD3	120
EH3-AD4	131

Table 2.4: Typical trigger rates for the ADs, as given in [21]. Although there is not a readily available exact breakdown into the different trigger types (ESUM, NHIT, and prescaled), it is straightforward to approximate it: First, a small fraction (10 Hz) comes from prescaled triggers. Of the remaining triggers, a brief investigation using the Offline Data Monitor indicates that about 70% satisfy both the ESUM and NHIT conditions. The other 30% are roughly evenly split between ESUM-only and NHIT-only triggers.

Thus, the raw information collected from each AD consists of a set of triggers. Each trigger, in turn, contains a timestamp and a collection of *hits*; each hit describes the PMT ID, the TDC count, and the high- and low-gain ADC values reported by the peak-finding circuitry. In order to be made useful for physics analysis, the raw ADC values must first be converted into photoelectron counts for each PMT, as described in Chapter 3, and then the individual PMTs must be combined and corrected to produce the amount of energy deposited in the scintillator, as elaborated in Chapter 4.

2.2.2 Detection principle

The vast majority of events recorded by the ADs are unrelated to antineutrinos. Most events come from natural radioactivity in the detector and scintillator materials, as well as from cosmic-ray muons and the byproducts of muon-nucleon reactions. Fortunately, it is possible to exploit the double-trigger nature of antineutrino events in order to effectively extract them from the data, as we explain here.

Antineutrinos can interact with the GdLS target in a number of ways. Their interactions can be mediated either by the W^- boson (corresponding to a *charged current*, or CC, interaction), in which case the electron antineutrino becomes a positron, or they can be mediated by the Z boson (corresponding to a *neutral current*, or NC, interaction), in which case the antineutrino escapes from the detector after depositing some recoil energy in the target. Furthermore, the interaction may take place between the antineutrino and either an electron, a proton, a neutron, or (coherently) an entire nucleus. Among these many possibilities, however, only one channel provides a signature that allows for efficient discrimination from background: inverse beta decay (IBD),

$$\bar{\nu}_e + p \rightarrow e^+ + n. \tag{2.1}$$

This interaction is illustrated in Figure 2.4. An initial *prompt* scintillation signal is produced by the positron (first from direct ionization, then from the gamma rays produced by annihilation). Meanwhile, the free neutron thermalizes and is then captured by a nucleus, which quickly de-excites, emitting gamma rays that provide a second, *delayed,* signal, closely correlated in time with the prompt signal. This "double-pulse" signature effectively distinguishes IBDs from events produced by natural radioactivity, as the latter largely consists of single pulses. Meanwhile, muon-induced double-pulse events can be eliminated by simply vetoing for a sufficient length of time after each muon, as described in Chapter 5. The use of gadolinium further improves background separation: Without gadolinium, most neutrons would be captured on hydrogen, with a time constant of $\sim$200 μs and a release of 2.2 MeV in gamma ray energy. However, in the GdLS, some 85% of the neutrons are captured on gadolinium at 0.1% concentration [25]; compared to hydrogen captures, gadolinium captures have a much shorter time constant of $\sim$30μs and a much higher delayed gamma ray energy of 8 MeV. These two properties significantly reduce the probability of accidentally identifying a pair of uncorrelated events as an IBD, especially since the rate of uncorrelated events drops dramatically above 5 MeV. Thanks to the use of gadolinium captures, such "accidental" backgrounds make up only 1-2% of the selected IBD sample. Further details regarding background measurement and subtraction can be found in Chapter 6.

The kinematics of the IBD reaction are fairly straightforward. Essentially all of the antineutrino's energy $E_{\bar{\nu}}$ comes from its kinetic energy. During the interaction, some of this energy, equal to the neutron-proton mass difference $m_n - m_p$ (1.3 MeV) goes toward the conversion of the target proton into a neutron. An additional small amount of energy, on the order of 10 keV [19], provides the recoil kinetic energy K_n of the neutron. The conversion of the antineutrino into a positron consumes additional energy equal to the positron mass m_e (511 keV). All of the remaining energy takes the form of the positron's kinetic energy K_{e^+}. Thus,

$$K_{e^+} = E_{\bar{\nu}} - (m_n - m_p) - m_e - K_n \tag{2.2}$$

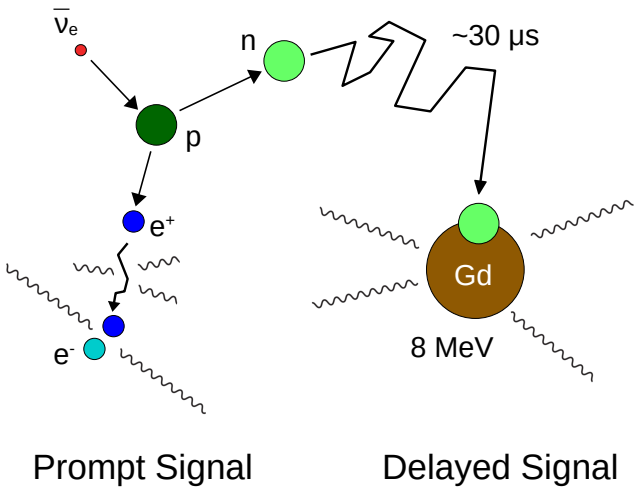

Figure 2.4: An illustration of the inverse beta decay reaction. Unlike a water Cherenkov detector, a Daya Bay AD cannot discern the direction of the positron.

After depositing its kinetic energy via ionization of the scintillator, the positron annihilates with an electron, releasing a pair of gamma rays with a total energy of $2m_e$. These gamma rays then undergo Compton scattering and/or photoelectric absorption, and the scintillator is further ionized by the kinetic energy of the resulting electrons. The total deposited energy E_{dep} is then

$$\begin{aligned} E_{\text{dep}} &= K_{e^+} + 2m_e \\ &= E_{\bar{\nu}} + m_e - (m_n - m_p) - K_n. \end{aligned}$$

Since the energy resolution of the AD is, in the worst case (i.e. around 1 MeV), about 10% (or 100 keV) [19], the O(10 keV) kinetic energy of the neutron can be safely neglected. Using the fact that

$$(m_n - m_p) - m_e \approx 0.8\,\text{MeV}, \tag{2.3}$$

we finally find that

$$E_{\text{dep}} \approx E_{\bar{\nu}} - 0.8\,\text{MeV}. \tag{2.4}$$

The energy threshold for the IBD reaction is equal to the amount of energy needed to produce the positron and to convert the proton into a neutron:

$$\begin{aligned} E_{\bar{\nu},\text{min}} &= (m_n - m_p) + m_e \\ &\approx 1.3 + 0.5\,\text{MeV} \\ &= 1.8\,\text{MeV}. \end{aligned}$$

Using Equation 2.4, it then follows that

$$E_{\text{dep,min}} = 1.0\,\text{MeV}. \tag{2.5}$$

Of course, this is also apparent from the simple fact that a threshold positron, with zero kinetic energy, will deposit an energy equal to $2m_e$, or ∼1.0 MeV. Given the finite energy resolution of the detector, some of these events will be reconstructed below 1.0 MeV. As such, a standard prompt-energy cut of 0.7 MeV is used in the official Daya Bay oscillation analyses, corresponding to about 3σ below the 1.0-MeV peak. In Chapter 9, we explore the effects of modifying this 0.7-MeV cut.

2.3 Water pools and muon system

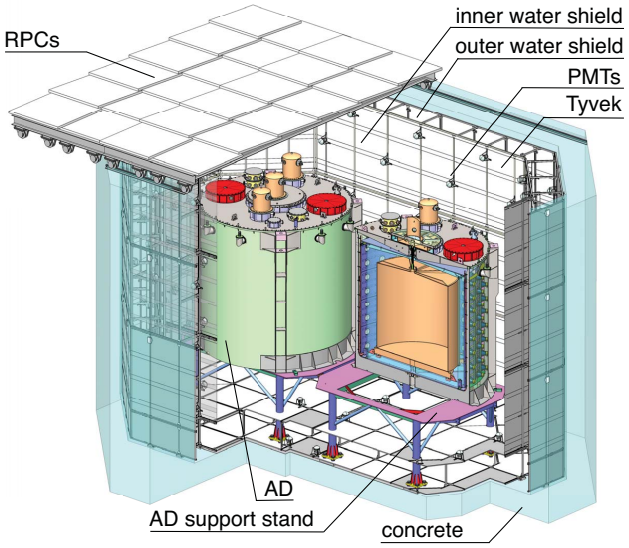

Figure 2.5: Water pool (including ADs) as configured in the near halls. The far hall is similar, with four ADs instead of two. From [19].

Within each experimental hall, the ADs are immersed inside an instrumented pool of ultra-pure water. The water pools serve two purposes: First, to shield the ADs against both ambient γ radioactivity and muon-induced neutrons from the surrounding rock, and second, to detect cosmic-ray muons that pass within the vicinity of the ADs. With respect to shielding, the ∼2.5 m-thick layer of water surrounding the ADs provides a $\sim 10^6$ reduction in the rate of PMT hits from rock radioactivity [26]; without this suppression, antineutrino detection would be impossible.

Meanwhile, muon detection was accomplished by using PMTs[7] to observe the Cherenkov light produced when muons traverse the water. In order to allow for cross-checking of the muon system [26], the water pools were divided by Tyvek sheets into two optically isolated zones, the *inner* and *outer water pools* (IWP/OWP). The IWP is instrumented by inward-facing PMTs protruding from the Tyvek, while the OWP is instrumented by both outward-facing PMTs on the Tyvek and inward-facing PMTs on the walls. The EH1 and EH2 water pools are essentially identical, while the EH3 pool was twice as wide in order to accommodate four detectors (in a 2×2 arrangement) instead of two. The number of PMTs in each water pool is listed in Table 2.5 The FEE and trigger systems for the water pools are the same as those for the ADs, but the trigger configuration is different: ESUM is not used, and the NHIT thresholds are $\geq$6 for the IWP and $\geq$7 (8) for the near (far) hall OWP. Offline, in software, more stringent cuts (e.g., NHIT $\geq$ 12) were used as the definition of a muon event (see Table 2.6 for typical trigger rates in the muon system); AD events were then ignored in the immediate aftermath of such muons, greatly reducing muon-induced IBD-like backgrounds (especially neutrons).

Hall	IWS PMTs	OWS PMTs
EH1	121	167
EH2	121	167
EH3	160	224

Table 2.5: The number of PMTs in each water pool [21].

Additional muon detection was provided by an assembly of modular resistive plate chambers (RPCs) mounted on a rolling frame on top of each water pool. Each module, approximately 2.2 m square and 8 cm thick, contained four layers of Bakelite sheets instrumented by eight readout strips, oriented alternately (between layers) in the x and y directions. The RPCs thus provide the 2D coordinate, at around 10 cm resolution, of each muon track that intersects the RPC plane. Two additional *telescope* modules were installed in each hall along the center of opposing long edges of the water pool, to allow for high-angular-resolution tracking of a subset of muons in muon-related studies. The RPCs used HV, FEE, and trigger electronics distinct from those used by the AD and WP PMTs, with a trigger being issued whenever three out of four layers in a module are above threshold. Although the RPCs proved to be extremely useful in studies of muons and muon-induced backgrounds, they are not used in the analysis

[7]The water pool PMTs consisted of 619 Hamamatsu R5912 PMTs, as used in the ADs, as well as 341 EMI 9350KA and D642KB PMTs recycled from the MACRO experiment. The MACRO PMTs ultimately proved to be somewhat failure-prone, but not to the point of degrading the overall muon detection efficiency.

Detector	Trigger rate (Hz)
EH1-IWP	220
EH1-OWP	325
EH1-RPC	215
EH2-IWP	192
EH2-OWP	245
EH2-RPC	103
EH3-IWP	39
EH3-OWP	54
EH3-RPC	36

Table 2.6: Typical trigger rates for the muon system, as given in [21].

described in this thesis, as the water pools alone were sufficient to detect muons with effectively 100% efficiency.

2.4 Operations

The operational procedures at Daya Bay are intended to ensure that data is recorded continuously, with minimal interruptions. As such, the majority of time is spent with the DAQ recording so-called *Physics* runs, with the standard NHIT, ESUM, and prescaled trigger settings discussed in Section 2.2.1. Each hall has a separate DAQ, so in general, three Physics runs were being taken in parallel at any given time. A typical Physics run lasts from 1–7 days, with a trend toward longer runs later in the experiment. Physics runs are generally ended either to perform maintenance or repairs, or to conduct calibration runs (discussed shortly). Due to the large volume of data collected in each run, the DAQ does not emit a single data file for the run, but instead outputs a new file after roughly 1 GB of data has been accumulated, corresponding to 10–15 (30–40) minutes in the near (far) halls. Individual data files are thus identified by the *run number* (taken from an incrementing counter common among all halls) and the *file number* (which, for each run, ranges from 1 up to the total number of files in the run).

Every time the onsite DAQ emits a data file, it is stored temporarily on disk, where it is detected by a custom data transfer service known as SPADE. At this point, SPADE then initiates a transfer of the file to an offsite permanent storage facility at IHEP in China, where the file is categorized and stored. A second instance of SPADE, running at IHEP, then transfers the file to permanent storage at NERSC in the US. As a result, two redundant long-term copies of each data file are stored shortly after the data has been recorded.

Each Friday morning, ongoing Physics runs are stopped in order to perform a weekly set of calibration, or *ADCalib* runs. These runs make use of the three automated calibration units (ACUs) located at the top of each AD. ACU A is located above the center, ACU B lies near the edge of the GdLS, and ACU C is situated above the LS. Each ACU features a turntable containing multiple calibration sources. The turntable rotates in order to select a particular source, which is then lowered by a cable, at an operator-specified height, into the liquid volume.

The sources deployed by each ACU include an LED (contained within a diffuser ball), a ^{60}Co source (producing 1.17 and 1.33 MeV gamma rays), a ^{68}Ge source (producing positrons), and a ^{241}AmC-^{13}C source (producing low-rate neutrons). The ^{241}AmC-^{13}C sources were always co-deployed with the ^{60}Co and ^{68}Ge sources, and were removed from ACUs B and C of every AD during commissioning of EH2-AD2 and EH3-AD4, so as to reduce correlated backgrounds from the AmC sources (Section A.3).

A typical Friday calibration campaign, lasting for about three hours, consisted of LED, ^{60}Co (plus ^{241}AmC-^{13}C), and ^{68}Ge (plus ^{241}AmC-^{13}C) runs, using all three ACUs at $\sim$8 vertical positions ranging from the top to the bottom of the AD. Of these runs, only the ^{60}Co runs were used in regular calibration procedures, namely, the energy scale calibration of the so-called *AdScaled* reconstruction algorithm, which we do not employ in this analysis. The other types of calibrations were useful for various studies of detector response (and, in the case of LED runs, for recalibrating the timing characteristics of each channel following a change in the electronics, as discussed in Section 3.2). In this analysis, the calibration runs are not used directly; as detailed in Chapter 3, channel gains are calibrated using *in situ* dark noise, and the energy scale of our chosen reconstruction (*AdSimple*) uses spallation neutrons recorded during ordinary Physics runs.

Additional non-Physics runs, taken briefly and occasionally, included *Pedestal* runs, used for monitoring the baselines of the ADCs, *FEEDiag* runs, used for collecting diagnostic data from the electronics, and *MOMonitor* runs, for monitoring the clarity of the mineral oil. None of these run types are relevant for this analysis.

In contrast to the "fast" operations of the DAQ, an auxiliary *slow control* system (the Detector Control System, or DCS) was used for monitoring and controlling the environment in which the experiment operated. Most notably, the DCS was responsible for the high voltage (HV), allowing the HV to be enabled, disabled, and fine-tuned in order to achieve the desired gains. The DCS also incorporated a number of sensors to record temperature, humidity, barometric pressure, radon concentrations, water pressure, liquid levels in the overflow tanks, and various parameters of the gases above the liquids; cameras were also connected in order to visually monitor the interiors of the detectors. Any time the sensors would detect a deviation outside of the expected range, an alarm would be raised. These alarms would be noted by the weekly *shifter* (the individual drafted with monitoring the experiment) and, if necessary, the appropriate expert would be consulted in order to correct any issues. In this way, the experiment was kept operating under the conditions needed to ensure quality of

data. Control of temperature was particularly important, since any variations in the density of the GdLS target would affect the absolute antineutrino detection efficiency. Fortunately, the DCS enabled the temperature to be maintained within an adequately tight range, eliminating this source of potential bias.

Chapter 3

Channel calibration

3.1 Overview

For the antineutrino detectors and the water pools' Cherenkov detectors, the Daya Bay DAQ system outputs timestamped PMT hits, grouped into readout windows and tagged with trigger information. As discussed in Section 2.2, these hits take the form of ADC readings from the peak-finding electronics, along with the TDC count at the time that each PMT waveform crossed the discriminator threshold. In order to carry out any sort of physics analysis, it is necessary to first convert this raw ADC and TDC data into the higher-level quantities of photoelectron count and photon arrival time, and then to combine individual channels into overall event parameters such as energy and position.

This chapter describes the first half of that process; namely, the detector calibration and data processing involved in the channel-by-channel calculation of hit time and charge. (The second half is discussed in Chapter 4.) These calibrated quantities are fundamental in that they are used in all further analysis stages and by all reconstruction algorithms. Accuracy is vital, as any bias in the channel charge will be reflected in the total event energy and in any (charge-based) reconstructed vertex; likewise, accurate times are important for time-based vertex reconstructions. Furthermore, it is necessary to identify, exclude, and compensate for any misbehaving channels, a process that will also be discussed here.

3.2 Timing calibration

The timing calibration takes each hit's TDC count[1] and converts it into an estimate of the time at which the photon struck the photocathode. The absolute time is unimportant, but for the purpose of time-based vertex reconstruction (or track

[1] I.e., the number of ticks that elapsed between the hit and the trigger.

reconstruction, as applied in muon studies), the *relative* times between channels must be accurately determined. These calibrated times, in addition to their use in vertex reconstruction, are also used in defining the time window for hit selection (performed during the event-wide charge calculation, described in the next chapter as the first stage of the energy reconstruction).[2]

This process involves subtracting a channel-specific offset (largely corresponding to cable length) from each channel's hit times, with an additional correction for the charge-dependent *timewalk effect*, in which smaller pulses take longer than larger ones to cross the FEE discriminator threshold. The offset, as well as a parameterization of the timewalk curve, is stored in the calibration database and applied to the TDC readings during data processing. We now discuss the preparation of these calibration constants.

3.2.1 Calibration constant preparation

To measure each channel's offset and timewalk profile, we require a well-defined event vertex and an external source of T_0 (true event time) information.[3] Toward that end, Daya Bay uses LED calibration runs in which the LED is positioned at the center of the detector. When a pulse is sent to fire the LED, a "hit" is also sent to a "fake" T_0 channel in the FEE, providing a reference from which other TDCs can be subtracted. In order to ensure recording of this T_0 hit as well as all of the true PMT hits, an external trigger signal is sent to the trigger board, triggering a readout. For each PMT hit on channel i, we take its TDC count N_i and calculate a corrected time

$$t_i = \frac{N_0 - N_i}{f_{\mathrm{TDC}}} - \frac{nr_i}{c}, \tag{3.1}$$

where N_0 is the TDC count of the T_0 channel, f_{TDC} is the TDC frequency (640 MHz), and nr_i/c gives the time of flight (TOF) to PMT channel i, located a distance r_i from the AD's center. The TOF subtraction ensures that we can directly compare channels from different rings without any further geometric considerations. In principle, the resulting t_i should be equal for all channels (modulo the timewalk effects caused by the variation of photon flux for different rings), but in reality, discrete offsets are observed between rings, due to the discretization resulting from the TDC clock's 1.5625 ns

[2]Without a hit selection window, the event's total charge could include hits, such as those from dark noise, uncorrelated with the underlying physical event. To be fair, given that this window is 400 ns wide, and any calibration corrections are on the order of a few ns, the raw times would actually suffice for hit selection.

[3]There is a natural variance, on the order of a dozen ns, in the timing of triggers relative to the true event time (even for identical calibration events), smearing the TDC measurements between different events. Knowledge of T_0 effectively provides knowledge of this trigger "jitter" for each event, allowing it to be subtracted out. Without this information, it is difficult to obtain a useful timewalk curve due to the aforementioned smearing.

period, which is on the same scale as the difference in propagation times between rings. This discretization is not explicitly corrected for, as it does not prevent time-based vertex reconstructions from achieving equal or superior resolution compared to charge-based reconstructions.

In the next step, a 2D histogram is constructed for each channel by taking all of the channel's hits (within a reasonable time window, loosely constructed on the scale of the propagation time between the AD's center and the PMTs) across all events, and plotting each hit's corrected time t against its ADC count q. Sampling across q is ensured both by variation of the LED voltage and by the underlying Poisson statistics of photon observation. This histogram's profile is then fit to the six-parameter functional form

$$t(q) = a_1 + a_2 \exp(-a_3 q) + a_4 \exp(-a_5 q) + a_6 \log q, \tag{3.2}$$

which was empirically found to produce good fits under appropriate restrictions on the parameters. Figure 3.2 shows an example of a timewalk fit.

After a manual verification of the fit quality, the six parameters for each channel were uploaded to the database and marked with a suitable validity period. Whenever the electronics had been modified in a way that could affect the channel timing (for example, a cabling change or a board replacement), a new set of constants was prepared for the affected AD, using this same procedure.

It is worth noting that the a_1 parameter can be expressed as $\overline{a}_1 + \delta a_1$, where $\overline{a}_1$ is the average value of a_1 across all channels, and δa_1 is the channel-specific deviation. In this case, δa_1 represents intrinsic (and significant, for such purposes of vertex reconstruction) differences in timing offsets between channels, due to such factors as variation in transit time of the PMTs. Meanwhile, $\overline{a}_1$ has no effect on vertex reconstruction and thus can be removed, meaning that we apply the transformation $a_1 \mapsto a_1 - \overline{a}_1$. Doing so ensures that, for ordinary physical triggers, the physical hits always fall within the same well-defined band of timewalk-corrected times. If $\overline{a}_1$ were not removed in this way, then this band would shift as $\overline{a}_1$ does, which could bias the calculation of the total charge, since the charge calculation uses a fixed acceptance window for hit selection. Removing $\overline{a}_1$ also ensures that each stored timewalk curve is nearly zero for large charges (modulo the a_6 term), with the deviations from zero reflecting the differences in timing offsets between channels.

Due to this last property, the expected time for a "large" pulse (for which the timewalk correction is nearly zero), with TDC count N, is approximately $-N/f_{\mathrm{TDC}}$, where $f_{\mathrm{TDC}} = 1/(1.625\ \mathrm{ns})$. As shown in Figure 3.1, typical physical hits have TDC counts of 950–1000, corresponding (respectively) to times of -1545 to -1625 ns (where $t = 0$ is the time of the trigger). The details of selecting physical hits, including the exact range of times used, are discussed in Section 4.2.1.

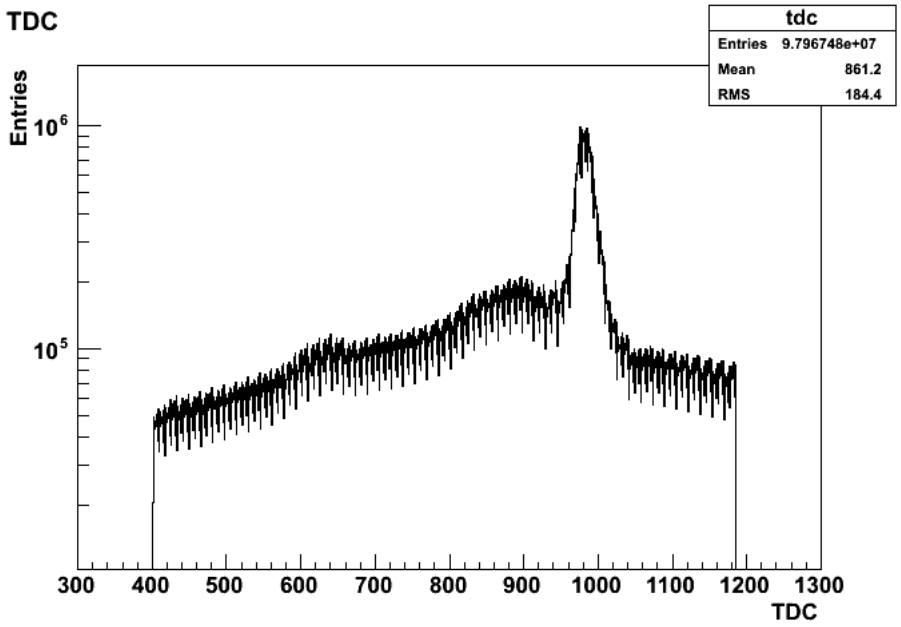

Figure 3.1: A typical distribution of TDC values for hits in a single PMT during a physics run. The spikes originate from the fact that the bin spacing does not exactly equal the period of the TDC clock. From the Daya Bay Offline Data Monitor.

3.2.2 Calculation of corrected times

Once the calibration parameters are in the database, applying them to the raw data is straightforward. For each hit in a given channel, we take the ADC count q and the raw TDC count N, and then calculate the calibrated time t_c (in nanoseconds):

$$t_c = -\frac{N}{f_{\text{TDC}}} - t(q), \tag{3.3}$$

This calibrated time is independent of the intrinsic differences between channels, thus meeting the requirements for time-based vertex and track reconstruction.

3.3 Charge calibration

At a given operating voltage, every PMT will produce a unique amount of charge per photoelectron. Furthermore, the Daya Bay FEE channels have an intrinsic 3% variation in the number of ADC counts output per unit of charge. As such, the response of each channel to a single photoelectron (PE), that is, the *total gain*[4] (in

[4]Here, the term *gain* will refer to the total gain (in ADC/PE), and we will use *PMT gain* to refer to the (unitless) gain of the PMT itself.

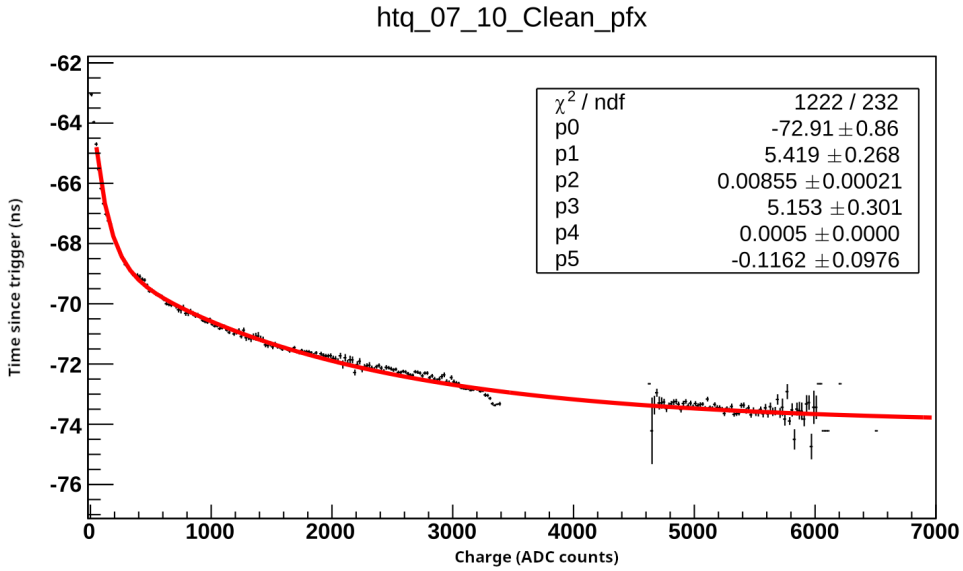

Figure 3.2: An example of a timewalk fit. The time axis is defined relative to the trigger. Since all hits occur before a trigger is issued, the time values are negative; the more negative, the earlier. The large gap around 4000 ADC counts is due to the nonlinear behavior of the calibration LED. That is, the LED voltage increment was not small enough to cover this region.

ADC/PE), cannot be precisely predicted based on operating voltage, but instead must be measured *in situ*. Since environmental conditions and the passage of time can alter this response, the gain measurement must be repeated regularly. Once the gain is known, it is possible to determine the total number of photoelectrons observed in an event, regardless (within reason) of any time variation of PMT gains. The gain measurement is described in this section.

Prior to data taking, each PMT underwent benchtop tests in order to estimate an operating voltage that would deliver a PMT gain of $1 \times 10^7 \pm 5\%$. At this gain, the peak voltage of the shaped pulse corresponds to approximately 19 ADC counts, based on the inherent (and fixed) calibration of the ADC. This voltage then remained fixed as long as the gain didn't deviate excessively from the intended range. Minor deviations were acceptable and expected, and were effectively compensated by the gain calibration, provided that the drift occurred over time scales longer than the typical calibration period of ∼6 hours. This was indeed the case: Gain drifts were found to occur on a scale of a few percent per year (Figure 3.3).

To measure the gain, Daya Bay uses "dark noise", the thermal emission of electrons from the photocathode and the dynodes. Dark noise can be measured "for free" during physics data taking by taking advantage of the size of the Daya Bay readout window[5].

[5]It bears mentioning that in the past, Daya Bay has in fact employed a separate, redundant gain

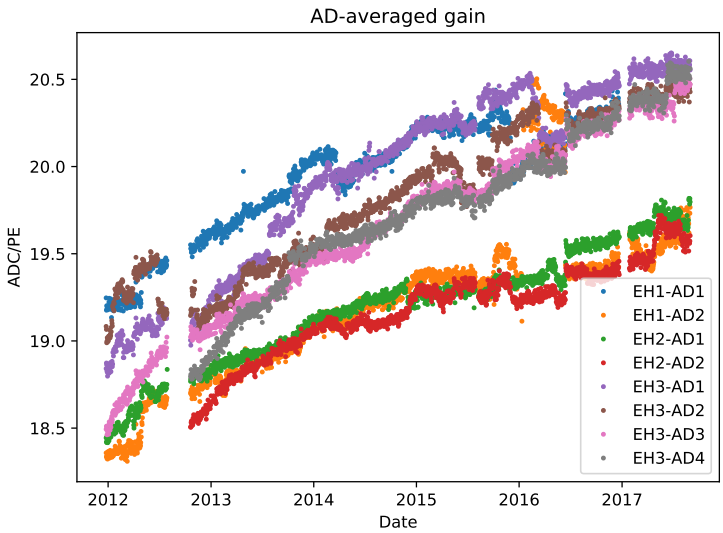

Figure 3.3: PMT gain over time, averaged over 8" PMTs (excluding bad channels) in each AD. The jump in 2016 for EH1-AD2 was due to a miscabling issue which caused certain channels to receive the incorrect HV.

Within a single event, there will be a spread of a few dozen nanoseconds between the hits that correspond to the actual physical event. This timescale is determined by the time it takes light to propagate within the detector volume. Since each readout window is about 1.3 μs wide, and a trigger is issued about 1 μs after a physical event takes place, the first few hundred nanoseconds of a readout window serve as an observation of each PMT during detector quiescence[6]. Daya Bay's gain calibration software extracts the hits that lie in this part of each readout. After subtraction of the ADC pedestal for each hit, a histogram of ADC counts for each channel is then obtained.

Once the dark noise histogram has accumulated sufficient statistics (six hours' worth) to provide a stable fit, it is fit to a model that includes the total gain as one of the parameters. Thermal emission is modeled as a Poisson process (in terms of the number of photoelectrons), and, in turn, the collected charge per photoelectron is modeled as a Gaussian distribution. Convoluting the two then gives the probability density function used in the fit:

$$P(Q) = \sum_{n=1}^{\infty} \frac{\mu^n e^{-\mu}}{n!} \frac{1}{\sigma_{\mathrm{SPE}}\sqrt{2n\pi}} \exp\left(-\frac{(Q - n\overline{Q}_{\mathrm{SPE}})^2}{2n\sigma_{\mathrm{SPE}}^2}\right) \tag{3.4}$$

calibration using LED runs, which gave results consistent with this "rolling gain" method.

[6]Dark hits can also be found in the lengthy portion of each readout window following the "physics" hits, but since the charge readings can be biased by ringing in the electronics, they are not used in the gain calibration.

where Q is the observed ADC count, μ (always close to 1) is the mean number of photoelectrons per thermal emission event, $\overline{Q}_{\mathrm{SPE}}$ is the gain of a given channel, and σ_{SPE} is the width of the Gaussian peak. Due to the extremely low probability of observing multiple thermal electrons in a single hit, the sum was restricted to $n \leq 2$ in practice, with negligible effect. This model was found to fit the data stably and precisely (e.g., Figure 3.4) as long the fit region was restricted to lie above 10 ADC, below which noise fluctuations had the ability to destabilize the fit.

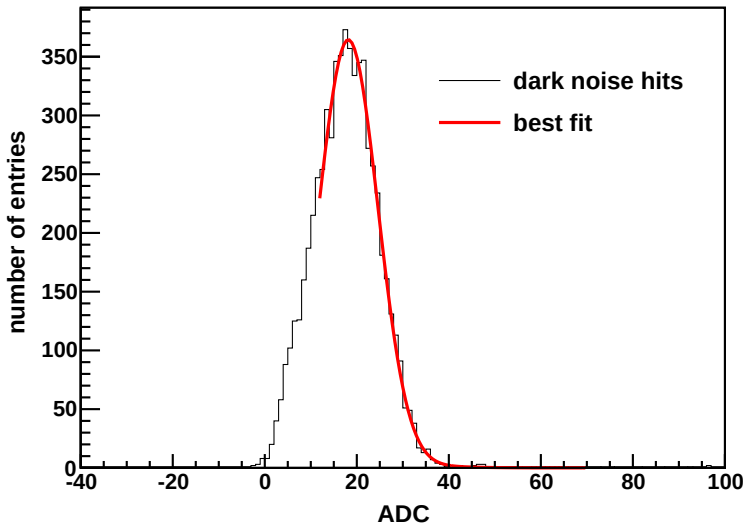

Figure 3.4: Example of fitting Equation 3.4 to a histogram of ADC counts for dark-noise hits in a single channel. From [19].

The results of this fitting procedure were stored in the offline database for use during data production, In cases where a poor fit quality was obtained, a "default" value of 19 ADC/SPE was stored in the database, making it possible to flag problematic channels by scanning for those that have exactly this value in the database. Similarly, excessively small or large fit gains are used in the determination of channel quality, as discussed in Section 3.4.

3.3.1 Hit charge calculation

In the simplest case, the calibrated charge q (in PE) of a hit is related to the raw charge Q (in ADC counts) as

$$q = \frac{Q - Q_{\mathrm{pre}}}{\overline{Q}_{\mathrm{SPE}}}. \tag{3.5}$$

Here, Q_{pre} is the pre-ADC value described in Section 2.2.1 (i.e. the average of the four ADC samples preceding the over-threshold instant of the discriminator), and $\overline{Q}^{\mathrm{SPE}}$ is

the calibration constant retrieved from the offline database for the particular PMT and time period in question. This calculation is fairly reliable as long as a hit is not preceded by any other hits within ∼300 ns.

When hits are more closely spaced in time, however, the calculation is made more complicated by the peculiar behavior[7] of the electronics, as illustrated in Figure 3.5. This peculiarity can be boiled down to two facts. First of all, the peak-finding algorithm always takes the largest sample within a continuous series of above-baseline samples, even when this causes the same peak to be reported for more than one hit (see the first three columns of Figure 3.5). And secondly, the pre-ADC of a hit can be biased by the preceding hit (see the last two columns of Figure 3.5).

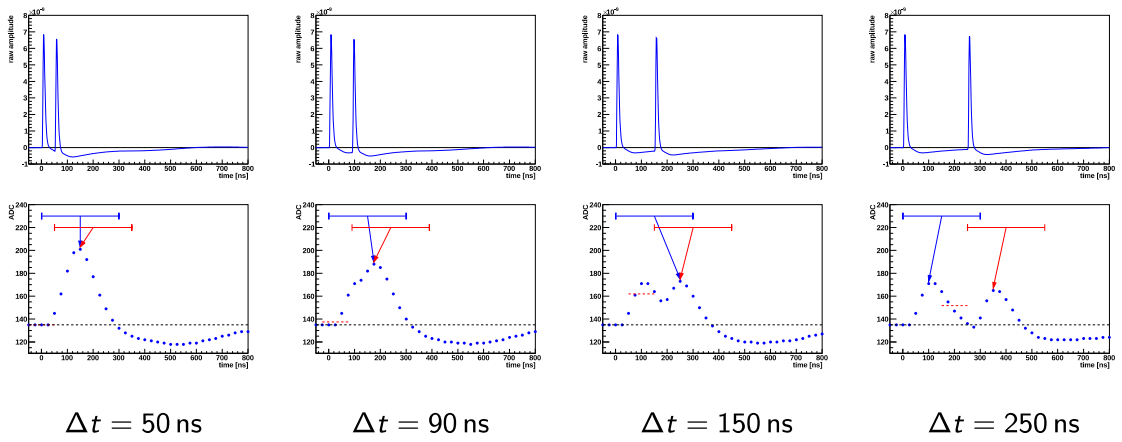

Figure 3.5: An illustration of the different scenarios that can occur when dealing with closely spaced hits. In this case, each hit is one photoelectron. The red dashed lines are the pre-ADC values *for the second hit*. From [27]

Simply using Equation 3.5 for all hits will result in over or underestimating the total charge, depending on the time separation between hits, as illustrated by the green/gray dots in Figure 3.6. For separations of less than ∼100 ns, the peak finder will approximately report the correct *total* charge, but this charge will be assigned to two hits and will thus be double counted. This is shown in the first two columns of Figure 3.5. For separations of ∼100–250 ns, the samples will not reach baseline and thus a single peak value will still be reported twice, but in this case the peak's amplitude is low enough that double-counting it will more closely approximate the total charge than single-counting it. However, the pre-ADC for the second hit is biased by the samples from the first hit, and so this pre-ADC should be discarded in favor of the one from the first hit. Finally, for separations greater than ∼250 ns, the samples will hit the baseline in between the peaks, so two different peak values will be reported, and both should be used, but once again, the pre-ADC for the second

[7] Largely due to the AC-coupling of the PMT signal into the front-end.

hit may be biased by the first hit (whose shaped pulse includes an undershoot that extends for several hundred ns beyond the positive pulse).

On account of this behavior, an algorithm was designed with the goal of minimizing the overall bias in the total charge (as studied under the scenario of two SPE hits) [27]. In what follows, Δt is the time since the previous hit in the same readout:

1. If the hit is the first one in the readout, treat it simply using Equation 3.5.
2. If $\Delta t < 100\,\text{ns}$, set the hit's calibrated charge to zero. In this case the aim is to avoid double-counting the total charge, as illustrated in the first two columns of Figure 3.5.
3. If $\Delta t \geq 100\,\text{ns}$, and the peak value is the same as that of the previous hit, use the pre-ADC of the *previous* hit. This avoids the bias on the second pre-ADC value, as shown in the third column of Figure 3.5.
4. If $\Delta t \geq 100\,\text{ns}$ and a *different* peak values is reported, ignore the hit's pre-ADC, and instead subtract the parameterized tail shape of the previous pulse (as this tail is essentially the "true" pedestal of the hit). This method, dubbed *charge correction*, was developed by the author.

However, studies showed that the overall effect (on resolution and bias of reconstructed energy) of the undershoot correction was negligible, at the cost of considerable operational complexity (in maintaining the parameterization of the pulse shapes). In the end, the algorithm as implemented goes:

1. If the hit is the first one in the readout, again treat it simply using Equation 3.5.
2. If $\Delta t < 100\,\text{ns}$, *and the peak value is the same as that of the previous hit*, set the hit's calibrated charge to zero.
3. If $\Delta t \geq 100\,\text{ns}$, *whether the peak value is the same as or different from the previous one*, use the pre-ADC from the *previous* hit.
4. *(Special case)* If $\Delta t < 100\,\text{ns}$, and the peak value is *different*, treat the hit simply using Equation 3.5. The reasoning for this special case does not seem to be documented, and it likely only applies to the (rare) scenario of two closely spaced sub-SPE hits, for which it may be possible for the ADC to hit baseline in between the two peaks. In this scenario, keeping both hits may produce a more accurate charge measurement.

This algorithm is illustrated in Figure 3.6, which shows how this "piecewise" calculation reduces the overall bias compared to the naive application of Equation 3.5 to all hits. In practice, however, only the *first* hit (within a specified time window) is used in reconstructing the energy of an event. The reasons for this, and the details of the reconstruction, are discussed in Section 4.2.1.

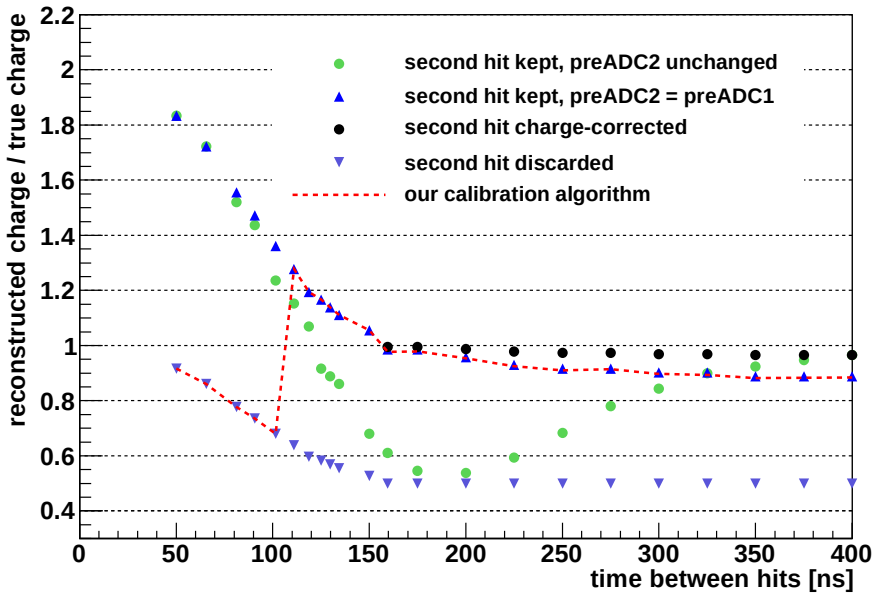

Figure 3.6: Biases induced by different methods of handling closely spaced hits. Modified from [27].

3.4 Channel quality

In order to prevent any biases from being introduced by problematic channels, a channel quality database is consulted during data production, and any tagged channels are excluded from the total event charge (see Section 4.2.1). Bad channels are identified and tagged during the regular "keep-up" processing of each new data file. Four criteria are used for each channel. For ordinary physics runs[8], these conditions are as follows (evaluated once for each data file, corresponding to ∼10 (30) minutes of DAQ time in a near (far) hall AD):

1. The occupancy (i.e., the percentage of events in which the channel recorded at least one hit) must be between 0.1 and 0.65/0.5 (near/far). Too low of an occupancy would imply a dead or dying channel, and too high would indicate a noisy channel.

2. The high voltage must be at least 1200 V. A low HV would indicate a problem with the HV supply, requiring the exclusion of the channel from the charge sum.

[8]Given that calibration runs feature a different population of events compared to physics runs, these runs use alternative acceptable ranges of occupancy and ADC, determined by observation. In all other respects, channel quality determination is the same for these runs.

3. The RMS variation of the HV, during the time span of the data file, must be less than 0.5 V. Anything higher would indicate an unstable HV supply and hence an unstable gain.

4. The baseline-subtracted ADC count, on average, must be between 10 and 45 ADC. This range is based on observations, and its relatively large spread is due to the differences in the average amount of light seen by PMTs near the top and bottom, relative to those near the center.

These four "ingredients", as well as the "good or bad" verdict, are stored in a dedicated channel quality (CQ) database during keep-up production[9]. In turn, an automated batch job takes the verdicts, compresses them into bit fields, and inserts them into the main offline database, from which they are taken during event charge calculation. If, at a later date, a decision is made to revise the cuts used in the channel quality determination, the "ingredients" can be taken from the CQ DB and a new verdict can be made, without any need to revisit the underlying data file. In practice, however, the Daya Bay CQ cuts have proven to be generally robust enough not to require revision. As a crosscheck, human inspection of the gains was periodically performed in order to identify channels that failed to be tagged by this automated procedure. In such rare cases, the database was updated manually to ensure exclusion of the bad channel(s) from the charge calculation.

[9]Keep-up production (KUP) refers to the near-real-time offline processing (calibration and reconstruction, using approximate calibration constants) of newly acquired data. Although KUP data isn't usable for precision physics analysis, due to this use of "stale" calibration constants and the lack of careful validation, it nevertheless serves various important purposes related to data quality, including this evaluation of channel quality.

Chapter 4

Event reconstruction

4.1 Introduction

Chapter 3 discussed the process of taking the raw ADC and TDC values of PMT hits, as measured by the front-end electronics, and converting those values, channel-by-channel, into the more useful quantities of hit charge (in photoelectrons) and hit time (in, e.g., nanoseconds). The next step is to combine information from all of the channels in order to derive properties of the event as a whole, such as the amount of deposited energy and the approximate location of the vertex. This is the purpose of *reconstruction*.

Reconstruction begins with the calculation of the total observed charge (i.e. photoelectron count) by summing hits across all channels, with a correction for the presence of any inactive channels. This *nominal charge* is then converted into *raw energy*, in MeV, according to an energy scale determined using regular (weekly or more) calibrations. At the same time, the distribution of charge across PMTs is used to estimate the location within the AD of the event. The position is then used to apply a *nonuniformity* adjustment to the raw energy, to correct for the position-dependent response of the detector. This gives the *reconstructed energy*, which is used in most subsequent analysis stages. The details of the energy reconstruction (as illustrated by the flow chart in Figure 4.1) are discussed in Section 4.2.

The reconstructed energy should not be regarded as the best estimate of the true energy deposited by the event, given the complexities involved in the nonlinearity of the scintillator and its varying responses to different particle types. Rather, reconstructed energy should be considered a *position-corrected measure of the total observed amount of light*, and hence should be regarded as proportional to the total amount of light produced in the scintillator. Due to the calibration methods used, reconstructed energy *does* agree (by construction) with deposited energy for the 8 MeV gamma-ray cascade from neutron-capture on gadolinium (nGd capture), but this is only a special case.

In parallel with the energy reconstruction, vertex reconstruction is carried out using the spatial distribution of recorded photoelectrons. This process is illustrated by the flow chart in Figure D.6 and detailed in Section D.3. Although the reconstructed vertex has historically not been *directly* used in the nGd oscillation analysis (e.g., in the selection of IBD candidates), vertex information plays an important role in the estimation of certain correlated backgrounds, particularly the so-called AmC background (Section A.3), as well as in alternative oscillation analyses using neutron capture on hydrogen [28]. In Section 9.5, we explore the effects of including a vertex requirement in our IBD selection, demonstrating that our analysis is robust against such modifications.

Daya Bay has developed multiple independent reconstruction algorithms. The two that have been widely used in published results are known as AdSimple and AdScaled. They differ primarily in their calibration procedures, their vertex reconstruction algorithms, and their methods of correcting for nonuniformity. Both give consistent results in the oscillation analysis, and both will be detailed in this chapter, but only AdSimple will be used in our analysis.

4.2 Energy reconstruction

The energy reconstruction begins with the total measured charge, and proceeds through a series of corrections to produce a best estimate of the true deposited energy, as illustrated in Figure 4.1. In the sections that follow, we will be repeatedly referring to various estimated energies:

- **E_{raw}:** The *raw* energy, i.e., the total measured charge from the PMTs, divided by the energy scale.
- **E_{cor}:** The *corrected* energy, calculated from E_{raw} by applying a *time-independent* correction for the geometric nonuniformity of the AD's response (Section 4.2.3).
- **E_{rec}:** The final *reconstructed energy*, obtained from E_{cor} by applying an additional correction for the time-dependence of the geometric nonuniformity[1] This is stored as the output of the AdSimple reconstruction. Two events which occur at different positions/times, but which are otherwise identical, will produce the same E_{rec}. The energy scale is defined such that $E_{rec} = 7.94\,\text{MeV}$ for spallation neutrons captured on gadolinium.
- **E_{vis}:** The *visible* energy, obtained from E_{rec} by applying a correction for the nonlinearity of the electronics (i.e., correcting the charge such that it is directly

[1] Caused by, e.g., the gradual increase in the attenuation length of the scintillator.

proportional to the number of photoelectrons). In physical terms, E_{vis} is proportional to the number of photons produced in the scintillator.[2]

- **$\mathbf{E_{dep}}$:** The *deposited* energy, obtained from E_{vis} by correcting for the (particle-dependent) nonlinearity of the scintillator. This is the final estimate of the actual energy physically deposited.
- **$\mathbf{E_\nu}$:** The *neutrino* energy, calculated from E_{dep} based on the kinematics of the IBD interaction.

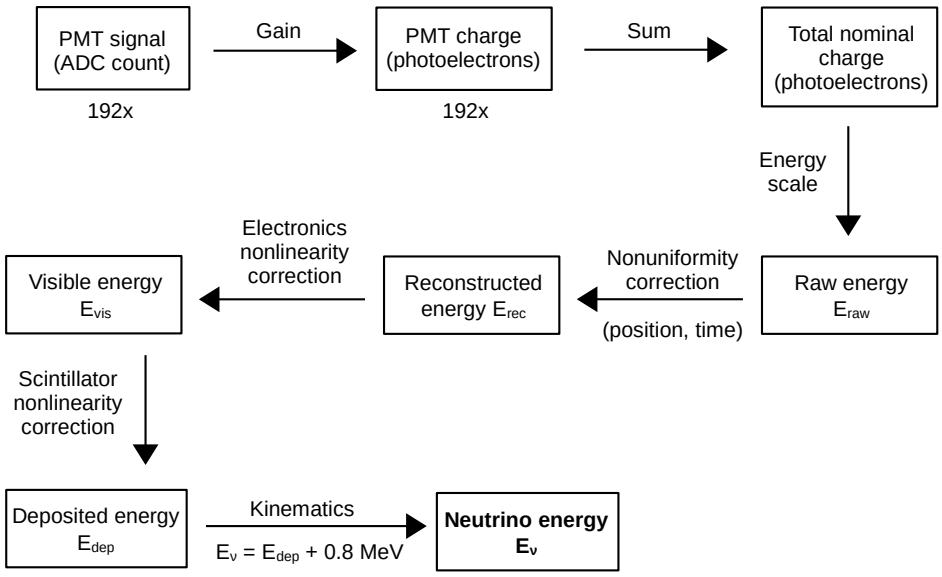

Figure 4.1: Conceptual flowchart of the AdSimple energy reconstruction process. In practice, the electronics and scintillator nonlinearity corrections are applied in a single step, using the correction function described in [29]. An alternative reconstruction, known as AdSimpleNL, corrects for the electronics nonlinearity at the level of individual channels, prior to summing of the charges to obtain the nonlinearity-corrected nominal charge ("NominalChargeNL"). Both methods produce consistent results. Here, we use the more "traditional" AdSimple algorithm.

4.2.1 Event charge determination

[2] In practice, E_{rec} is converted into E_{dep} in a single step using the combined (scintillator + electronics) nonlinearity model described in Section 4.2.4. Hence, E_{vis} never explicitly appears.

4.2.1.1 Hit selection

The first step in the energy reconstruction is to estimate the total *charge*, i.e., number of photoelectrons, observed from the underlying interaction. Here, the main consideration is the choice of hits to include in the sum. Based on the design of the trigger electronics, a trigger will be issued about (1550 ± 50) ns after the accumulation of a sufficient number of hits and/or total charge. Including an additional spread of 50 ns to account for the time-of-flight of the photons, one would infer that a window of around [-1650, -1450] ns would be reasonable.[3] In practice, Daya Bay actually uses a window of [-1650, -1250] ns. The justification for this wider window is related to the properties of the liquid scintillator itself.

When an interaction deposits energy in the LS, various molecular excited states decay stochastically, emitting light in the process. In the Daya Bay LS, the light emission can be accurately modeled with three components: a fast one (∼5 ns time constant), a medium one (∼30 ns), and a slow one (∼150 ns). The time for light to propagate, directly or via reflections, adds a position-dependent delay of a few dozen ns. Altogether, 5% of the PMT hits occur some 50-150 ns after the primary peak [30]. In order to include this "late" light, and thereby hopefully improve the energy resolution, Daya Bay uses the widened hit selection window of [-1650, -1250] ns.

With a window defined for hit selection, the next question is which hits to use from inside this window. Based on the measures discussed in Section 3.3.1 for correcting the biases in closely-spaced hits, in principle every hit should be trustworthy. In practice, hits that arrive within 100 ns of each other will produce a single shaped peak, and hence only the first hit will have a nonzero calibrated charge. Since most primary light hits *do* in fact arrive within 100 ns of each other, there is usually no difference between taking all hits and taking only the first hit. The *default* or *nominal* charge is accordingly defined as *the sum across channels of the earliest hit in the time window of [-1650, -1250] ns (relative to the trigger time).*

The nominal charge will generally account for all of the fast/medium light, but will omit *some* of the slow light *unless there is no fast/medium light seen by the channel.* As such, high-energy events will miss a greater proportion of slow light compared to low-energy events, since in the latter case there will be more channels seeing no fast/medium light. This introduces a degree of nonlinearity in the overall detector response. If, instead, one were to take *all* hits in [-1650, -1250] ns, instead of just the earliest hit, the sum would in principle accurately include all of the components, without the aforementioned nonlinearity. This does not appear to have ever been proposed; the reasons are unknown, but may be related to the fact that this method is more sensitive to the details of the corrections for photon multiplicity per PMT (Section 3.3).[4]

[3] Here, as in previous discussions, the origin of time ($t = 0$) is the instant at which the trigger is issued. Since all of the recorded hits occur prior to the trigger, their time values are negative.

[4] As an alternative to the nominal charge, the *peak charge* is defined as the sum across channels of

4.2.1.2 Active channel correction

At any given time, there may be dead or malfunctioning channels in an AD. As described in Section 3.4, these are recorded in the channel quality (CQ) database according to a number of criteria. If, at the time of a given trigger, a channel is marked as "bad", then its charge is *not* included in the total nominal charge. This, naturally, will result in a downward bias on the total. In principle, the size of the effect depends on the position of the event: The effect is larger if the event is closer to the PMT, and vice versa. In practice, however, Daya Bay uses a simple, position-independent correction of $192/N$, where N is the number of active channels. Given that the Daya Bay ADs almost always have fewer than two bad channels, this correction was found to reliably correct the bias, with negligible impact on the resolution.

4.2.1.3 Summary

In summary, the nominal charge is computed as follows: For every active channel, take the calibrated charge of the earliest hit in the pre-trigger window of [-1650, -1250] ns. Sum these up, and then apply a correction of $192/N$, where N is the number of active channels. In subsequent stages of the energy reconstruction, the nominal charge (in PE) is scaled by a time-dependent energy scale to give the *raw energy* (in MeV), then adjusted by a time- and position-dependent nonuniformity correction to give the *reconstructed energy* and, finally, at the highest levels of analysis, adjusted again to correct for electronics nonlinearity, scintillator nonlinearity, and IBD kinematics to give the *true neutrino energy*. These steps are discussed below.

4.2.2 Energy scale calibration

The nominal charge produced by a given interaction can vary over time due to, for instance, degradation or contamination of the scintillator. Furthermore, for the purpose of physics analysis, we would prefer to speak of the energy (in, e.g., MeV) deposited in the scintillator, rather than the amount of light observed. Accordingly, the object of the energy scale calibration is to fix the definition of a "visible" MeV, and to ensure that any given event will yield the same reconstructed energy in every AD, regardless of changes over time in the behavior of the scintillator.

In what follows, repeated references will be made to the so-called *Crystal Ball (CB) function* [31]. This empirical function was initially developed by the Crystal

the earliest hit in [-1650, -1480] ns. This effectively excludes the late light, mitigating the associated nonlinearity found in the nominal charge. One downside is that slightly less light is included, but late light accounts for only 5% of the total, and the nominal charge misses some of it anyway, so overall, only a couple percent of photons are lost compared to the nominal charge. In any case, it is possible to measure and correct for any nonlinearity inherent in the charge calculation, as discussed in Section 4.2.4. We use the nominal charge, since that is what the nonlinearity has been measured against.

Ball collaboration, which operated a neutral-particle detector (containing an inner spark chamber surrounded by a sphere of scintillating crystals) at SLAC around the early 1980s. The CB function is designed to model "lossy" processes, such as energy deposition in a detector where some energy can escape detection. At Daya Bay, neutron capture on gadolinium provide an example of such a process, as gamma rays may exit the scintillating volume before depositing all of their energy. To account for both fully and partially contained events, the CB function includes a Gaussian "core" and a power-law "tail", respectively:

$$f(x;\alpha,n,\bar{x},\sigma) = N \cdot \begin{cases} \exp(-\frac{(x-\bar{x})^2}{2\sigma^2}), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\bar{x}}{\sigma})^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leqslant -\alpha, \end{cases} \tag{4.1}$$

where

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right), \qquad B = \frac{n}{|\alpha|} - |\alpha|\,.$$

Here, α, n, $\bar{x}$, and σ are fit parameters, and N is a normalization factor. In the case of fitting the energy spectrum of nGd captures, we will be discussing the use of a *double Crystal Ball function*, that is, the sum of two CB functions, one of which fits the 7.937 MeV peak from capture on ^{157}Gd, and the other of which fits the 8.536 MeV peak from ^{155}Gd. Among the isotopes of Gd with significant neutron capture cross sections, these two are the most abundant in natural Gd.

Given that the response of the ADs (i.e. the nominal charge) depends on the type of interaction and is nonlinear with respect to the deposited energy, the energy scale (in charge per MeV) will depend on the choice of interaction used to calibrate the scale. Daya Bay's two main reconstruction algorithms, AdSimple and AdScaled, both define the energy scale such that a neutron capture on gadolinium will yield approximately 8 MeV[5]. However, there are significant differences between the methodology of the two calibrations.

For AdSimple, the calibration uses Gd captures of spallation neutrons produced by high-energy cosmic-ray muons traversing the AD. Since this analysis is based on AdSimple, we give a detailed description of its calibration procedure in the section that follows. One of the advantages of using spallation neutrons is that they are distributed uniformly throughout the target volume, much like IBD neutrons. A disadvantage is that the ensuing energy scale is slightly biased (upward), relative to that of IBD neutrons, due to PMT afterpulsing resulting from the large amount of

[5]More precisely, the energy scale is defined such that the nGd capture spectrum contains two peaks (as fit by a double Crystal Ball function) at reconstructed energies of 7.937 and 8.536 MeV. In principle, the nonlinearity of the detector could mean that if the first peak is fixed at 7.94 MeV, the second might not lie exactly at 8.54 MeV, suggesting that the spacing between the two peaks should be allowed to float in the fit. In practice, however, at such high energies, the degree of nonlinearity (relative to the energy resolution) is insufficient to compromise the fit, and so a fixed peak spacing is used.

charge produced by the parent muon. In the end, however, this is accounted for in the nonlinearity model (Section 4.2.4); as long as the energy scale calibration provides consistency in time, space, and between ADs, it is sufficient.

In comparison to AdSimple, AdScaled uses a significantly different method of calibrating the energy scale. We only discuss it briefly, since AdScaled is not used in this analysis. Essentially, the method is based on using weekly ^{60}Co calibrations to monitor the time variation of the light yield, and occasional ∼40-hour AmC calibrations (which produce nGd captures) to measure the nonlinearity between the ^{60}Co and nGd peaks. Every Friday, the ^{60}Co source is deployed from ACU A to the center of each AD for 10 minutes. From this data, a histogram containing the total nominal charge of each event is extracted. In the vicinity of the ^{60}Co peak, this histogram is fit to a Gaussian plus Crystal Ball function [32]. The nominal charge (at the peak of the fit function) is then multiplied by the ratio between the nGd and ^{60}Co charge peaks[6], as determined by the nearest long AmC run, and this scaled light yield is stored in the database for use by the reconstruction. This method works because the *ratio* of the nGd and ^{60}Co peaks is quite stable, even when the peaks themselves are varying. (Omitting ^{60}Co, and using AmC alone, would avoid the need for this scaling, but the rate of neutrons from the AmC source is insufficient to provide the necessary statistics.) It is worth noting that the resulting energy scale is defined in terms of events at the *center* of the AD, rather than uniformly distributed throughout the GdLS (as in AdSimple). This leads to a consistent ∼5% difference in the energy scale calibration constants between the two algorithms. Essentially, this is only a difference in conventions (i.e., defining the energy scale based on uniformly distributed vs. centered events), which is accounted for at the event-by-event level by the nonuniformity correction, as discussed in Section 4.2.3.

4.2.2.1 AdSimple calibration procedure

The AdSimple energy calibration begins with the selection of a sample of spallation neutron candidates. These are defined based on their proximity in time to a preceding AD muon, where an AD muon is regarded as any event that produces more than 3,000 photoelectrons of nominal charge. Non-muon events are filtered through a simplified cut to remove instrumental backgrounds ("flashers"); specifically, the *ellipse*

[6] This ratio, which is relatively stable, is determined from runs in which the ^{60}Co and ^{241}Am-^{13}C sources are deployed together at the center of the AD. In the nominal charge spectrum from such a run, the ^{60}Co peak is fit, as above, to a Gaussian plus Crystal Ball function. Meanwhile, the nGd peak is fit to a simple Gaussian function [32]; since the nGd captures occur at the detector's center, the gamma-ray leakage tail is very small, allowing the use of a Gaussian function instead of a (double) Crystal Ball function. (Of course, the same reasoning would permit the use of a simple Gaussian function for ^{60}Co as well. The authors of AdScaled nonetheless chose the more complicated function in production, even though they found a simple Gaussian function to work well during testing.) Once both peaks have been fit, the ratio in question is simply the ratio of the two peaks as defined by the best fits.

cut described in Section A.1 is employed (Equation A.4). For any surviving event with a nominal charge of more than 100 PE (roughly 0.6 MeV), the charge (after correcting for any dead PMTs or high-voltage channels, as described in Section 4.2.1.2) is added either to a *signal* histogram, if the time since the previous muon is between 20 and 1000 µs, or to a *background* histogram if Δt is between 1020 and 2000 µs[7]. Given that the characteristic nGd capture time is ∼30 µs, the latter histogram provides a clean sideband measurement of the background spectrum.

These histograms are stored in files that correspond one-to-one with Daya Bay DAQ files (each spanning roughly ten minutes in one hall). The files are processed sequentially, and for each AD, a new energy scale constant is calculated once 10,000 entries have been accumulated in the background[8]-subtracted histogram of spallation neutron charges. The constant is determined by fitting the charge spectrum to a double CB function, whose two components, as discussed previously, correspond to the peaks from neutron capture on ^{155}Gd and ^{157}Gd. The relationship between the two CB functions ("peaks") is constrained as follows:

1. The shape parameters α and n are the same, and constrained to lie within (0, 5) and (0, 1), respectively.

2. The amplitude of the ^{155}Gd peak is constrained to be 0.227 of the ^{157}Gd amplitude, according to the product of the relative abundances (14.80% and 15.65%, respectively) and neutron capture cross sections (60,700 and 257,000 barns, respectively [33]) of the two isotopes.

3. The location of the ^{155}Gd peak is constrained to be 1.0755 of that of the ^{157}Gd peak, based on the total gamma-ray energies of 8.536 and 7.937 MeV emitted after neutron capture on the two isotopes.

4. The two σ (width) parameters are related by the square root of the aforementioned ratio of peak locations.

After the fit is performed (as illustrated in Figure 4.2), the location parameter μ of the first peak (generally between 1200 and 1350 PE) is assumed to correspond to 7.937 MeV (from ^{157}Gd), and so, in this convention, the energy scale constant is simply μ/7.937 PE/MeV. However, due to the PMT afterpulsing that occurs after a high-energy muon event, this value is biased upward compared to the energy scale for the IBD nGd captures. This would not be an issue if the bias were the same size in all halls (since it could then simply be absorbed into a common nonlinearity correction), but because the muon rate and spectrum differ between the halls, so does this bias. Such a systematic difference in energy scales could bias the oscillation fit. Accordingly,

[7] The 20 µs gap between the two windows ensures that both windows are of the same length. Alternatively, a background window of 1000 to 1980 µs could have been used, etc.

[8] From the sideband.

the energy scales are corrected by an AD-specific factor (Table 4.1), empirically determined in order to match the IBD-nGd energy scale in the extrapolated limit of zero muon energy [34]. These corrected energy scales are stored in the database for use by the AdSimple reconstruction. In Figure 4.3 we plot the energy scale for each AD as a function of time.

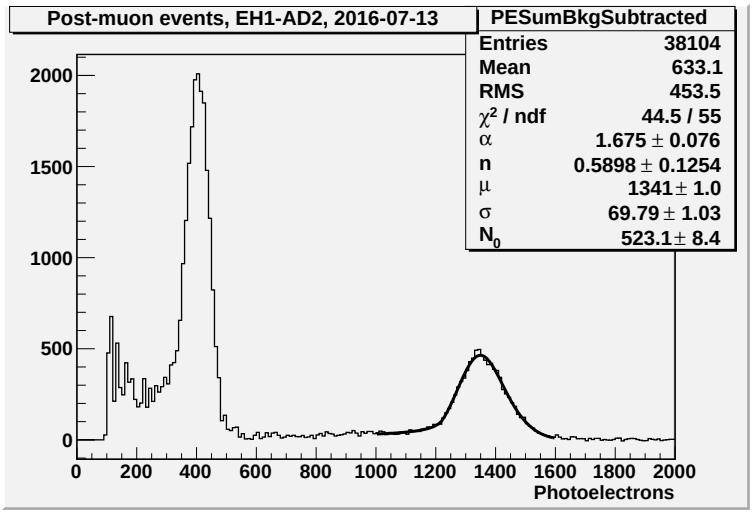

Figure 4.2: Example of a fit used for determining the AdSimple energy scale.

AD	Correction
EH1-AD1	0.9927
EH1-AD2	0.9934
EH2-AD1	0.9921
EH2-AD2	0.9922
EH3-AD1	0.9901
EH3-AD2	0.9904
EH3-AD3	0.9899
EH3-AD4	0.9895

Table 4.1: Correction factors ($E_{\mathrm{IBD}}/E_{\mathrm{spall}}$) relating the energies of nGd captures from IBDs to those from spallation neutrons [34].

4.2.3 Nonuniformity correction

Given the nominal charge Q and the energy scale S, we define the *raw visible energy* E_{raw} as simply Q/S. By construction, E_{raw} is anchored to zero and to 7.937 MeV (for

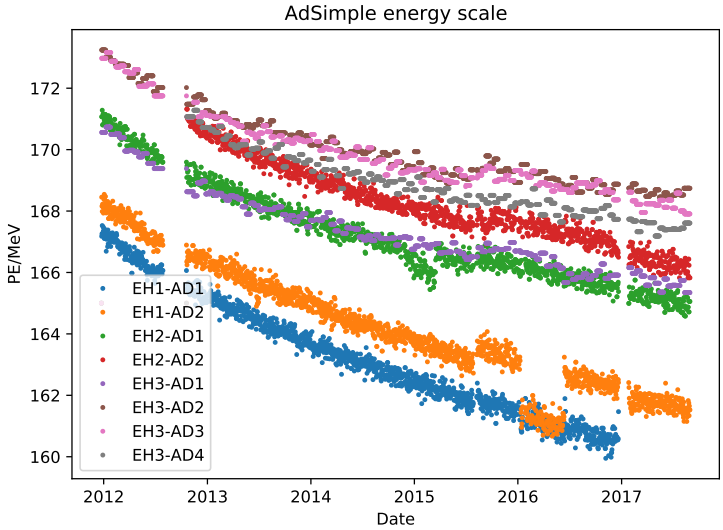

Figure 4.3: AdSimple energy scale in each AD as a function of time. Deviations from the trend (most notable in EH1-AD2) were caused by application of the incorrect HV (see Figure 3.3); although the effects of this were canceled to first-order by the gain calibration, some channels with low gain were often unable to exceed the threshold of the discriminator, leading to an effective decrease in the energy scale. This decrease had no significant detrimental impact on the energy reconstruction.

nGd captures uniformly distributed in the GdLS). At other energies, E_{raw} is biased due to the nonlinearity of the scintillator and electronics. The correction for this nonlinearity is discussed in Section 4.2.4. However, since the light collection efficiency varies as a function of position within the AD, we must first correct for this geometric *nonuniformity*. Otherwise, the energy resolution would be degraded. Fortunately, this nonuniformity can be measured and corrected for.

In the case of AdSimple, the current nonuniformity map was produced from three years of data by dividing the AD into pixels (10 in R^2 $\times$ 10 in Z) and selecting spallation neutron captures on both gadolinium and hydrogen within each pixel. For each pixel i within the GdLS, the nGd spectra was fit to a double Crystal Ball function (along with an additional exponential tail to improve fit quality), and the location of the peak was divided by the mean nGd peak among all N^{GdLS} GdLS pixels, giving the correction factor f_i^{GdLS}:

$$\begin{aligned} \overline{E}_{\text{raw}}^{\text{nGd}} &= \frac{1}{N^{\text{GdLS}}} \sum_{i}^{\text{GdLS}} E_{\text{raw},i}^{\text{nGd}} \\ f_i^{\text{GdLS}} &= \frac{E_{\text{raw},i}^{\text{nGd}}}{\overline{E}^{\text{nGd}}} \end{aligned} \tag{4.2}$$

(This choice of denominator reflects the fact that the energy scale is determined using events uniformly distributed within the GdLS). Meanwhile, for each LS pixel, the nH peak was fit to the "Daya Bay Function" (a specific case of the general *Calorimeter Function* developed by members of the collaboration [35]):

$$\begin{aligned} f_{\mathrm{DYB}}(E) = N_{\mathrm{peak}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(E-\mu)^2}{2\sigma^2}} \\ + N_{\mathrm{tail}} \cdot \left\{ \frac{\lambda}{2} e^{\frac{\sigma^2\lambda^2}{2}} e^{\lambda E} \left[\mathrm{erf}\left(\frac{\mu - (E+\sigma^2\lambda)}{\sqrt{2}\sigma} \right) - \mathrm{erf}\left(\frac{0 - (E+\sigma^2\lambda)}{\sqrt{2}\sigma} \right) \right] \right. \\ \left. + \frac{a}{2} \left[\mathrm{erf}\left(\frac{\mu - E}{\sigma\sqrt{2}} \right) - \mathrm{erf}\left(\frac{0 - E}{\sigma\sqrt{2}} \right) \right] \right\}, \end{aligned} \tag{4.3}$$

where $E \equiv E_{\mathrm{raw}}$. This function has six parameters: The normalizations N_{peak} and N_{tail}, the peak location μ, the peak width σ, and the two tail shape parameters λ and a. For a given pixel i, the fitted peak μ was divided by the mean nH peak among all *GdLS* pixels [36, p. 20] to obtain the pixel's nonuniformity factor:

$$\begin{aligned} \overline{E}^{\mathrm{nH}}_{\mathrm{raw}} &= \frac{1}{N^{\mathrm{GdLS}}} \sum_{i}^{\mathrm{GdLS}} E^{\mathrm{nH}}_{\mathrm{raw},i} \\ f_i^{\mathrm{LS}} &= \frac{E^{\mathrm{nH}}_{\mathrm{raw},i}}{\overline{E}^{\mathrm{nH}}_{\mathrm{raw}}} \end{aligned} \tag{4.4}$$

Although it may seem strange that the GdLS-wide average is used for the nH-based correction (which only gets applied in the LS), this choice is in fact essential to the validity of the correction. As was discussed earlier, spallation neutron nGd captures are used for defining the AdSimple energy scale, and these events occur only in the GdLS (where their distribution is essentially uniform). Since each pixel's correction factor represents a correction to the energy scale for events inside that pixel, the pixel must be compared to the region used for defining the energy scale, that is, the GdLS.

Each pixel is finally assigned its correction factor based on whether it lies in the GdLS or the LS. For those pixels that contain the boundary of the IAV, the average of the two factors is used:

$$f_i = \begin{cases} f_i^{\mathrm{GdLS}}, & \text{inside IAV}, \\ f_i^{\mathrm{LS}}, & \text{outside IAV}, \\ (f^{\mathrm{GdLS}} + f^{\mathrm{LS}})/2, & \text{IAV boundary}. \end{cases} \tag{4.5}$$

E_{cor} is then obtained by applying the correction factor:

$$E_{\mathrm{cor}} = f \cdot E_{\mathrm{raw}}, \tag{4.6}$$

where f is calculated by performing bilinear interpolation between the four values of f_i located nearest to the reconstructed position of the event (Section D.3).

An additional complication is the fact that the nonuniformity of the AD changes over time (along with the energy scale), presumably because decreases in the attenuation length of the scintillator will disproportionately affect events located near the edge of the AD, compared with events near the center. This behavior was found to be consistent between all ADs, and adequately captured by the simple analytical expression [37, p. 16]:

$$\frac{\Delta E_{\text{cor}}}{E_{\text{cor},t_0}} = (a + bR^2)\,(t - t_0), \tag{4.7}$$

where R is the radial coordinate (in meters) of the event, t is the time of the event (in units of years, relative to an atbitrary reference point), t_0 is the (livetime–weighted) average time for the dataset used in constructing the static nonuniformity corrections f_i, and

$$a = -0.00149 \pm 0.00030,$$
$$b = 0.00109 \pm 0.00013,$$

Computing the final reconstructed energy E_{rec} from E_{cor} then involves simply undoing this energy shift:

$$E_{\text{rec}} = \left[1 - (a + bR^2)(t - t_0)\right] \cdot E_{\text{cor}} \tag{4.8}$$

After applying the full (static $\times$ time-dependent) nonuniformity correction, the resulting reconstructed energy E_{rec} was stored in the processed data file for use in analysis.

4.2.4 Nonlinearity correction

Ideally, E_{rec} would be directly proportional to the true deposited energy, across all energies. Unfortunately, this is not the case, due to nonlinear effects produced both by the scintillator and by the electronics.

Within the scintillator, there are two primary sources of nonlinearity: *quenching* [38] and *Cherenkov radiation* [39]. Quenching occurs when the local ionization density is high, allowing fluorescently excited molecules to be "quenched" by excited neighbors, preventing light emission. Ionization density is highest when a particle is moving slowly (especially when it is near the end of its range), so, for a low-energy particle, a greater fraction of light will be quenched in comparison to a higher-energy particle, leading to nonlinear light emission as a function of energy.

Meanwhile, the production of Cherenkov light is a complicated and nonlinear function of a particle's initial energy; below the *Cherenkov threshold* E_{thr}, there is no Cherenkov emission at all. The Cherenkov threshold is the energy at which a particle's velocity $\beta_{\text{thr}}c$ is equal to the speed of light in the scintillator, c/n:

$$E_{\text{thr}} = \gamma_{\text{thr}} m = \frac{1}{\sqrt{1 - \beta_{\text{thr}}^2}} m = \frac{1}{\sqrt{1 - n^{-2}}} m \tag{4.9}$$

The Daya Bay scintillator has an index of refraction of approximately 1.5 [40], giving, for positrons, a threshold of 0.7 MeV; adding the additional ∼0.5 MeV from the annihilated electron, this implies that Cherenkov light becomes relevant for reconstructed energies above ∼ 1.2 MeV, i.e., the vast majority of the IBD positron spectrum.

As for the electronics, the nonlinearity arises largely from the fact that only the *first* hit is used from each PMT in the charge calculation, leading to the potential exclusion of delayed light from the scintillator. All three of these effects (quenching, Cherenkov emission, and electronics nonlinearity) are detailed in Section D.2. Here we simply present the total nonlinearity model used in this analysis, as illustrated in Figure 4.4. This model was developed by the Collaboration's energy-response experts and numerically evaluated by them at 458 evenly-spaced values of the true positron energy, from 1.022 to 11.99 MeV, to predict the corresponding values of the prompt reconstructed energy. The results were distributed within the collaboration in the form of a `ROOT TGraph` containing the 458 points. During the oscillation fit, for each IBD event, this `TGraph` is evaluated (using linear interpolation) in order to convert the prompt reconstructed energy into the corresponding antineutrino energy (see Figure 8.1 and Section C.3.1). For those rare events in which the true energy lies below 1.022 MeV or above 11.99 MeV, the `TGraph` is evaluated at the nearest endpoint.

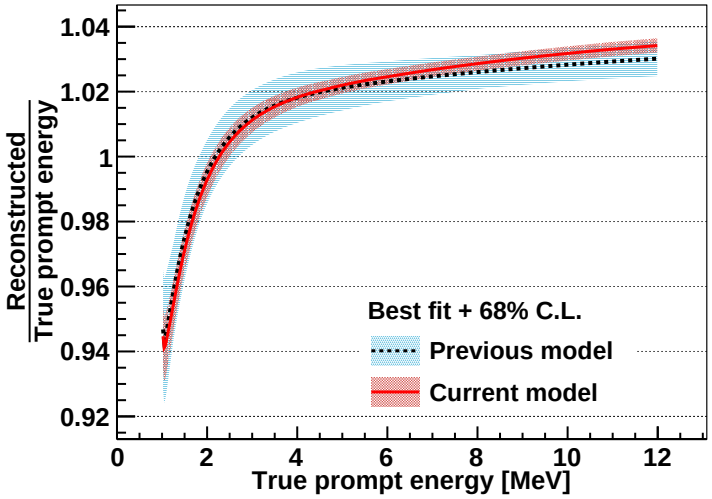

Figure 4.4: Best-fit nonlinearity model for positrons. From [29].

Chapter 5

Event selection

From the sequence of reconstructed triggers in the ADs, we are primarily interested in extracting IBD candidates, in order to obtain the antineutrino rate and spectrum. The tight time correlation of the prompt and delayed triggers, as well as the relatively high 8 MeV energy released from the neutron capture on gadolinium (nGd capture), enable the extraction of a ∼98% pure sample of IBDs, from which the independently estimated backgrounds can then be subtracted.

Aside from the IBD selection, this analysis also employs an extraction of *singles,* that is, those events that produce only a single trigger, uncorrelated in time with any others. The purpose of the singles sample is to enable determination, firstly, the rate and spectrum of backgrounds produced by accidental coincidences, and secondly, the efficiency of the multiplicity cut (discussed in Section 5.2.2.2).

Both selections are implemented using a two-stage approach. In the first stage, the *pre-selection,* processed (i.e., calibrated and reconstructed) Daya Bay data files are scanned, unimportant events are ignored, and of the remaining events, only the minimum required data fields are stored in the output. This process reduces ∼600,000 input files (each representing ∼10 minutes of data collection), totaling some 600 TB, down to about 5,500 files (each representing one hall × day), totaling one terabyte. In the second stage, the *final selection,* the full set of selection criteria are applied to the pre-selected data, producing samples of IBDs and singles for use in the oscillation analysis. This two-stage approach significantly reduces the amount of time needed to generate new IBD/singles samples after modifying the selection criteria, since the pre-selection does not need to be re-run. When the NERSC cluster is not under severe disk I/O load, the two-stage approach provides a speed improvement of 3 to 4; during disk overload, the improvement can be greater still.

5.1 Data production

The IBD and singles selection do not proceed directly from the raw Daya Bay data files, but instead from the processed files produced during *data production*, in which the previously-discussed steps of calibration and reconstruction are applied. We briefly describe the production process here. For the data sample used in this analysis, production was largely conducted and monitored by the author. This was the first time that production was performed on the high-performance computing (HPC) systems at the National Energy Research Scientific Computing Center (NERSC). To achieve this milestone, nontrivial changes to the production infrastructure were required, and new scripts and tools had to be developed.

Production begins with the raw data files produced by Daya Bay's onsite data acquisition (DAQ) software. We will refer to these as "DAQ files". The DAQ files, which use a specialized binary file format, contain the ADC count (and pedestal) and TDC count for each PMT hit in a given trigger, as well as the global details of each trigger, such as the detector, timestamp, and trigger type. Although each hall as its own DAQ system, the subdetectors within each hall (ADs, WPs, and RPCs) are all read out by the same DAQ. Each DAQ file thus contains all of the data generated by a given hall within a given time period. The DAQs are configured to output a new file after approximately 1 GB of data has been accumulated, corresponding to about 10 minutes in the near halls and 20 minutes in EH3. Each DAQ file is identified by its *run number*[1] and *file number* (starting at 1 and increasing sequentially within each run).

During production, each DAQ file is independently processed by Daya Bay's offline software framework, NuWa, which is based on the Gaudi [41] framework used by various other experiments, including the LHC's ATLAS. NuWa was developed by extending Gaudi with various algorithms, services, and tools that serve such purposes as parsing the DAQ file format, outputting the ROOT file format, and performing calibration, reconstruction, and event tagging. For each DAQ file, NuWa produces a corresponding ROOT file containing, for each trigger the calibrated charge and time of each PMT hit, the total nominal charge, the reconstructed energy and vertex, and numerous other useful high-level quantities. In addition, the output file contains tags that identify potential muons, coincidence pairs, isolated single events, etc., although these tags are generally not employed in this analysis.[2]

Previous data productions ran on NERSC's older (non-HPC) Parallel Distributed Systems Facility (PDSF), on which each file was processed by a separate batch job (using Univa Grid Engine, now Altair Grid Engine [42]) on the cluster. However, for the HPC systems (Cori and Edison), a considerably more complex scheme was necessary

[1] Run numbering is sequential, and all three halls share the same global counter; that is, there is no possibility of two runs in different halls sharing the same run number.

[2] However, tagged muon-like events are used in our characterization of the ^{9}Li/^{8}He background, as discussed in Section 6.3.2.

in order to harness the full performance potential of the facilities. In particular, due to the large number of parallel jobs, it was no longer possible to point them all toward a single instance of the Daya Bay offline database (containing calibration constants, etc.), as doing so was guaranteed to overload the database. Furthermore, effective use of the HPC system required using long-running jobs in which each allocated core would need to process multiple files sequentially, a significant deviation from the previous technique of pre-assigning a single DAQ file to each job.

Accordingly, for the HPC systems (based on Slurm [43] instead of Univa Grid Engine), the production system was completely overhauled. First, the list of DAQ files was loaded into a RabbitMQ [44] queue. Then, instead of having each job process a single file on a single core, with a single central database, we opted to submit Slurm jobs which each requested a few dozen compute nodes (with the exact number depending on the particular system used). Using Shifter container technology [45], a clone of the offline database was provisioned (on one node) for each job. On the remaining nodes for each job, a few dozen parallel *pilot tasks* were launched (again using Shifter to provide the correct software environment for NuWa), limited by the available amount of memory.[3] Each pilot task then repeatedly queried the RabbitMQ queue for DAQ files to process, continuing until the job's time limit approached.

This system was successful in scaling up Daya Bay's data production in order to harness the power of NERSC's HPC facilities. Using a small fraction of the total system capacity, this data production was completed in approximately a week, with a shorter turnaround easily obtainable had a larger allocation been given. The previous production system would have taken more than a month. The infrastructure developed for this production continues to be used for high-throughput processing of Daya Bay's data.

Once each file has been processed and validated, the resulting ROOT file is stored and cataloged for use in analysis. It is these files that we use in selecting IBDs and singles for this analysis.

5.2 IBD selection

We begin by discussing the IBD selection. The singles selection proceeds similarly, with minor differences in the final steps, as discussed in Section 5.3.

5.2.1 Pre-selection

[3] Due to NuWa's high memory requirements, memory was the bottleneck, rather than the number of CPU cores.

5.2.1.1 Input data

The processed (i.e. reconstructed) Daya Bay files (in ROOT format) serve as the input to the pre-selection. Although these files contain basic taggings of muon-like events and coincidence clusters, this information is not used here; our event selection is a completely independent implementation.

Two ROOT TTrees are read in parallel: the `AdSimple` tree, which contains the reconstructed energy, and the `CalibStats` tree, which contains the nominal charge (used, in some cases, for pre-muon[4] identification in the ADs), the number of hit PMTs (ignored for AD events, but used for identifying pre-muons in the water pool), and various quantities that are used for removing instrumental backgrounds. Both trees are of the same length, with one entry per trigger, including triggers in the water pools and RPCs (for which AD-specific quantities are left blank). Being the same length, the two trees can be "friended" together (in ROOT parlance) and scanned as one. Other fields loaded from this combined tree are the detector ID, the trigger type, the trigger ID, and the trigger time.

5.2.1.2 Trigger type restriction

The very first criterion applied in the pre-selection is a restriction on the type of triggers saved. In particular, six types of triggers are excluded: manual triggers, cross triggers (issued when a trigger occurs in a different detector subsystem), periodic and random triggers (used for sub-threshold and noise measurements), pedestal triggers (used in special runs to measure the ADC pedestals), and calibration triggers (issued, for instance, when a calibration LED is pulsed). The remaining events consist solely of `NHit` and `ESum` triggers, issued (respectively) when the number of hit channels or the total measured charge, within the previous µs, is above a configured threshold. The NHit threshold is 45, while the ESum threshold (in units roughly, but not exactly, analogous to the nominal charge) is 100, 107, and 130 in EH1, EH2, and EH3 (corresponding roughly to 0.63, 0.67, and 0.81 MeV, respectively[5].). These thresholds were determined during commissioning in order to ensure perfect trigger efficiency at the IBD prompt energy threshold of 0.7 MeV, without overwhelming the trigger rate.

5.2.1.3 Pre-muon selection

After removing unwanted trigger types, the next step in the pre-selection is to extract muon-like events to an output tree of *pre-muons.* The actual definition of a muon (for the purpose of applying the muon veto) is applied in stage two; the pre-muon criteria

[4]The meaning of *pre-muon* is explained below, in the section labeled *Pre-muon selection.*

[5]Even though the EH3 ESum threshold may appear to lack perfect efficiency at 0.7 MeV, the NHit condition, along with the subtle differences between the (hardware) ESum and the (software) nominal charge, ensure near-100% efficiency at 0.7 MeV. Indeed, the NHit trigger alone reaches 100% efficiency at 0.5 MeV [46]

are thus designed to be loose enough to encompass any final muon definition. An event in the water pool is considered a pre-muon if the number of hit PMTs is more than 12,[6]and an AD pre-muon is defined as having a nominal charge of more than 3,000 photoelectrons. Each pre-muon was stored with its trigger time and its *strength*, either its nominal charge (for AD triggers) or its hit multiplicity (for WP triggers).

5.2.1.4 Flasher removal

Among the remaining non-pre-muon events, some are *flashers,* instrumental backgrounds produced by arcing within the base of the PMTs. The light from these arcs can illuminate the detector, resulting in a trigger. As detailed in Section A.1, these events can be easily distinguished by their characteristic conical pattern of light emission. The three cuts described in Section A.1 (the *ellipse, PSD,* and *2" PMT* cuts) are used to identify and remove flashers from the output.

5.2.1.5 Saving and merging

Finally, the non-pre-muon, non-flasher triggers are saved in their own tree (one for each AD), separate from the pre-muon tree. For each event, the run number, file number, trigger time, trigger number, and energy are saved. A minimum reconstructed energy of 0.7 MeV (the threshold for prompt-like triggers) is applied here to further reduce the data volume, since lower-energy events are not considered in this analysis.

The pre-selection files are initially produced in one-to-one correspondence with the reconstructed DAQ files, resulting in ∼600,000 small files. To reduce the file count and improve IO performance, these files are merged (using ROOT's `hadd`) into files that each represent one calendar day in one hall, a total of ∼5,500 files. Finally, these daily files are pre-loaded into a solid-state device *burst buffer* at the National Energy Research Scientific Computing Center (NERSC), to ensure that the performance of the final selection will be minimally impacted by disk load conditions at the facility.

5.2.2 Final selection

The specific thresholds for the cuts described in this section are *nominal*; they are defined somewhat arbitrarily based on qualitative observations and notions of reasonableness, intended to give a satisfactory ratio of signal to background. Later, in Chapter 9, we study the effects of varying these cuts, with the aim of jointly optimizing both the uncertainty on the oscillation parameters as well as the stability of the fit with respect to variations in the cuts. Doing so will eliminate the arbitrariness inherent in the cuts described here.

[6]Any PMT with at least one hit recorded in the readout, regardless of location in time relative to the trigger, is included in the `CalibStats` count of hit PMTs, for both the water pools and the ADs, although we do not use the hit count in the latter case.

5.2.2.1 Muon veto

When a muon passes through or near the AD, it can produce triggers in the aftermath. These can include instrumentally-induced triggers (caused by PMT afterpulsing and electronics ringing in the 20 µs following the muon [47]), as well as physical events. The physical events can include spallation neutrons, whose thermalization and subsequent capture can mimic an IBD, isotopes such as ^{9}Li, and ^{8}He, which produce neutrons when they decay, and various uncorrelated decays that can form accidental IBD-like pairs.

For this reason, it is essential to veto the time period immediately following a muon. Although the mean neutron capture time in the GdLS is only 28 µs, neutrons can be produced outside the GdLS and slowly diffuse into it, necessitating a significantly longer veto window. In the case of muons passing through the water pools ("WP muons"), a veto time of 600 µs was shown to effectively remove all such neutrons. Only relatively energetic neutrons, i.e. fast neutrons, would have the ability to reach the GdLS; diffusion from the WP to the GdLS, meanwhile, is not a significant possibility for thermal neutrons produced by WP muons. Meanwhile, for muons passing through the ADs ("AD muons"), neutrons *can* diffuse slowly from the LS or mineral oil, leading to the requirement of a veto window closer to a millisecond. The nominal window for this case is 1.4 ms, a factor of ∼7 larger than the mean neutron capture time (mainly on hydrogen) in the LS region. AD muons are nominally defined as those triggers having a nominal charge of at least 3,000 photoelectrons.

Muons that deposit an especially high amount of energy in the AD are termed *shower* muons. Compared to lower-energy (i.e., minimum ionizing) muons, shower muons have a much higher probability of producing the two cosmogenic isotopes ^{9}Li and ^{8}Be, discussed further in Section 6.3. These isotopes undergo beta decay, producing a prompt-like trigger, and then break up to produce neutrons which are captured, producing a delayed-like trigger. Given that the lifetime of these isotopes is of order 100 ms, the ordinary AD muon veto of around one millisecond would fail to significantly reduce these backgrounds. Accordingly, a much longer veto window is needed after a shower muon. The nominal window is 1 s, with shower muons defined as having an energy of at least 2 GeV. At this threshold energy, the rate of ^{9}Li/^{8}He production is low enough to avoid backgrounds from sub-threshold muons, while the rate of such muons themselves is also low enough to avoid too large of a loss in effective detector livetime. These qualitative statements are refined more quantitatively in Chapter 9, where we explore the effects of modifying the muon veto thresholds and windows.

An additional veto window is applied, spanning 2 µs *before* each muon, common to the WP, AD, and shower muons. Given that trigger latencies can vary, it is possible for a WP muon trigger to receive a trigger timestamp that comes *after* the timestamps for muon-induced events in the AD. The 2 µs pre-veto eliminates this possibility. There is no particular need to veto the 2 µs preceding an *AD* muon, but the original

LBNL IBD selection did so anyway, and we honor its legacy by following suit.[7]

During application of the muon veto, the total amount of vetoed time is tracked, accounting for overlaps. This value is used in order to calculate the muon veto efficiency, determined (on a daily basis) simply as the ratio of unvetoed DAQ livetime to total DAQ livetime. It should be noted that the muon veto is only applied to the *delayed* trigger of a coincidence pair, and this is what enables the efficiency to be calculated so simply. Otherwise, we would require a complex calculation involving the prompt-delayed time distribution and the muon rate. Given (as described below) that the maximum time difference between the prompt and delayed event is 200 µs, it is possible for the prompt trigger to lie 200 µs before the end of the veto window. The windows are thus made large enough to ensure that this time period is free of muon-correlated activity. An illustration of the three muon vetoes is provided in Figure 5.1.

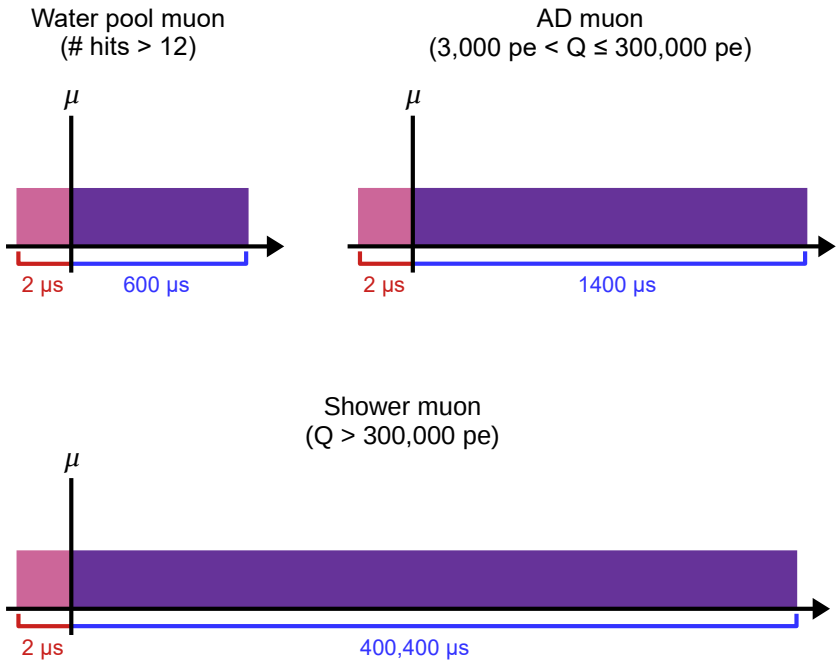

Figure 5.1: Illustration of the muon veto scheme. The purple (pink) shaded region is the vetoed time after (before) the muon. Not to scale.

[7] As discussed in Chapter 7, an enlarged pre-veto of 200 µs provides a more rigorous mathematical decoupling of the muon veto and multiplicity cut efficiencies, but the practical advantage of this is negligible.

5.2.2.2 Pair selection

A potential IBD candidate may be lurking whenever there is a non-vetoed delayed-like trigger (i.e., one lying between 6 and 12 MeV). Specifically, the following conditions (known as the *decoupled*[8] *multiplicity cut*, or DMC) determine the existence of an IBD candidate (Figure 5.2):

1. There is a prompt-like (i.e. 0.7–12 MeV) trigger between 1 and 200 µs before this delayed-like trigger
2. There are no other triggers of more than 0.7 MeV between 1 and 400 µs before this delayed-like trigger.[9]
3. There are no delayed-like triggers within 200 µs after this one.

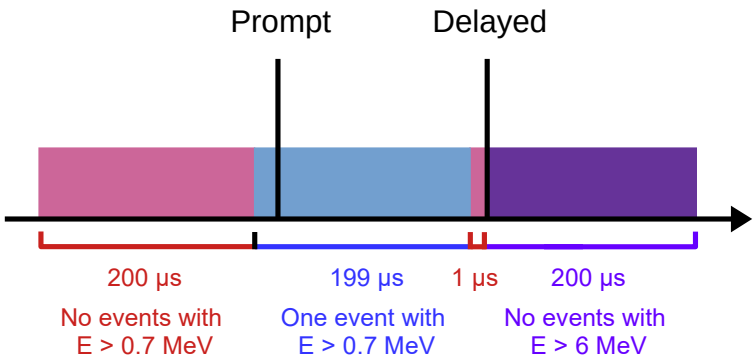

Figure 5.2: Illustration of the decoupled multiplicity cut. The blue region must contain exactly one prompt-like event, the pink regions must contain no prompt-like events, and the purple region must contain no delayed-like events. Not to scale.

The purpose of the latter two conditions is to avoid the ambiguity that can arise, for instance, in the contrived example of three 7 MeV events spaced 100 µs apart. Here there are three possible ways to form a pair. There are other possible ways to define cuts that would avoid this ambiguity, for instance, by defining "empty" windows relative to the prompt trigger, but the DMC allows for a simple calculation of the efficiency as well as an avoidance of correlations with the muon veto efficiency, as described in Chapter 7.

[8] The meaning of "decoupled" is explained further in Chapter 7.

[9] Originally, this condition was framed in terms of *prompt-like*, rather than >0.7 MeV triggers, but the permitted "extra" events above 12 MeV (i.e. low-energy muons) were found to be correlated with backgrounds consisting of pairs of neutron captures. The modified condition eliminates this background.

IBD candidates that pass the DMC are stored in an output tree containing the run and file number, the prompt-delayed time difference, and the IDs and energies of the two triggers.

5.2.2.3 Merging and post-processing

After the ∼5,500 hall-daily files have been fully produced, they are merged with `hadd` into nine files, the product of three halls and three periods (the 6AD, 8AD, and 7AD periods, whose names reference the number of operating ADs). The splitting into periods is done for convenience in preparing the input files required by the fitter, which expects separate files for each period. The fitter's input files, which consist of the IBD spectra, the accidentals spectra, and textual tables of rates, efficiencies, backgrounds, uncertainties, etc., are prepared by a simple script from these nine hall-period files.

5.2.2.4 Extracted prompt spectra

Altogether, these steps produce a set of IBD candidates (including backgrounds) for each AD. The prompt energies of these candidates are then used to generate the prompt spectra (in the form of histograms) Figure 5.3 shows the total prompt spectrum for each hall obtained using the nominal IBD selection cuts. Although backgrounds have not been subtracted, the total relative background rate at Daya Bay is only 1–2%, so these spectra provide a reasonable measurement of the true IBD spectral shape, and will be referenced for this purpose throughout this text.

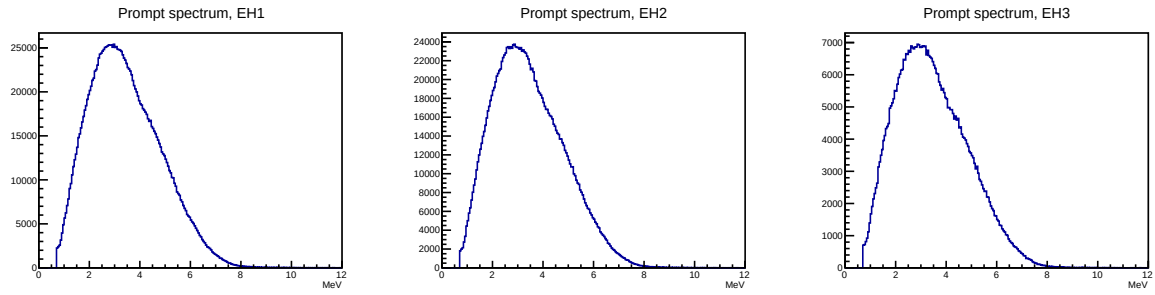

Figure 5.3: Reconstructed prompt energy of the nominal set of IBD candidates for each hall.

5.3 Singles selection

The singles selection proceeds in a similar manner to the IBD selection, except that pairs are no longer being selected. Instead, when a non-muon trigger of at least 0.7 MeV is found, an *isolation cut* is applied, eliminating those events for which

another 0.7 MeV trigger lies within specified windows before and after the "singles candidate". As implemented, the windows used are 400 µs before the trigger and 200 µs after, chosen to resemble the DMC as closely as possible. In principle, the exact sizes of these windows should not matter (provided they are wide enough to eliminate correlated multiplets), as the efficiency of the isolation cut is corrected for in calculating the DMC efficiency and accidental background rate. In practice, due to correlated low-energy processes such as alpha-alpha and beta-alpha cascades (from, e.g., Bi-Po decays in the decay chains of actinides, as considered in Section A.4), the singles sample does not fully consists of "true" singles, so the choice of DMC-like time windows is a naive attempt to minimize any resulting biases. Ultimately, the uncertainty of the accidental background rate is inflated beyond the statistical uncertainty in order to account for this problem.

5.4 Data quality

In order to avoid potential biases on the extracted oscillation parameters, it is essential that data only be used from periods when the detectors are behaving as designed. From time to time, hardware might malfunction, runs might be misconfigured, or calibrations may be erroneous, resulting in questionable data quality. Accordingly, Daya Bay features a comprehensive program of data quality monitoring, review, and record-keeping. These activities, carried out by the Data Quality Working Group (DQWG), culminate in the publication of "good run lists" (technically, good *file* lists) which specify those data files that are suitable for use in physics analysis.

Daya Bay's first line of defense against bad data is made up of "shifters", ordinary collaborators carrying out their obligation to occasionally take 8-hour shifts monitoring the experiment. At any given time, there is at least one shifter on duty, who periodically carries out a checklist to verify that the experiment is operating normally. This procedure includes checking for any alarms from the Detector Control System, which monitors environmental parameters (temperature, humidity, etc.), liquid levels, PMT high voltages, and various mechanical and electronic systems. Aside from the alarms, a number of plots are also checked by the shifter. These plots show various quantities (such as PMT hit rates and average pulse sizes, trigger rates and trigger types, and overall data rates), comparing the current values to those from "normal" historical data.

If any possible issues are noticed, the shifter records a note in the logbook and notifies the designated experts for the subsystem in question. The experts may then contact the DQWG to report a potential data quality issue. Alternatively, the shifter may submit a report directly to the DQWG either via email or a web form. In addition to receiving such reports, the DQWG also monitors the shift logbook to see whether the shifters have noted anything worthy of further investigation.

Data quality records, such as the aforementioned reports from shifters, are maintained in a Data Quality Database (DQDB). The DQDB also stores the quality rating (i.e., good, bad, or to-be-determined) of each data file, along with a reference to a comment record that describes the reason why a file has been "tagged" as bad. Finally, the DQDB stores various physics-driven metrics for each file (such as event rates and peak energies), which are utilized in the procedures carried about by the DQ shifter, as described later.

The DQ metrics are calculated and recorded at IHEP during Keep Up Production (KUP), in which Daya Bay's data production software (which carries out calibration, reconstruction, tagging, etc.) is automatically run on each new raw data file received from onsite. During KUP, calibration is performed using the "old" constants from the end of the previous data production, so the results are not suitable for precise analysis, but they suffice for monitoring detector stability and sanity.

For each detector, the KUP software calculates the rates of various event classes (overall triggers, prompt-like events, ^{40}K events, etc.), the energies of various peaks (^{40}K, nGd capture, etc.), and the rate of infrequent blocked triggers (caused by excessive pileup, typically by noise in the electronics). These quantities are then uploaded to the DQDB for use by external tools employed by DQ shifters and the DQWG.

The review of these metrics is the responsibility of the DQ shifter, Daya Bay's second line of defense. The DQ shifter uses the DQ shift website to view time series plots of one or more selected metrics (Figure 5.4). All metrics, and all detectors in a given hall, are shown on the same page. Under general circumstances, the DQ shifter typically looks at the prompt-like rate, the ^{40}K peak energy, and the AD and WP blocked trigger rates, since the vast majority of data quality issues will manifest themselves in at least one of these plots.

Any abnormal data points (i.e. files) can be clicked on by the DQ shifter; this will "locally" tag the file as bad (or vice versa, if the file was previously tagged) in the personal "session" of the DQ shifter. A comment is also recorded, indicating which metric the shifter clicked on, and whether the tagged file was above or below the local average. Locally tagged files are shown in a different color from the rest of the files. There are also specific colors for "officially" tagged files (i.e. those marked as bad in the DQDB), and for officially tagged files that the shifter has clicked on (to indicate that the official tagging should be reversed). Since the colors are synchronized between all metrics and halls, the shifter is able to view possible correlations. It is also possible to draw a rectangle around a collection of points, which can be used both for viewing correlations (since the selected files are highlighted in all plots) and for (un)tagging files in bulk.

The work of the DQ shifter is subsequently reviewed by a member of the DQWG, who logs into the shifter's session to view and potentially alter the shifter's tagging decisions. Once the session has been thus finalized, it is exported from the website in the form of a text file listing each file to be (un)tagged and its associated comment.

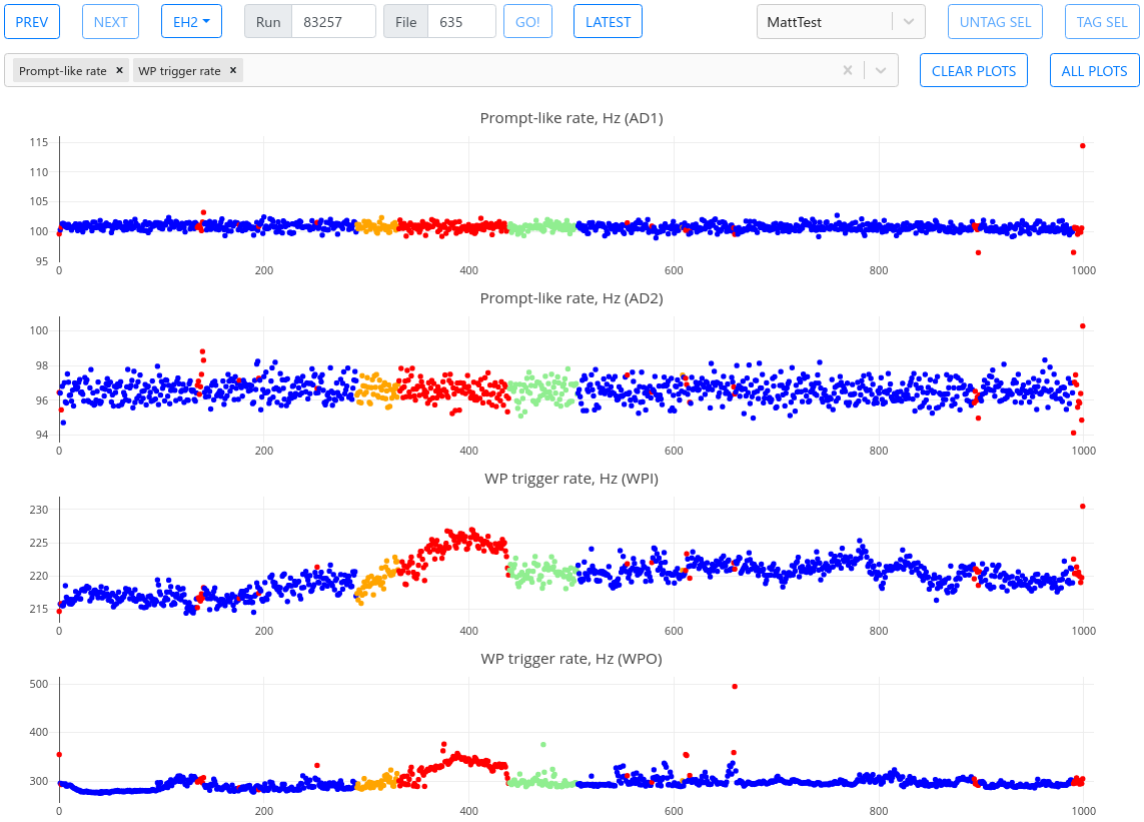

Figure 5.4: Screenshot of the Daya Bay data quality shift website, developed by the author [3]. Each point represents a data file; each plot represents a quantity of interest. Blue points are good (untagged) files; red points are files tagged as bad; orange points are files proposed for tagging as bad; green points are tagged files proposed for untagging.

A script is then run on this file to insert the comment and tagging decision in the DQDB.

The regular shift and DQ shift both take place in "real time" (with a potential modest delay in the case of the DQ shift), as data is being taken. An additional (i.e. third) defense against data quality is executed much later, after data has been processed (with correct calibrations) in an official production campaign. After such a production, a preliminary good run list is issued by the DQWG, based on the information contained at the time in the DQDB. Then, a number of Daya Bay's analysis groups run a series of independent data quality checks, producing plots in which outlying points serve as an indication of possible DQ issues. Such outliers are reported and investigated, and any files deemed bad are tagged as such in the DQDB. Once this work is complete, a new good run list is issued, and the checks are repeated

to ensure that all outliers are gone or understood to be harmless.

Through these redundant and complementary procedures, Daya Bay is able to ensure that any bad data is identified and removed from the analysis, enabling full confidence in the final result.

Chapter 6

Background subtraction

While Daya Bay's design ensures a very pure sample of antineutrino events, a small contamination of backgrounds (at the percent level) is unavoidable. These backgrounds can be subdivided into *correlated* and *uncorrelated* backgrounds. The uncorrelated backgrounds consist entirely of accidental coincidences between singles events, and the rate and spectrum can be easily estimated from that of the singles (see Section 7.7). On the other hand, the correlated backgrounds are so named because the prompt and delayed pulses are correlated in time and space, as they both originate from a single underlying process. The correlated background in Daya Bay consists of four distinct processes, and each one requires its own technique for determining the rate and energy spectrum.

In this chapter, we discuss the measurement of each of these backgrounds. Once they have been measured, their scaled prompt spectra can be subtracted from that of the IBD candidates, allowing the oscillation fit to proceed with a purer prompt spectrum, albeit one with an additional uncertainty stemming from the background measurements. We carry out our own prediction of the rates of the two dominant backgrounds: accidental coincidences and ^{9}Li/^{8}He decays. For the three other backgrounds (fast neutrons, AmC events, and $^{13}\mathrm{C}(\alpha,\mathrm{n})^{16}\mathrm{O}$), we use the results of the detailed studies that were carried out by members of the collaboration (see Appendix A). Those background studies were considerably complex and carefully validated analyses in their own right, and we make no attempt to improve upon them.

Before considering these double-coincidence backgrounds, we begin by discussing a background defined at the level of *individual* triggers, namely, the so-called "flashing" of PMTs. Reduction of these "flashers" is necessary in order to minimize the rate of the uncorrelated backgrounds. This reduction occurs at an early stage in the analysis (the "pre-selection", as described in Section 5.2.1), eliminating flashers entirely from further consideration.

6.1 PMT light emission ("flashers")

The PMTs used at Daya Bay have the tendency to occasionally "flash", emitting light due to arcing in their bases. The rate and intensity of this flashing changes over time for each PMT. At any given moment, some 5% of the PMTs in each AD will have the tendency to flash brightly enough to trigger the detector [17], in some cases producing as much as 100 MeV of reconstructed energy. Within the delayed energy region of 6-12 MeV, the flasher rate has averaged at around 0.7 Hz for each AD. These "delayed-like" flashers, if included in the analysis, would significantly increase the rate of backgrounds caused by the accidental coincidence of two uncorrelated signals. As discussed in Chapter 7, the rate of such "accidentals" is proportional to the rate of delayed-like signals, and this rate (excluding flashers) ranges from around 0.05 Hz at EH3 to 1 Hz at EH1. While the flashers would merely (roughly) double the 1% accidental background in the near halls, in the far hall they would increase this background by an order of magnitude to the 10% level, counter to Daya Bay's goal of percent-level background contamination.

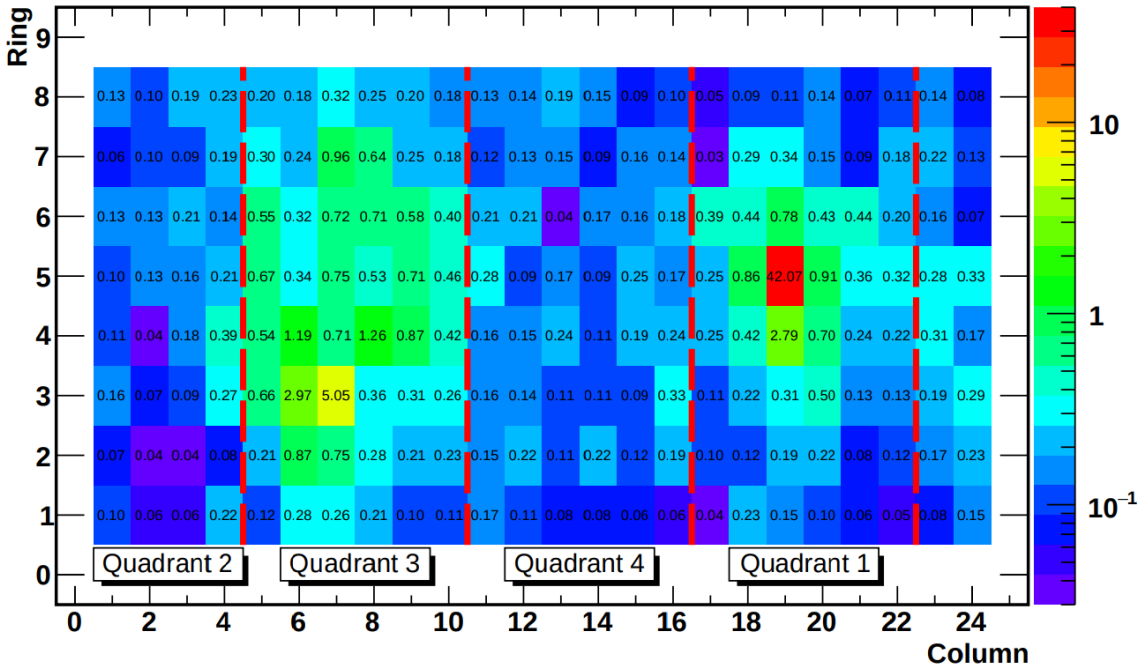

Figure 6.1: PMT charge distribution of a flasher candidate, illustrating the division of the AD into four azimuthal quadrants, with Quadrant 1 being centered around the PMT with the highest charge. From [19].

Fortunately, flashers are easily distinguished from "physical" events due to their unique conical pattern of light emission (Figure 6.1), enabling them to be removed

from the analysis with high efficiency while minimally affecting true IBDs. This light pattern is characterized by two "hot spots" on opposite sides of the AD. As described in Section A.1, a highly effective discriminator can be used to reject flashers based on their charge and time patterns, respectively. This discriminator, denoted by f_{ID}, is based on two variables. The first, f_{max}, is the ratio of Q_{max} (the maximum individual PMT charge across all PMTs) over the total charge Q_{tot}:

$$f_{\mathrm{max}} = \frac{Q_{\mathrm{max}}}{Q_{\mathrm{tot}}}. \tag{6.1}$$

The second, f_{quad}, is based on dividing the AD into four quadrants (Figure A.1): "Quadrant 1" (q1) is the one that is centered on the highest-charge PMT, q3 is the one across from q1, and q2 and q4 are the two "to the side." Then, f_{quad} captures the conical nature of the light emission:

$$f_{\mathrm{quad}} = \frac{Q_{\mathrm{q3}}}{Q_{\mathrm{q2}} + Q_{\mathrm{q4}}}. \tag{6.2}$$

Individually, f_{max} and f_{quad} do not cleanly distinguish flashers from physical events. Their combination, however, does, leading to the definition of f_{ID}:

$$f_{\mathrm{ID}} = \log_{10}\left[f_{\mathrm{quad}}^2 + \left(\frac{f_{\mathrm{max}}}{0.45}\right)^2\right]. \tag{6.3}$$

Flashers can then be identified as those events having $f_{\mathrm{ID}} > 0$. As shown in Figure 6.2, this discriminator effectively eliminates flashers from the analysis. Complete elimination is unnecessary; all that matters is that the residual flashers do not significantly increase the rate of delayed-like singles (and hence accidentals). This is easily the case for the f_{ID} discriminator. Two additional discriminators, neither of which are strictly necessary, are described in Section A.1.

The exact rejection factor for flashers is unimportant, as long as it is high enough to avoid a significant rate of flasher-associated accidentals. Any residual flashers will automatically be counted in the singles rate, and thus so will their contribution to the accidental background rate. However, it *is* important to study the *signal inefficiency* of the flasher rejection, that is, the proportion of non-flashers that incorrectly get identified as flashers. As discussed in Section A.4.1, this inefficiency has been found to be $0.039\% \pm 0.006\%$, with no significant AD-to-AD variation. This 0.006% uncertainty is rounded up to 0.01% and included in the detection efficiency uncertainty (Table D.1).

6.2 Accidental coincidences

IBD-like pairs can be formed when two uncorrelated triggers (*singles*) "accidentally" occur closely together in time. Given that the prompt-like singles rate (after flasher

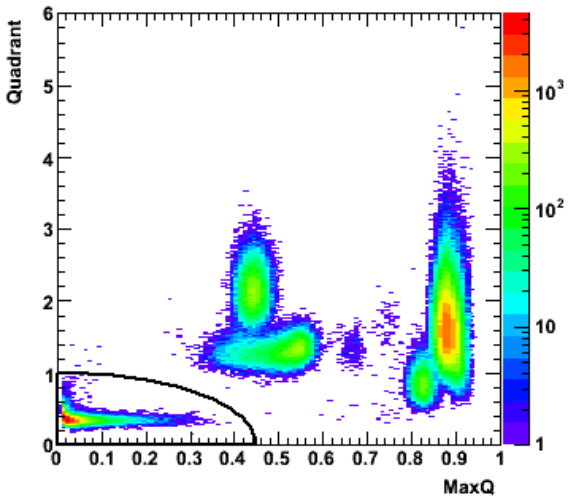

Figure 6.2: Two-dimensional distribution of f_{max} ("MaxQ") and f_{quad} ("Quadrant") for physical triggers in EH1-AD2. The black "ellipse" corresponds to $f_{\text{ID}} = 0$; events outside the ellipse are identified as flashers. High flasher rejection and negligible signal inefficiency are apparent. From [48].

rejection) is approximately 50 Hz in each AD,[1] the characteristics of singles (and thus of accidentals) can be measured with extremely high statistical precision. Once the rates of prompt-like and delayed-like singles have been determined, the accidentals rate can be calculated via a straightforward application of Poisson statistics. This procedure is detailed in Section 7.7. The spectrum, meanwhile, is equal to that of the singles sample, whose extraction is described in Section 5.3. In Section 7.7.2 and Section 7.7.3 we discuss the statistical and systematic uncertainty on the accidentals rate; in total, a conservative uncertainty of 2% is assigned to the rate for all ADs and data periods.

6.3 Cosmogenic ^{9}Li/^{8}He

The dominant correlated background for Daya Bay comes from the isotopes ^{9}Li and ^{8}He, which are produced as spallation products of carbon when energetic muons traverse the AD. These two isotopes have relatively long lifetimes of 257 ms for ^{9}Li and 172 ms for ^{8}He [49] (see Table 6.1); thus, while the majority of them will decay within the O(1 s) veto window that follows showering muons, a non-negligible fraction will survive past it. As shown in Figure 6.3, when one of these isotopes undergoes beta decay into the relevant excited states of the daughter nucleus, the daughter will

[1] Meanwhile, the delayed-like singles rate ranges from ∼0.001 Hz at the far site to ∼0.1 Hz at the near sites; this is still large enough to provide ample statistics from Daya Bay's multi-year dataset.

immediately emit a neutron (and, in the case of ^{9}Li, will further disintegrate into two alpha particles). The relevant β decay endpoints extend up to 12 MeV, placing these decays squarely within the IBD prompt-energy region. The combination of this β decay and the subsequent neutron-capture on gadolinium (nGd capture) produces the characteristic double-coincidence signature of an IBD event. In what follows, we will collectively refer to these two isotopes as ^{9}Li, given that it is believed to be the predominant of the two. (For instance, the KamLAND collaboration's FLUKA simulations [50] indicated a 10:1 ratio of ^{9}Li to ^{8}He production.[2]) Later we will propagate the uncertainty on this ratio into the background estimation.

Isotope	Lifetime (ms)	β decay endpoint (MeV)	Final products
^{9}Li	257.2	11.17	$e^- + \alpha + \alpha + \text{n}\ (+\ \bar{\nu}_e)$
^{8}He	171.6	7.44	$e^- + {}^7\text{Li} + \gamma + \text{n}\ (+\ \bar{\nu}_e)$

Table 6.1: Properties of the cosmogenic isotopes ^{9}Li and ^{8}He [49]. The quoted β decay endpoint is the endpoint of the highest-energy transition *that produces a final neutron.*

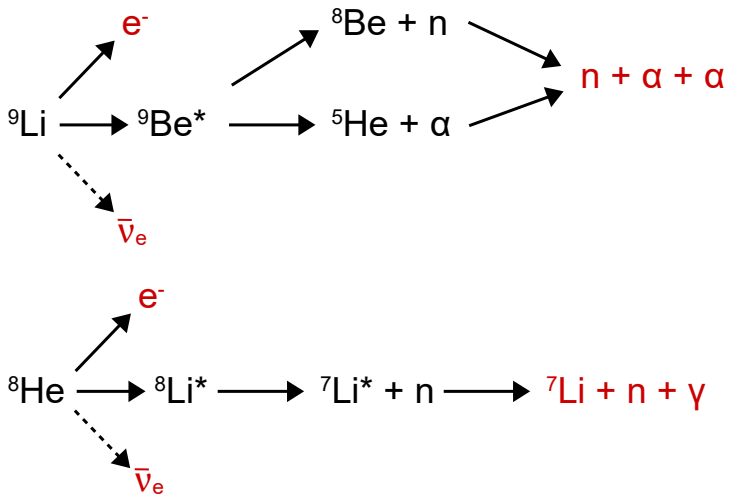

Figure 6.3: Decay cascades of ^{9}Li and ^{8}He with excited daughter nuclei after β decays. Final products are highlighted in red. Extended from [51].

[2]Given KamLAND's higher $\langle E_\mu \rangle$ of 260 GeV, compared to Daya Bay's ∼50 (100) GeV at the near (far) site(s), KamLAND's measured ratio of ^{9}Li to ^{8}He may not be completely representative, but in the absence of additional data, it is an acceptable starting point. In practice, as discussed later, we use a nominal (and somewhat arbitrary) ^{8}He fraction of 5.5%, with other values trialed as part of the overall uncertainty estimation.

Although at Daya Bay there is no feasible way of distinguishing ^{9}Li decays from IBDs at the level of individual events, it is still possible to statistically estimate the total ^{9}Li rate by exploiting the correlation in time between ^{9}Li events and preceding muons. To see this, let the muon rate be denoted by R_μ, and the ^{9}Li livetime by τ. Since IBD events are uncorrelated with muons, the time between an IBD and the most recent preceding muon is given, according to Poisson statistics, by the probability density function (PDF)

$$f_{\text{IBD}}(t) = R_\mu e^{-R_\mu t}. \tag{6.4}$$

Meanwhile, a ^{9}Li decay is correlated in time with the parent muon:

$$f_{\text{Li, Parent}}(t) = \frac{1}{\tau} e^{-t/\tau}. \tag{6.5}$$

However, in the time between a muon shower and an associated ^{9}Li decay, additional muons may be detected, and these may closely resemble the true parent muon. As such, the quantity that we can reliably measure is not the time between the ^{9}Li decay and the parent muon, but (as with IBDs) the time between the decay and the *last* muon. This detail has the effect of modifying the time constant in Equation 6.5. To derive this modification, let us consider a ^{9}Li event with a time-to-last-muon of t. Either the last muon was the parent muon, or it wasn't. Defining our time axis with the decay at the origin, these two possibilities can be stated quantitatively as:

1. The parent muon occurred at time $-t$, with no intervening uncorrelated muons, or

2. The parent muon occurred at some (unknown) time prior to $-t$, and the most recent uncorrelated muon occurred at $-t$.

The parent muon's time PDF is given by Equation 6.5, and that of the most recent uncorrelated muon by Equation 6.4. Meanwhile, the Poisson probability of observing zero uncorrelated muons in a time window of t is given by $e^{-R_\mu t}$. Thus, letting P_1 and P_2 be the probabilities[3] of the two cases above, we have

$$P_1 = \frac{1}{\tau} e^{-t/\tau} \cdot e^{-R_\mu t} = \frac{1}{\tau} e^{-(R_\mu + 1/\tau)t}$$

and

$$P_2 = \int_t^\infty \frac{1}{\tau} e^{-t'/\tau}\, dt' \cdot R_\mu e^{-R_\mu t} = e^{-t/\tau} \cdot R_\mu e^{-R_\mu t} = R_\mu e^{-(R_\mu + 1/\tau)t}.$$

Letting

$$\lambda = R_\mu + 1/\tau, \tag{6.6}$$

[3]Here we will say "probability" when we really mean "probability density".

we finally add P_1 and P_2 to obtain the PDF of time-to-last-muon for ^{9}Li decays:

$$f_{\mathrm{Li}}(t) = \lambda e^{-\lambda t}. \tag{6.7}$$

Comparing Equation 6.7 and Equation 6.4, we see that, if λ is sufficiently different from R_μ (in other words, if R_μ is sufficiently small relative to $1/\tau$), then there will be a measurable difference between the time-to-last-muon PDFs of ^{9}Li and IBD events. In this case, the separate event counts can be obtained by constructing a histogram of time-to-last-muon from a mixed sample of IBDs and ^{9}Li events. This histogram can then be fit to a weighted sum of Equation 6.7 and Equation 6.4, with the weights (as determined by the fit) corresponding to the number of events in the two categories:

$$f(t) = N_{\mathrm{Bkg}}\lambda e^{-\lambda t} + N_{\mathrm{IBD}}R_\mu e^{-R_\mu t}. \tag{6.8}$$

If the fit is allowed to extend down to t of a couple dozen ms, an additional correlated component must be considered resulting from accidental double coincidences of ^{12}B and/or ^{12}N decays. These two isotopes, with respective lifetimes of 29 and 16 ms [49], and $\beta^\pm$ endpoints of 13.7 and 16.3 MeV [49], can be produced multiple times by a single muon shower, and their decays extend well into the delayed-energy region. Their double coincidences are already considered as part of the accidental background estimation[4], so we are not concerned with measuring them here as a separate background; however, they do distort the time-to-last-muon histogram at low times, and thus they must be considered in order to extract the ^{9}Li component accurately. Past fits [29] to the spectrum of muon-correlated singles have indicated that ^{12}B is by far the dominant of these isotopes (comprising some 97% of the events), so in what follows we treat it as the only one of relevance.

When fitting the time-to-last-muon for the ^{12}B double coincidences (BB events, for short), the time constant must be considered carefully. We first consider the time to the *parent* muon, rather than to the *last* muon. For the single ^{12}B events, the corresponding time constant is simply the ^{12}B lifetime τ_{B}. However, if a muon produces two ^{12}B nuclei, then a timespan of t will correspond to an $e^{-t/\tau_{\mathrm{B}}}$ survival probability of one nucleus, and likewise for the other. The observation of one decay, of either nucleus, constitutes an observation of the whole pair, since we are recording the time between muons and the *prompt* event in an IBD candidate. The probability of not observing the pair, then, is the probability of seeing *neither* nucleus decay, which is $(e^{-t/\tau_{\mathrm{B}}})^2 = e^{-t/(\tau_{\mathrm{B}}/2)}$. That is, the time constant for the BB time-to-parent-muon distribution is $\tau_{\mathrm{B}}/2$. For the time to the *last* muon, rather than the parent, the arguments preceding Equation 6.7 then imply a rate constant of $\lambda_{\mathrm{BB}} = R_\mu + 2/\tau_{\mathrm{B}}$.

[4]There is some bias owing to the fact that the prompt and delayed ^{12}B/^{12}N are not uncorrelated in time (as assumed in the accidentals calculation), but are in fact correlated by virtue of their usually coming from the same parent muon. However, the rate of such events is extremely low (relative to the total accidentals rate), as can be seen in the fits shown later in this chapter, so there is no need to attempt a correction to the accidentals rate.

To finish deriving the full expression used in fitting the time to last muon, we simply add a factor r corresponding to the ^{8}He fraction, giving

$$f(t) = N_{\text{Bkg}}\left[(1-r)\lambda_{\text{Li}}e^{-\lambda_{\text{Li}}t} + r\lambda_{\text{He}}e^{-\lambda_{\text{He}}t}\right] + N_{\text{BB}}\lambda_{\text{BB}}e^{-\lambda_{\text{BB}}t} + N_{\text{IBD}}R_\mu e^{-R_\mu t}, \quad (6.9)$$

where

$$\begin{aligned}\lambda_{\text{Li}} &= R_\mu + 1/\tau_{\text{Li}}\\ \lambda_{\text{He}} &= R_\mu + 1/\tau_{\text{He}}\\ \lambda_{\text{BB}} &= R_\mu + 2/\tau_{\text{B}}.\end{aligned}$$

Although this method is simple in theory, a significant challenge arises from the fact that a minimum muon energy must be defined when calculating the time between each event and its most recent preceding muon. If this cut is too low, then the time between muons will be comparable to (or smaller than) the ^{9}Li livetime, and with finite statistics, it will be difficult to reliably distinguish between the two components in the fit. Conversely, if the muon cut is too high, then some fraction of ^{9}Li-producing muons will be discarded, and those ^{9}Li events will not appear to be muon-correlated, leading to an underestimation of the rate.

This issue is mitigated somewhat by the fact that low-energy muons produce a relatively small portion of the total ^{9}Li rate.[5] However, an accurate assessment must still attempt to quantify the contribution from low-energy muon events. To enhance the ^{9}Li signal in the fit, either the muon sample or the ^{9}Li candidate sample (or both) must be somewhat purified. Purifying the muon sample will enhance the difference in time constants between the ^{9}Li and IBD components, while purifying the ^{9}Li sample will enhance the difference in amplitudes. We use both methods in what follows.

For the muon sample, the goal is to remove muons that are unlikely to produce a ^{9}Li event. This will reduce the muon rate R_μ, thus increasing the difference between the time constants of the fit components. However, the cost of this *muon reduction* is that some fraction of ^{9}Li-producing muons will be discarded. The associated ^{9}Li events will have a time-to-last-muon rate constant of R_μ rather than λ, so they will not contribute to the measured ^{9}Li rate, leading to an inefficiency (and corresponding uncertainty) in the total rate estimate. Here, muon reduction is achieved by using *neutron tagging*, in which we only consider those muons for which a neutron capture candidate is observed in the immediate aftermath. Given the sizable uncertainty on the efficiency of the tagging, this method is only used where absolutely necessary, i.e., in the bin of the lowest muon energy. The details are discussed later.

As for increasing the purity of the ^{9}Li sample, the method used here is simply to apply a cut on the prompt energy. As can be seen by comparing Figure 6.10 to

[5] More precisely, at the near (far) sites, some 80% (90%) of the ^{9}Li in the IBD sample comes from muons of visible energy above 1 GeV. This can be deduced from the results shown in Table 6.3 (where 1 GeV $\sim$ 160,000 photoelectrons).

Figure 5.3, the ^{9}Li and IBD spectra differ substantially, with the former being much harder. As such, a higher prompt-energy cut will increase the ratio of ^{9}Li to IBD events. This benefit, however, must be weighed against the loss in the total statistics arising from an aggressive prompt-energy cut. Here, heuristically chosen cuts are used in order to obtain acceptable fits. As will be detailed later, more aggressive cuts are used in the near hall (where the IBD "background" fraction is larger) and for lower muon energies (where a greater ^{9}Li purity is needed due to the proximity of λ to R_μ). These prompt cuts, together with neutron tagging (for low muon energies), enable the ^{9}Li fit to be performed for muon energies that extend all the way down to the "AD muon" threshold of 3,000 photoelectrons (PE).

In the sections that follow, we describe the ^{9}Li rate estimation in detail, beginning with the selection of ^{9}Li candidates, followed by the muon selection, the fitting procedure, its results, the various selection efficiencies, and the error budget. Our method is a repeated and slightly modified version of the one described by Chris Marshall in [4], which in turn was based on a 2014 analysis performed by the author in [52].

6.3.1 ^{9}Li candidate selection

The ^{9}Li selection is largely identical to the standard IBD selection, as described in Chapter 5, but with two key differences: First, the prompt-energy cut is higher, as described above (although this cut is applied after the initial selection, for the sake of flexibility), and, secondly, the "shower muon" veto is disabled. That is, rather than vetoing for 0.4004 s after a muon of >300,000 PE (as in our nominal IBD selection), only the "AD muon" veto of 1.4 ms is applied. This is necessary because the shower veto would otherwise remove a large fraction of the ^{9}Li events, greatly reducing the already-limited statistics of the sample. More precisely, for a given ^{9}Li event produced by a shower muon, the probability of its falling within the veto window is

$$1 - \exp\left(-\frac{400.4}{257}\right) \approx 80\%, \tag{6.10}$$

where the numerator is the veto time and the denominator is the ^{9}Li lifetime (in milliseconds, in both cases). As shown in Table 6.3, some 90% of ^{9}Li events are produced by shower muons. Multiplying this percentage by that from Equation 6.10 implies that some 70% of ^{9}Li events would be lost if the shower veto were applied. A similar argument applies for ^{8}He.

When producing the final ^{9}Li rate estimation, we must correct for the efficiency of this modified muon veto. Using the muon toy MC described in Section 9.2.1, these efficiencies were determined to be 87%, 90%, and 99% in EH1, EH2, and EH3, respectively. (Compare to 82%, 85%, and 98% for the standard muon veto.)

6.3.2 Muon selection

For the purposes of applying the muon veto and constructing the time-to-last-muon distributions, we use the muon tree generated by the Daya Bay software framework (`NuWa`) during data production, as discussed in Section 5.1. In this tree, muons that trigger multiple detectors are merged into a single muon, and prompt retriggers are discarded. This represents a minor difference from our IBD selection, where retriggers are *not* discarded, leading to a slight reduction in our muon veto efficiency compared to what we'd obtain from using the `NuWa` muon tree. (This difference is correctly accounted for in the summation of veto windows, and thus there is no resulting bias.) For the ^{9}Li analysis, the removal of retriggers helps to prevent any biasing of the time constant for ^{9}Li events. The `NuWa` muon selection's cuts are designed to be looser than any conceivable muon definition used in a physics analysis (which would in turn select a subset of `NuWa's` "muons"). Accordingly, they are as follows:

- A trigger in the inner (outer) water pool, with at least 6 (8) fired PMTs, is tagged as a WP muon.
- A trigger in an AD of more than 3000 PE is tagged as an AD muon.
- Tagged muons that occur within a 300 ns window are merged into a single muon (with the WP hit counts and AD charges individually stored), with the timestamp taken to be the time of the earliest trigger.
- Any subsequent "muons" that occur within 10 µs are deemed to be retriggers and are thus discarded.

During the ^{9}Li selection, a given muon object will result in the 1.4-ms AD muon veto if the object contains an AD muon for the detector in question. If not, but if there is an IWP or OWP muon present with *more than 12 hit PMTs*, then the 600-µs WP muon veto is applied instead.

If a ^{9}Li candidate is found, then all preceding muons within a 10-second window are stored in an array specific to that candidate. This array can then be looped over when constructing the time-to-last-muon histogram, skipping over muons outside of a chosen energy range until the most recent one within the energy range is found.

6.3.2.1 Neutron tagging

In order to enable the neutron tagging technique (as needed for calculating the rate of ^{9}Li produced by low-energy muons), an additional flag is stored for each muon, indicating whether the muon occurred in coincidence with a neutron-capture candidate. Specifically, the neutron tag is applied if any trigger ranging from 1.8 to 12 MeV occurs between 20 and 200 µs after the muon. Although these cuts are heuristic, they are reasonable in principle: The energy range is wide enough to include captures

on both gadolinium and hydrogen, and the time window includes most of the Gd captures and a decent fraction of H captures, while avoiding spurious signals that can arise from retriggers occurring in the first dozen or so µs after a muon.

The primary challenge in using the neutron-tagging technique is accounting for the efficiency of the tagging. That is, what percentage of ^{9}Li-producing AD muons (within a specified energy range) will be neutron-tagged? For sufficiently high-energy muons (roughly above 160,000 PE), the ^{9}Li rate can be measured both with and without neutron tagging, and thus the efficiency can be extracted directly. However, for muons below 160,000 PE (where neutron tagging is unavoidable), this is not possible, and the efficiency must be extrapolated. We discuss this efficiency, and its uncertainty, in Section 6.3.4.2.

6.3.3 Time-to-last-muon fit

Once a sample of ^{9}Li candidates has been collected, along with the history of prior muons for each such candidate, the next step is to construct the histogram of times since the most recent muon (within a given muon energy range). We use a histogram spanning from 2 ms to 1 s, with a nominal binning consisting of 24 variable-width bins. The lowest bin covers 2 to 20 ms; the next four bins (up to 100 ms) are 20-ms-wide; the next four (up to 0.5 s) are 0.1-s-wide; the next six (up to 2 s) are 0.25-s-wide; the next two (up to 3 s) are 0.5-s-wide; and the final seven are 1-s-wide. This variable binning scheme captures the fine structure at low times while keeping the number of bins at a manageable level. Later we discuss the systematic uncertainty arising from variation of the binning.

For each hall, three histograms were constructed, corresponding to three muon visible energy ranges of <160,000, 160,000-300,000, and >300,000 PE.[6] Different cuts are used in the three ranges, as detailed in Table 6.2. To fill each histogram, we loop over all ^{9}Li candidates for the hall in question, and if the candidate is accepted by the prompt-energy cut, we then find the most recent preceding muon (within the specified energy range, and with a neutron tag when required). The time between the candidate and the muon is then used to fill the histogram.

After constructing the histogram, we fit it to Equation 6.9. The results are shown in Figures 6.4–6.6. For the nominal result, we fix the ^{8}He fraction r to be 5.5% (an inherited "reasonable guess", as discussed further in Section 6.3.9), and include the ^{12}B component. Later we discuss the effects of varying the ^{8}He fraction and the inclusion of ^{12}B. The best-fit value of N_{Bkg}, as determined by `MINUIT`, corresponds to the raw number of (muon-correlated) ^{9}Li/^{8}He events present in the sample used for filling the histogram. As discussed in [4], the fit does not suffer from any appreciable correlation between N_{Bkg} and N_{IBD}, as the latter is strongly constrained by the tail

[6] As detailed in Chapter 9, we later vary the definitions of these ranges in order to capture the dependence of the final ^{9}Li rate on the muon energy used in the definition of the shower muon veto. The ranges described here apply to the shower veto used in the nominal IBD selection.

Hall	PE range ($\times 10^3$)	Prompt cut (MeV)	Neutron tag?
EH1/2	5-160	8.0	Yes
	160-300	8.0	No
	300-inf	3.5	No
EH3	5-160	6.0	Yes
	160-300	6.0	No
	300-inf	3.5	No

Table 6.2: Cuts used in constructing the three separate time-to-last-muon histograms for each hall.

at high time-to-last-muon. This was verified in [4] by hand-scanning the variation of χ^2 as a function of N_{Bkg}. Thus, `MINUIT's` reported statistical uncertainty on N_{Bkg} can be safely taken at face value and propagated into our final uncertainty. Our fit results are summarized in Table 6.3. As discussed further in Chapter 9, the fit was also performed using alternative muon energy ranges, in order to characterize the dependence of the ^{9}Li rate on the choice of shower-muon threshold.

Hall	PE range ($\times 10^3$)	N_{Bkg}
EH1	Low (5-160)	164.0 ± 51.4
	Mid (160-300)	554.9 ± 200.5
	High (300-inf)	6969.2 ± 675.4
EH2	Low (5-160)	95.8 ± 27.4
	Mid (160-300)	301.7 ± 118.5
	High (300-inf)	6282.7 ± 733.3
EH3	Low (5-160)	66.8 ± 25.5
	Mid (160-300)	280.1 ± 89.4
	High (300-inf)	1915.1 ± 182.4

Table 6.3: Measured number of ^{9}Li+^{8}He events, as determined by the time-to-last-muon fit, for each hall and range of muon energies. Quoted errors are the statistical uncertainties reported by the fitter. No efficiency corrections have been applied.

6.3.4 Selection efficiencies and their uncertainties

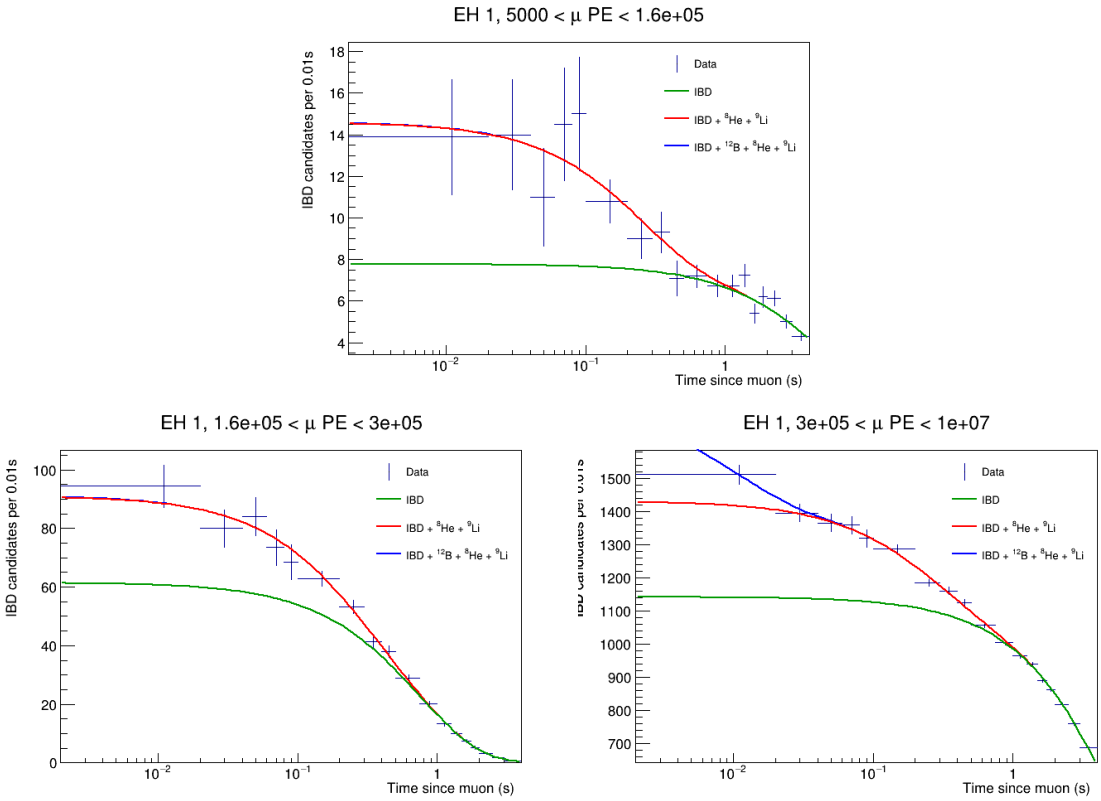

Figure 6.4: Time-to-last-muon fits for IBD-like events in EH1.

6.3.4.1 Prompt-energy cut

To determine the efficiencies (and the corresponding uncertainties) of the prompt-energy cuts, we replicate the procedure developed in [4].[7] These efficiencies can simply be calculated by integrating the normalized ^{9}Li spectrum (expressed in reconstructed energy) over the energies above each cut. The difficulty, however, lies in obtaining this spectrum. When subtracting the ^{9}Li background from the spectrum of IBD candidates, we follow the tradition of using Ochoa's theoretical calculation of the spectrum, as described in Section 6.3.9. However, for the prompt-energy cut efficiencies, we extract the ^{9}Li reconstructed spectrum directly from data.

The spectrum extraction proceeds in two steps. First, a spectrum is obtained from a sample that is enriched in ^{9}Li, and then a ^{9}Li-deficient sample (consisting mainly of IBD "background") is acquired, rescaled, and subtracted from the ^{9}Li-rich sample. This procedure produces a spectrum that contains very little contamination and can thereby give a reliable measurement of the prompt-energy cut efficiency.

To obtain the ^{9}Li-rich sample, the ^{9}Li candidates from all three halls were combined,

[7] We deviate from [4] in using a muon cut of 300,000 PE, rather than 400,000 PE, when selecting the ^{9}Li-enriched sample used in the efficiency determination.

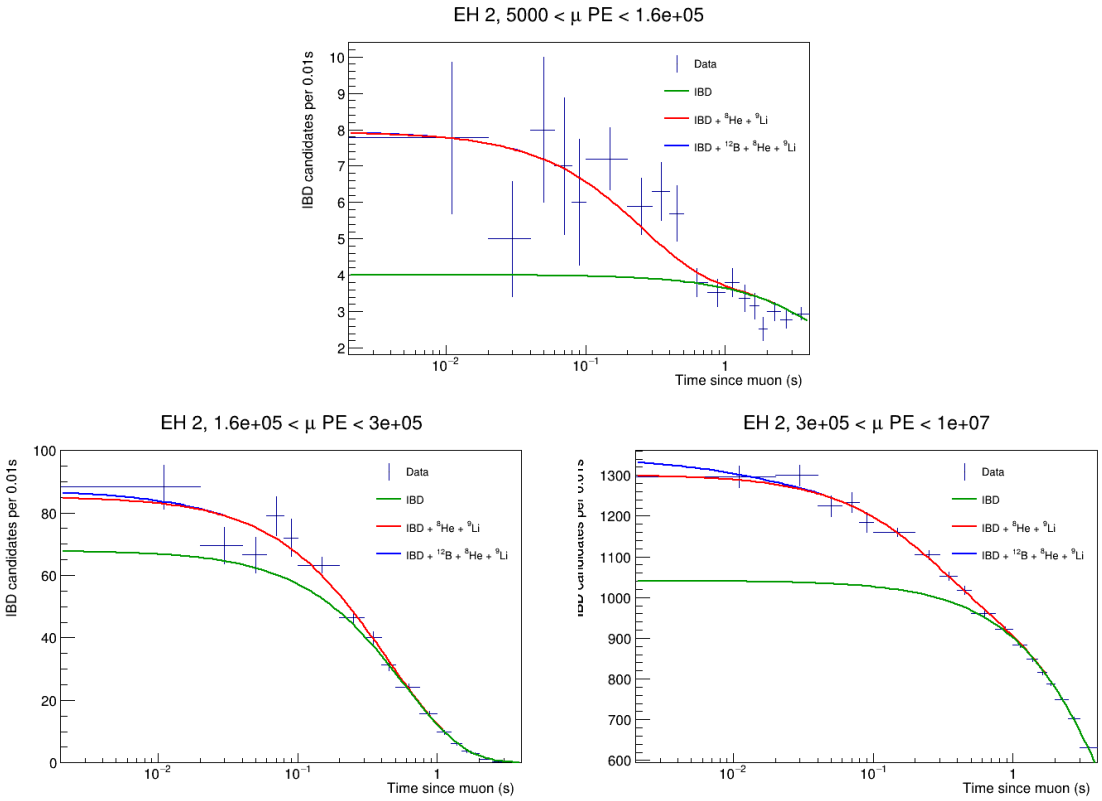

Figure 6.5: Time-to-last-muon fits for IBD-like events in EH2.

and a subset was taken by selecting events that fell between 2 and 200 ms after a high-energy shower muon, defined here as a muon that produced more then 300,000 PE of visible energy. The time-to-last-muon fit of Equation 6.9 was then performed, giving the relative fraction of the sample comprised by *shower-correlated* ^{9}Li. The remaining, shower-uncorrelated, events consisted largely of non-^{9}Li "background" (predominantly true IBDs).

Within the shower-uncorrelated fraction, we expect the presence of some ^{9}Li events produced by sub-300,000 p.e. muons, but this is a small contribution. As shown by the fit results in Table 6.3, some $f = 90\%$ of the total ^{9}Li sample (before applying any shower-muon veto) is produced by showers of at least 300,000 PE. Meanwhile (as can be seen from visually inspecting the fits in Figures 6.4–6.6), in this muon energy range (and time window), the correlated events take up approximately $R = 1/3$ of the candidate sample. As such, the number of shower-uncorrelated ^{9}Li events in this subsample, relative to the total number of uncorrelated events, is roughly

$$\begin{aligned} F &= \frac{(1-f)R}{1-R} = \frac{(1-0.9)\cdot 0.33}{1-0.33} \\ &\approx 5\% \end{aligned} \tag{6.11}$$

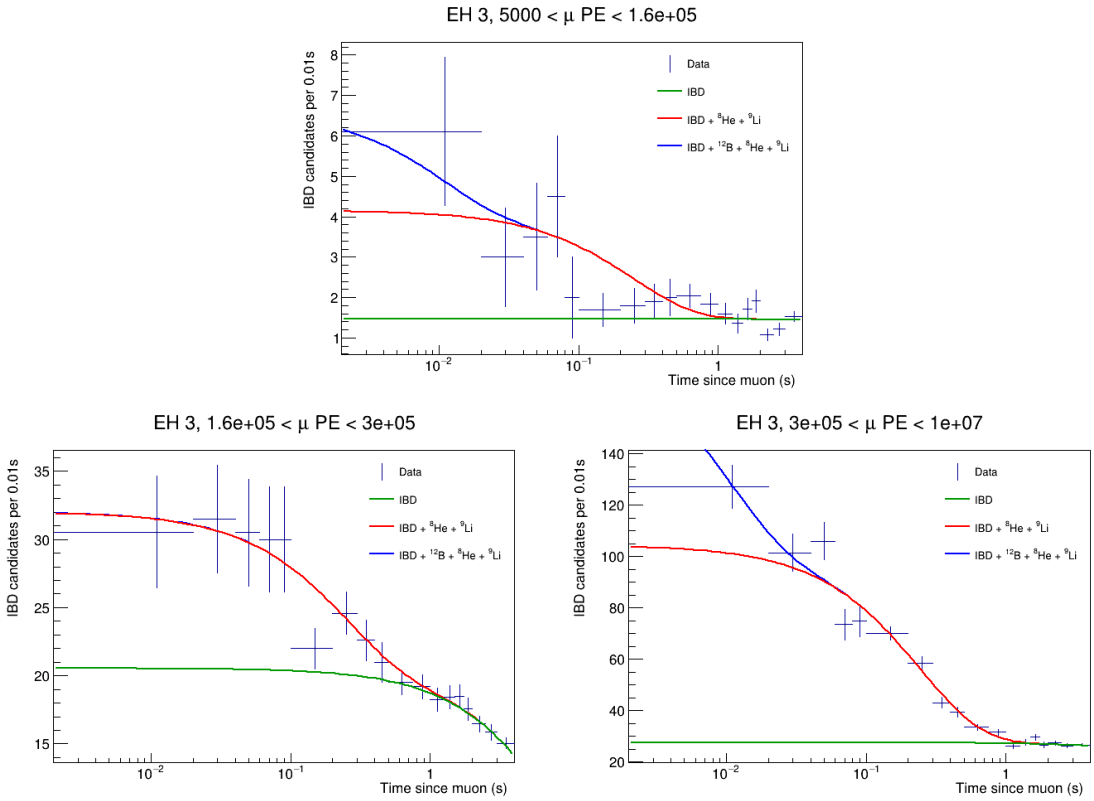

Figure 6.6: Time-to-last-muon fits for IBD-like events in EH3.

That is, the uncorrelated rate, as reported by the fit, describes a set that is about 95% non-^{9}Li. Following [4], in what follows we assume that 95% $\approx$ 100%. It would be more correct to scale the non-^{9}Li "background" spectrum (described later) by $\sim$0.95 before subtracting it. In practice, we choose to simply assign an additional 5% systematic uncertainty to account for this issue.

Given that around 2/3 of the events in this enriched sample are in fact non-^{9}Li "background" (mainly IBDs), it is imperative to obtain a clean background sample for subtraction. To do so, we use events from the ^{9}Li selection which were preceded by at least 1.5 s of isolation from muons of greater than 200,000 PE. This cut excludes virtually all (99.8%, based on the ^{9}Li lifetime) ^{9}Li events produced by >200,000 PE muons. The remaining ^{9}Li events are a very small contribution: About 0.4% of the (full) ^{9}Li candidate sample consists of actual ^{9}Li (according to the fits), and from Table 6.3, it can be seen that, at most, 15% of these events are produced by muons of below 300,000 PE, and hence an even smaller percentage are produced by sub-200,000 PE muons. Conservatively taking 15%, and multiplying it by 0.4%, we derive an upper bound of 0.06% on the level of ^{9}Li contamination in the "background" sample. This is more than acceptable. Meanwhile, the overall rate of ^{12}B events is less than 10% of the ^{9}Li rate (Figures 6.4–6.6), so they are even less of a concern.

Once the "signal"[8] (^{9}Li) and "background" (IBD) samples were extracted, the background spectrum was scaled such that its integral equaled the number of shower-uncorrelated events in the signal sample (as reported by the time-to-last-muon fit), and then subtracted from the signal spectrum. The result was our final extracted ^{9}Li spectrum (Figure 6.10). The uncertainty in each bin, for both the signal and background spectra, were calculated based on Poisson statistics, and combined in quadrature to give the error on each bin of the spectrum. Finally, the efficiency of each prompt-energy cut was calculated by summing the counts of each accepted bin, and dividing this by the total count in all bins. The statistical uncertainty, meanwhile, was simply the result of adding the error on each bin in quadrature, and propagating the combined errors through the calculation of the ratio. As a cross-check, the calculation was performed separately for each hall, with the results found to be consistent with each other, validating the combination of all three halls in calculating a single efficiency.

Two potential sources of systematic uncertainty are considered in the calculation of the prompt-energy cut efficiency.[9] The first is the choice of time binning used. In [4], the efficiency analysis was repeated for the two alternative binnings described in Section 6.3.5.2, revealing a variation of less than 1%. The second potential uncertainty arises from the choice of muon energy cut used in extracting the ^{9}Li-deficient sample. When the cut was increased to 400,000 PE, the result was again a sub-1% variation in the calculated efficiency [4]. Based on these studies, we assign a total systematic uncertainty of 1% (absolute) to the prompt-energy cut efficiency.

The efficiencies and uncertainties for the three prompt-energy cuts are summarized in Table 6.4. For each hall, the total uncertainty arising from the cut was determined from a weighted sum of the efficiencies that correspond to each muon PE range:

$$\sigma_p = \frac{w^{\text{low}}\sigma_p^{\text{low}} + w^{\text{mid}}\sigma_p^{\text{mid}} + w^{\text{high}}\sigma_p^{\text{high}}}{w^{\text{low}} + w^{\text{mid}} + w^{\text{high}}}, \tag{6.12}$$

where σ_p^{low} is the uncertainty (statistical + systematic, in quadrature) of the efficiency of the prompt-energy cut used for the low muon PE range (and similarly for "mid" and "high"). Meanwhile, the weights w are proportional to the contribution of each range to the predicted daily rate (compare to Equation 6.15):

$$\begin{aligned} w^{\text{low}} &= \frac{N^{\text{low}}}{\epsilon_p^{\text{low}}\epsilon_n} \\ w^{\text{mid}} &= \frac{N^{\text{mid}}}{\epsilon_p^{\text{mid}}} \\ w^{\text{high}} &= \left(\frac{f_{\text{He}}}{e^{t_{\text{sh}}/\tau_{\text{He}}}} + \frac{1-f_{\text{He}}}{e^{t_{\text{sh}}/\tau_{\text{Li}}}}\right)\frac{N^{\text{high}}}{\epsilon_p^{\text{high}}}. \end{aligned} \tag{6.13}$$

[8] Signal + background, more precisely.

[9] We directly use the results from [4], without repeating the procedure.

Here, N^{low} etc. are the values of N_{bkg} for the three PE ranges in Table 6.3, ϵ_p^{low} etc. are the corresponding prompt efficiencies (Tables 6.2 and 6.4), ϵ_n is the neutron tagging efficiency (discussed in Section 6.3.4.2), f_{He} is the nominal ^{8}He fraction of 5.5%, t_{sh} is the shower veto time (0.4004 s), and τ_{Li} (τ_{He}) is the ^{9}Li (^{8}He) lifetime. The results are shown in Table 6.5.

Prompt energy cut	Efficiency
3.5 MeV	$78.3\% \pm 5.4\%$ (stat.) $\pm 1\%$ (syst.)
6 MeV	$43.0\% \pm 2.9\%$ (stat.) $\pm 1\%$ (syst.)
8 MeV	$16.8\% \pm 1.4\%$ (stat.) $\pm 1\%$ (syst.)

Table 6.4: Prompt-energy cut efficiencies for the ^{9}Li selection.

6.3.4.2 Neutron tagging

Historically, the largest contribution to the overall ^{9}Li rate uncertainty has come from the neutron-tagging efficiency. In previous analyses, the prompt-energy cut was 3.5 MeV, for which the IBD "background" was so large as to require neutron tagging for all muons below 300,000 PE. Given that some 60-70% of the total ^{9}Li rate[10] comes from muons below 300,000 PE, this uncertainty applied to that entire subsample. Making matters worse, this uncertainty is quite large, as the efficiency can only be directly measured for higher-energy muons (where neutron tagging is not necessary in order to obtain a valid time-to-last-muon fit). The efficiency must thus be extrapolated (with considerable uncertainty) for lower-energy muons. Owing to our use of a stricter prompt-energy cut (of 8 MeV and 6 MeV in the near and far halls, respectively) for muons below 300,000 PE, the neutron-tagging requirement has been eliminated in the range of 160,000–300,000 PE, significantly reducing the total uncertainty from neutron tagging. However, for 5,000–160,000 PE (which contributes 12–16% of the final ^{9}Li rates), where neutron tagging remains unavoidable, the extrapolated efficiency must still be used.

In [53], the efficiency was found to range from about 60 to 100% within the energy ranges where it could be measured directly. As such, an uncertainty of 45% was assigned in [4] to the efficiency, with a nominal value of unity. This latter choice can be post-hoc justified by the results in [54], where an efficiency of 93.9% was found for 160,000 to 400,000 PE in EH1; nevertheless, values closer to 70% were found in the other two halls, and the efficiency is expected to decrease for lower muon energies. Although a lower efficiency might be more accurate, we use unity for the sake of consistency with [4]. In any case, the 45% uncertainty encompasses the range of values observed in [54].

[10] In the oscillation analysis, i.e., with the shower-muon veto.

The uncertainty on the final ^{9}Li rate resulting from the neutron-tagging efficiency, for each hall, is equal to 45% scaled by the proportion of ^{9}Li events that come from muons of 5,000–160,000 PE (the "low" range):

$$\sigma_n = \frac{w^{\text{low}}}{w^{\text{low}} + w^{\text{mid}} + w^{\text{high}}} \cdot 0.45, \tag{6.14}$$

where the weights w were given previously in Equation 6.13. The values for each hall are shown in Table 6.5.

6.3.5 Other systematics

6.3.5.1 Variation of neutron-tagging cutoff

In our ^{9}Li fits, neutron tagging is applied for muons below 160,000 PE. To study the effects of varying this dividing line, the analysis was repeated in [4] for various values, ranging from 150,000 to 180,000 PE. The resulting variation in the final estimated rate of ^{9}Li was found to be of order 10%. We assign this value verbatim as an additional systematic uncertainty on the final rate.

6.3.5.2 Binning

To study the effect of varying the time binning used in the fit,[11] two alternative binnings were tested in [4]. The first was finer, with 40 total bins and 10-ms bins below 100 ms. The second was coarser, with 20 total bins and 50-ms bins below 100 ms. These alternative binnings produced less than a 5% shift in the final result. Accordingly, we assign a 5% systematic uncertainty to account for the effect of binning.

6.3.5.3 ^{12}B and ^{8}He components

Our fits assume a 5.5% nominal ^{8}He fraction and include a component correspond to ^{12}B. To characterize the sensitivity of the analysis to these choices, the fit was repeated after disabling the ^{12}B component and after varying the ^{8}He fraction from 0 to 15%. These variations shifted the final rate by some 10%, which we assign as an additional systematic uncertainty.

The ^{8}He fraction, in addition to affecting the result of the fit, also affects the conversion into the final daily rate (described in the next section). This is because a correction factor must be applied to account for the shower-muon veto used in the IBD selection. The efficiency of this veto is a function of the isotope livetime, which of course varies between ^{9}Li and ^{8}He.[12]

[11] As described above, 25 variable-width bins from 0.002 to 10 seconds, with 20 ms bins below 100 ms

[12] This systematic uncertainty in this shower-muon veto survival probability, resulting from the uncertainty in the ^{8}He fraction, was not considered in [4].

We consider three cases: 0% ^{8}He, 5.5% ^{8}He (the nominal fraction), and 15% ^{8}He. For the nominal shower-muon veto of ∼0.4 s, the three efficiencies are, respectively,

$$\begin{aligned} \exp(-0.4/0.257) &= 0.211 \\ 0.945 \times \exp(-0.4/0.257) + 0.055 \times \exp(-0.4/0.172) &= 0.205 \\ 0.85 \times \exp(-0.4/0.257) + 0.15 \times \exp(-0.4/0.172) &= 0.194. \end{aligned}$$

Taking the range of these three values (0.017), and dividing by the mean (0.203), we find a relative spread of 0.084, corresponding to an uncertainty of roughly 4% in the final rate for ^{9}Li events produced by muons above 300,000 PE (our nominal shower-muon cut in the oscillation IBD selection). Given that these high-energy muons produce some 30-40% of the final ^{9}Li events in the sample (after the shower-muon veto), we conservatively assign 2% (half of 4%) as an additional systematic uncertainty on the total ^{9}Li rate.

6.3.6 Summary of uncertainties

The uncertainty budget of the ^{9}Li rate is shown in Table 6.5. The various systematic uncertainties were discussed in the preceding sections. To obtain the (relative) statistical uncertainties, first, the absolute statistical uncertainties were calculated by inserting the uncertainties in the various N_{Bkg} (from Table 6.3) into Equation 6.15, and then these were divided by the rates (calculated by inserting the N_{Bkg} values themselves into Equation 6.15, as described in Section 6.3.7). The individual uncertainties were added in quadrature to obtain each hall's total uncertainty.

Uncertainty	EH1 (%)	EH2 (%)	EH3 (%)
Statistical	27.5	26.5	24.1
Spectrum extraction (see Equation 6.11)	5.0	5.0	5.0
Prompt-energy cut efficiency	2.8	3.3	4.0
Neutron tagging efficiency	7.2	6.4	5.3
Neutron tagging cutoff	10.0	10.0	10.0
Binning	5.0	5.0	5.0
^{12}B component and ^{8}He fraction (fitting)	10.0	10.0	10.0
^{8}He component (veto survival probability)	2.0	2.0	2.0
Total	32.7	31.7	29.6

Table 6.5: Uncertainty budget for the ^{9}Li/^{8}He rate.

6.3.7 Calculation of daily rates

Based on the results of fitting the three muon energy ranges, the final daily ^{9}Li rate per AD in each hall is calculated as

$$\frac{1}{\epsilon_\mu \epsilon_m T N_{\text{det}}^{\text{eff}}} \left[\frac{N^{\text{low}}}{\epsilon_p^{\text{low}} \epsilon_n} + \frac{N^{\text{mid}}}{\epsilon_p^{\text{mid}}} + \left(\frac{f_{\text{He}}}{e^{t_{\text{sh}}/\tau_{\text{He}}}} + \frac{1 - f_{\text{He}}}{e^{t_{\text{sh}}/\tau_{\text{Li}}}} \right) \frac{N^{\text{high}}}{\epsilon_p^{\text{high}}} \right] \tag{6.15}$$

where ϵ_μ is the efficiency of the muon veto (without the shower-muon veto), e_m is the multiplicity cut efficiency, T is the total livetime, $N_{\text{det}}^{\text{eff}}$ is the livetime-weighted number of detectors, $N^{\text{low/mid/high}}$ are the raw rates from the fits (as listed in Table 6.3), $\epsilon_n = 1$ is the neutron-tagging efficiency, ϵ_p are the prompt-energy cut efficiencies (Table 6.4), $f_{\text{He}} = 5.5\%$ is the nominal ^{8}He fraction, τ_{Li} (τ_{He}) is the ^{9}Li (^{8}He) lifetime, and t_{sh} is the shower-muon veto time of 0.4004 s. As mentioned earlier, the muon-veto efficiencies are calculated using the toy MC described in Section 9.2.1, and the multiplicity cut efficiencies (which are relatively stable) are taken from the IBD selection.[13] The relative uncertainties, as calculated in Section 6.3.6, were scaled by the rates to give the absolute uncertainties quoted in the table.

Hall	ϵ_μ	ϵ_m	T (days)	$N_{\text{det}}^{\text{eff}}$	Events/AD/day
EH1	0.8711	0.9772	1737	1.8888	2.18 ± 0.71
EH2	0.9015	0.9782	1729	1.8925	1.39 ± 0.44
EH3	0.9883	0.9829	1737	3.8921	0.199 ± 0.059

Table 6.6: Final estimates of the daily ^{9}Li rate (per AD) in each hall, according to Equation 6.15. The uncertainties of ϵ_μ, ϵ_m, and T (and consequently $N_{\text{det}}^{\text{eff}}$) are negligible. Note that, since the efficiencies of the muon veto and multiplicity cut (in the ^{9}Li+^{8}He selection) were divided out in Equation 6.15, we must multiply these rates by the corresponding efficiencies of our IBD selection in order to obtain the predicted number of ^{9}Li+^{8}He events in our raw IBD sample.

6.3.8 Linear regression

When we experiment with modifying the charge threshold and veto time of the shower-muon veto, as described in Chapter 9, the ^{9}Li rate must be recalculated using Equation 6.15 for each case. Although the modified value of τ can simply be inserted into the equation, there is also an implicit dependence on the charge threshold, which determines the division of muons between the mid- and high-energy bins. Although it would be possible to repeat the ^{9}Li selection for each value of the threshold, this would

[13] The multiplicity efficiencies are 97-98%; the 2-3% inefficiency is negligible in comparison to the overall ^{9}Li uncertainty, so we need not dwell on it any further.

be a rigid and time-consuming approach. Instead, we perform the selection at 44 evenly-spaced values of the mid/high dividing line, ranging from 1.7×10^5 to 6.0×10^5 PE. When our threshold lies in between these grid points, we must interpolate.

Our initial approach was to simply perform a linear interpolation between the two points on each side of the threshold. This would be done once for the mid-energy bin, and again for the high-energy bin. However, the results of the ^{9}Li selection/fit demonstrate a fair amount of random jitter between adjacent values of the mid/high threshold, as shown in Figures 6.7–6.9. Although this variation is well below the systematic uncertainty on the ^{9}Li rate, it does introduce artificial structure to the results of the shower-muon veto variation. Fortunately, the ^{9}Li rates show a fairly linear dependence on the threshold, as shown in the same plots. Therefore, our approach is to calculate the ^{9}Li rate using Equation 6.15 at each of the 44 grid points for the charge threshold (using the same modified veto time in each case), fit the results to a line, and then evaluate this line at the desired value for the threshold. This procedure produces the ^{9}Li rate that is used in the oscillation fit for arbitrary shower-muon veto parameters in Chapter 9.

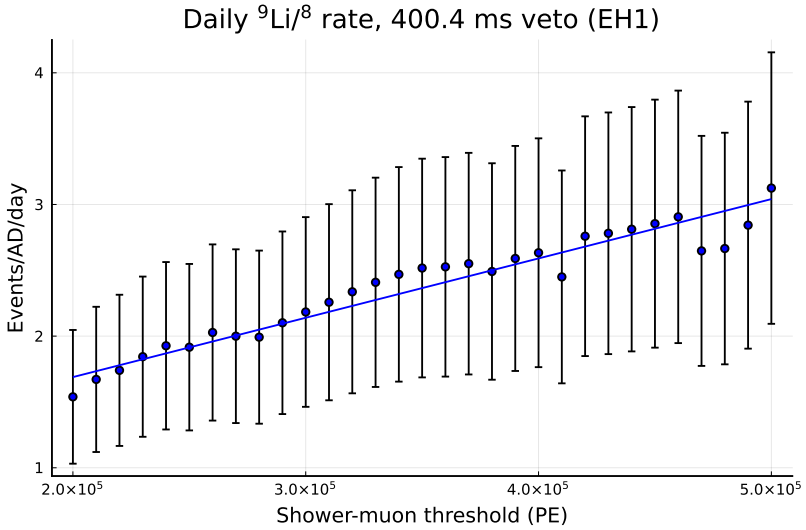

Figure 6.7: Dependence of the measured ^{9}Li rate (for the nominal shower-muon veto time of 400.4 ms) on the shower-muon threshold in EH1. The line-of-best-fit is used in determining the ^{9}Li rate for arbitrary shower-muon thresholds in Chapter 9.

6.3.9 Spectrum

The ^{9}Li/^{8}He spectrum can be either extracted from data or predicted from nuclear tables. Both methods give consistent results (Figure 6.10). The extracted spectrum (used for determining the efficiency of the prompt-energy cut in the ^{9}Li selection) was

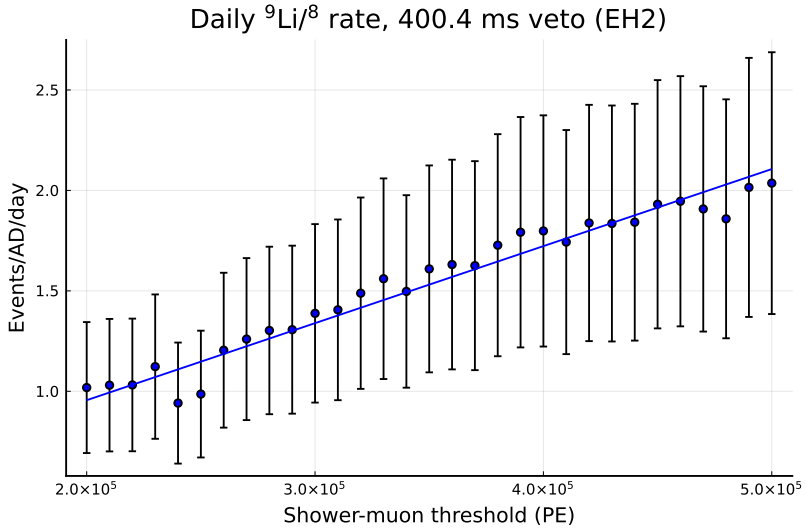

Figure 6.8: Dependence of the measured ^{9}Li rate (for the nominal shower-muon veto time of 400.4 ms) on the shower-muon threshold in EH2. The line-of-best-fit is used in determining the ^{9}Li rate for arbitrary shower-muon thresholds in Chapter 9.

described in Section 6.3.4.1. Largely for historical reasons, our background subtraction does not employ the extracted spectrum, instead using the predicted one. Here we briefly describe this prediction [51, 55].

For the prediction, three types of reference tables were consulted: nuclear structure, branching ratios, and measured spectra (of neutrons, alpha particles, and gamma-rays). Nuclear levels and decay schemes are shown in Figures 6.11 and 6.12. Given the number of decay pathways involved, a purely analytic approach was infeasible (at least for ^{9}Li), so a toy Monte Carlo was used to produce random decay events. The discussion here uses the example of ^{9}Li, but ^{8}He was essentially treated the same way. The initial β decay (into any of the four ^{9}Be* states) was simulated using the Fermi theory [8] and the published energy levels and branching fractions. For the decays that produce a $2\alpha + n$ final state (i.e. the only decays of interest to us), ^{9}Be* disintegrates via two consecutive two-body decays (via either ^{8}Be or ^{5}He). The angular distribution is assumed to be uniform. The disintegration was treated using basic kinematics, with the width of each state (Table 6.7) modeled with a Breit-Wigner function. The result was a collection of simulated events, each one recording the true energies of the electron, the neutron, and the two alphas (or, in the case of ^{8}He, the electron, the neutron, and the gamma-ray).

In order to benchmark this simulation, its output was compared against published measurements of ^{9}Li/^{8}He spectra of neutrons, alpha particles, and gamma-rays. Based on this comparison, one particularly broad level of ^{8}He had to be augmented with a Gaussian density of states. Since the published branching ratios were relatively

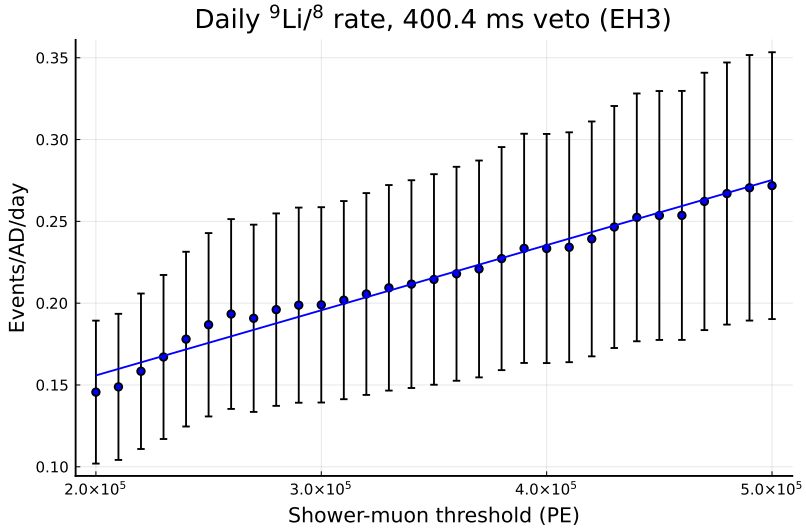

Figure 6.9: Dependence of the measured ^{9}Li rate (for the nominal shower-muon veto time of 400.4 ms) on the shower-muon threshold in EH3. The line-of-best-fit is used in determining the ^{9}Li rate for arbitrary shower-muon thresholds in Chapter 9.

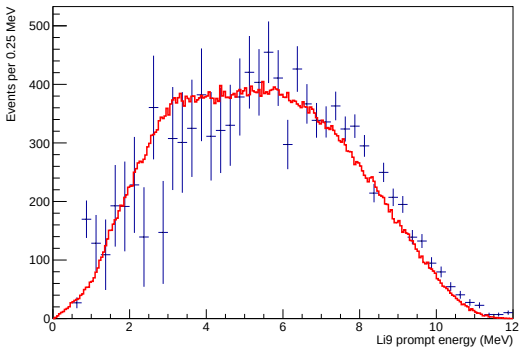

Figure 6.10: Comparison of the extracted ^{9}Li spectrum (blue points) with the predicted spectrum from [51, 55] (red line).

imprecise, they were hand-tuned (within statistical limits) in the simulation so as to achieve satisfactory agreement with the published spectra.

Finally, to obtain the predicted spectrum in terms of prompt energy (i.e. reconstructed energy), the simulated events were passed through a model of the detector nonlinearity. Given that the ^{9}Li/^{8}He spectrum prediction was performed in 2013, it is based on an older nonlinearity model than the one discussed in Section C.2.3.[14] Crucially, this model includes nonlinearity curves for the alpha particle and neutron,

[14]The differences between models are not significant enough to warrant concern, especially in

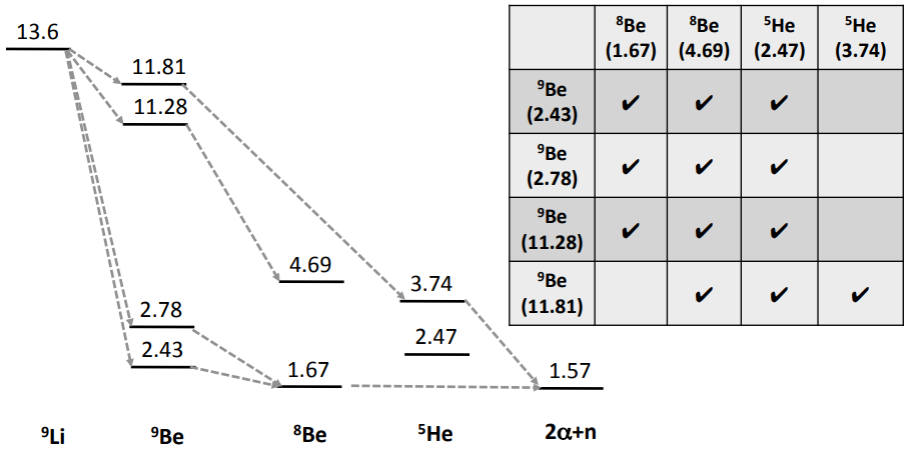

Figure 6.11: The decay of ^{9}Li. Not all decays are shown with arrows. Energy levels are in MeV, relative to the ground state of ^{9}Be. From [51].

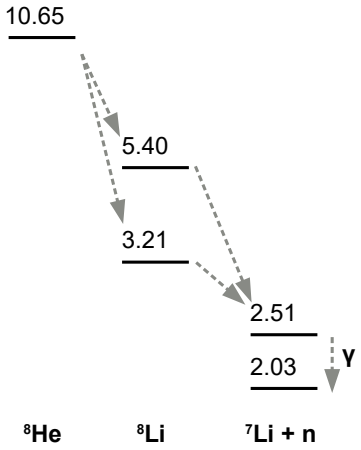

Figure 6.12: The decay of ^{8}He. Energy levels are in MeV, relative to the ground state of ^{8}Li.

whose kinetic energies were also considered when determining the prompt energy for each event [57]. The resulting prompt-energy spectra could then be combined according to the best estimate of the relative proportions of ^{9}Li and ^{8}He. For the sake of this analysis, a nominal 5.5% fraction of ^{8}He was used.[15] Section C.2.4 discusses

light of the uncertainty we assign to the ^{9}Li/^{8}He spectrum, as discussed in [56].

[15]The measured spectrum and the rate fit both give results that are consistent with zero ^{8}He, while rough predictions indicate that the ^{8}He proportion should not exceed 20%. (See the discussion of KamLAND on p. 65.) Meanwhile, the predicted spectrum does not change very significantly when the fraction is varied from 0 to 15%, in comparison to the other sources of uncertainty (neutron and

Parent	Daughter	Level (MeV)	Width (MeV)
^{9}Li	^{9}Be	11.81	0.400
		11.28	0.575
		2.78	0.110
		2.43	0.780×10^{-3}
	^{8}Be	4.69	5.57×10^{-6}
		1.67	1.51
	^{5}He	3.74	5.57
		1.67	0.648
^{8}He	^{8}Li	5.40	0.650
		3.21	1.00
	^{7}Li	2.51	9.02×10^{-9}
		2.03	Stable

Table 6.7: Widths of the nuclear excited states relevant for the prediction of the ^{9}Li and ^{8}He spectra [49]. For the daughters of ^{9}Li (^{8}He), energy levels are expressed relative to the ground state of ^{9}Be (^{8}Li).

the uncertainty assigned to the ^{9}Li/^{8}He spectrum.

6.4 Cosmogenic fast neutrons

As cosmic-ray muons travel through the rock and other materials surrounding the ADs, they can eject energetic neutrons from the medium. If a fast neutron of the appropriate energy thermalizes and stops inside the AD, it will produce an IBD-like coincidence pair, in which the prompt signal consists largely of scintillation from the recoiled protons, and the delayed signal results from nGd capture. This process leads to a background amounting to some 20-30% of that produced by ^{9}Li/^{8}He. This background can be estimated by two complimentary methods, the so-called *extrapolation* and *scaling* methods.

In the extrapolation method, the prompt-energy cut of the IBD selection is extended past 12 MeV to 100 MeV or beyond, where true IBDs are completely absent and the spectrum consists almost entirely of fast-neutrons events (Figure 6.13, black). The high-energy part of the spectrum (above 12 MeV) is then extrapolated below 12 MeV to estimate the fast-neutron component of the IBD sample, using a fit to a well-motivated model of the fast-neutron spectrum.

In the scaling method, a search is performed for the "muon-tagged" IBD-like events in the immediate aftermath of "peripheral" muons that only trigger the outer water

alpha quenching). Hence, 5.5%, an "inherited" feature of this analysis, is as good a guess as any.

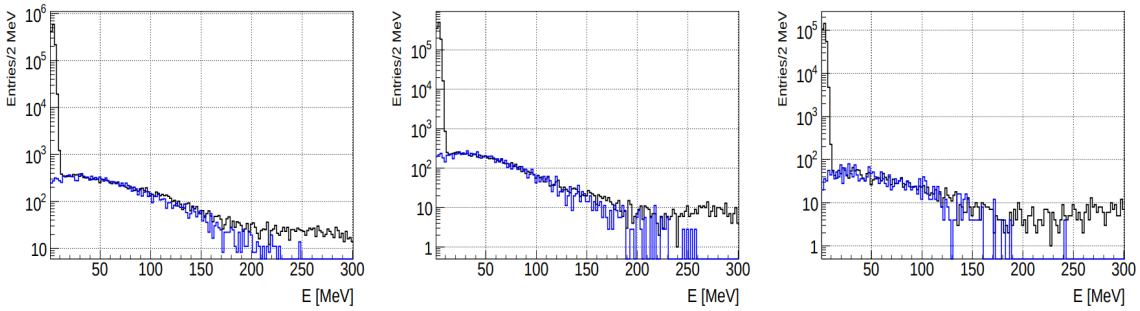

Figure 6.13: Prompt spectra of the muon-tagged (blue) and untagged (black) samples obtained using an IBD selection with an extended prompt-energy cut (so as to include fast neutrons). The tagged spectrum has been rescaled to match the normalization of the untagged spectrum above 12 MeV. From [58].

pool (Figure 6.13, blue). As with the extrapolation method, the prompt-energy cut is significantly extended. In the region above 12 MeV, where true IBDs are absent, a scaling factor can be determined from the ratio of muon-tagged to untagged events. Then, below 12 MeV, the tagged spectrum (which contains very few true IBDs due to the short post-muon time searched) is rescaled according to the scaling factor, yielding an estimate of the fast-neutron spectrum in the sub-12 MeV region.

The two methods are consistent to within 1–3% (an order of magnitude smaller than the estimated uncertainty of each method), providing a high level of confidence in the estimation of the fast-neutron background. This analysis makes use of the fast-neutron rates and shapes determined in [58] and reproduced in Table 6.8. In Section A.2 we provide a more detailed review of the rate estimation, the spectral shape, and their uncertainties.

EH1	EH2	EH3
0.843 ± 0.083	0.638 ± 0.062	0.053 ± 0.009

Table 6.8: Final estimated fast-neutron rates (per AD per day) [58].

6.5 AmC source

To study the response of the detectors to neutrons, each AD was initially configured with a low-intensity (∼0.7 Hz) ^{241}Am-^{13}C neutron source in each of the three automated calibration units (ACUs) housed on the detector's lid. When not in use, the location and shielding of the sources ensures that neutrons do not infiltrate the GdLS region,

protecting against correlated backgrounds from proton recoils followed by neutron capture.

In spite of these precautions, a rare mechanism can still produce correlated backgrounds: A neutron may scatter inelastically against Fe, Cr, Mn, or Ni in the stainless steel, producing prompt gamma rays (generally totaling less than 3 MeV of reconstructed energy), before thermalizing and being captured either by one of these four elements or by Gd in the GdLS overflow tank (producing a signal between 6 and 12 MeV). A few of these gamma rays reach the scintillator. When this happens for both the prompt and delayed gamma rays, the resulting pair can sometimes pass the IBD selection criteria.

The measurements of this background's rate and shape were published in [59]. The key to the measurement is the fact that the rate of AmC correlated backgrounds is proportional to the rate of delayed-like *uncorrelated* events originating from the AmC source:

$$R_{\text{corr}} = R_{\text{uncorr}} \times \xi \tag{6.16}$$

The rate R_{uncorr} can be measured, with high statistics, from the top/bottom asymmetry in the rate of delayed-like singles. Meanwhile, the proportionality constant ξ can be determined from MC simulations. As described in [59], the simulations were benchmarked using a high-activity AmC source (HAS) placed on top of EH3-AD4. For the HAS, ξ could be directly measured (as well as the prompt spectrum of the correlated events); comparison to the MC's predicted HAS ξ then produced a correction for the MC-derived ξ of the ordinary AmC source. Insertion of this corrected ξ, together with the measured R_{uncorr}, then gives the rate of the background.

In this analysis, we use the AmC background rates from the analysis in [60], where the methods of [59] were repeated using the full P17B dataset. The rates and their uncertainties are summarized in Table 6.9. A more detailed review of this background can be found in Section A.3.

	EH1		EH2		EH3			
Period	AD1	AD2	AD1	AD2	AD1	AD2	AD3	AD4
6AD	0.29 ± 0.13	0.27 ± 0.12	0.30 ± 0.14		0.24 ± 0.11	0.23 ± 0.10	0.23 ± 0.10	
8AD	0.15 ± 0.07	0.15 ± 0.07	0.12 ± 0.06	0.14 ± 0.06	0.04 ± 0.02	0.03 ± 0.01	0.03 ± 0.02	0.04 ± 0.02
7AD		0.11 ± 0.05	0.09 ± 0.04	0.08 ± 0.04	0.02 ± 0.01	0.02 ± 0.01	0.03 ± 0.01	0.02 ± 0.01

Table 6.9: AmC background rates for the P17B data set [60].

6.6 $^{13}C(\alpha, n)^{16}O$

A final and relatively minor ($\lesssim 0.01\%/0.07\%$ near/far[16]) correlated background arises from (α, n) reactions initiated by natural radioactivity within the detector. In these reactions, an alpha particle from natural radioactivity is captured by a nucleus, which then emits a neutron. A prompt signal arises from a number of sources of energy deposition, including the kinetic energy of the alpha particle, gamma rays from nuclear deexcitation (including potentially those from inelastic scattering of the neutron), and proton recoils caused by the neutron. This prompt signal is then followed by capture of the neutron, mimicking the signature of an IBD.

This background can be estimated based on the techniques described in [61] and [19]. The first step in this estimation is to measure the rates of the decay chains responsible for the alpha-particle emission; namely, the uranium, thorium, and actinium decay chains, as well as the decay of ^{210}Po. With the exception of ^{210}Po, this can be done by searching for the double-coincidence signature of the Bi-Po decay cascade within each chain. The decay rate of ^{210}Po, meanwhile, can be measured via the 5.3-MeV alpha particles produced by the decay, visible (after quenching) as a peak around 0.5 MeV in the singles spectrum.

Along with the rates and energies of these alpha-particle decays, there is another ingredient needed to estimate the $^{13}C(\alpha, n)^{16}O$ background rate, namely, probabilities and spectra (of reaction products) of $^{13}C(\alpha, n)^{16}O$ reactions for each individual alpha-particle peak. This can be determined from Monte Carlo. Combining the measured alpha-decay rates with the simulated reaction probabilities (and spectra) then gives the $^{13}C(\alpha, n)^{16}O$ background prediction. In this analysis, we use the updated estimates from [60], which employ the alpha-decay rates measured from the entire P17B dataset. These rates are listed in Table 6.10. A more detailed review is provided in Section A.4.

	EH1		EH2		EH3			
Period	AD1	AD2	AD1	AD2	AD1	AD2	AD3	AD4
6AD	0.09 ± 0.04	0.07 ± 0.04	0.05 ± 0.02		0.05 ± 0.02	0.04 ± 0.02	0.04 ± 0.02	
8AD	0.08 ± 0.04	0.06 ± 0.03	0.04 ± 0.02	0.06 ± 0.03	0.04 ± 0.02	0.04 ± 0.02	0.03 ± 0.02	0.04 ± 0.02
7AD		0.05 ± 0.03	0.03 ± 0.02	0.06 ± 0.03	0.03 ± 0.02	0.03 ± 0.02	0.03 ± 0.01	0.03 ± 0.02

Table 6.10: $^{13}C(\alpha, n)^{16}O$ background rates for the P17B data set [60].

[16] As with the AmC background, the choice of denominator here (between signal and signal+background) makes no difference at the level of one significant figure.

Chapter 7

Accidental-background rate and multiplicity-cut efficiency

We combine our discussion of these two quantities because they both depend on the rate of uncorrelated physics events, or *singles.* The accidental-background rate ("accidentals rate") is determined by the probability of two singles occurring in the same coincidence window, while the efficiency of the decoupled multiplicity cut (DMC) is similarly based on the chance of one or more singles occurring within a certain distance in time from the delayed event.[1]

Let $R_{\mathrm{s}}(E_{\mathrm{min}}, E_{\mathrm{max}})$ be the rate of singles whose reconstructed energy lies between E_{min} and E_{max} (in MeV). To be precise, R_{s} is the *true physical rate* of all *muon-uncorrelated* processes that produce such singles. Naively, one could attempt to calculate R_{s} by counting all non-muon-vetoed triggers in $(E_{\mathrm{min}}, E_{\mathrm{max}})$ and dividing by the veto-corrected DAQ livetime. However, the rate will then be overestimated due to the inclusion of triggers from correlated events, and, likewise, the predicted accidentals spectrum will be distorted.

Instead, the correct approach is to apply an isolation cut (in time) to ensure a clean sample of true singles. A correction must then be applied for the efficiency of this cut. Once R_{s} has been obtained in this way, calculation of the accidentals rate and DMC efficiency is a straightforward application of Poisson statistics.

7.1 Event classes

Let us define a *muon-like* event as an AD trigger with charge of at least 3,000 p.e. (about 18 MeV), corresponding to the charge threshold of the AD-muon veto in our IBD selection (Section 5.2.2.1) Then a *sub-muon* event is one with reconstructed energy of at least 12 MeV, but not enough charge to be muon-like. A *prompt-like* event has

[1] IBDs and correlated backgrounds also contribute to the inefficiency of the DMC, but the effect is negligible given the vast disparity in rates between singles and correlated events.

energy of 0.7–12 MeV, a *delayed-like* event 6–12 MeV, and, finally, a *prompt-plus* event is one that is either prompt-like or a sub-muon. The sample described in Section 7.2 consists of prompt-plus events.

Similarly, we define the *prompt-like rate* R_{p} as $R_{\mathrm{s}}(0.7, 12)$, the *delayed-like rate* R_{d} as $R_{\mathrm{s}}(6, 12)$, the *sub-muon rate*[2] R_{λ} as $R_{\mathrm{s}}(12, \sim 18)$,[3] and the *prompt-plus rate* R_{+} as $R_{\mathrm{p}} + R_{\lambda}$. Likewise, the event counts in our sample are N_{p}, etc. To complete this round of definitions, let ϵ_{i} be the isolation cut efficiency, ϵ_{μ} the muon cut efficiency (for the singles selection), and T be the raw DAQ livetime of the sample.

7.2 Singles selection

We begin with a sample of singles. As described in Section 5.3, these are events that meet the following criteria:

1. Not a flasher or forced trigger.
2. Not in a muon veto window.
3. AdSimple energy of at least 0.7 MeV.
4. Charge of less than 3,000 p.e. (i.e., not muon-like)
5. No other such events within 400 µs before the event
6. No other such events *between 6 and 12 MeV* within 200 µs after the event

The muon veto conditions need not be the same as those used in the IBD selection, as long as they are sufficiently stringent (as explained in Section 7.3.1). Likewise, the final two conditions (the *isolation cut*) can be chosen somewhat arbitrarily, provided that the window is large enough to remove correlated triggers and small enough to provide sufficient statistics. The actual employed isolation cut is designed to mimic the IBD selection's multiplicity cut as closely as possible, as discussed in Section 5.3.

7.3 Muon veto considerations

7.3.1 Avoiding muon-correlated events

In the selection criteria described in Section 7.2, the final one (the isolation cut) requires "no other events passing the above cuts within time window $\pm t$". Since

[2] λ precedes μ in the Greek alphabet. Right?

[3] Technically, we define a sub-muon as an event that has a reconstructed energy above 12 MeV and that is *not classified as muon-like.* Thus an event with an energy slightly above 18 MeV, but a reconstructed charge below the 3,000-PE muon threshold, will be classified as a sub-muon.

"the above cuts" includes "not in a muon veto window", this implies that an extra *muon-correlated* trigger will *not* lead to event rejection by the isolation cut.

This formulation of the isolation cut is necessary for the mathematical consistency of the calculation of its efficiency: We will be assuming that the events we are measuring—muon-uncorrelated singles—belong to the exact same class as the "extra" events that enter the definition of the isolation cut.

However, at first glance this poses a problem: Our goal is to measure muon-uncorrelated events, but if we allow a potential "single" to be preceded by a muon-correlated trigger, then the "single" may itself be the product of muon activity.

Fortunately, this problem can be easily eliminated by an appropriate choice of the muon veto window. In particular, we need to ensure that, if there is a muon-correlated event lying inside the isolation window preceding a singles candidate, the same muon *will veto our candidate*. Previous studies have shown that the majority of muon-induced activity occurs within 400 μs of the muon. Therefore, if we veto at least $t + 400$ μs after a muon (where t is the length of the isolation cut window), the presence of a muon-correlated trigger guarantees that the candidate will get rejected, albeit by the muon veto, not by the isolation cut. During singles selection, therefore, we are careful to use muon veto windows that are sufficiently long.

7.3.2 Decoupling of efficiencies

In the calculations that follow, we assume that

$$\epsilon_{\text{tot}} = \epsilon_\mu \epsilon_{\text{i}}, \tag{7.1}$$

that is, that we can calculate the muon-veto and isolation-cut efficiencies independently, and get the total efficiency by multiplying the two. This assumption is valid only if the two occurrences—either a muon, or an "extra" single—are statistically independent. As was explained above, extra muon-correlated triggers are *not* considered when applying the isolation cut. Therefore, the presence of a muon has no effect on the probability of passing the isolation cut, and, conversely, the presence of an extra uncorrelated single has no effect on the probability of being inside a muon veto window. The requirements for statistical independence are thus satisfied.

As a final verification of Equation 7.1, let us divide rejected events into three non-overlapping categories: (a), events rejected only by the muon veto, (b), events rejected by the isolation cut, and (ab), events rejected by both. Since the two cuts are independent, $P_{\text{ab}} = P_{\text{a}} P_{\text{b}}$. We have:

$$\begin{aligned}
\epsilon_{\text{tot}} &= 1 - P_{\text{a}} - P_{\text{b}} - P_{\text{ab}} \\
&= 1 - (1 - \epsilon_\mu)\,\epsilon_{\text{i}} - \epsilon_\mu\,(1 - \epsilon_{\text{i}}) - (1 - \epsilon_\mu)(1 - \epsilon_{\text{i}}) \\
&= \epsilon_\mu \epsilon_{\text{i}}
\end{aligned}$$

Clearly, Equation 7.1 indeed applies. The muon-veto efficiency is determined simply by keeping track of the total vetoed livetime, while the isolation-cut efficiency is described next.

7.4 Isolation cut efficiency

Let "L" and "R" (i.e. left and right) refer to the two classes of events that would lead to rejection of a singles candidate, should such an event be measured within a time window of (respectively) t_{L} before or t_{R} after the candidate. Let N_{L} and R_{L} denote, respectively, the raw measured count and the true rate of L-class events, and likewise for N_{R} and R_{R}. Also let T denote the total livetime.

As a starting point, we have the following identities:

$$\begin{aligned} N_{\mathrm{L}} &= \epsilon_\mu \epsilon_{\mathrm{i}} R_{\mathrm{L}} T \\ N_{\mathrm{R}} &= \epsilon_\mu \epsilon_{\mathrm{i}} R_{\mathrm{R}} T \end{aligned} \tag{7.2}$$

in which ϵ_μ and T and each N are known, while the unknowns are ϵ_{i} and each R. In addition, the Poisson distribution implies that

$$\epsilon_{\mathrm{i}} = e^{-R_{\mathrm{L}} t_{\mathrm{L}} - R_{\mathrm{R}} t_{\mathrm{R}}}. \tag{7.3}$$

This is simply the probability of observing zero L-class events in a time window of length t_{L}, and zero R-class events in a t_{R}-length window. Combining Equation 7.2 and Equation 7.3, we have:

$$\begin{aligned} N_{\mathrm{L}} &= \epsilon_\mu e^{-R_{\mathrm{L}} t_{\mathrm{L}} - R_{\mathrm{R}} t_{\mathrm{R}}} R_{\mathrm{L}} T \\ N_{\mathrm{R}} &= \epsilon_\mu e^{-R_{\mathrm{L}} t_{\mathrm{L}} - R_{\mathrm{R}} t_{\mathrm{R}}} R_{\mathrm{R}} T \end{aligned}$$

or, after rearranging and multiplying the top and bottom equations by t_{L} and t_{R}, respectively,

$$\begin{aligned} e^{R_{\mathrm{L}} t_{\mathrm{L}} + R_{\mathrm{R}} t_{\mathrm{R}}} N_{\mathrm{L}} t_{\mathrm{L}} &= \epsilon_\mu R_{\mathrm{L}} t_{\mathrm{L}} T \\ e^{R_{\mathrm{L}} t_{\mathrm{L}} + R_{\mathrm{R}} t_{\mathrm{R}}} N_{\mathrm{R}} t_{\mathrm{R}} &= \epsilon_\mu R_{\mathrm{R}} t_{\mathrm{R}} T. \end{aligned}$$

Adding these two coupled equations gives

$$e^{R_{\mathrm{L}} t_{\mathrm{L}} + R_{\mathrm{R}} t_{\mathrm{R}}} (N_{\mathrm{L}} t_{\mathrm{L}} + N_{\mathrm{R}} t_{\mathrm{R}}) = \epsilon_\mu (R_{\mathrm{L}} t_{\mathrm{L}} + R_{\mathrm{R}} t_{\mathrm{R}}) T \tag{7.4}$$

which, after some rearrangement, gives:

$$(-R_{\mathrm{L}} t_{\mathrm{L}} - R_{\mathrm{R}} t_{\mathrm{R}}) e^{-R_{\mathrm{L}} t_{\mathrm{L}} - R_{\mathrm{R}} t_{\mathrm{R}}} = -\frac{N_{\mathrm{L}} t_{\mathrm{L}} + N_{\mathrm{R}} t_{\mathrm{R}}}{\epsilon_\mu T}. \tag{7.5}$$

This equation takes the form $we^w = z$, which cannot be solved for w using elementary functions. Instead we employ the Lambert W function, defined as the inverse of the

map $w \mapsto we^w$ (that is, $W(we^w) = w$). As shown in Figure 7.1, the W function has two branches. We know that, for an O(1 ms) isolation window, $R_\mathrm{L} t_\mathrm{L} \ll 1$ (and likewise for R_R), implying that the correct choice is the upper branch W_0. Then

$$R_\mathrm{L} t_\mathrm{L} + R_\mathrm{R} t_\mathrm{R} = -W_0\left(-\frac{N_\mathrm{L} t_\mathrm{L} + N_\mathrm{R} t_\mathrm{R}}{\epsilon_\mu T}\right) \tag{7.6}$$

Finally, inserting this into Equation 7.3, we obtain the isolation-cut efficiency:

$$\epsilon_\mathrm{i} = \exp W_0\left(-\frac{N_\mathrm{L} t_\mathrm{L} + N_\mathrm{R} t_\mathrm{R}}{\epsilon_\mu T}\right). \tag{7.7}$$

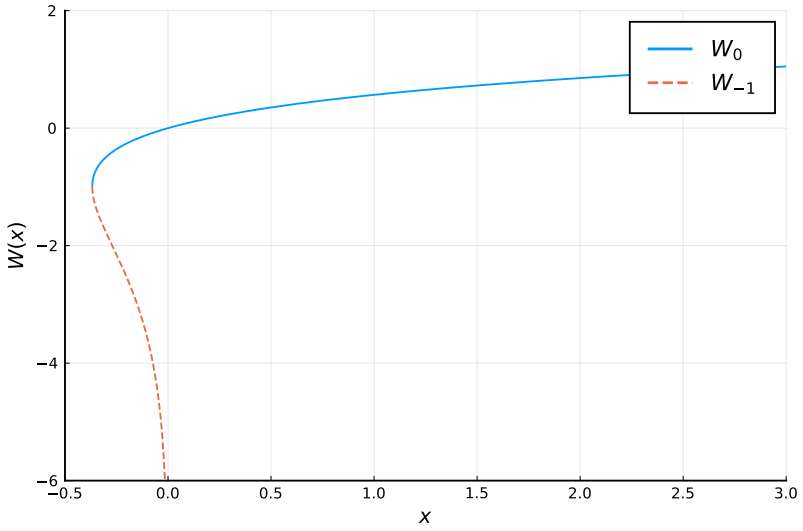

Figure 7.1: The two branches of the Lambert W function.

7.5 Singles rates

Once the isolation-cut efficiency is known, it is trivial to calculate the prompt- and delayed-like singles rates:

$$R_\mathrm{p} = \frac{N_\mathrm{p}}{\epsilon_\mathrm{i}\epsilon_\mu T}, \qquad R_\mathrm{d} = \frac{N_\mathrm{d}}{\epsilon_\mathrm{i}\epsilon_\mu T}, \tag{7.8}$$

and so on. These quantities are essential for calculating the DMC efficiency and accidentals rate, as described below.

7.6 DMC efficiency

In order for an IBD candidate to satisfy the decoupled multiplicity cut, there must be no delayed-like triggers in the 200 μs following the delayed event, and no extra prompt-plus events[4] in the 400 μs prior to the delayed event. More formally,

$$\begin{aligned} \epsilon_{\mathrm{m}} &= P(0; R_{+} \cdot 2\tau)\ P(0; R_{\mathrm{d}} \cdot \tau) \\ &= \exp[-(2R_{+}\tau + R_{\mathrm{d}}\tau)], \end{aligned} \tag{7.9}$$

where $\tau = 200$ μs.

7.7 Accidentals rate

7.7.1 Rate calculation

An accidental event, in order to enter the IBD selection, must pass the DMC, as with any other IBD candidate. Since the DMC is "centered" on the delayed event of the pair, it is simplest to calculate the accidentals rate by similarly centering the calculation on the delayed event.

Given an uncorrelated delayed-like trigger, we can calculate the probability that it forms the delayed half of an accidental IBD candidate. Letting all time intervals be defined in relation to the delayed-like trigger, there are four conditions:

1. Exactly one prompt-like event in [-200, 0] μs.
2. No sub-muon events in [-200, 0] μs.
3. No prompt-plus (prompt-like or sub-muon) events in [-400, -200] μs.
4. No delayed-like events in [0, 200] μs.

The accidentals rate is then simply the product of these probabilities multiplied by the delayed-like singles rate:

$$\begin{aligned} R_{\mathrm{acc}} &= R_{\mathrm{d}} \cdot P(1; R_{\mathrm{p}}\tau) \cdot P(0; R_{\lambda}\tau) \cdot P(0; R_{+}\tau) \cdot P(0; R_{\mathrm{d}}\tau) \\ &= R_{\mathrm{d}} \cdot R_{\mathrm{p}}\tau e^{-R_{\mathrm{p}}\tau} \cdot e^{-R_{\lambda}\tau} \cdot e^{-R_{+}\tau} \cdot e^{-R_{\mathrm{d}}\tau} \end{aligned} \tag{7.10}$$

Note that this calculation incorporates the inefficiency of the DMC (but not of the muon veto cut). By convention, the oscillation fitter expects background rates to be provided as "theoretical" rates, that is, the rate expected if all cuts were perfectly efficient. Internally, the fitter multiplies these rates by the DMC and veto efficiencies.

[4] Formerly, in the LBNL analysis, only extra *prompt*-like events would lead to a DMC rejection. In order to eliminate the double neutron background, "prompt-like" was later changed to "prompt-plus".

Therefore, we must divide out the DMC efficiency when providing the accidentals rate to the fitter:

$$R'_{\mathrm{acc}} = \frac{R_{\mathrm{acc}}}{\epsilon_{\mathrm{m}}}. \tag{7.11}$$

7.7.2 Statistical uncertainty

The accidentals rate, as calculated using Equation 7.10, carries a statistical uncertainty which derives from that of the rates that enter the calculation: the prompt-like rate, delayed-like rate, etc. In turn, these rates, as specified by Equation 7.8, derive their own uncertainties from those of the event counts in each classification. The uncertainties of the counts are trivial, being merely the Poisson uncertainty of $\sqrt{N}$. However, once we consider the *rates* rather than the *counts*, we begin to encounter complications. Most notably, the isolation cut efficiency ϵ_{i} in Equation 7.8 is *also* a function of the event counts, as indicated by Equation 7.7. This nonlinear dependence on the event counts is only made more complicated when the rates are inserted into Equation 7.10. Furthermore, the event classifications in question are not always statistically independent: For instance, the prompt-like and delayed-like counts are correlated by virtue of the fact that they both include events in the 6–12 MeV region.

In spite of these complications, it is still possible to undertake a theoretical calculation of the statistical uncertainty on the accidentals rate. This would involve expanding Equation 7.10 in terms of the various event counts N_{p}, N_{d}, etc., and then re-expressing the event counts (such as the prompt-like and delayed-like counts) as sums of counts of *disjoint* event classes N_1, N_2, etc. (e.g., an 0.7–6 MeV class and a 6–12 MeV class). This complicated function $R_{\mathrm{acc}}(N_1, N_2, \ldots)$ could then be expanded in a power series using computer algebra. From that, an expression for the variance $\langle R_{\mathrm{acc}}^2 \rangle - \langle R_{\mathrm{acc}} \rangle^2$ could be obtained up to any desired order, and evaluated using the known cumulants of the Poisson distribution.

Given the complexities of this procedure, and the fact that previous studies have established this statistical uncertainty to be significantly smaller than the systematic uncertainty, the author has no desire to pioneer such an undertaking. Nevertheless, it remains desirable to have some way of evaluating the statistical uncertainty in order to verify its relative insignificance. For this reason, a simple Monte Carlo procedure was developed for calculating this uncertainty. We proceed to discuss its implementation and results.

7.7.2.1 Procedure

To recap, only three inputs are required to calculate the accidentals rate:

- The livetime.
- The muon veto efficiency (for correcting the livetime).

- The energy histogram of all events selected by the singles selection.

In turn, to study the statistical uncertainty σ of the calculation, we can apply statistical fluctuations to the energy histogram. These fluctuations will automatically propagate to the event counts N_p, etc., and then to the final result. Repeating this procedure will produce a sample of accidentals rates, from which σ can be determined by taking the standard deviation or by fitting a Gaussian function.

For each bin i of the singles' energy histogram, with measured count $\overline{N}_i$, let N_i denote the corresponding random variable describing a hypothetical ensemble of experiments. The PDF $f(N_i)$ is then given by the Poisson distribution:

$$f(N_i) = \frac{\overline{N}_i^{N_i} e^{-\overline{N}_i}}{N_i!}. \tag{7.12}$$

This distribution is sampled for each bin, resulting in a fluctuated singles histogram, which is then used as the input for the calculation of the accidentals rate described in Section 7.7.1. We repeat this procedure 10,000 times for each AD (and data period) in order to obtain a sample of accidentals rates that approximates their statistical distribution, as illustrated in Figure 7.2. Finally, the statistical uncertainty is determined from this distribution in two ways: First, by directly calculating the standard deviation, and second, by performing a fit to a Gaussian function. Both methods produce consistent results.

As shown in Figures 7.3–7.6, the statistical uncertainty of the accidentals rate is relatively small, remaining below the 1% level even in the lowest-statistics case (EH3 in the 7AD period). Rather than using these uncertainties directly, we absorb them into a conservative 2% total uncertainty in order to account for possible systematic effects, as discussed in Section 7.7.3.

7.7.3 Systematic and total uncertainty

The systematic uncertainty on the accidentals rate can be estimated by comparing our method to an independent technique. Rather than carrying this out ourselves, we refer to previous studies that performed such comparisons [19]. In those studies, the alternative calculations used the so-called *offset-window* method, in which the IBD selection is repeated with the requirement of a time offset (ranging from 1 to 20 ms) between the prompt and delayed triggers. This offset has the effect of largely eliminating correlated events, producing a sample enriched in accidentals. From this sample, one can obtain the distribution of the spatial separations between the prompt and delayed triggers in each pair. Finally, for the true IBD sample, the prompt-delayed distribution can be plotted, and the distribution for the accidental sample can be normalized to fit it, giving an estimate of the total number of accidentals in the true sample (Figure 7.7). This method was shown to give consistent results with methods based on measuring the singles rates (such as our method), with a slightly

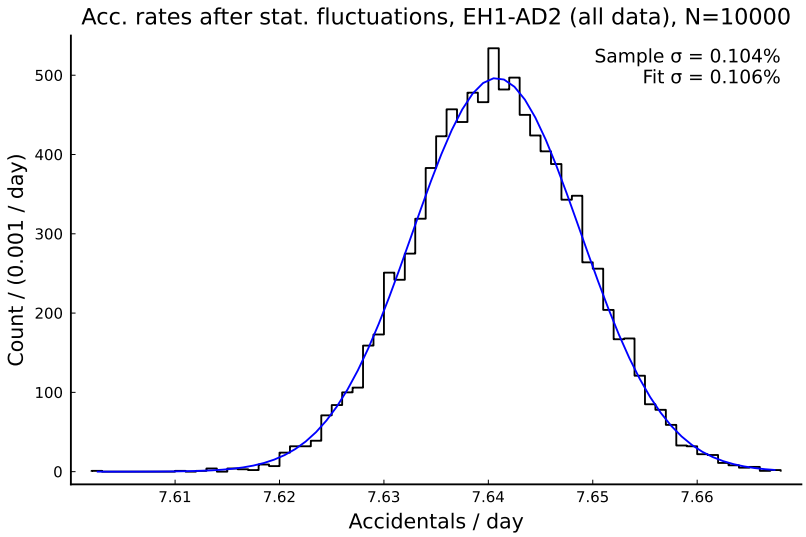

Figure 7.2: Distribution of full-dataset accidentals rates in EH1-AD2 calculated under statistical fluctuations of the singles rate and spectrum. Consistent estimates of the spread are obtained from directly calculating the standard deviation and from performing a fit to a Gaussian function.

higher statistical uncertainty of about 1%. To obtain our own final uncertainty on the accidentals rate, we take this 1% and conservatively add it linearly to our own 1% worst-case statistical uncertainty, giving 2%. Although this is an overestimate, it has a minimal effect on the final uncertainty of the oscillation parameters, given that the total background uncertainty remains dominated by ^{9}Li/^{8}He.

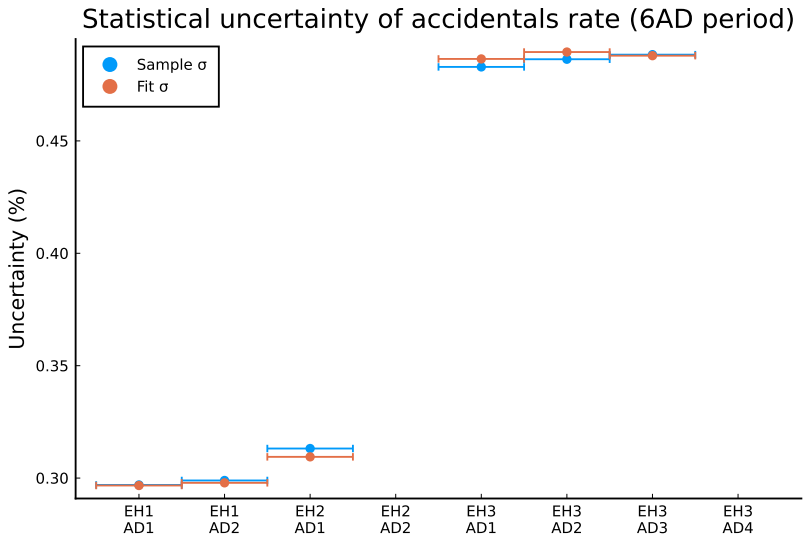

Figure 7.3: Statistical uncertainty of the accidentals rates in each AD for the 6AD period. "Samples σ" refers to the directly-calculated standard deviation, while "Fit σ" refers to the results of fitting to a Gaussian function.

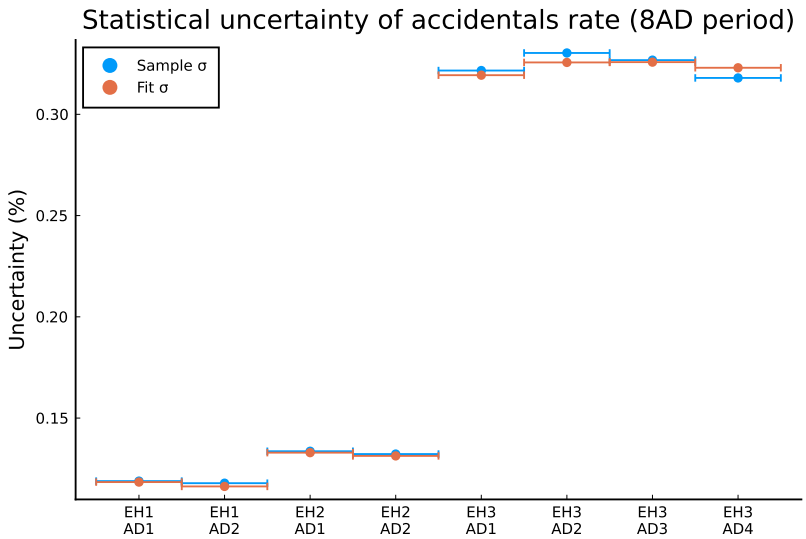

Figure 7.4: Statistical uncertainty of the accidentals rates in each AD for the 8AD period. "Samples σ" refers to the directly-calculated standard deviation, while "Fit σ" refers to the results of fitting to a Gaussian function.

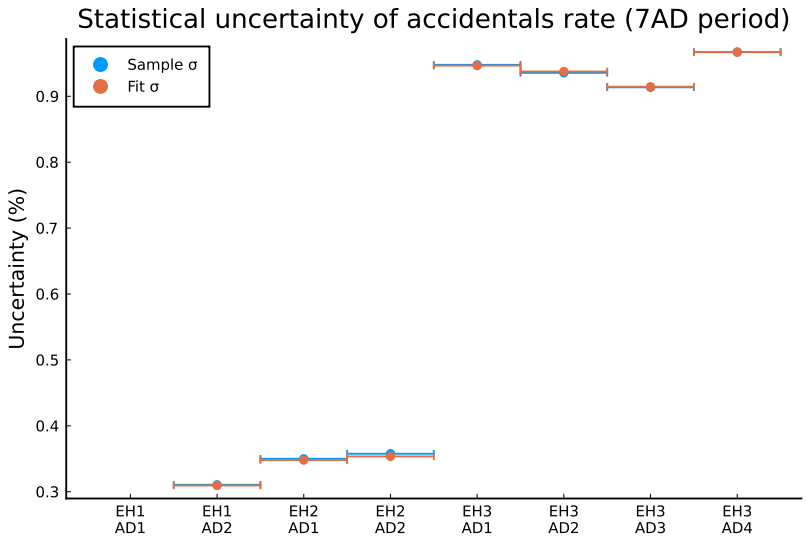

Figure 7.5: Statistical uncertainty of the accidentals rates in each AD for the 7AD period. "Samples σ" refers to the directly-calculated standard deviation, while "Fit σ" refers to the results of fitting to a Gaussian function.

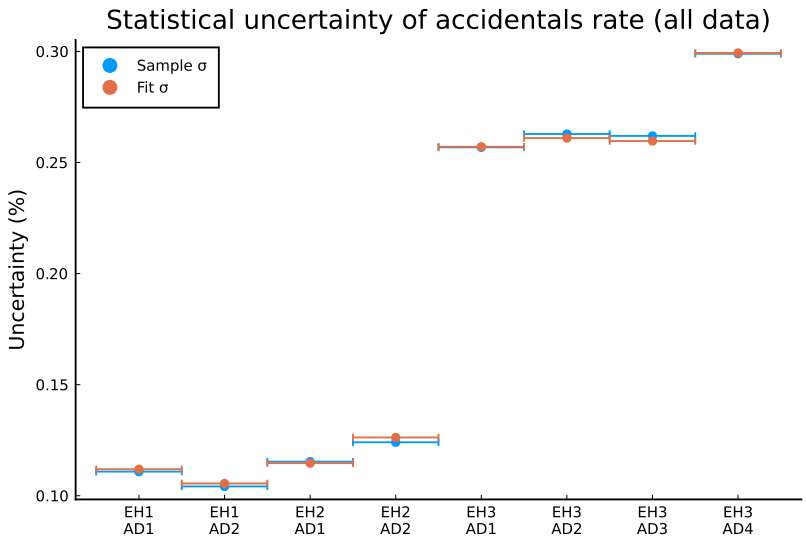

Figure 7.6: Statistical uncertainty of the accidentals rates in each AD for the full dataset. "Samples σ" refers to the directly-calculated standard deviation, while "Fit σ" refers to the results of fitting to a Gaussian function.

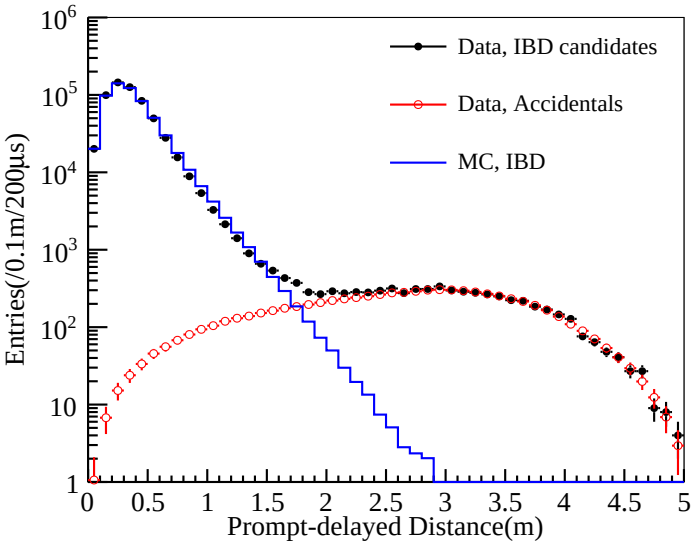

Figure 7.7: Distribution of prompt-delayed distances for a true IBD sample and for an accidentals-enriched sample, as obtained when using the offset-window method of measuring the accidentals rate. The particular AD shown is unspecified. From [19].

Chapter 8

Oscillation fit

8.1 Introduction

At this stage in the analysis, we have a sample of IBD candidates, the background predictions, and the calculated efficiency corrections (for the muon veto and the multiplicity cut). The next step is to take the background-subtracted and efficiency-corrected IBD prompt spectra at the eight ADs, and fit them to the 3-flavor oscillation model in order to extract $\sin^2 2\theta_{13}$ and Δm^2_{ee}.

Our analysis is based on the Daya Bay oscillation fitter developed at Lawrence Berkeley National Laboratory (LBNL) [56, 62], which is one of a half-dozen independent fitters used in the production and cross-checking of Daya Bay's official oscillation fits. While it is relatively straightforward to find the oscillation parameters that best fit Daya Bay's data, determining their uncertainties is significantly more complicated, given the need to consider a wide range of systematics (e.g., those related to the reactors, the detectors, and the backgrounds). The modeling and methods used by the LBNL fitter have been carefully developed and validated over the years, allowing our analysis to benefit from these efforts. Furthermore, our use of the LBNL fitter ensures that our cut-variation studies are directly applicable to Daya Bay's past and future results.

The LBNL fitter is a *relative* fitter: That is, instead of simultaneously comparing all eight ADs to a model of the reactor spectra and oscillation probability (an *absolute* fit), it uses the measurements of the near ADs to predict the far AD spectra (including oscillation effects). This approach ensures that the results are independent (to first-order) of the modeling used for the reactor spectra. A reactor model is still required in order to decompose each near AD's spectrum into the components from individual cores, but the final result is negligibly affected by the details of the model. Among the Daya Bay fitters developed outside of LBNL, some take the absolute approach, and they too are able to produce results that are stable across reactor models, but this requires using explicit nuisance parameters (or "pull terms") to represent the

uncertainties in the reactor modeling.

Aside from its use of a relative approach, a second defining characteristic of the LBNL fitter is its use of a covariance matrix to encode all systematic (as well as statistical) uncertainties. This ensures that finding the best-fit oscillation parameters is a simple two-dimensional minimization problem, with an objective function of the form

$$\chi^2 = (F_i - \overline{F}_i)V_{ij}^{-1}(F_j - \overline{F}_j). \tag{8.1}$$

where the index i runs over the energy bins in the far hall, the F_i are the measured far-hall spectra, the $\overline{F}_i$ are the predicted far-hall spectra based on the near-hall observations, and V_{ij} is the total covariance matrix (including both systematic and statistical components). All of the complexity lies in the determination of V_{ij} (which we discuss shortly).

In contrast to the covariance matrix approach, it is possible to parameterize each systematic with a pull term η_j (and a corresponding nominal value $\bar{\eta}_j$ and variance ζ_j^2), resulting in an objective function akin to

$$\chi^2 = \sum_i \frac{(F_i - \overline{F}_i)^2}{\sigma_i^2} + \sum_j \frac{(\eta_j - \bar{\eta}_j)^2}{\zeta_j^2}. \tag{8.2}$$

In this case, the predictions $\overline{F}_i$ are implicitly dependent on the pull terms η_j. Performing the fit then involves finding both the oscillation parameters *and the pull terms* that minimize χ^2. This is a high-dimensional minimization problem, with an attendant computational cost and need to avoid false minima, but it avoids the calculation of a complex covariance matrix. Both methods, covariance-matrix-based and pull-term-based, can be used in either a relative or an absolute approach, leading to a total of four options. Each option has benefits and drawbacks from the standpoint of implementation complexity; however, assuming the use of consistent modeling, all four should give consistent final results, as has indeed been demonstrated repeatedly during the Collaboration's internal cross-checks. We do not make any claim of superiority for the LBNL fitter's relative, covariance-matrix approach; it is simply what we have chosen to use. Further discussion of these different methodologies can be found in Section C.1.

As was mentioned, the price of a simple two-dimensional minimization is that we must first calculate the covariance matrix V_{ij}. This matrix encodes the variance in each energy bin, and its covariance with other bins, as the combination of contributions from *all* sources of uncertainty, systematic and statistical. For any reasonably complex experiment such as Daya Bay, this cannot be done analytically, or even (deterministically) numerically. Instead, Monte Carlo (MC) must be used. The LBNL fitter includes a toy MC for this purpose. The toy MC generates sets of predicted prompt spectra at the eight ADs. Each set ("toy sample") is generated under randomized assumptions regarding reactor, detector, and background uncertainties. From a large

ensemble of toy samples, the covariance between each pair of energy bins can be calculated, giving a partial covariance matrix that accounts only for the detector and background systematics. This is then combined with an analytically calculated statistical covariance matrix, giving the full matrix used in this fit.

In the remainder of this chapter, we provide basic overviews of the reactor model, the toy MC, and the fitter, as well as a summary of our inputs and results. Further details can be found in Appendices B and C.

8.2 Reactor model

Although the details of the reactor model have minimal bearing on the final result, a model must still be chosen in order to decompose each near AD's spectra into individual reactor contributions, for extrapolation to the far hall. In keeping with the official practice of the Collaboration, we use the results of Huber [63] and Mueller [64]. In particular, for fissions of ^{235}U, ^{239}Pu, and ^{241}Pu, we use Huber's conversions of the total β spectra measured at Institut Laue-Langevin (ILL), while for ^{238}U, we use Mueller's *ab initio* calculations based on nuclear data. These predictions include time-dependent corrections for each reactor's departure from equilibrium, as well as additional contributions from the spent nuclear fuel (SNF) pools near each core.

The four individual isotope spectra must then be appropriately scaled and combined to obtain the total prediction for each core. This is accomplished using data provided by the power plant on each core's weekly average power and fission fractions. This allows the total nominal spectrum $\Phi^0_{c,t}(E_\nu)$ (per unit time) at each core c to be determined, for week t, as:

$$\Phi^0_{c,t}(E_\nu) = \sum_k \left(\frac{W_{c,t}\, f_{k,t}}{\sum_k f_{k,t}\, e_k} S_k(E)\, c^{\mathrm{ne}}_{k,t}(E) \right) + S^{\mathrm{snf}}_t(E), \tag{8.3}$$

where the index k runs over the four fission isotopes, W_t is the average thermal power, the $f_{k,t}$ are the fission fractions, the e_k are the average energy released per fission, the $S_k(E)$ are the Huber-Mueller isotope spectra, the $c^{\mathrm{ne}}_{k,t}(E)$ are the off-equilibrium corrections, and $S^{\mathrm{snf}}_t(E)$ is the SNF contribution.

Livetime-weighted summing of these weekly spectra then gives the predictions for the entire data period:

$$\Phi^0_c(E_\nu) = \frac{\sum_t T_t \Phi^0_{c,t}(E_\nu)}{\sum_t T_t}, \tag{8.4}$$

where the T_t are the weekly livetimes. The toy MC can then feed these spectra into its model of propagation/oscillation, the IBD cross-section, and the detector response, producing a toy sample of the predicted prompt spectrum at each AD.

Aside from the predicted core spectra, the reactor model is also responsible for producing the reactor-related covariance matrix. This is done by quantifying all

systematic uncertainties, include those on the weekly power, the fission fractions, the isotope spectra, and the off-equilibrium correction [65]. The model is then used to generate samples of core spectra under variations of these systematics. Letting Φ_a denote the vector containing the binned spectra from all the cores (such that the index a runs over bins and cores), the reactor covariance matrix V_{reac} is

$$(V_{\text{reac}})_{ab} = \frac{1}{M}\sum_{n=1}^{M}(\Phi_a^n - \Phi_a^0)(\Phi_b^n - \Phi_b^0), \tag{8.5}$$

where the index n runs over the M simulated spectra. The Cholesky decomposition U of V^{reac}, where

$$V_{\text{reac}} = U^T U, \tag{8.6}$$

is later used by the LBNL toy MC to generate a fluctuated core spectrum for each toy sample:

$$\Phi_c(E_\nu) = \Phi_c^0(E_\nu) + \sum_b U_{[a_c(E_\nu)]b} y_b, \tag{8.7}$$

where $a_c(E_\nu)$ is the index corresponding to core c and energy E_ν, b runs over all cores and energy bins, and y_b is a vector of standard normal random variables.

8.3 Toy Monte Carlo

Given the livetime-normalized antineutrino spectrum $\Phi_c(E_\nu)$ at each core c (including any random fluctuations applied using V^{reac}), the LBNL toy MC begins by calculating the antineutrino spectrum $R_i(E_\nu)$ at each AD i:

$$R_i(E_\nu) = T_i\, N_i\, \epsilon_i\, \sigma(E_\nu) \sum_c \Phi_c(E_\nu) \frac{1}{4\pi L_{ci}^2} P_{\text{osc}}(E_\nu, L_{ci}), \tag{8.8}$$

where T_i is the livetime, N_i is the number of target protons, L_{ci} is the baseline, ϵ_i is the detection efficiency (comprising the muon-veto and multiplicity-cut efficiencies), and $\sigma(E_\nu)$ is the IBD cross-section (Equation C.4). P_{osc} depends on the assumption of nominal oscillation parameters; however, the final result is stable against reasonable variations in these assumptions. The R_i may be freely multiplied by a common constant factor without affecting the oscillation fit; as such, the efficiency ϵ_i does not account for factors that are equal among all ADs. Similarly, there is no impact from any constant error in the absolute normalization of the core spectra. It is only necessary that the R_i correctly capture the *relative* rates and shapes between the ADs.

The antineutrino spectra $R_i(E_\nu)$ must be converted by the toy MC into reconstructed prompt-energy spectra. The first step in this process is to convert antineutrino energy E_ν into positron energy E_e (including rest mass) using Equation C.14. Next, a detector response matrix (Section C.2.3.2) is used to convert E_e into the total

scintillator-deposited energy E_{dep}, which accounts for energy loss in the acrylic. E_{dep} is then converted to the average corresponding reconstructed energy using the detector nonlinearity model (Section C.2.3.3). Finally, the reconstructed energy is smeared using the detector resolution model (Section C.2.3.4). These various aspects of the detector response model can be randomly fluctuated during the generation of each toy sample.

The final step in generating each toy sample is to add backgrounds to the prompt spectra. This is accomplished trivially, by taking the nominal shape of each background, scaling it according to the predicted rate, and adding it to the IBD prompt spectrum. Both the rate and the shape can be fluctuated in accordance with the background uncertainties (Section C.2.4).

In practice, we generate two sets of toy samples, corresponding to signal-related (i.e., reactor and detector) and background-related fluctuations, respectively. The former matrix is proportional to the toy MC's assumptions on the absolute normalization (reactor flux and detection efficiency); it should thus be rescaled according to the measured normalization (Section C.2.5.3). Meanwhile, the backgrounds are insensitive to the flux normalization, and their data-driven estimations already account for the detection efficiency, so no rescaling is necessary for the background matrix. Since the two are treated differently, they must be generated separately, from separate sets of toy samples.

Before presenting the calculation of the covariance matrices, we must pause to discuss the binning scheme employed. As was explained, the LBNL fitter compares the far-site data to the far-site predictions obtained from the near-site data. In practice, as discussed in Section C.3.3, we combine the data from all the ADs in a given hall, and we consider *two* predictions for EH3: one from EH1, and one from EH2. These predictions are further divided among 26 energy bins (Section C.2.1) and three time bins (the 6, 8, and 7-AD periods), giving a total of

$$2 \times 26 \times 3 = 156 \tag{8.9}$$

predictions, which we represent by the vector $\overline{F}_i$, where the index i runs over the 156 individual predictions. Meanwhile, the number of far-site *observations* is equal to 78, half of the number of predictions, since we have separate predictions from EH1 and EH2. We represent the background-subtracted and efficiency-corrected observations in the vector F_i, in which each observation is repeated twice, in alignment with $\overline{F}_i$.

Given M toy samples, indexed by the variable t, the background covariance matrix V_{bkg} is calculated according to

$$(V_{\text{bkg}})_{ij} = \frac{1}{M} \sum_{t}^{M} (F_i^t - \overline{F}_i^t)(F_j^t - \overline{F}_j^t). \tag{8.10}$$

The signal covariance matrix V_{sig}, on the other hand, must be rescaled to account for

the measured absolute normalization:

$$(V_{\text{sig}})_{ij} = \frac{1}{M}\sum_{t}^{M} \frac{F_i^t \cdot F_j^t}{\overline{F}_i^t \cdot \overline{F}_j^t}(F_i^t - \overline{F}_i^t)(F_j^t - \overline{F}_j^t). \tag{8.11}$$

The statistical covariance matrix V_{stat} is not calculated from the toy samples, but rather from the measured data. As discussed in Section C.3.2, for a given prediction-bin i, the corresponding diagonal element of V_{stat} is

$$(V_{\text{stat}})_{ii} = (\sigma_{\text{near}})_i^2 + (\sigma_{\text{far}})_i^2, \tag{8.12}$$

where

$$\begin{aligned}(\sigma_{\text{near}})_i &= \frac{\overline{F}_i}{N_i}\sqrt{N_i + N_i^{\text{bkg}}} \\ (\sigma_{\text{far}})_i &= \sqrt{\overline{F}_i + F_i^{\text{bkg}}}.\end{aligned} \tag{8.13}$$

Here, N is the (background-subtracted) near-site observation, and N^{bkg} (F^{bkg}) is the expected near-site (far-site) background. Given our method of binning and combining the data, an off-diagonal element $(V_{\text{stat}})_{ij}$ is nonzero only when i and j correspond to the same far-site energy bin and data period (and thus differ only when it comes to the near-site used for the prediction). More precisely,

$$(V_{\text{stat}})_{ij}|_{i\neq j} = K_{ij} \cdot (\sigma_{\text{far}})_i \cdot (\sigma_{\text{far}})_j, \tag{8.14}$$

where $K_{ij} = 1$ if prediction-bins i and j correspond to the same far-site observation, and zero otherwise.

The total covariance matrix V is then calculated as the sum of the three above:

$$V = V_{\text{sig}} + V_{\text{bkg}} + V_{\text{stat}}. \tag{8.15}$$

8.4 Fitter

Compared to the toy MC, whose main purpose is to generate the covariance matrix V, the fitter's role is to find the oscillation parameters that minimize the χ^2 given by Equation 8.1. Most of the fitter's complexity comes not from the minimization procedure, which is conducted using the well-established numerical library MINUIT [66]. Rather, the complexity arises from the calculation of the far-site predictions $\overline{F}_i$. As illustrated in Figure 8.1, this process involves using reactor knowledge (and a given set of oscillation parameters) to decompose each near-site observation into the sum of components from individual cores. Each component is then separately extrapolated to the far site, accounting for the difference in baselines. The components are then summed to give the prediction $\overline{F}_i$. The technical details of this procedure can be found in Section C.3.1.

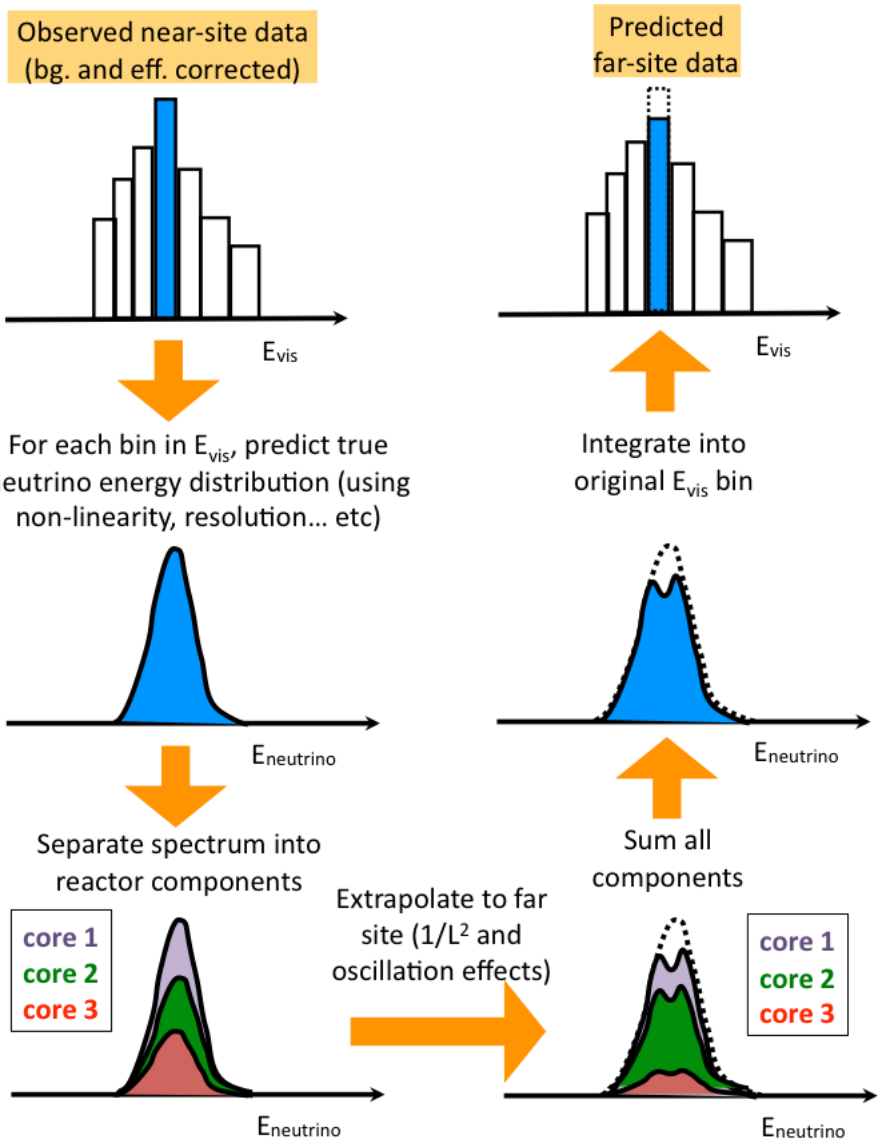

Figure 8.1: Conceptual illustration of the procedure for extrapolating a near AD measurement to a far AD. In the bottom-right panel, the dashed curve represents the no-oscillation prediction. From [62].

Given the observations F_i, the covariance matrix V_{ij}, and the ability to calculate $\overline{F}_i$ as a function of the oscillation parameters, the remainder of the fit procedure is straightforward. We provide the function $\chi^2(\sin^2 2\theta_{13}, \Delta m^2_{ee})$ (Equation 8.1) to MINUIT, which then returns the best-fit values of the oscillation parameters. Beyond reporting the best-fit parameters, the fitter also generates a grid of χ^2 values in the $\sin^2 2\theta_{13}$–Δm^2_{ee} parameter space, enabling the calculation of two-dimensional contours

at any desired confidence level (Section C.3.5).

8.5 Results

Figure 8.2 shows the results of applying the fitter to our sample of IBD candidates. For this sample, AD-specific quantities (signal and background rates, efficiencies, etc.) are summarized in Tables 8.1–8.3. We obtain the following best-fit oscillation parameters:

$$\sin^2 2\theta_{13} = 0.0850 \pm 0.0029 \qquad \Delta m^2_{ee} = (2.5010 \pm 0.0069) \times 10^{-3}\ \text{eV}^2, \tag{8.16}$$

where the errors include both statistical and systematic uncertainties (as encoded and combined in the covariance matrix, Equation C.51). The minimum χ^2 is given by

$$\chi^2/\nu = 150.843/154 = 0.9795 \qquad (p = 0.5569). \tag{8.17}$$

Here, the number ν of degrees of freedom is[1]

$$\begin{aligned} \nu &= N_{\text{periods}} \times N_{\text{predictions}} \times N_{E\ \text{bins}} - N_{\text{osc pars}} \\ &= 3 \times 2 \times 26 - 2 \\ &= 154. \end{aligned} \tag{8.18}$$

These oscillation parameters are consistent with those published in [19]. This is particularly unsurprising in light of our use of practically the same IBD selection criteria as Selection B in that publication. In Chapter 9, we proceed to explore the effects of changing these criteria, with the aim of demonstrating that the result remains consistent.

	EH1		EH2		EH3			
	AD1	AD2	AD1	AD2	AD1	AD2	AD3	AD4
IBD candidates	99099	100547	91655		13814	13722	13554	
Livetime (days)	186.786	186.786	187.190		187.666	187.666	187.666	
Target mass (kg)	19941	19966	19891		19913	19991	19892	
Veto eff.	0.8163	0.8140	0.8480		0.9808	0.9803	0.9802	
Mult. eff.	0.9765	0.9768	0.9778		0.9769	0.9767	0.9766	
Accidentals (day^{-1})	9.12 ± 0.18	8.89 ± 0.18	7.20 ± 0.14		2.80 ± 0.06	2.76 ± 0.06	2.71 ± 0.05	
^{9}Li/^{8}He (day^{-1})	2.35 ± 0.64	2.35 ± 0.64	1.50 ± 0.43		0.21 ± 0.08	0.21 ± 0.08	0.21 ± 0.08	
Fast neutrons (day^{-1})	0.84 ± 0.08	0.84 ± 0.08	0.64 ± 0.06		0.05 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	
^{241}AmC-^{13}C (day^{-1})	0.29 ± 0.13	0.27 ± 0.12	0.30 ± 0.14		0.24 ± 0.11	0.23 ± 0.10	0.23 ± 0.10	
^{13}C(α, n)-^{16}O (day^{-1})	0.09 ± 0.04	0.07 ± 0.04	0.05 ± 0.02		0.05 ± 0.02	0.04 ± 0.02	0.04 ± 0.02	
IBD rate (day^{-1})	652.91 ± 2.22	663.71 ± 2.24	582.30 ± 2.01		73.58 ± 0.67	72.88 ± 0.67	72.39 ± 0.66	

Table 8.1: Summary of inputs used in the oscillation fit (6AD period).

[1] Recall, as discussed in Section C.3.3, that we have two predictions per data period and energy bin: One prediction of EH3 from EH1, and one of EH3 from EH2, where all ADs within a given hall are combined.

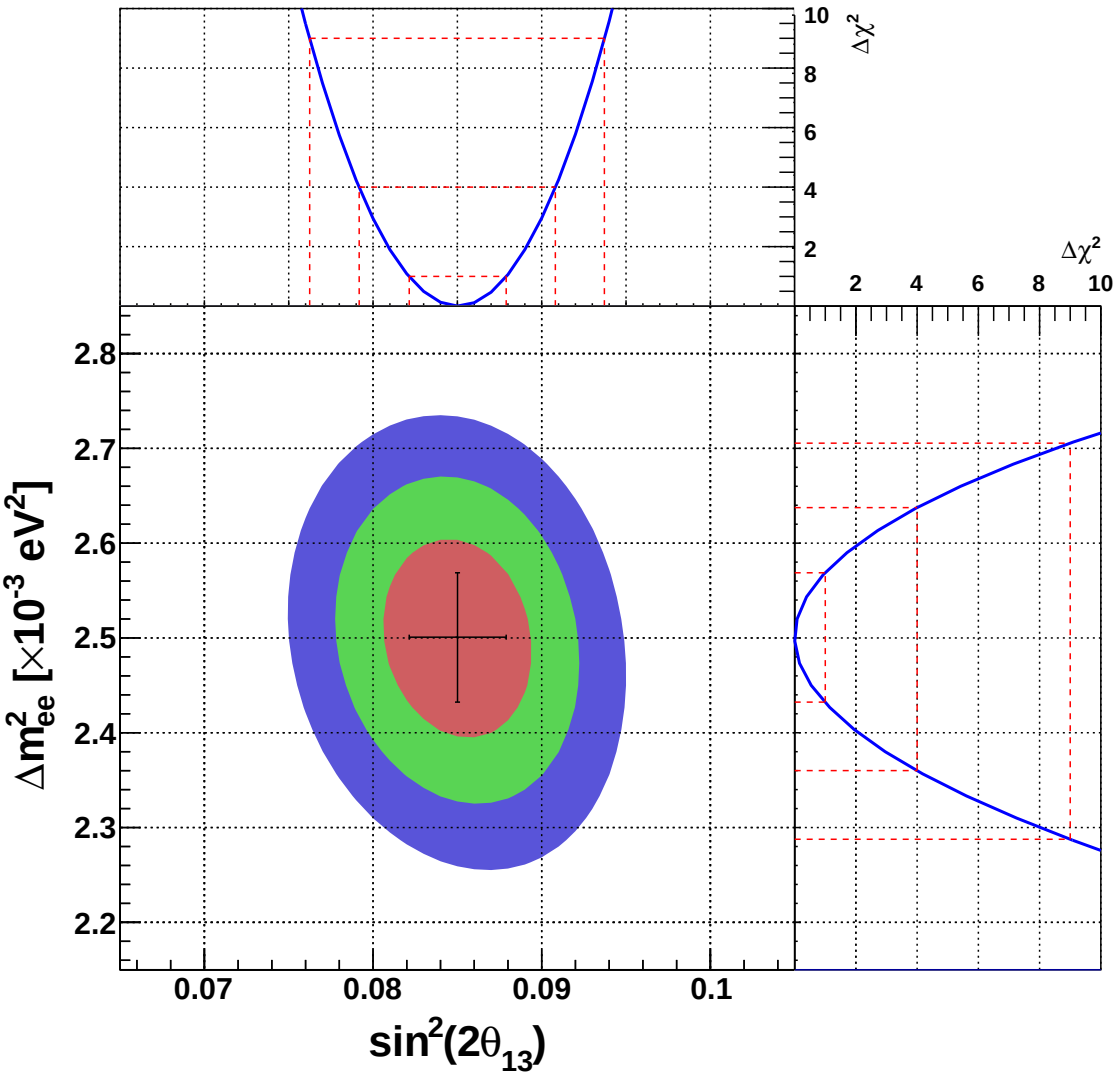

Figure 8.2: 68%, 95%, and 99% C.L. contours obtained from running the fitter on our IBD sample. The black cross is the best-fit point.

	EH1		EH2		EH3			
	AD1	AD2	AD1	AD2	AD1	AD2	AD3	AD4
IBD candidates	727648	737307	688655	677720	98495	98905	98147	99155
Livetime (days)	1349.831	1349.831	1353.589	1353.589	1351.873	1351.873	1351.873	1351.873
Target mass (kg)	19941	19966	19891	19945	19913	19991	19892	19931
Veto eff.	0.8235	0.8197	0.8517	0.8502	0.9832	0.9830	0.9829	0.9831
Mult. eff.	0.9770	0.9769	0.9783	0.9781	0.9786	0.9787	0.9785	0.9784
Accidentals (day^{-1})	7.64 ± 0.15	7.55 ± 0.15	5.56 ± 0.11	5.61 ± 0.11	0.82 ± 0.02	0.76 ± 0.02	0.78 ± 0.02	0.84 ± 0.02
^{9}Li/^{8}He (day^{-1})	2.35 ± 0.64	2.35 ± 0.64	1.50 ± 0.43	1.50 ± 0.43	0.21 ± 0.08	0.21 ± 0.08	0.21 ± 0.08	0.21 ± 0.08
Fast neutrons (day^{-1})	0.84 ± 0.08	0.84 ± 0.08	0.64 ± 0.06	0.64 ± 0.06	0.05 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	0.05 ± 0.01
^{241}AmC-^{13}C (day^{-1})	0.15 ± 0.07	0.15 ± 0.07	0.12 ± 0.06	0.14 ± 0.06	0.04 ± 0.02	0.03 ± 0.01	0.03 ± 0.02	0.04 ± 0.02
^{13}C(α, n)-^{16}O (day^{-1})	0.08 ± 0.04	0.06 ± 0.03	0.04 ± 0.02	0.06 ± 0.03	0.04 ± 0.02	0.04 ± 0.02	0.03 ± 0.02	0.04 ± 0.02
IBD rate (day^{-1})	658.93 ± 1.03	670.29 ± 1.03	604.30 ± 0.87	594.01 ± 0.86	74.67 ± 0.26	74.77 ± 0.26	74.57 ± 0.26	75.11 ± 0.26

Table 8.2: Summary of inputs used in the oscillation fit (8AD period).

	EH1		EH2		EH3			
	AD1	AD2	AD1	AD2	AD1	AD2	AD3	AD4
IBD candidates		122645	102357	100994	14894	15083	14987	14927
Livetime (days)		200.995	200.440	200.440	200.071	200.071	200.071	200.071
Target mass (kg)		19966	19891	19945	19913	19991	19892	19931
Veto eff.		0.8218	0.8550	0.8532	0.9849	0.9847	0.9846	0.9848
Mult. eff.		0.9740	0.9785	0.9783	0.9787	0.9787	0.9788	0.9785
Accidentals (day^{-1})		7.26 ± 0.15	5.26 ± 0.11	5.10 ± 0.10	0.63 ± 0.01	0.64 ± 0.01	0.69 ± 0.01	0.61 ± 0.01
^{9}Li/^{8}He (day^{-1})		2.35 ± 0.64	1.50 ± 0.43	1.50 ± 0.43	0.21 ± 0.08	0.21 ± 0.08	0.21 ± 0.08	0.21 ± 0.08
Fast neutrons (day^{-1})		0.84 ± 0.08	0.64 ± 0.06	0.64 ± 0.06	0.05 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	0.05 ± 0.01
^{241}AmC-^{13}C (day^{-1})		0.11 ± 0.05	0.09 ± 0.04	0.08 ± 0.04	0.02 ± 0.01	0.02 ± 0.01	0.03 ± 0.01	0.02 ± 0.01
^{13}C(α, n)-^{16}O (day^{-1})		0.05 ± 0.03	0.03 ± 0.02	0.06 ± 0.03	0.03 ± 0.02	0.03 ± 0.02	0.03 ± 0.01	0.03 ± 0.02
IBD rate (day^{-1})		750.78 ± 2.27	604.44 ± 1.97	596.17 ± 1.95	76.40 ± 0.64	77.07 ± 0.64	76.91 ± 0.64	76.54 ± 0.64

Table 8.3: Summary of inputs used in the oscillation fit (7AD period).

Chapter 9

Variation of IBD selection requirements

9.1 Introduction

The IBD selection described in Chapter 5 involves a number of numerical parameters, or *cuts,* which must be specified. For the official IBD selection, these cuts were essentially chosen arbitrarily at an early stage in the experiment's history, although heuristic arguments were used to justify the claim that the chosen values would provide reasonable signal statistics and background levels. Some of the cuts differed between different analyses, which nonetheless produced consistent results, thus providing some indication of the robustness of the oscillation analysis with respect to variations in the cuts. Still, prior to this work, a systematic investigation of cut variations had never been undertaken, leaving open the question: Are Daya Bay's reported values of $\sin^2 2\theta_{13}$ and Δm^2_{ee} sensitive to the precise values of the IBD selection cuts, and if so, to what extent? Naturally, any such sensitivity would need to be included as an additional source of systematic uncertainty. In this chapter we demonstrate[1] that Daya Bay's result is indeed insensitive to reasonable variations of the cuts.

Aside from the stability of the best-fit oscillation parameters as functions of the cuts, an additional consideration is the size of the uncertainty reported by the fitter. In principle, as the cuts are varied, there will be changes in the balance of efficiencies, raw statistics, and background rates, all of which can influence the size of the final errors. It is therefore possible that the official IBD selection may be suboptimal in terms of minimizing the final uncertainty, and thus a secondary goal of this study is to determine whether an alternative "optimal" set of cuts can be found. If so, the final uncertainty will be improved, and if not, we will have shown a further aspect of the analysis's robustness.

[1] Largely for the first time; to date, there have been no similarly detailed studies within the collaboration.

In total, there are 12 cut parameters that enter the IBD selection:

- Minimum and maximum prompt energy
- Minimum and maximum delayed energy
- Minimum and maximum prompt-delayed time separation
- Water-pool-muon charge threshold and veto time
- AD-muon charge threshold and veto time
- Shower-muon charge threshold and veto time

We do not separately consider the parameters of the decoupled multiplicity cut, as these are fixed by the prompt/delayed energy and time-separation cuts. Among the 12 cuts listed above, most will, upon varying, have minimal impact on the analysis. The maximum prompt/delayed energy of 12 MeV is well above the endpoint of both spectra. Likewise, there are very few events at the time-separation limits of 1 and 200 µs. The shower-muon veto window is three orders of magnitude longer than that of the WP-muon and AD-muon vetoes; thus, as long as the latter two vetoes can identify muons with ∼100% efficiency while vetoing sufficiently long to remove prompt cosmogenic events, they are sufficient. This leaves four cuts which may have a significant effect upon the analysis: The shower-muon threshold and window, the minimum delayed energy, and the minimum prompt energy. In subsequent sections, we discuss the reasons to expect these four cuts to be potentially impactful. Aside from these four cuts, we will also explore the effects of applying a vertex (i.e., position) cut, a nonstandard addition to the analysis which can provide a further demonstration of its overall robustness.

9.1.1 Technical requirements

In order to carry out this study, significant modifications needed to be made to the LBNL oscillation analysis. The original LBNL IBD selection was intended to be run just once per data set, without any variation of the cut parameters, which were hardcoded and scattered throughout the code. As such, a new, general-purpose event processing framework was written from the ground up. The IBD selection was then implemented on top of this framework, with all cut parameters specified just once in an external text file, eliminating any need to recompile the code for each new set of cuts. As opposed to the previous IBD selection, which ran in a single pass, the new one takes a two-stage approach (as discussed in Section 5.2), with an initial, cut-independent pre-selection stage that performs a data reduction on the full processed Daya Bay data files, extracting only those events and variables of potential interest to the IBD selection. (Flashers and nonphysical triggers are removed at this

stage, along with the vast majority of event variables, leaving only the reconstructed charge and energy, trigger time, and reconstructed vertex.) The second stage then applies a given set of IBD cuts to perform the selection. Computational efficiency was a primary consideration in the development of this code, in order to enable the reprocessing of data with many different cuts.

In addition to the IBD selection, the fitter also required a significant amount of work, although in this case a ground-up rewrite was not necessary. The only consideration here was computational efficiency, since each IBD selection must be followed by its own fit. Originally, the full fitting chain (including the toy Monte Carlo) took a couple of hours to run on NERSC's Cori cluster. After aggressive parallelization of the code, both by running independent steps in parallel processes, and by using OpenMP to harness multi-threaded data parallelism within each process, the full chain was reduced to a runtime of around ten minutes. These changes, combined with the rewrite of the IBD selection, made this study possible.

9.1.2 Strategy

Our overall strategy here is fairly straightforward: We simply run the IBD selection and fitter for a variety of different cut values, observing the behavior of both the best-fit points as well as the final reported uncertainties. As an additional sanity check, when feasible, we generate toy MC samples assuming modified cut values, to verify that our understanding of the experiment, as encapsulated in the toy MC, does not predict any variation of the best fit as a function of the cuts. We do not expect any such variation in the best fit; however, it is possible that the toy MC will predict a variation in the size of the error, which can be compared to the results from fitting to data. Each cut will be thus studied individually, using both real data and the toy MC. Having drawn conclusions (regarding analysis robustness and, potentially, cut optimization) from these individual studies, we will then perform an ultimate study in which these cuts are randomly varied jointly and applied to our dataset, with the aim of providing final confirmation that the analysis is not biased by our particular choice of cut values.

Unfortunately, running the IBD selection with modified cuts requires more than simply specifying the new cuts; depending on the cut in question, there may also be the need to account for changes in efficiencies and background rates. In some cases, such as the accidentals rate, the multiplicity cut efficiency, and the veto efficiency, the correct value is automatically calculated by the IBD selection. However, other values, such as the correlated background rates, were externally calculated under the assumption of the nominal cut values, and so we must derive corrections for each and apply them during the process of converting the output of the IBD selection into the input for the fitter. We detail these calculations in the discussion of the corresponding cuts.

9.2 Shower-muon threshold and veto time

The shower-muon veto parameters are important for two reasons: They significantly affect the overall veto efficiency (and thus the signal statistics), and they determine the rate of ^{9}Li, which is a major source of background uncertainty, contributing more than 50% (30%) of the total near-site (far-site) background uncertainty. Altering these parameters can thus potentially affect both the statistical and systematic uncertainty of the oscillation fit. Furthermore, variation of these parameters can serve to demonstrate that the analysis gives consistent results under the ensuing variations in the event sample, and that the ^{9}Li background subtraction is performed properly.

In what follows, we will be referring to two nominal sets of shower-muon veto parameters, termed the "LBNL" and "IHEP" vetoes, named after the institutions from which the respective analyses originate. These two vetoes are applied in the official IBD selections used in Daya Bay's publications, e.g., Selections B and A, respectively, in [19]. They are defined as follows:

- LBNL veto: Shower-muon threshold of 3×10^5 p.e.; veto window of 0.4004 s.
- IHEP veto: Shower-muon threshold of 2.5 GeV ($\sim 4 \times 10^5$ p.e.); veto window of 1 s.

A particular goal, then, of this study is to confirm that the LBNL and IHEP vetoes both lie within a region of veto parameter space in which the best-fit oscillation parameters remain stable.[2]

As described in Section 6.3.8, our ^{9}Li calculation correctly captures the dependence of the rate upon the veto parameters. Likewise, the veto-efficiency calculation simply sums up the unvetoed time windows, and is thus valid for any set of veto parameters. Both of these calculations are performed by the IBD selector. No additional steps need to be taken when reprocessing the data.

On the other hand, when we investigate the toy MC's predictions of the effects of varying the cut, we are not re-running the IBD selector, so we must explicitly adjust the veto efficiency and ^{9}Li rate. Even though the toy MC produces its own sample of "data", it does rely on an input, the so-called "`Theta13`" text file, which is produced by the IBD selector. This file lists, for each AD, the livetime, target mass, efficiencies (veto, multiplicity, delayed energy), and background rates and uncertainties. These values are then used by the toy MC to generate the predicted prompt spectra at each AD.

[2] In the grid of cut parameters used in this study, the LBNL veto's closest approximation has a veto time of 0.375 s, rather than 0.4004 s. In what follows, then, whenever this grid is under consideration, "nominal" (or "LBNL") will refer to this grid point. Given the small difference between it and the true nominal cut, we will mix the two meanings of "nominal" (or "LBNL") freely. An analogous situation exists for the IHEP veto, which is nominally defined in terms of energy, rather than photoelectrons.

An additional input from the IBD selection is the singles spectrum (which gives the spectral shape of the accidental background). The AD-muon veto window of O(1 ms) (to say nothing of the O(1 s) shower window) is more than long enough to remove all prompt, isolated, muon-induced triggers, so we expect no variation in the shape of the singles spectrum as we vary the shower veto.

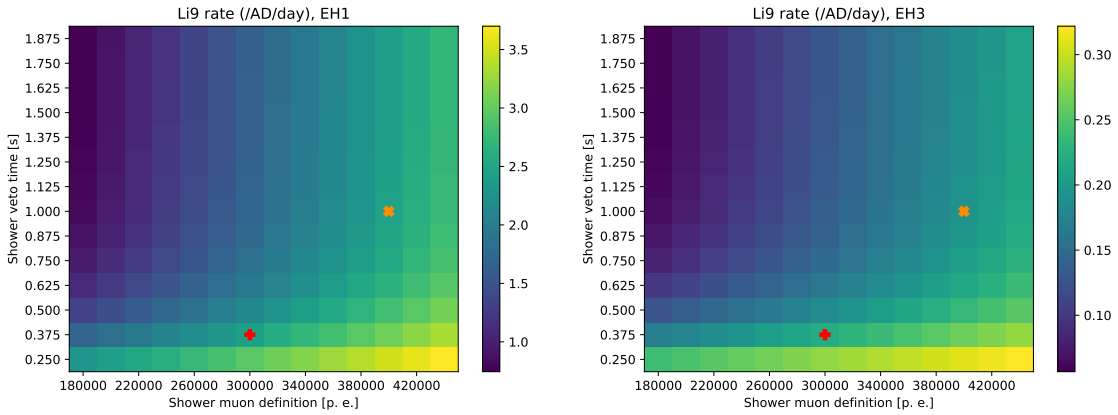

Figure 9.1: ^{9}Li rate as a function of the shower veto parameters, for EH1 (left) and EH3 (right). The LBNL and IHEP parameters are indicated by the red "+" and orange "×", respectively.

The toy MC implementation of this study, then, requires taking the `Theta13` file produced by the nominal selection, and adjusting the veto efficiency, the ^{9}Li rate, and its uncertainty. The veto-efficiency correction is described in the following section. For the ^{9}Li rate and uncertainty, we simply load the `Li9Calc` class from the IBD selector, as described in Section 6.3 and feed it the modified veto parameters, producing the ^{9}Li rates exemplified in Figure 9.1. Once the `Theta13` file has been thus modified, the toy MC will generate toy spectra that properly account for the expected effects on the veto efficiency and ^{9}Li rate.

9.2.1 Veto efficiency

As was just mentioned, we must explicitly provide the overall muon veto efficiency as an input to the toy MC (and subsequently to the fitter). Since, again, the veto efficiency is computed as one of the outputs of the IBD selection, we could in principle simply run the IBD selection on real data, with the desired veto parameters, and then feed the resulting veto efficiency to the toy MC. However, this is computationally expensive, and more importantly, we would prefer for the toy MC cross check to be as decoupled as possible from the IBD selection and its associated data. Thus, we seek an independent method of determining the muon veto efficiency.

The veto efficiency is, naturally, a function of the muon rate. In particular, it depends on the rates of the three types of muons: water pool, AD, and shower. The first step, then, is to measure these rates, while avoiding double counting due to retriggers and multi-detector muons. For this purpose, a "muon selector" was written to extract all of the muon events in our dataset. This is the one place where actual data enters this process, but the code is independent of the IBD selection's second stage, where IBD cuts (including the veto) are normally applied. Once the muon sample has been obtained, the rates of the three classes can be calculated. The final step is to calculate the efficiency, which requires a proper treatment of the possible overlap of veto windows between closely spaced muons. In what follows we describe these steps in detail and demonstrate that the results agree well with the veto efficiency obtained from the IBD selection.

9.2.1.1 Muon-rate measurement

The objective of the muon-rate measurement is to obtain, for each AD, two results: First, the rate of *water pool-only* muons (i.e., those muons that produced a "WP muon" trigger without an associated "AD muon" trigger), and second, the spectrum (in terms of charge or energy) of those muons that triggered the AD. Given that the AD/shower veto windows are longer than the WP window, any muon that triggers an AD *as well as* one or both WPs should should be regarded as a single AD muon.[3] The measurement of the AD-muon *spectrum*, rather than the total rate, is important because it enables a breakdown into "shower" and "non-shower" AD muons, which carry different veto windows, and it makes this breakdown possible while varying the shower-muon definition, without requiring a re-run of the muon selection.

The muon selector runs on the output of the first stage of the IBD selection, i.e., the pre-selection, which does not depend on the IBD selection cuts. In particular, the muon selector reads the contents of the `muons` tree, which simply contains basic information (time, detector, hit multiplicity, and charge) for each WP trigger with more than 12 above-threshold PMTs, and each AD trigger with more than 3000 PE of charge. We collectively refer to these events as muon-like triggers.

If we were to naively count the number of events of each type in the `muons` tree, we would overestimate the muon rates, for two reasons: Muons that trigger more than one detector must still be only counted once, and muon-induced retriggers must be discarded. The strategy for handling these subtleties is similar to the one described in Section 6.3.2 for the muon selection used in the ^{9}Li rate study:

- No distinction is made between the inner and outer water pools.

[3] Since the muon rates are measured separately for each AD, this muon would instead be counted as a WP-only muon, rather than as an AD muon, with respect to any AD that did not record an associated AD muon event.

- If a WP muon occurs less than 15 µs after an AD muon, it is discarded in favor of the AD muon.
- If a WP muon occurs less than 5 µs after another WP muon (which, in practice, will always be in the other pool), and the number of hit PMTs is greater than the previous one, this muon replaces the previous one.[4]
- If a WP muon occurs between 5 and 15 µs after another WP muon, it is discarded as a retrigger.
- If an AD muon occurs less than 15 µs after another muon in the same AD, it is discarded as a retrigger.
- If an AD muon occurs less than 5 µs after a WP muon, the WP muon is discarded in favor of the AD muon.

Each final "merged" muon produced by this procedure is stored in a histogram binned according to the number of hits (for WP muons) or the charge (for AD muons). It should be noted that there is a separate histogram of the WP muons for each AD, even though all ADs in a hall share the same water pools. This is because an AD muon will override a WP muon, but a given muon will not necessarily pass through every AD, so the WP muon will still be counted for those ADs that do not see an associated AD muon.

Using the histograms of the WP and AD muons, the rates of the three muon classes can be calculated: First, the number of WP-only muons is determined by integrating the WP histogram for `nHit`> 12, while the number of non-shower AD muons is the integral of the AD histogram from 3000 PE to the shower threshold, and the number of shower muons is the above-threshold integral. These counts are then divided by the total livetime to obtain the three rates.

9.2.1.2 Efficiency from the rates

The final step is to calculate the veto efficiency from the three rates. Although the three muon classes are simple independent Poisson processes, it is difficult to calculate the veto efficiency analytically, given the possibility of overlapping veto windows between and within the three classes. On the other hand, it is trivial to simulate the processes, and the simulated veto windows can then simply be added (after removing overlaps) to determine the total vetoed time and hence the efficiency. For the sake of validation, three different approaches were tested, all with consistent results:

[4] This reflects the fact that the IBD selection's muon veto simply requires one pool, and not necessarily both, to be above-threshold.

- A "shotgun" toy MC that generates N (rate $\times$ time) random muons (of each type) distributed in time according to the uniform distribution, before sorting and merging them
- A "sequential" toy MC that uses the relative rates to determine the type of the "next" muon, and the exponential distribution to determine the time to it
- A "parallel" toy MC in which the three processes independently generate random samples of the time-to-next (i.e. exponential) distributions, with the events then sorted and merged. This version generated output files in the same format as the pre-selector, so the actual second-stage IBD selection could be run on them; this cross-check indeed confirmed that our window-adding calculation agreed with that in the IBD selector.

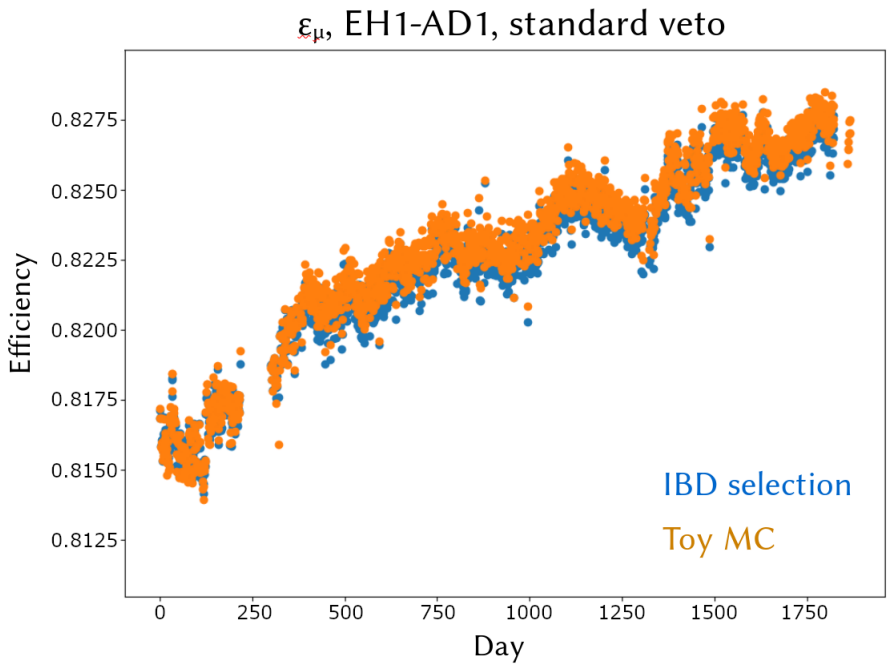

Figure 9.2: Daily muon veto efficiency in EH1-AD1, both as reported by the IBD selection (by summing veto windows) and as calculated using the measured rates in the toy simulation.

Figure 9.2 shows the daily efficiency of the nominal muon veto, both as reported by the IBD selection, and by performing the calculation above using the rates measured from data. As can be seen, there is a very small bias (on the order of 0.02%) inherent in this approach. This is unsurprising, given that the IBD selection does not discard muon retriggers, which can therefore extend the veto window by O(10 µs). In any

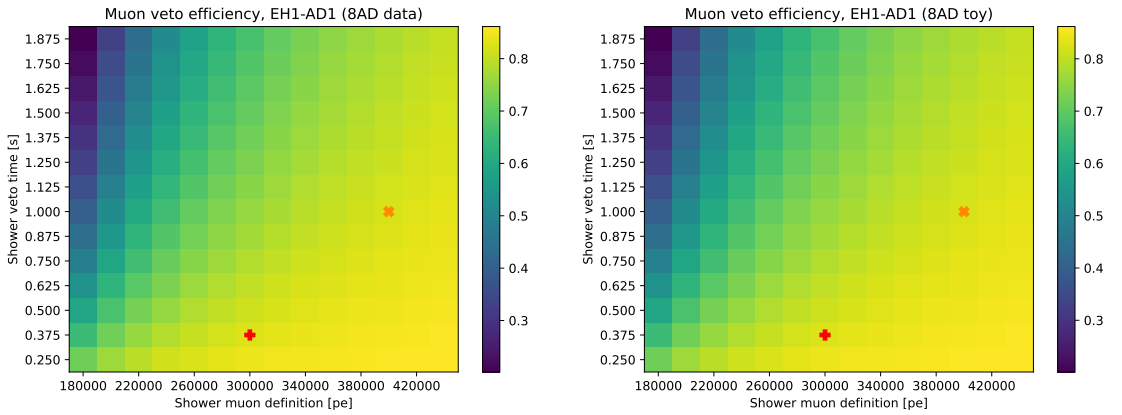

Figure 9.3: Veto efficiency as a function of shower-muon veto parameters in EH1-AD1 as determined by the IBD selection (left) and the rate measurement and simulation (right). The LBNL and IHEP parameters are indicated by the red "+" and orange "×", respectively.

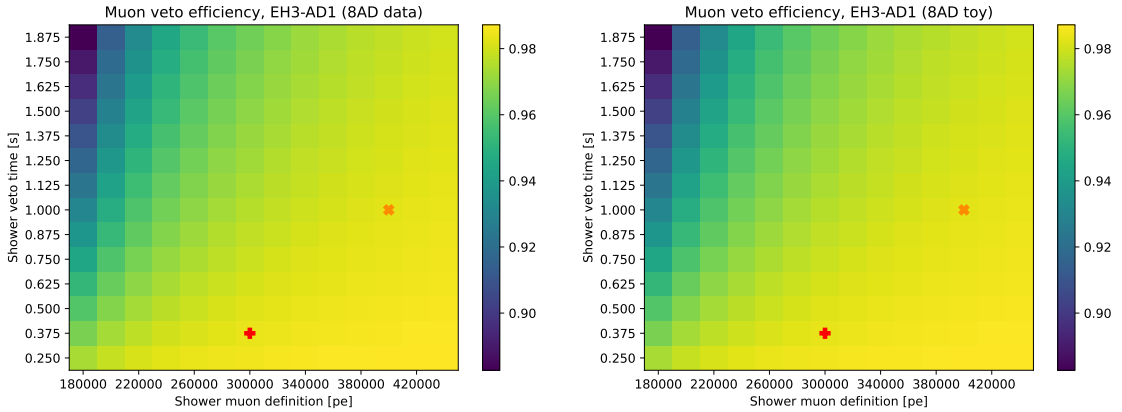

Figure 9.4: Veto efficiency as a function of shower-muon veto parameters in EH3-AD1 as determined by the IBD selection (left) and the rate measurement and simulation (right). The LBNL and IHEP parameters are indicated by the red "+" and orange "×", respectively.

case, this bias is not large enough to meaningfully impact the toy MC study of the shower veto[5]. Figures 9.3 and 9.4 compare (for a near and a far AD, respectively) the dependence of the efficiency upon the two shower-muon veto parameters, as determined both by running the full IBD selection, and by using the above calculation on the muon rates measured by the nominal selection. Excellent agreement can be

[5]And the whole purpose of the toy MC study is merely to provide a sanity check; ultimately, the results from data are of primary interest.

seen.

Having determined the expected veto efficiency for a modified shower veto, this value can be substituted into the `Theta13` file, along with the predicted ^{9}Li rate. The toy MC are then run as usual in order to generate the covariance and detector-response matrices, etc., and then the Asimov toy spectra are fed to the fitter in order to extract the oscillation parameters. Meanwhile, for fits to data, the `Theta13` file is taken directly from the output of the IBD selection, and the fitter receives the prompt spectra from the data sample. In both cases, the toy samples (and hence the covariance matrices) are generated under the assumption of "nominal" oscillation parameters, namely $\sin^2 2\theta_{13} = 0.0850$ and $\Delta m^2_{ee} = 0.002501$[6]. In the following sections we discuss the fit results for both the toy and the real data samples.

9.2.2 Fit results

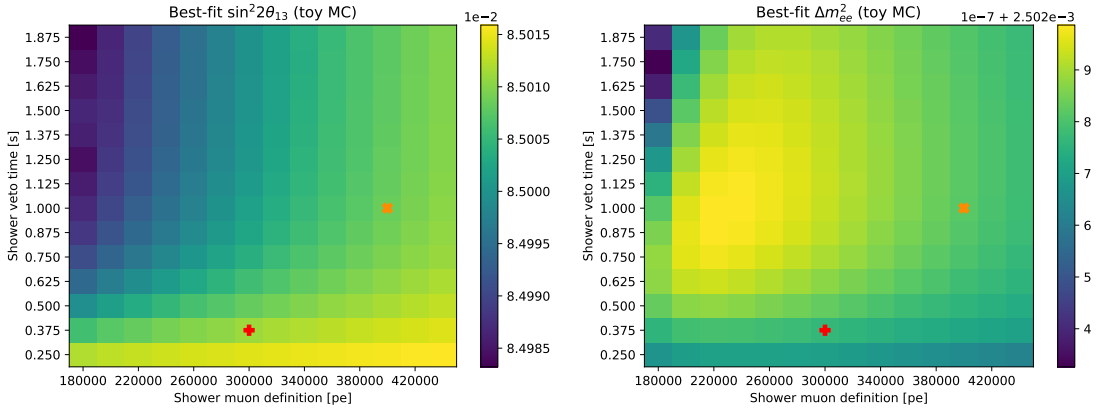

Figure 9.5: Results of oscillation fit to toy samples as a function of the shower-muon veto parameters. On the left is $\sin^2 2\theta_{13}$ and on the right is Δm^2_{ee}. The LBNL and IHEP parameters are indicated by the red "+" and orange "×", respectively.

Toy samples were produced and fit in a uniform 14×14 grid of shower veto parameters, with the charge threshold ranging from 1.8×10^5 to 4.4×10^5 photoelectrons, and the veto time ranging from 0.25 to 1.875 seconds. Figure 9.5 shows the results of the fits. The variation is on the order of 0.01%, which is far below the scale of the analysis's uncertainty, and is likely attributable to rounding error. Thus, as expected,

[6] As the covariance matrix has been shown to vary little under changes of the assumed oscillation parameters, we do not attempt to make the process more "self-consistent", such as by using an iterative procedure in which the covariance matrices are regenerated using the best-fit parameters, the fit is repeated, and so forth until convergence is reached. We have chosen the toy MC's oscillation parameters to be equal to our best-fit parameters from P17B data when the nominal IBD selection is applied.

the toy MC predicts essentially no variation in the results of the oscillation analysis when the shower-muon veto is varied.

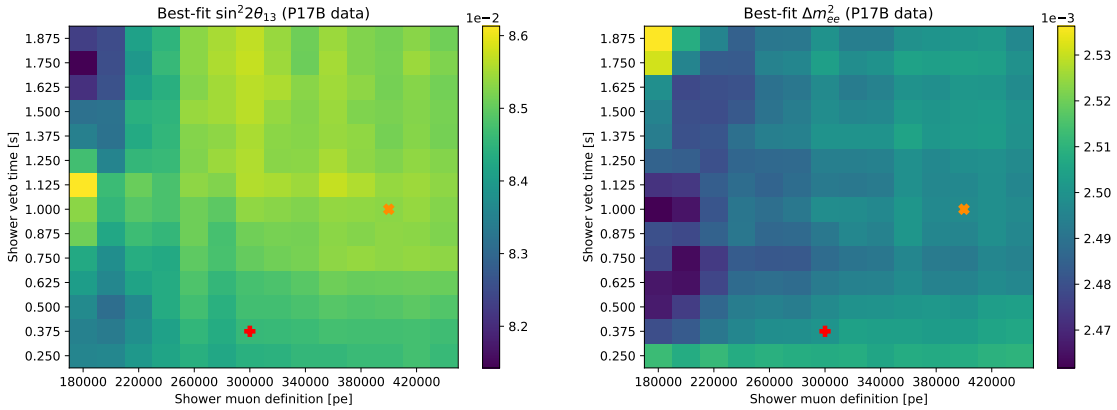

Figure 9.6: Results of oscillation fit to P17B data as a function of the shower-muon veto parameters. Note different scale compared to Figure 9.5. On the left is $\sin^2 2\theta_{13}$ and on the right is Δm^2_{ee}. The LBNL and IHEP parameters are indicated by the red "+" and orange "×", respectively.

Figure 9.6 shows the results of running the IBD selection and fitter on the P17B dataset, on the same grid of veto parameters as used in the toy MC study. In this case we see a greater degree of variation, which is unsurprising given that, in data, modifying the shower-muon veto will produce statistical fluctuations, while this effect is absent for the toy MC. Nonetheless, in a large region of the parameter space, including both the LBNL and IHEP vetoes, the fit is acceptably stable. Significant deviations are observed only on the left edge, where the loose definition of a shower muon results in an unreasonably low veto efficiency.

Although the toy simulations and the data fits showed very disparate scales of variation in the best-fit oscillation parameters, the variation in the uncertainty was more consistent between the two, as shown in Figure 9.7 and Figure 9.8. As expected, the toy MC predicts a small but non-negligible amount of variation in the uncertainty, due to the aforementioned effect of the shower-muon veto on both the veto efficiency (i.e. signal statistics) and the ^{9}Li rate. The fits to data indicate a slightly higher uncertainty overall, possibly due to statistical fluctuations in the data sample, but with the same general pattern of increasing uncertainty toward the upper left of the grid (the region of smaller shower-muon energy threshold and longer veto-time window).

Ultimately, the key conclusion from both the toy samples and the data is that the LBNL and IHEP vetoes indeed lie within a large, flat valley in the parameter space. We are free to vary the parameters within this region without consequence, but there is also nothing to be gained from it. The LBNL and IHEP vetoes can therefore both be

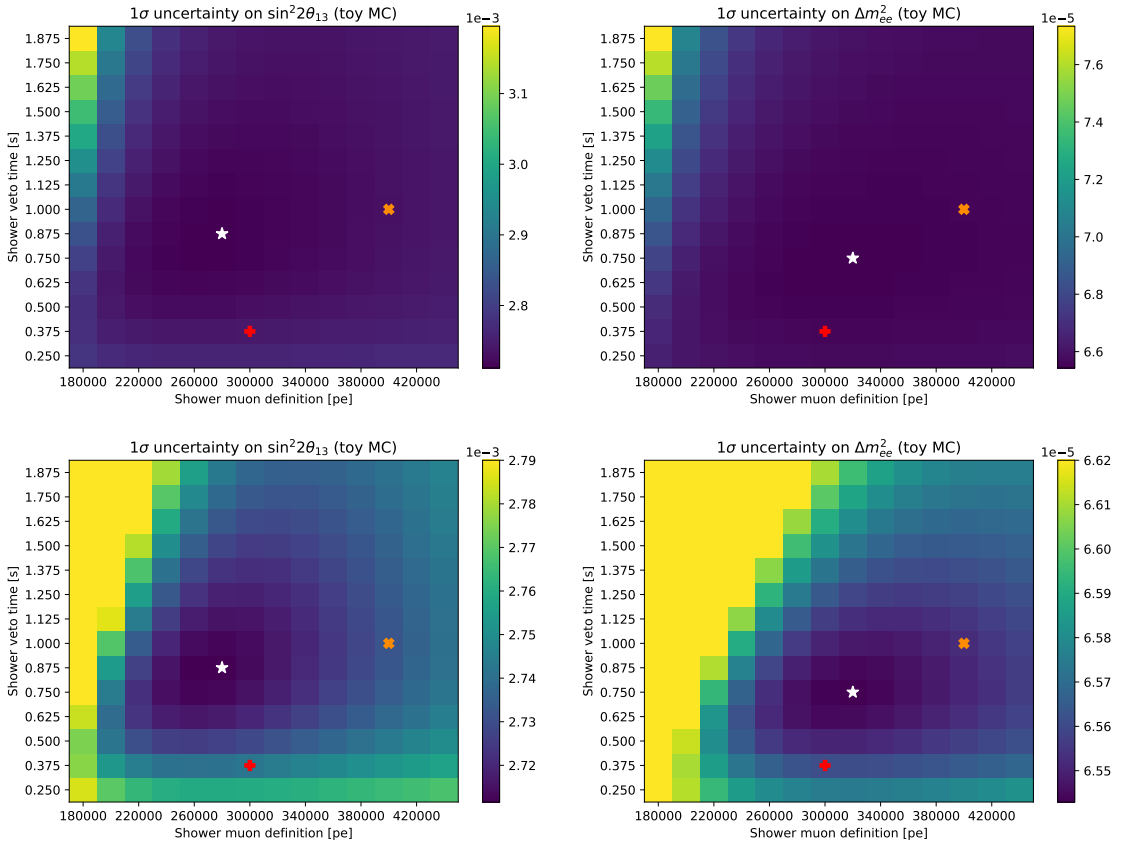

Figure 9.7: Uncertainty on $\sin^2 2\theta_{13}$ (left) and Δm^2_{ee} (right) for fits to toy MC samples. The optimal (lowest) uncertainty is marked with a white star. The LBNL and IHEP parameters are indicated by the red "+" and orange "×", respectively. The bottom plots use alternative color scales chosen to highlight details near the nominal and optimum cuts.

considered "optimal" in the sense that they both achieve near-minimum uncertainty. It is worth noting that there is an O(0.5%) difference between the best-fit values of $\sin^2 2\theta_{13}$ for the LBNL and IHEP parameters (holding the remaining cuts constant). However, given that the oscillation fit's statistical uncertainty is on the order of 3%, this result is unsurprising.

9.3 Minimum delayed energy

In the reconstructed energy spectrum of neutron captures on gadolinium (nGd captures), there is a long, non-negligible tail below the nominal cut of 6 MeV (Figure 9.9). This tail is populated by events in which one or more gamma-rays "leak" beyond the

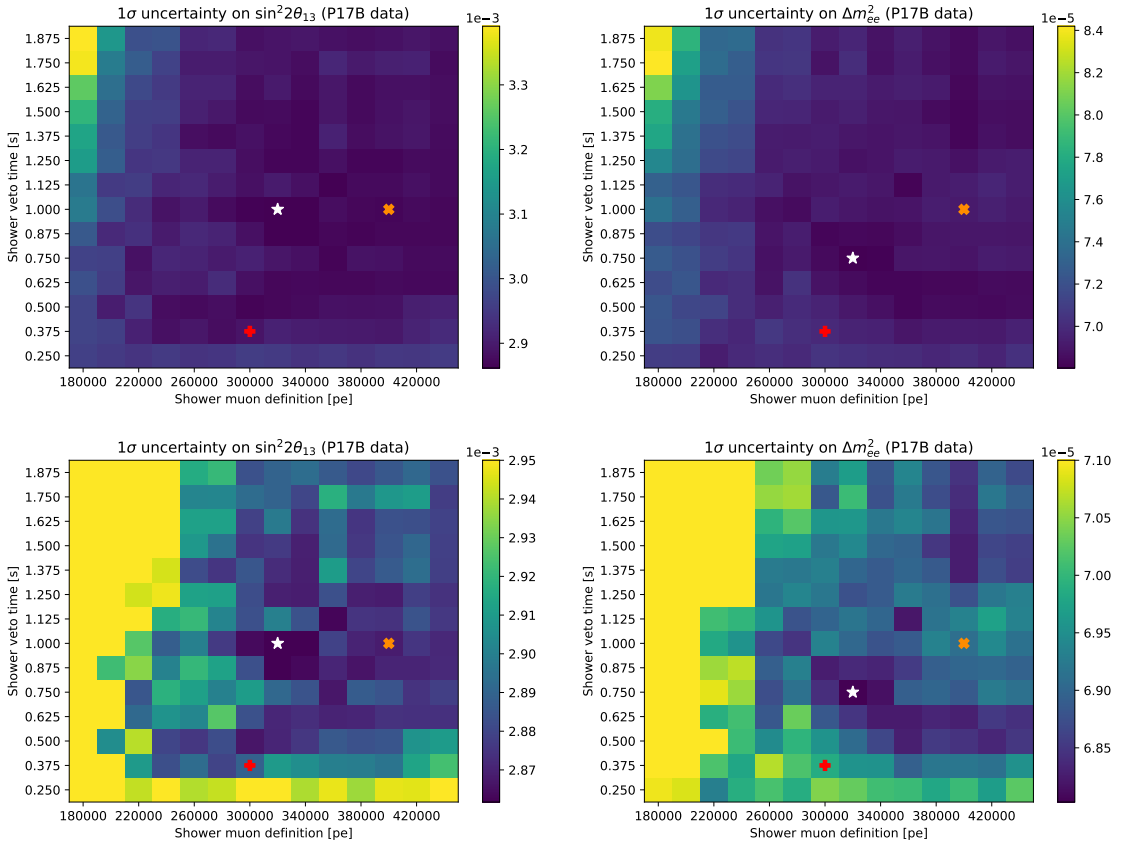

Figure 9.8: Uncertainty on $\sin^2 2\theta_{13}$ (left) and Δm^2_{ee} (right) for fits to P17B data. The optimal (lowest) uncertainty is marked with a white star. The LBNL and IHEP parameters are indicated by the red "+" and orange "×", respectively. The bottom plots use alternative color scales chosen to highlight details near the nominal and optimum cuts.

active volume of the AD, leaving some of the initial energy unreconstructed. Fitting the delayed-energy spectrum to a double Crystal Ball function (Equation 4.1) suggests that some 10% of nGd-capture events fall below the 6 MeV cut. Due to this potential of gaining O(10%) in signal statistics, it is worthwhile to explore the variation of the 6-MeV cut. Here we study the modified delayed-energy cuts ranging from 4 to 7 MeV.

Of course, there is a trade-off: With an increase in signal comes an increase in backgrounds. For those correlated backgrounds that involve an actual delayed nGd capture (^{9}Li, fast neutrons, and α-n), the rate will increase according to the same scaling factor as the signal rate.

The AmC background possesses a different delayed energy-spectrum (dominated by neutron captures in the AD's stainless steel vessel); unfortunately, none of the available internal studies of the AmC background provide the delayed-energy spectrum,

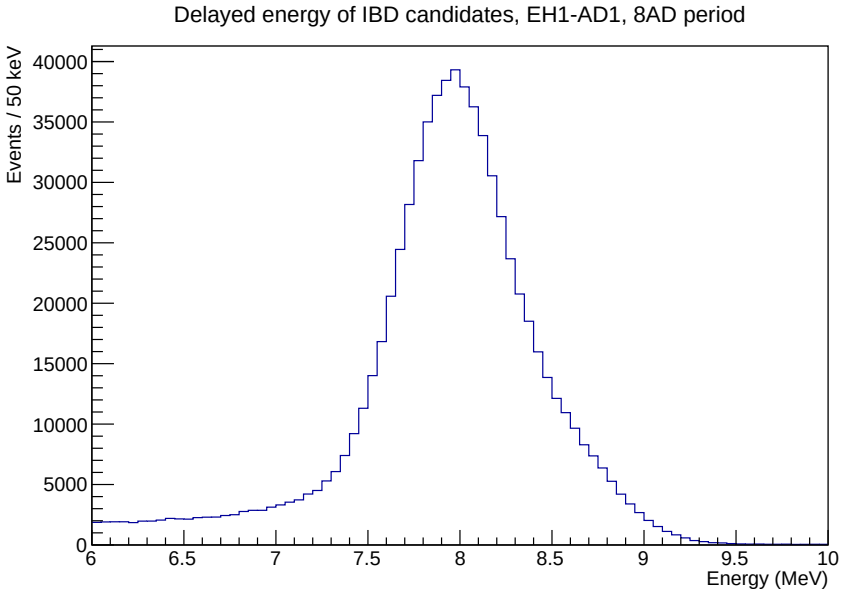

Figure 9.9: The delayed-energy spectrum of IBD candidates in EH1-AD1 during the 8AD period. Note the long tail to the left of the peak.

neither from simulations nor from the runs with the strong AmC source. Due to the small size of this background, we simply ignore any potential dependence of its rate on the delayed-energy cut. We will later see that, at least in the neighborhood of a 6 MeV delayed-energy cut, the oscillation analysis is remarkably stable, in spite of this simplification.

Although these correlated backgrounds do increase as the delayed-energy cut is loosened, the effect on the uncertainty is outweighed by the increase in signal statistics. Letting k denote this scaling factor (i.e. the ratio of the new delayed-energy efficiency to the old one):

$$N_{\text{corr}} \equiv N_{\text{obs}} - N_{\text{bkg}} \tag{9.1}$$

$$N'_{\text{corr}} = kN_{\text{corr}} \tag{9.2}$$

$$\implies \frac{\sigma(N'_{\text{corr}})}{N'_{\text{corr}}} = \frac{1}{\sqrt{k}} \frac{\sigma(N_{\text{corr}})}{N_{\text{corr}}}, \tag{9.3}$$

that is, even though there is an increase in the absolute uncertainty on the background-subtracted rate, the *relative* uncertainty actually decreases. Therefore, if these were the only backgrounds at Daya Bay, it would be advantageous to use a delayed-energy cut that is as low as possible.

Unfortunately, the accidental background is not as well-behaved. As shown in Figure 9.10, the spectrum of the delayed-like singles rises much more steeply below 5 MeV

compared to the neutron-capture spectrum, implying that the accidental background increases faster (percentage-wise) than the IBD rate (as confirmed in Figure 9.13). When this effect is large enough, the overall uncertainty on the background-subtracted IBD rate no longer decreases (as illustrated in Equation 9.1); instead, it *increases*. In principle, then, it is the accidental background that serves to penalize overly loose definitions of the delayed-energy cut.[7]

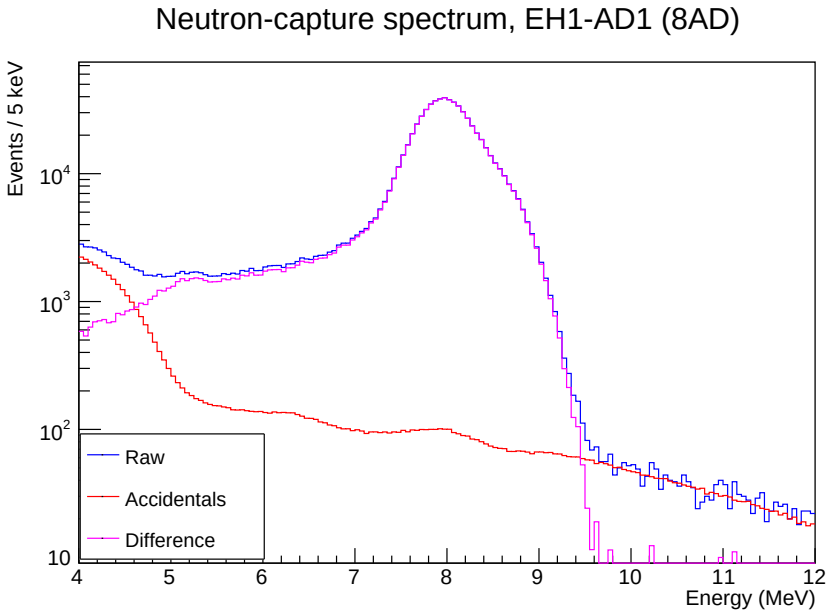

Figure 9.10: Extracted neutron-capture spectrum for EH1-AD1 in the 8AD period.

In theory, the toy MC should be able to capture and quantify the *known* effects of varying the delayed-energy cut, that is, the effects on the delayed-energy cut efficiency (and by extension the signal statistics), on the accidentals, and on the nominal set of correlated backgrounds. We expect the toy MC to show no variation in the best fit (again, as a confirmation of the self-consistency of the toy MC / fitter system), but we do expect the uncertainty on the oscillation parameters to vary, based on the balance of the competing effects described above.

[7] In saying this, we are ignoring the possibility of new, unaccounted-for *correlated* backgrounds (or increased AmC backgrounds) appearing when we loosen the delayed-energy cut. As we discuss later, this is in fact a likely issue. Increases in the rates of the main correlated backgrounds are, however, accounted for, as discussed in Section 9.3.1, but this is a minor effect compared to the increase in accidental backgrounds. It also bears mentioning that a delayed-energy cut below ∼5.5 MeV will accept IBDs in which the neutron was captured on carbon ("nC IBDs"), as illustrated by the small peak above 5 MeV in Figure 9.10. Since, aside from their delayed energies, these IBDs are no different from nGd IBDs, their presence should not bias the oscillation fit, provided that the ADs all possess consistent nC capture fractions (as they were designed to do).

As with the shower-muon veto, we carry out the toy MC study by taking the `Theta13` file from the nominal (i.e. 6 MeV) IBD selection, and altering those quantities that vary with the delayed-energy cut. However, the toy MC does not have any built-in knowledge of how the signal efficiency and background rates vary with the delayed-energy cut. As such, we must externally provide this knowledge, which we obtain from measurements down to 4 MeV of the neutron-capture spectrum and the delayed-like singles spectrum. These measurements are described in Section 9.3.2.1.

When we perform fits to the actual data using a modified delayed-energy cut, the correct accidentals rate is automatically calculated by the IBD selector, and of course there is no need to apply (as in the toy MC case) a scaling factor to the raw-signal rate. However, the correlated-background rates must still be corrected using the same formalism as that applied for the toy MC. Furthermore, an additional complication arises from the fact that the ADs differ among themselves in the shape of their neutron-capture spectra (Figure 9.11), due to residual (i.e., post-calibration) small differences in the optical and electronic characteristics of the ADs. As a result, when we modify the delayed-energy cut, the efficiency of this cut will not necessarily vary in unison among the ADs, possibly altering the measured near-far ratio. There are multiple ways to deal with this issue, and we discuss them further in Section 9.3.2. First, however, we discuss our treatment of the backgrounds.

9.3.1 Treatment of backgrounds

9.3.1.1 Accidentals background

For fits to the actual data, the accidentals background is calculated by the IBD selector as described in Section 7.7. This calculation is valid for any delayed-energy cut, so no further correction is necessary.

For fits to the toy spectra, we use information from a special IBD selection run on the full P17B dataset. This special selection is simply the nominal selection with a loose delayed-energy cut of 4 MeV. From the spectrum of delayed-like singles measured with this selection, we can calculate the accidentals rate for any delayed-energy cut of 4 MeV or greater. We do this by manually instantiating the relevant class from the IBD selector. The calculated rates are then substituted into the `Theta13` file.

9.3.1.2 Correlated backgrounds

For both toy and data fits, we use the same procedure. Since ^{9}Li, fast neutron, and α-n events all include a delayed neutron capture, we can employ the neutron-capture spectrum $S_n(E)$ obtained from the IBDs, whose extraction we describe in Section 9.3.2.1. For a delayed-energy cut of q MeV, the backgrounds are then simply

scaled according to the ratio

$$\frac{\int_q^{12} S_n(E)\,dE}{\int_6^{12} S_n(E)\,dE} \tag{9.4}$$

9.3.2 Delayed-energy cut efficiency

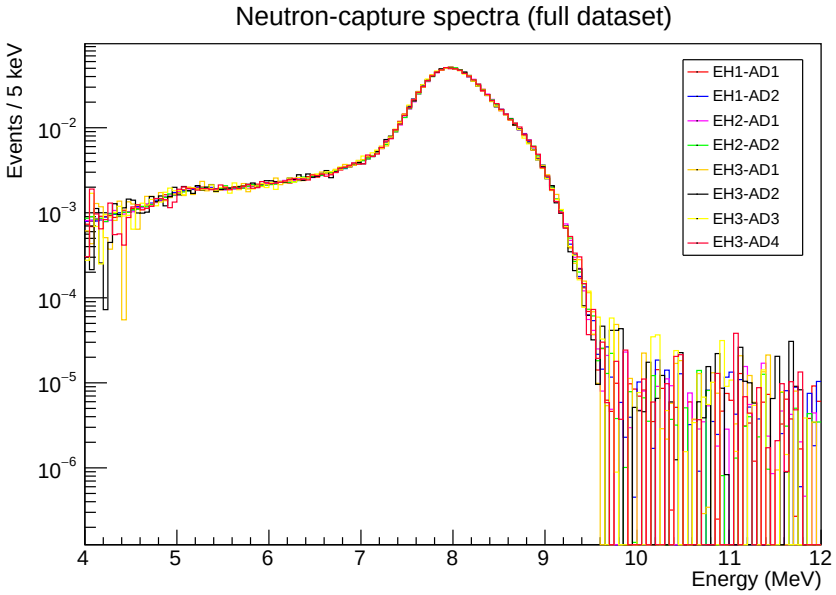

Figure 9.11: Background-subtracted nGd-capture spectra for the 8 ADs. Visually, the shape differences are practically indiscernible, but these small variations are nevertheless sufficient to produce an AD-dependent efficiency variation when the delayed-energy cut is varied.

As we show later (Section 9.3.3), naively varying the delayed-energy cut will cause a shift in the best-fit oscillation parameters (particularly $\sin^2 2\theta_{13}$). This shift comes from the slight AD-to-AD differences in the shape of the IBD delayed spectrum. In what follows, we will refer to this spectrum as the "neutron capture" or "nGd" spectrum.[8] Due to these shape discrepancies, varying the delayed-energy cut will cause its efficiency to vary by different amounts between the ADs.

In the published Daya Bay analyses, we have always assumed that all 8 ADs have the same delayed-energy cut efficiency, to which we assign a nominal (and irrelevant) value of 0.88 in the `Theta13` file. However, from the observed differences in the nGd shape, it is clear that this assumption is inaccurate. Obviously, the efficiency of the

[8] Although the bulk of delayed triggers are indeed nGd captures, there is in fact a small peak around 5 MeV (Figure 9.10) from neutron captures on carbon.

6 MeV cut cannot be measured by a direct comparison of the IBD rates between ADs, since this wouldn't account for all of the other effects (e.g., oscillation!) which affect the measured rate. However, the delayed-energy cut efficiency[9] *could*, in principle, be measured by a fitter with 8 pull terms for the individual efficiencies and an additional pull term for the "reactor antineutrino anomaly" (the last of which already is included in the pull-based fitters used in the collaboration). Of course, including these 8 additional pulls could lead to a shift in the best-fit $\sin^2 2\theta_{13}$, reflecting the fact that a particularly unfortunate combination of delayed-energy cut efficiencies could be mistaken for an oscillation signal.

Our fitter, however, does not support pull terms. Instead, we could take the measured IBD rates, correct them for baseline and oscillation effects, and then take the ratios to, say, AD1, giving the 8 *relative* efficiencies, normalized to AD1. Again, though, such an approach would suffer from the ambiguity between the effects of oscillation and unequal efficiencies. Thus, we make no attempt to perform a corrected rate-based measurement of the 6 MeV cut's absolute efficiency. Instead, our approach is to measure the shape of the neutron-capture spectrum as best as possible (as described in Section 9.3.2.1), and use that as our sole source of information on the efficiency of the delayed-energy cut. Since our study extends down to delayed-energy cuts of 4 MeV, so does our measurement of the neutron-capture spectrum. The soft end of this spectrum includes events from the nGd tail as well as nC captures.

Once the neutron-capture spectrum has been measured, there are three ways it can be used (or not) to apply a correction factor to the IBD rate. We name and define these approaches as follows:

- `flat`: No correction. As in the official analysis, we simply assume that all ADs have the same delayed-energy cut efficiency, for all values of the cut.
- `relative`: We integrate the spectrum from E_{cut} to 12 MeV, and divide by the integral from 6 to 12 MeV. Thus, when $E_{\text{cut}} = 12$ MeV, we reproduce the official analysis.
- `absolute`: We again integrate from E_{cut} to 12 MeV, but this time we divide by the total integral of the measured spectrum (i.e., from 4 to 12 MeV).

The `flat` method is obviously the simplest, and as we shall show (Section 9.3.3), it produces the aforementioned shifts in $\sin^2 2\theta_{13}$, demonstrating that the delayed-energy cut efficiency varies among ADs. Meanwhile, the advantage of the `relative` correction is that it allows us to show that we can obtain consistent results as we move the delayed-energy cut away from 6 MeV, provided that we correct for differences in the neutron-capture spectral shape. This proves that we understand and can account

[9] More accurately, the total detection efficiency, divided by those efficiencies (veto and multiplicity) that we treat individually.

for the implications of varying the cut. However, the `relative` method says nothing about the size of the bias caused by the assumption of equal efficiencies at 6 MeV. This is where the `absolute` method provides insight. From the difference in the best-fit values of $\sin^2 2\theta_{13}$ at 6 MeV between the `relative` and `absolute` methods, we can infer the scale of the aforementioned bias. This measurement is still uncertain, since there is a degree of arbitrariness from the choice of the lower cutoff (in our case, 4 MeV), but as shown in Figure 9.10, there are not many neutron captures with such low energy, so that variations around 4 MeV can be expected to have minimal effect.

We will show (Section 9.3.3) the results of applying all three methods, ultimately demonstrating excellent stability in $\sin^2 2\theta_{13}$ when the relative method is applied between 5 and 7 MeV.

9.3.2.1 Measurement of neutron-capture spectra

Before we can integrate the neutron-capture spectrum, we must, of course, measure it for each AD. It would be incorrect to use the raw delayed-energy spectrum from the IBD candidates, since this includes contributions from backgrounds. Fortunately, all of the correlated backgrounds (except AmC, which we neglect due to its small rate) feature a neutron capture as the delayed event. On the other hand, accidental backgrounds will be drawn from the spectrum of delayed-like singles, which has a different shape from the neutron-capture spectrum. Fortunately, both this spectrum and the accidentals rate can be measured quite accurately, so that we can obtain our final measurement simply by taking the delayed-energy spectrum of the IBD candidates, and subtracting the delayed-like singles spectrum (scaled by the accidentals rate).

Although we did not diverge from this overall method, we implemented it in three different ways:

- `original`: We performed a "reference" IBD selection using a delayed-energy cut of 4 MeV, extracting the neutron-capture spectrum as described above. For any subsequent IBD selection with arbitrary cuts, this reference spectrum was used to obtain the (`relative` or `absolute`) efficiency.

- `calc-then-add`: Within each IBD selection, we perform a "parallel" IBD selection using the same cuts, except with a 4 MeV delayed-energy cut. The *daily* spectrum, and then the efficiency (both `relative` and `absolute`) are obtained from the parallel selection. For the efficiency in the total sample, we use the weighted mean of the daily efficiencies.

- `add-then-calc`: Again we use the "parallel" 4-MeV selection, but we do not calculate the efficiency on a daily basis. Rather, we add the daily neutron-capture spectra in order to obtain the spectra for the full dataset, and then calculate the efficiency from these total spectra.

For all cut variations we study *except for the application of a vertex cut*, these three methods generally give equivalent results. However, when we apply a vertex cut (as discussed in Section 9.5), there can be some variation in the shape of the neutron-capture spectrum. The latter two methods were implemented so as to enable direct measurement of the efficiency for each vertex cut, without recourse to a reference spectrum that may have the wrong shape. Although `calc-then-add` and `add-then-calc` are equivalent in principle, the former can suffer from division-by-zero in EH3 on days where there is no event in a given bin, so in practice we reject it in favor of `add-then-calc`.

9.3.3 Results

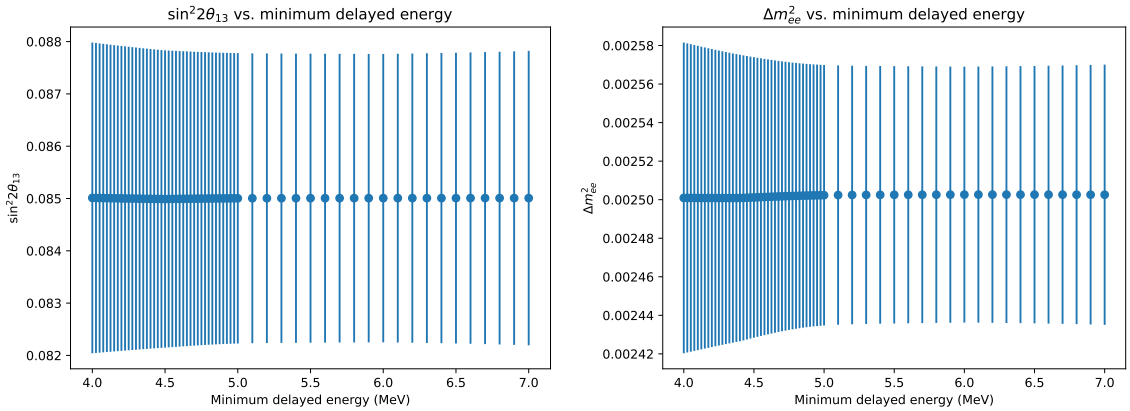

Figure 9.12: Fit results from simulating the variation of the delayed-energy cut in the toy MC.

As a sanity check, we begin by fitting the toy MC samples. The toy MC does not simulate the delayed-energy spectrum, but we can still adjust the efficiencies and background rates in the `Theta13` file to verify that the oscillation fit remains stable. In particular, for the results shown in Figure 9.12, we inserted the delayed-energy efficiencies calculated using the neutron capture-spectra obtained from the nominal IBD selection (i.e., the `original` implementation), integrated according to the `relative` approach described previously. The background rates were scaled as described in Equation 9.4.

The toy MC results are shown in Figure 9.12. They demonstrate the desired level of internal consistency, as both $\sin^2 2\theta_{13}$ and Δm^2_{ee} are practically constant as the delayed-energy cut is varied, with only a negligible increase in $\sin^2 2\theta_{13}$ for the extreme case of a delayed-energy cut near 4 MeV. We observe a $\sim$20% increase in the uncertainty of both parameters near a delayed-energy cut of 4 MeV, which can be attributed to the steep increase in the accidental background, as shown in Figure 9.13.

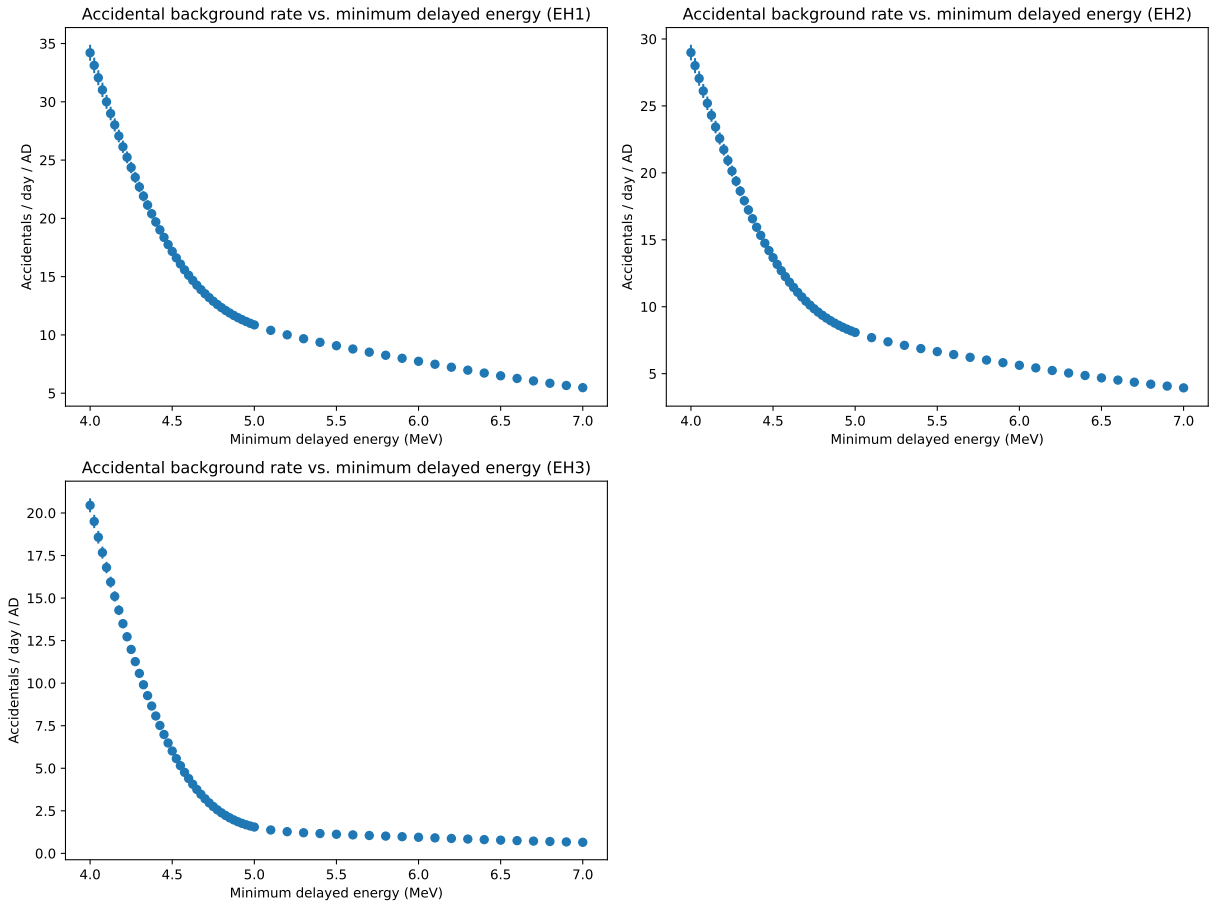

Figure 9.13: Accidental background rates (livetime-averaged across ADs in each hall) as a function of the delayed-energy cut. The full singles spectrum (from 0.7 to 12 MeV), taken from the nominal IBD selection, was used for the calculation.

In Figure 9.14 we show the results of varying the delayed-energy cut for the P17B data sample. In the legend, `old` refers to the `original` method of measuring the nGd spectrum, while `new` refers to the `add-then-calc` method, as described previously. Meanwhile, `flat`, `rel` and `abs` refer, unsurprisingly, to the `flat`, `relative` and `absolute` approaches for integrating the nGd spectrum.

The first thing to note is that there is almost no appreciable difference between `old` and `new`[10]. Thus we focus on the differences between `flat`, `relative`, and `absolute`. In this discussion, we focus on $\sin^2 2\theta_{13}$, since all five methods give essentially the same results for Δm^2_{ee}. The `flat` approach shows by far the most variation in $\sin^2 2\theta_{13}$, affirming the value of attempting to correct for AD-dependent variations in the delayed-energy cut efficiency. Both the `relative` and `absolute` approaches are generally

[10]This will not be the case when we apply a vertex cut, as shown in Section 9.5.5.

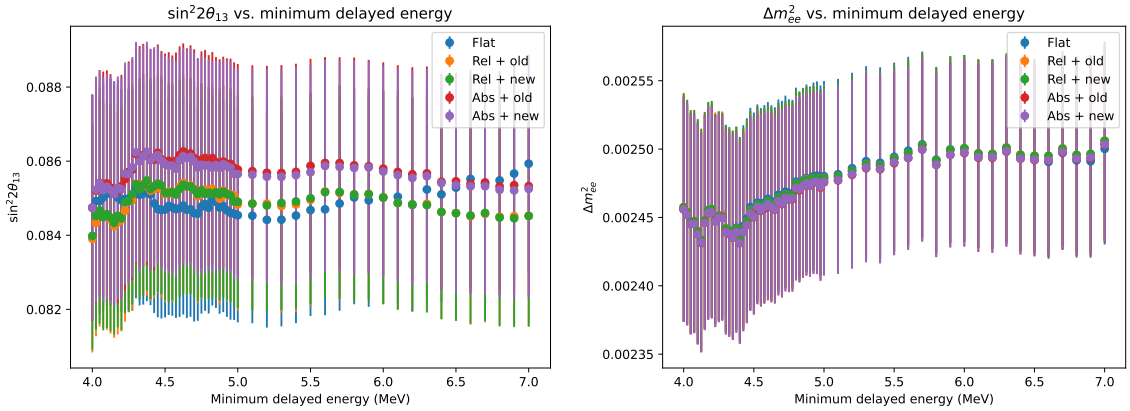

Figure 9.14: Fit results from actual data when varying the delayed-energy cut. Shown are five different methods of measuring and correcting for the efficiency of the cut.

fairly stable; although some unexpected behavior is observed below 5 MeV, this amounts to only 10-20% of the uncertainty. Unsurprisingly, an offset exists between the two approaches, as expected given that the `absolute` approach abandons the (arbitrary and unfounded) assumption that the ADs all have the same delayed-energy cut efficiency at 6 MeV. The uncertainty associated with this assumption is encoded in the 0.11% relative detection efficiency uncertainty (Section C.2.2) used by the toy MC in generating the systematic covariance matrix, so it would be incorrect to assign an additional systematic based on these results.

As a sanity check, suppose that the far hall has a detection efficiency that is 0.11% lower than that of the near halls. Since the near-to-far disappearance probability is about 5%, this 0.11% efficiency bias corresponds to a bias on the disappearance probability of 0.11/5, or about 2%. The fitter's reported total uncertainty on $\sin^2 2\theta_{13}$ is at the 5% level, so we crudely estimate an 0.4σ shift in $\sin^2 2\theta_{13}$ from this fluctuation in efficiencies. The actual shift we observe is consistent with this (and in fact smaller by a factor of 2). Thus, it is no cause for alarm, and it stands in agreement with our uncertainty budget.

The behavior of Δm^2_{ee} below 5 MeV is surprising: There is a significant downtrend, culminating at approximately a third of the uncertainty. Moreover, two "bumps" are visible. These results suggest an unaccounted-for shape distortion for delayed-energy cuts below 5 MeV. The most likely explanation is that this comes from background. Although the total background *rate* calculation appears to remain generally correct, given the stability of $\sin^2 2\theta_{13}$, it is possible that the background shape is being treated incorrectly. We further explore possible explanations of this effect in Section 9.5. In order to confirm that this behavior is not an artifact of the energy binning we employ, in Figure 9.16 we compare the fit results between our "BCW" binning and the alternative "LBNL" binning (see Section C.2.1). From this comparison, it is

clear that the binning is not the issue. For the time being, we simply note that Δm^2_{ee} remains extremely stable between 5 and 7 MeV. Below 5 MeV, we are entering "dangerous" territory, where there was never any intention for the nGd analysis to tread, so anomalous behavior here is both unsurprising and unconcerning.

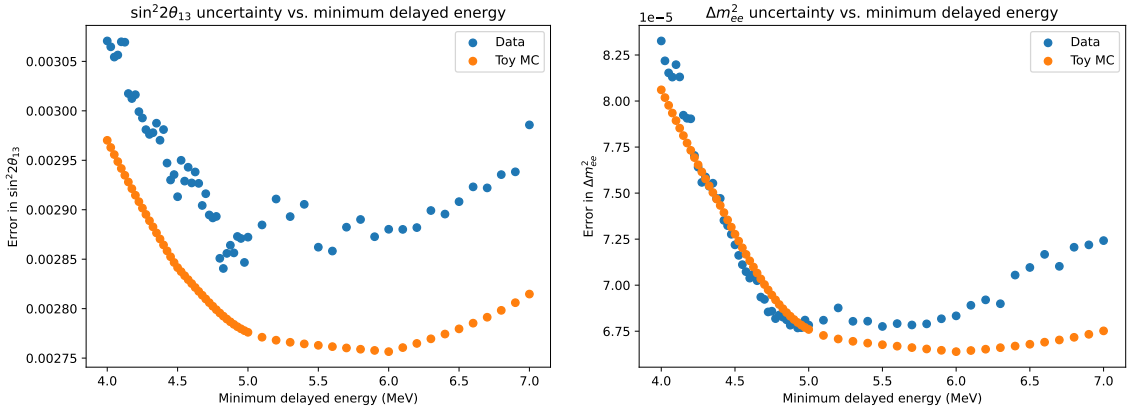

Figure 9.15: Uncertainties in $\sin^2 2\theta_{13}$ and Δm^2_{ee} under variations of the delayed-energy cut for both the toy MC and P17B data. The increase for low delayed-energy cuts is due to the accidental background, while the increase for high cuts arises from the loss in statistics. The toy MC has a minimum at 6 MeV; rather than being a coincidence, this is attributable to the fact that our artificial adjustments to the toy MC's signal and background rates are anchored at 6 MeV. The fits to data, while possessing more scatter, are in qualitative agreement with the toy MC's prediction of a largely flat region between 5 and 6.5 MeV.

We end this section by showing the fitter's reported uncertainty under variations of the delayed-energy cut (Figure 9.15). We do not expect perfect agreement here between the toy MC and the data, as the latter includes statistical fluctuations and systematic deviations, while the former uses the "Asimov" sample in which there is no statistical fluctuation and nominal assumptions are used for all systematics. Nevertheless, the overall scale of the uncertainty is in general agreement, as is the general form of a lopsided "U" shape. From these plots we can conclude that there is essentially no advantage, in terms of the fit uncertainty, to varying the standard 6 MeV.

9.4 Minimum prompt energy

Unlike the delayed-energy spectrum, the prompt-energy spectrum of the IBDs does not contain a long tail, so that varying the prompt-energy cut is likely to have a smaller effect on the analysis, with minimal potential statistical gain. Although the minimum

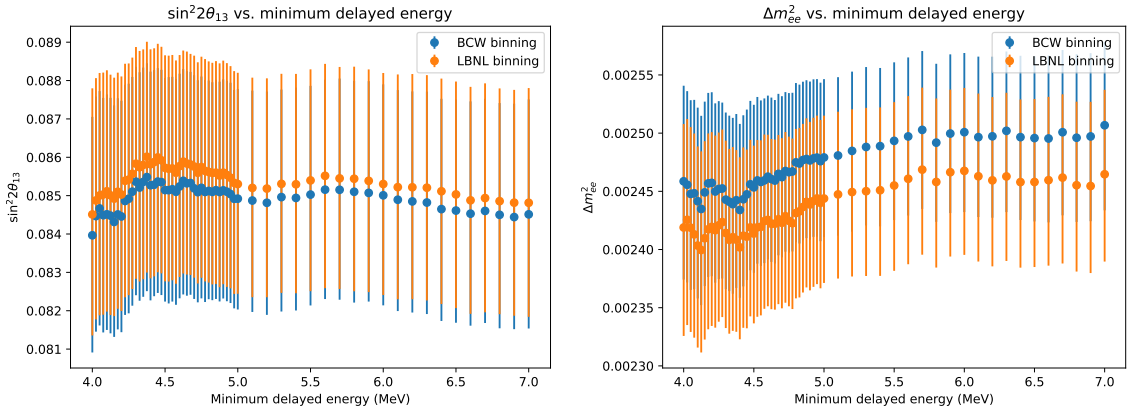

Figure 9.16: Comparison of delayed-energy cut variation (when fitting P17B data) between LBNL and BCW binning. The offset in Δm^2_{ee} between the two binnings is a known consequence of unresolved biases in the modeling of the low-energy response of the ADs (Section C.2.1). The BCW binning, which we employ, uses coarser bins, particularly at low energies, and is therefore less sensitive to these biases.

IBD energy deposition is 1.0 MeV (from positron annihilation), the finite detector resolution leads to the reconstruction of some events below this minimum energy. Since we can see that the prompt-energy spectrum doesn't fall to zero at 0.7 MeV (Figure 9.17), it is clear that events will be lost or gained upon variation of the cut. It is therefore worth verifying that the analysis is stable under such circumstances. In particular, prompt-energy bins below 1 MeV are especially sensitive to the IAV effect (Section C.2.3.2), with its associated systematic uncertainty. Thus, there may be a prompt-energy cut that optimally balances the advantage of increased statistics with the disadvantage of increased IAV-related uncertainty.

A priori, there is no reason to assume that the prompt-energy cut has the same efficiency (and $d\epsilon/dE_{\text{cut}}$) across all ADs. Just as with the delayed-energy cut, it is possible that we may need to apply corrections for the AD-dependent prompt-energy cut efficiency. In that case, we could extract the prompt-energy spectrum analogously to the case of the delayed-energy spectrum, and then integrate it using either the `absolute` or `relative` methods. However, as shown later in Section 9.4.2, the `flat` method (in which we make no attempt to correct for the prompt-energy cut efficiency) produces remarkably stable results. There is therefore no need to extract the prompt-energy spectrum. We thus proceed directly to discussing the necessary corrections to the background rates for this study.

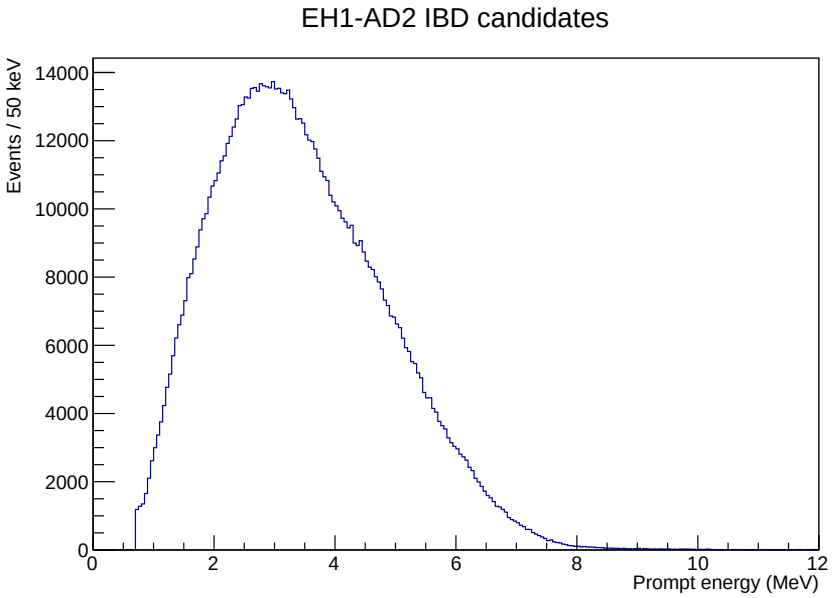

Figure 9.17: Prompt-energy spectrum of IBD candidates (including backgrounds) in EH1-AD2, showing the nonzero event count at the 0.7 MeV threshold.

9.4.1 Background rate correction

When varying the prompt-energy cut, the only adjustment we must make to the `Theta13` file is to correct the rates of the correlated backgrounds. Since we are in the possession of the predicted prompt-energy spectra $S_b(E)$ for each correlated background b (Figures 9.18–9.21), and since these spectra extend down to 0 MeV, we have all of the information needed. We simply take the nominal background rate (and uncertainty) and scale it according to the factor

$$F_b = \frac{\int_{E_{\text{cut}}}^{12} S_b(E)}{\int_{0.7}^{12} S_b(E)}. \tag{9.5}$$

The fitter, when subtracting the background, then takes $S_b(E)$, cuts it off at E_{cut}, normalizes it to the specified rate, and subtracts it from the raw prompt-energy spectrum. This adjustment of the correlated backgrounds is applied to both the cases of fitting data and toy MC samples. There are no further subtleties related to the variation of the prompt-energy cut.

9.4.2 Results

Figure 9.22 shows the results of varying the prompt-energy cut. There is essentially no variation in $\sin^2 2\theta_{13}$, and only minimal variation in Δm^2_{ee}. The uncertainties remain

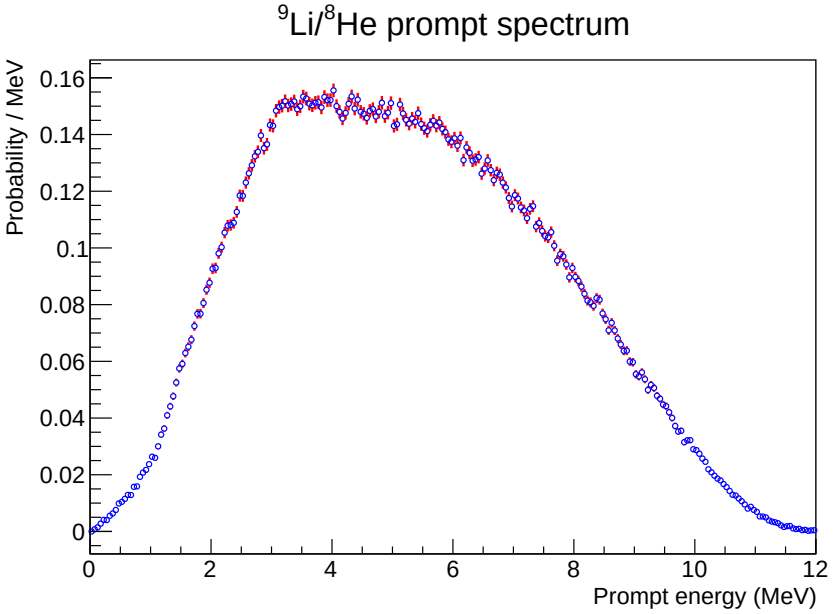

Figure 9.18: Prompt-energy spectrum of the ^{9}Li/^{8}He background. Generated from the input file used by the fitter.

stable within 2%, as shown in Figure 9.23. The increased scatter above 0.8 MeV is expected given the steepness of the prompt-energy spectrum and our lack of a correction for the prompt-energy cut efficiency. Although these results suggest that an "optimal" uncertainty may be achievable with a prompt-energy cut between 0.7 and 0.8 MeV, the gain would amount to only a $\sim$1% reduction in the final uncertainty on the oscillation parameters. Thus, with minimal effort, we have demonstrated that our analysis is stable against changes in the prompt-energy cut, and that there is nothing significantly suboptimal about the nominal cut of 0.7 MeV.

9.5 Vertex cut

Since our analysis is based on nGd-capture IBDs, the neutron captures (i.e. the delayed triggers of the IBD pairs) all take place within the GdLS volume. Given the small (cm-scale) neutron diffusion distance, most of the IBD interactions themselves (i.e. the prompt triggers, containing the positron energy) likewise occur in the GdLS. The situation is the same for the three correlated backgrounds (^{9}Li, fast neutrons, and α-n) that involve nGd captures. Meanwhile, for accidental backgrounds, there is no correlation between the prompt and delayed vertices. And finally, the AmC backgrounds have a tight prompt-delayed vertex correlation, but with a unique vertex distribution localized toward the top of the AD near the three ACUs.

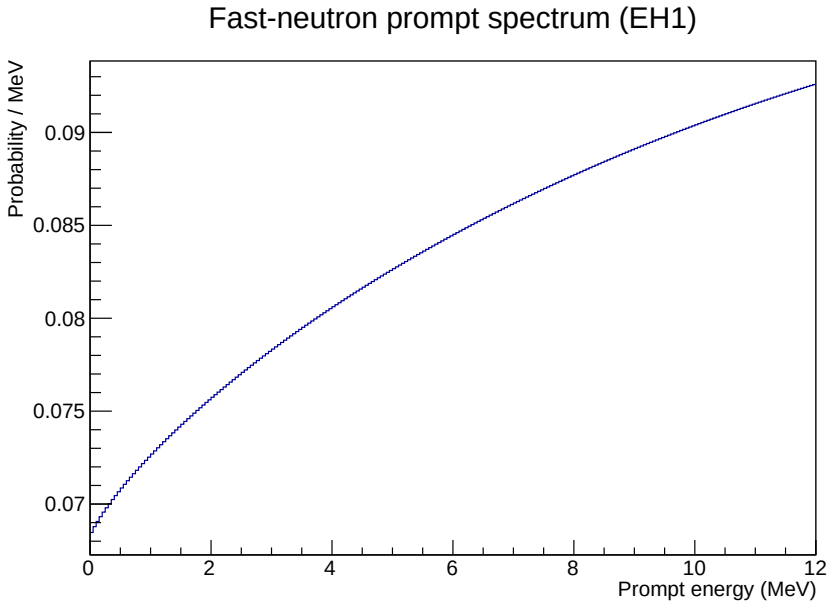

Figure 9.19: Prompt-energy spectrum of the fast-neutron background. Generated from the input file used by the fitter.

Based on these facts, it is obvious why the official nGd-based analysis does not employ any cuts on the event vertices. Although there is nothing preventing us from requiring reconstructed vertices to lie within the GdLS volume (or any other fiducial volume), such a cut will gain only a small reduction of the accidental and AmC backgrounds. This small gain is outweighed by the risk of introducing an AD-dependent efficiency of the vertex cut, which can bias $\sin^2 2\theta_{13}$ if it is not carefully measured and corrected for.

Nevertheless, it is worthwhile to explore the results of applying an arbitrary vertex cut *(with an efficiency correction)*, if only as a demonstration of the robustness of our analysis and the reliability of the vertex reconstruction. Going beyond this motivation, a vertex cut can be useful in exploring the behavior of *other* cuts. In particular, we recall that, as shown in Figure 9.14, the best-fit value of Δm^2_{ee} trends downward as we reduce the delayed-energy cut below 5 MeV. In Section 9.3.3, we hypothesized that an unaccounted-for correlated background, potentially originating from the AmC sources, may be responsible for this behavior. Given that the AmC-associated events occur closer to the top of the AD, the application of (vertical) vertex cuts can provide a means of testing this hypothesis: If the AmC sources are indeed responsible, then the downward trend in Δm^2_{ee} should be lessened when we select events closer to the bottom of the AD.

For these reasons, we choose to go beyond the official analysis and explore the use of vertex cuts. When applying a vertex cut, care must be taken in two areas:

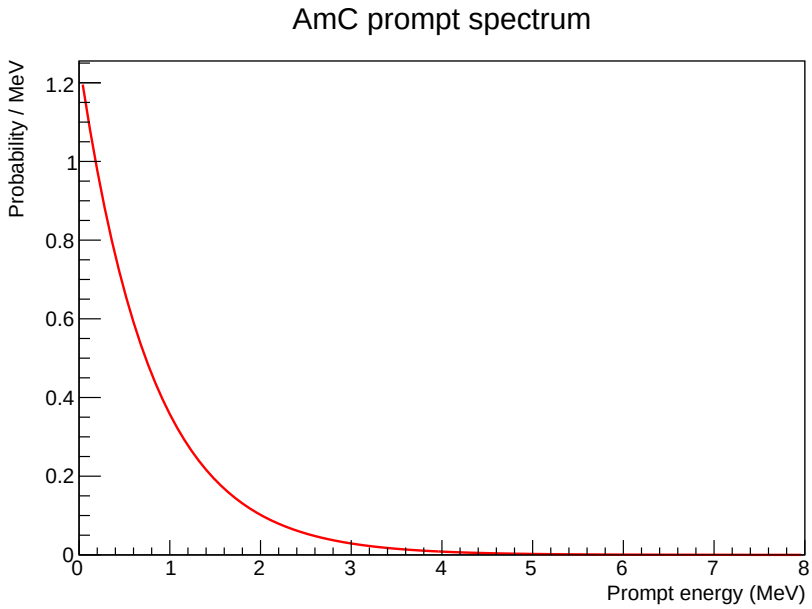

Figure 9.20: Prompt-energy spectrum of the AmC background. Generated from the input file used by the fitter.

the detection efficiencies and the background rates. As was mentioned, a vertex cut may not have the same efficiency among all ADs, so this efficiency must be measured and corrected for. The accidental-background rate calculation, being based on the singles selection (to which we apply the same vertex cut), accounts for the vertex cut automatically. As for the correlated backgrounds, we must assume a particular vertex distribution for each background, and then calculate the percentage of the distribution that lies within the fiducial volume.

Throughout this study, we apply a given vertex cut to both the prompt (positron) and the delayed (neutron) vertex. This is to be distinguished from, say, the muon veto, which considers only the time of the *delayed* trigger. We choose to cut on both vertices because it avoids the need to make an arbitrary choice between cutting on the prompt or on the delayed vertex, and it avoids the need to perform two separate singles selections with two different vertex cuts.

In the sections that follow, we first describe the measurement of the vertex cut efficiency and how we correct for it, followed by our adjustments of the correlated-background rates. We then proceed to explore the effects of applying five different vertex cuts (three vertical and two cylindrical; see Table 9.1) to the nominal selection, before extending to the case of modified delayed-energy cuts, with the goal of gaining insight into the trends observed in Δm^2_{ee}. We end by drawing what conclusions we can regarding those trends as well as the overall robustness of our analysis to vertex cuts.

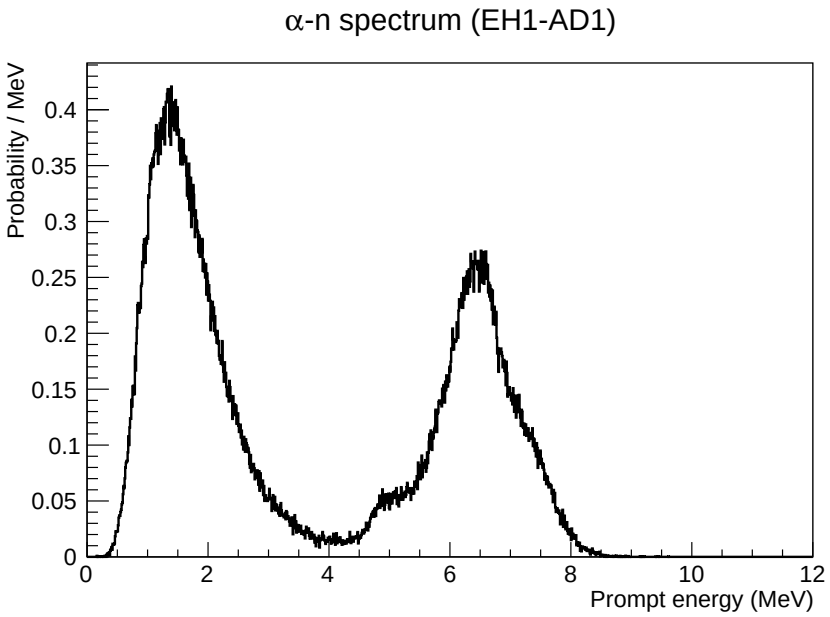

Figure 9.21: Prompt-energy spectrum of the α-n background. Generated from the input file used by the fitter.

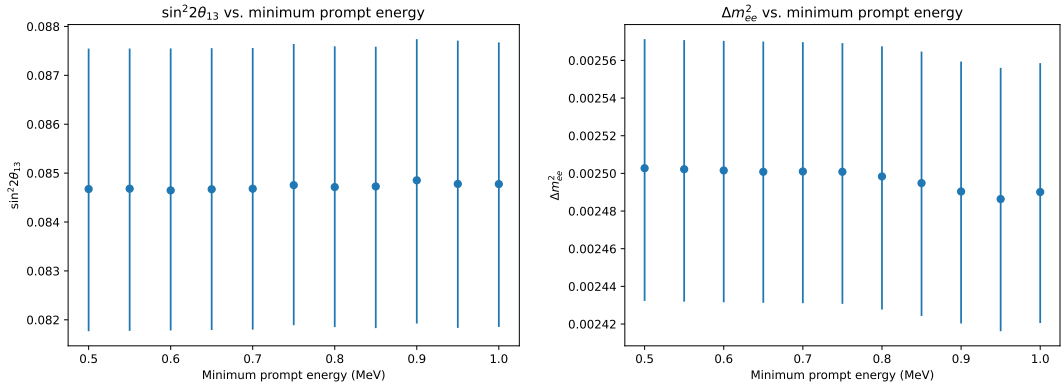

Figure 9.22: Results of oscillation fits to data under variations of the prompt-energy cut. On the left is $\sin^2 2\theta_{13}$, and on the right is Δm^2_{ee}.

9.5.1 Implementation details

Within the IBD selection, the vertex cut is implemented using the vertex provided by the `AdSimple` reconstruction (Section D.3), after conversion from Cartesian coordinates to the natural cylindrical coordinate system of the AD. The cut involves four parameters, corresponding to the minimum and maximum acceptable radial and vertical coordinates. Both the singles selection and the IBD selection apply the cut, in

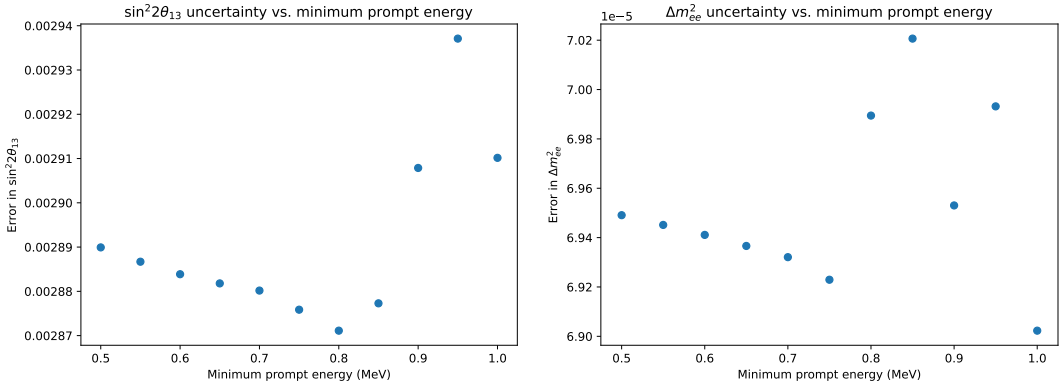

Figure 9.23: Uncertainty on the oscillation parameters from fits to data under variations of the prompt-energy cut. On the left is $\sin^2 2\theta_{13}$, and on the right is Δm^2_{ee}. Note the scales on the vertical axes; the variation in the uncertainty is minuscule. The small (but sharp) dip around 0.85 MeV is attributable to the balancing act between statistics and systematics (IAV effect) in the lowest-energy bins.

the latter case to both the prompt and delayed triggers. Furthermore, when applying the multiplicity cut, any "extra" triggers are ignored if they do not pass the vertex cut (i.e., such triggers will not cause the single/IBD candidate to fail the multiplicity cut). In effect, any (non-muon) trigger that fails to pass the vertex cut is completely excluded from the analysis.

9.5.2 Definitions of vertex cuts

For this study, five different vertex cuts were investigated: Three "vertical" cuts, and two "cylindrical" cuts. Their details are listed in Table 9.1, along with the names used in referring to them. It is worth noting that these cuts do *not* produce five statistically independent subsamples of the data. In particular, there is an overlap between each vertical sample and each cylindrical sample. However, the three vertical samples are independent from each other, as are the two cylindrical samples.

Name	Definition
`zTopThird`	$z \in [0.5,\ 1.5]\,\mathrm{m}$
`zMidThird`	$z \in [-0.5,\ 0.5]\,\mathrm{m}$
`zBotThird`	$z \in [-1.5,\ -0.5]\,\mathrm{m}$
`rInside1000`	$r < 1\,\mathrm{m}$
`rOutside1000`	$r > 1\,\mathrm{m}$

Table 9.1: The vertex cuts under investigation.

9.5.3 Efficiency calculation

The vertex-cut efficiency is determined by taking the corrected IBD rate obtained with the vertex cut, and comparing it to the corrected rate obtained from an IBD selection (the "reference" cut) which has no vertex cut but otherwise identical cuts. The meaning of "corrected" will be explained shortly. If the vertex cut is the only modification made to the nominal IBD selection, then we simply use the nominal selection as a reference; however, if we modify any other cuts (such as the delayed-energy cut) while applying the vertex cut, we must perform a reference IBD selection with those same modifications (minus the vertex cut). This ensures that the measurement is not biased by any efficiency variation caused by the other modified cuts, which could occur if we simply used the nominal selection as the reference in all cases.

The corrected IBD rate R_{corr} is calculated straightforwardly: We simply take the total number of IBD candidates N_{raw}, subtract the total predicted number of backgrounds, and divide by the livetime, the veto efficiency, and the multiplicity efficiency. Since the IBD selection's convention is to output the background prediction R_{bkg} as an efficiency- and livetime-corrected daily rate, we must undo these corrections before performing the subtraction. In summary:

$$\begin{aligned} N_{\text{bkg}} &= R_{\text{bkg}} \cdot T \cdot \epsilon_\mu \cdot \epsilon_m \\ R_{\text{corr}} &= (N_{\text{raw}} - N_{\text{bkg}}) \;/\; (T \cdot \epsilon_\mu \cdot \epsilon_m) \end{aligned} \tag{9.6}$$

The efficiency of the vertex cut is then obtained by simply taking the ratio with respect to the reference selection:

$$\epsilon_{\text{vtx}} = R^{\text{cut}}_{\text{corr}} \;/\; R^{\text{ref}}_{\text{corr}}. \tag{9.7}$$

9.5.4 Background rate adjustment

In the preceding discussion of the efficiency calculation, the background rate R_{bkg} is itself implicitly a function of the vertex cut, and is thus different between the selection in question and the reference selection. This modified value of R_{bkg} must also, of course, be the one we use when subtracting the backgrounds during the oscillation fit. Here we discuss how the five background rates are affected by the use of a vertex cut. In what follows, the parameters of any given vertex cut will be denoted by r_{min}, r_{max}, z_{min}, and z_{max}.

9.5.4.1 Accidentals

The accidental background is calculated, as usual, directly from the prompt- and delayed-like singles rates obtained from the singles selection. Since we apply the same vertex cut in the IBD and the singles selection, the calculated accidentals rate automatically takes the vertex cut into account. No adjustment is therefore necessary.

9.5.4.2 ^{9}Li, fast neutrons, and α-n

These backgrounds are assumed to be uniformly distributed throughout the GdLS. However, the fast neutrons have a slight preference to arrive from above and stop inside the AD; accordingly, some bias toward the top is known to exist ([67], p. 9), but for our present purposes, we make no attempt to model this effect. Meanwhile, for ^{9}Li/^{8}He, we know (using EH1 as an example) that the production rate of O(1) event/day is far lower than the shower-muon rate of O(0.1 Hz), implying that the interaction length for ^{9}Li/^{8}He production is much longer than the height of the ADs; as such, no vertical bias is expected for ^{9}Li/^{8}He. Likewise, α activity is uniformly distributed throughout the AD.

Based on this uniformity assumption, we define a "radial" factor f_r to account for the minimum and maximum radii accepted in the cut, and likewise a "vertical" factor f_z corresponding to the minimum and maximum acceptable vertical coordinates. In order to account for the finite resolution of the vertex reconstruction, we "extend" the dimensions of the IAV by 200 mm in every direction, leading to the following definitions of the nominal minimum and maximum coordinates:

$$\begin{aligned} r_{\text{min}}^{\text{nom}} &= 0\,\text{mm} \\ r_{\text{max}}^{\text{nom}} &= 1700\,\text{mm} \\ z_{\text{min}}^{\text{nom}} &= -1700\,\text{mm} \\ z_{\text{max}}^{\text{nom}} &= 1700\,\text{mm}. \end{aligned} \tag{9.8}$$

Since these backgrounds are not expected to occur outside the GdLS, we clamp the vertex cut parameters at the limits defined above:[11]

$$\begin{aligned} r'_{\text{min}} &= \max(r_{\text{min}}, r_{\text{min}}^{\text{nom}}) \\ r'_{\text{max}} &= \min(r_{\text{max}}, r_{\text{max}}^{\text{nom}}) \\ z'_{\text{min}} &= \max(z_{\text{min}}, z_{\text{min}}^{\text{nom}}) \\ z'_{\text{max}} &= \min(z_{\text{max}}, z_{\text{max}}^{\text{nom}}) \end{aligned} \tag{9.9}$$

The radial factor is then

$$f_r = \left(\frac{r'_{\text{max}}}{r_{\text{max}}^{\text{nom}}}\right)^2 - \left(\frac{r'_{\text{min}}}{r_{\text{max}}^{\text{nom}}}\right)^2. \tag{9.10}$$

Here, the first term is the fraction of the IAV we'd be using if r_{min} were zero. The second term then subtracts out the fraction excluded by nonzero r_{min}.

Meanwhile, the vertical factor is simply

$$f_z = \frac{z'_{\text{max}} - z'_{\text{min}}}{z_{\text{max}}^{\text{nom}} - z_{\text{min}}^{\text{nom}}} \tag{9.11}$$

[11] Note that $r'_{\text{min}} = r_{\text{min}}$ for any sensible (i.e. nonnegative) value of r_{min}.

Each of these three backgrounds is then adjusted according to

$$R = R^{\text{nom}} \cdot f_r \cdot f_z, \tag{9.12}$$

where R^{nom} is the background rate for the nominal IBD selection (without any vertex cut), and R is the predicted background rate after applying the vertex cut in question.

9.5.4.3 AmC background

Since the AmC background events are produced near the top of the AD, we assume a linearly rising vertical profile terminating at the center of the AD, in accordance with the results in [59]. For the sake of brevity, in what follows we let

$$Z \equiv z_{\text{max}}^{\text{nom}}. \tag{9.13}$$

Now letting $F(z)$ denote the fraction of events lying between the center of the AD and a vertical coordinate of z (above the center), we have

$$\frac{dF}{dz} = \begin{cases} \dfrac{kz}{Z}, & \text{for } 0 < z < Z, \\ 0 & \text{otherwise,} \end{cases} \tag{9.14}$$

where k is a normalization constant determined by the condition

$$F(Z) = 1. \tag{9.15}$$

Integrating dF/dz then gives

$$F(z) = \begin{cases} 0, & \text{for } z < 0, \\ \dfrac{kz^2}{2Z}, & \text{for } 0 < z < Z, \\ \dfrac{kZ}{2}, & \text{for } z > Z, \end{cases} \tag{9.16}$$

From Equation 9.15 it follows that

$$k = \frac{2}{Z}, \tag{9.17}$$

so that, finally,

$$F(z) = \begin{cases} 0, & \text{for } z < 0, \\ \dfrac{z^2}{Z^2}, & \text{for } 0 < z < Z, \\ 1 & \text{for } z > Z, \end{cases} \tag{9.18}$$

The AmC vertical factor f_z^{AmC} is then given by

$$f_z^{\mathrm{AmC}} = F(z'_{\mathrm{max}}) - F(z'_{\mathrm{min}}). \tag{9.19}$$

As for the radial distribution, even though the AmC sources are located at three discrete points, we assume that the inherent resolution of the vertex reconstruction, along with the scattering of the neutrons and gamma-rays in the stainless steel lid, together produce enough smearing to result in uniformity. Therefore, we use the same f_r as above. The AmC background is a small one, so any approximations here are not liable to significantly affect the overall analysis.

9.5.5 Results

Figure 9.24 shows the results obtained when a vertex cut is applied on top of the nominal IBD selection. For the most part, each cut produces a result that lies less than 1σ from the full-AD result. This is encouraging, especially given the simplifications employed in correcting the background rates (particularly for the fast-neutron background). Moreover, statistically independent subsamples are not overly biased toward either side of the nominal result, providing a reassuring sign of internal consistency. The deviations are not entirely symmetric about the nominal result, suggesting (unsurprisingly) the introduction of new systematics, likely from the efficiency measurement or background corrections.[12]

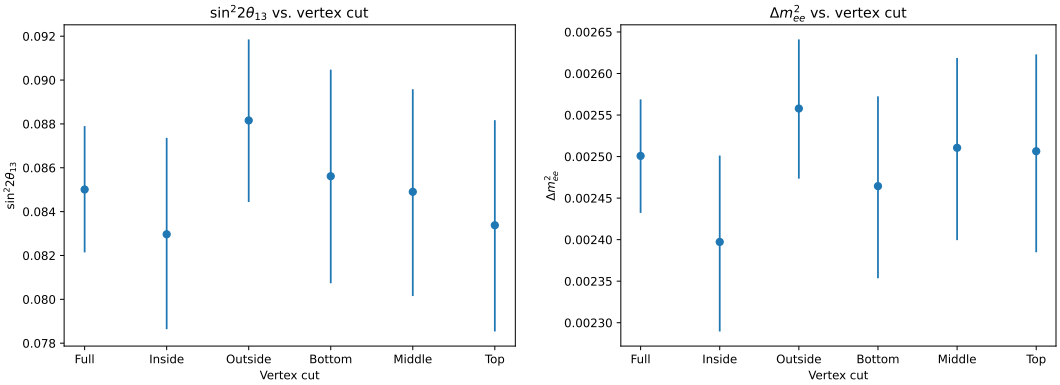

Figure 9.24: Variations in $\sin^2 2\theta_{13}$ and Δm^2_{ee} when applying a vertex cut to the otherwise-nominal selection.

In Figure 9.25, we show the results of applying these five vertex cuts in tandem with modifications of the delayed-energy cuts. One of the main goals of this study is

[12] Although, due to the complex nature of the fitter, it is not obvious that two statistically independent subsamples, even generated using identical IBD cuts, will necessarily give results symmetrically distributed around the nominal one.

to investigate possible explanations (particularly, localized correlated backgrounds) for the behavior of Δm^2_{ee} shown in Figure 9.14 for delayed-energy cuts below 5 MeV. Indeed, we observe that the `rInside1000` and `zMidThird` cuts produce remarkably flat behavior for both $\sin^2 2\theta_{13}$ and Δm^2_{ee}. (`zTopThird` is also flat for Δm^2_{ee}, but shows a trend for $\sin^2 2\theta_{13}$.) One possible conclusion is that the sub-5 MeV behavior of Δm^2_{ee} (for the full-AD sample) is simply a statistical fluke which goes away when fitting particular subsamples. However, based on these results, it is also quite possible that the behavior is caused by events localized near the edge and/or the bottom of the AD. We therefore conclude that it is unwise to apply a delayed-energy cut below 5 MeV due to the possibility of introducing uncharacterized backgrounds to the IBD sample; if, for some reason, such a low delayed-energy cut is ever warranted in a particular analysis, then it may be prudent to apply a vertex cut to only select events near the centers of the ADs.

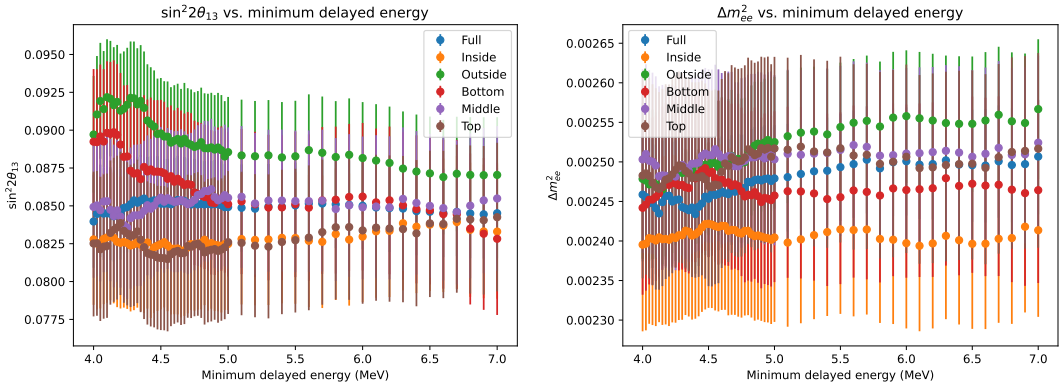

Figure 9.25: Variations in $\sin^2 2\theta_{13}$ and Δm^2_{ee} when applying a vertex cut in tandem with a modified delayed-energy cut.

9.6 Joint variation of cuts

Thus far, we have largely studied each cut independently of the others. (The exceptions are the shower-muon threshold and veto window, which we have varied jointly, as well as the simultaneous application of a vertex cut and a modified delayed-energy cut). The results have indicated that our analysis is acceptably stable under such variations. However, it is hypothetically possible that problems may arise when more than one cut is varied at the same time. As such, we conclude these studies by exploring the effects of jointly modifying all of these cuts (the two shower-muon veto parameters, the minimum delayed energy, and the minimum prompt energy).

To carry out this study, 100 sets of cut values were randomly generated from Gaussian distributions whose parameters were chosen to cover the same ranges of

values used when studying the cuts individually. Each distribution also included hard cutoffs at both ends in order to exclude extreme values. The parameters of these distributions are given in Table 9.2. Individual histograms of the generated values are shown in Figure 9.26.

Cut	Min	Max	Mean	Sigma
Shower muon threshold ($\times 10^5$ p.e.)	2.2	4.6	3.4	0.6
Shower veto time (s)	0.25	2	1.125	0.4375
Min delayed energy (MeV)	5	7	6	0.5
Min prompt energy (MeV)	0.5	1	0.75	0.125

Table 9.2: Parameters of the Gaussian distributions (with cutoffs) used in generating random cut values for the joint variation study.

Figure 9.27 shows the histograms of the best-fit values of $\sin^2 2\theta_{13}$ and Δm^2_{ee} for these 100 sets of cuts. The root-mean-square (RMS) spread of $\sin^2 2\theta_{13}$ is ~ 0.0006, significantly smaller than the nominal error on $\sin^2 2\theta_{13}$ of ~ 0.0029 (Section 8.5). Similarly, the RMS spread of Δm^2_{ee} is $\sim 1.9 \times 10^{-5}$, again a fraction of the nominal uncertainty of $\sim 6.9 \times 10^{-5}$. These relatively small spreads are expected, given that the fits are all highly correlated with each other, both by virtue of their shared statistics (i.e. IBD candidates) and their shared systematics (i.e. the same detectors and reactors). Some of the spread can be credited to the residual differences in the IBD samples, while the rest is attributable to systematic effects of varying the cuts (e.g. from imperfect corrections to efficiencies and background rates). We are interested in the second contribution, as it is a novel systematic that is not included in the covariance matrix. Isolating it would be equivalent to measuring (and subtracting) the statistical component. In principle, it is possible to estimate the statistical contribution, for instance by considering the number ΔN_i of non-overlapping IBDs between each sample i and some common reference sample (containing N IBDs), averaging ΔN_i across all samples, and then taking (for example)

$$\sigma = \sqrt{\frac{\langle \Delta N \rangle}{N}} \sigma_{\text{stat}} \tag{9.20}$$

to be the expected statistical spread (of the fit results for a particular oscillation parameter) in the set of fits. Here, σ_{stat} is the uncertainty reported by the fitter *when the systematic covariance matrix is disabled.* However, this procedure comes with its own questions: Is the result reproducible across different choices of the reference IBD sample? Would it be more correct to move the expectation-value operator outside the square-root in Equation 9.20? Rather than wrestle with such issues, we conservatively attribute *all* of the observed spread to the systematic effects of varying the cuts.

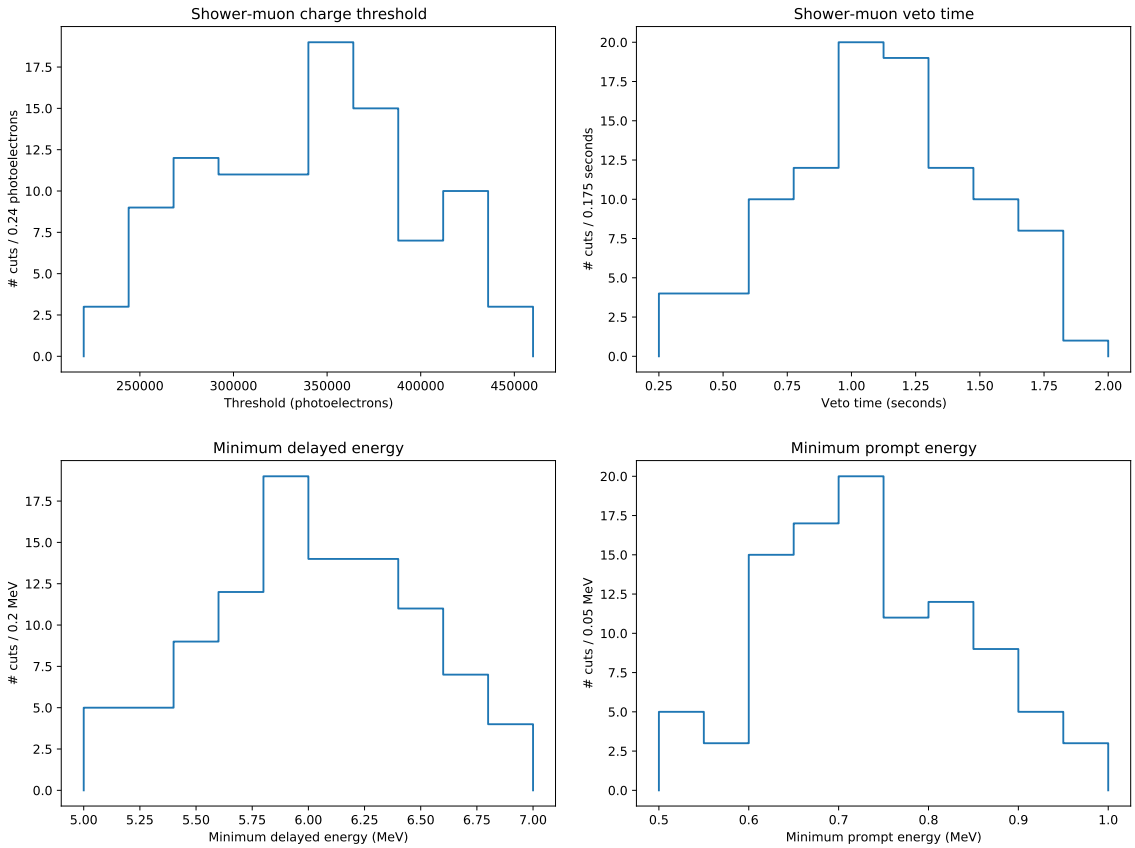

Figure 9.26: Projection histograms of the 100 random cuts used in the joint variation study. Each histogram corresponds to one of the four cut parameters. Each full cut (consisting of the four parameters, sampled from uncorrelated truncated-Gaussian distributions) is represented once in each histogram. The histograms have largely the expected shapes (based on the underlying distributions), and there are no notable outliers or other obvious sources of potential bias in this sample.

For $\sin^2 2\theta_{13}$ (Δm^2_{ee}), this means that we assign an additional absolute systematic of 0.0006 (1.9×10^{-5} eV^2).

In Figure 9.28, we simultaneously plot each best-fit oscillation parameter along with the corresponding cut values, illustrating again that the scale of variation is smaller than the error on each oscillation parameter, and that there are no obvious correlations between the cut values and the best-fit parameters. There are no new quantitative conclusions to be drawn from this plot, but the lack of pathological behavior[13] is reassuring, suggesting that we can conclude this study here, having succeeded in our goal of (conservatively) quantifying the systematic uncertainty associated with the

[13] Aside from some truncated error bars, which we speculate to arise from the MINUIT fitter's occasional failure to sample the parameter space sufficiently finely.

freedom to vary the IBD selection criteria.

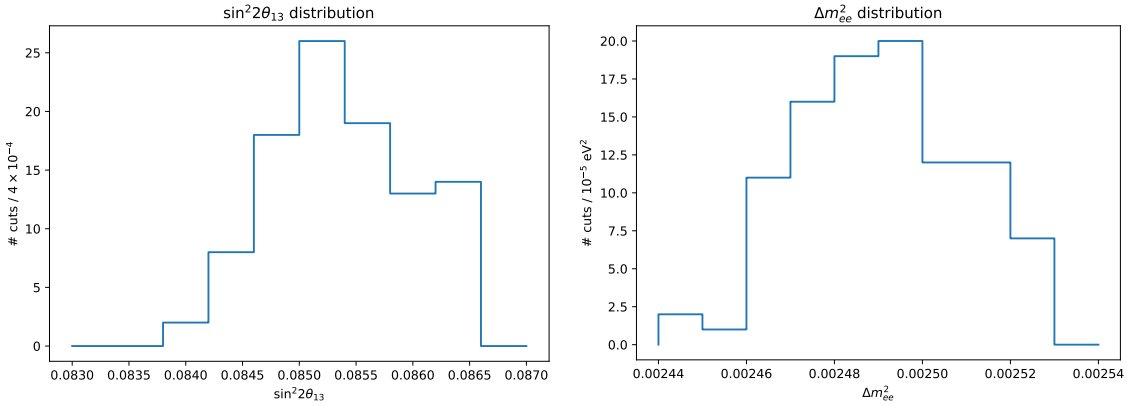

Figure 9.27: Projection histograms of the best-fit oscillation parameters for the 100 random cuts used in the joint variation study.

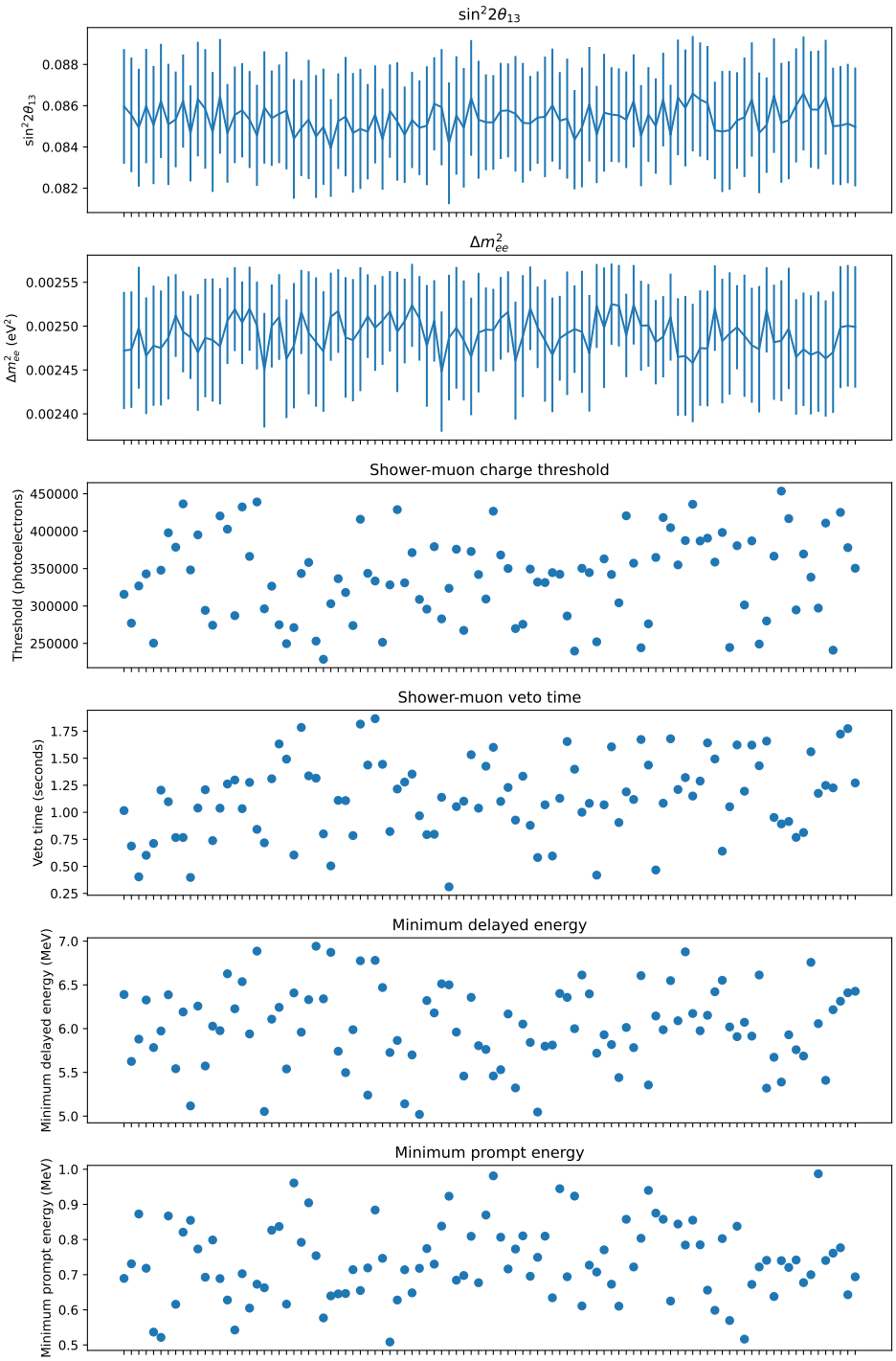

Figure 9.28: Best-fit $\sin^2 2\theta_{13}$ (top panel) and Δm^2_{ee} (second panel from top) versus the four cut parameters (bottom four panels) for the 100 random cuts in the joint variation study. The horizontal axis runs over the 100 cuts, so that each vertical column (of five points) corresponds to one cut.

Chapter 10

Conclusion

In Equation 8.16, we gave the best-fit oscillation parameters obtained when using the nominal IBD selection cuts. These values form the basis of our final result. However, in Section 9.6, we assessed an additional systematic of 0.0006 (1.9×10^{-5} eV^2) on $\sin^2 2\theta_{13}$ (Δm^2_{ee}), based on the degree of variation observed when the IBD selection cuts were randomly fluctuated. We therefore add this systematic (in quadrature[1]) to the fitter's reported (statistical + systematic) uncertainty of 0.0029 (6.9×10^{-5} eV) from Equation 8.16, giving our final result:

$$\sin^2 2\theta_{13} = 0.0850 \pm 0.0030 \qquad \Delta m^2_{ee} = (2.5010 \pm 0.0072) \times 10^{-3}\ \text{eV}^2, \tag{10.1}$$

We regard these uncertainties as comprehensive: They include the effects of statistics (from the toy MC's covariance matrix), signal/background systematics (again, from the toy MC), and the systematic uncertainty (added manually) arising from the cut-variation study. This work represents the most detailed assessment to-date of this final uncertainty, demonstrating that the Daya Bay experiment and oscillation analysis are of sufficiently high quality as to be largely insensitive to cut variations. In the worst case (Δm^2_{ee}), the total uncertainty is increased by only 4% when we account for cut variations. Nevertheless, it would be worthwhile for future Daya Bay results to be verified as stable against changes in IBD cuts; this work thus provides a foundation for further increasing the community's high confidence in the results of Daya Bay.

[1] In principle, this uncertainty may be partially correlated with the systematics encoded in the toy MC's covariance matrix (Section C.2). If this correlation were complete, and the total statistical uncertainty were nonexistent, then it would be appropriate to add this uncertainty linearly, rather than in quadrature. However, given that the P17B data sample induces a statistical uncertainty that is nearly equal to the total systematic uncertainty [68], and given that we have conservatively attributed the oscillation fit's cut-dependence entirely to systematics (thus ignoring its statistical component), the correlation between the cut-variation systematic and the nominal systematics is expected to be marginal. A fitter based on nuisance parameters, rather than on covariance matrices, may be better suited for studying any potential correlations between the cut-variation systematic and the nominal set of systematics.

In closing, we have presented a detailed description of the Daya Bay oscillation analysis, including original work in such areas as the IBD selection, the ^{9}Li and accidental backgrounds, the efficiencies of various IBD cuts, and the synthesis of these efforts in our study of the effects of varying the IBD cuts. These results are built upon our lower-level work in such areas as calibration and data quality, along with the work of the many others who have helped to design, build, run, and analyze the experiment. Having provided additional evidence of the robustness of Daya Bay's analysis, it is our hope and expectation that confident knowledge in the values of $\sin^2 2\theta_{13}$ and Δm^2_{ee} will help to enable the next generation of discoveries in neutrino physics. For the reader who has not yet had enough of neutrino physics, we explore some of these exciting possibilities in Section E.3.

Bibliography

[1] M. Kramer, https://github.com/mjkramer/SelectorFramework.

[2] M. Kramer, https://github.com/mjkramer/IbdSel.

[3] M. Kramer, https://github.com/mjkramer/dybvdq.

[4] C. Marshall, Daya Bay DocDB **11696** (2018).

[5] J. L. Rosner, in 55th Scottish Universities Summer School in Physics: Heavy Flavor Physics (Aug. 2001), arXiv:hep-ph/0108195.

[6] J. Chadwick, Verhandl. Dtsc. Phys. Ges. **16**, 383 (1914).

[7] W. Pauli, "Pauli letter collection: Letter to Lise Meitner," typed copy.

[8] E. Fermi, Il Nuovo Cimento (1924-1942) **11**, 1 (1934).

[9] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, Science **124**, 103 (1956).

[10] A. G. Cohen, S. L. Glashow, and Z. Ligeti, Physics Letters B **678**, 191 (2009).

[11] Daya Bay and MINOS Collaborations, European Physical Journal C **77** (2017).

[12] C. Giunti and C. W. Kim, *Fundamentals of neutrino physics and astrophysics* (Oxford University Press, Mar. 2007).

[13] P. Booth, R. Johnson, E. Williams, and J. Wormald, Physics Letters B **26**, 39 (1967).

[14] Particle Data Group, Progress of Theoretical and Experimental Physics **2020**, 083C01 (2020).

[15] JUNO Collaboration, Journal of Physics G **43**, 030401 (2016).

[16] D. Dwyer, K.-B. Luk, and Y. Nakajima, Daya Bay DocDB **10497** (2015).

[17] Daya Bay Collaboration, Nuclear Instruments and Methods A **685**, 78 (2012).

[18] Micro-Simulation Technology, http://www.microsimtech.com/pctran/CPR1000.html.

[19] Daya Bay Collaboration, Physical Review D **95** (2017).

[20] W. Beriguete *et al.*, Nuclear Instruments and Methods A **763**, 82 (2014).

[21] Daya Bay Collaboration, Nuclear Instruments and Methods A **811**, 133 (2016).

[22] H. R. Band *et al.*, Journal of Instrumentation **8**, P09015 (2013).

[23] P. DeVore *et al.*, Nuclear Instruments and Methods A **737**, 222 (2014).

[24] Daya Bay Collaboration, `https://physics.princeton.edu/~mcdonald/dayabay/tdr.pdf`, Jan. 2008.

[25] Daya Bay Collaboration, Chinese Physics C **41**, 013002 (2017).

[26] Daya Bay Collaboration, Nuclear Instruments and Methods A **773**, 8 (2015).

[27] D. Dwyer, B. Hu, S. Jetter, M. Kramer, C. Lin, Z. Wang, and Z. Yu, Daya Bay DocDB **6710** (2011).

[28] Daya Bay Collaboration, Physical Review D **93**, 072011 (2016).

[29] Daya Bay Collaboration, Nuclear Instruments and Methods A **940**, 230 (2019).

[30] S. Jetter, Daya Bay DocDB **8732** (2013).

[31] T. Skwarnicki, Ph.D Thesis, DESY F31-86-02 (1986), Appendix E; M.J. Oreglia, Ph.D Thesis, SLAC-236 (1980), Appendix D; J. E. Gaiser, Ph.D Thesis, SLAC-255 (1982), Appendix F.

[32] Z. Yu, L. Wen, L. Zhang, and Z. Wang, Daya Bay DocDB **7314** (2012).

[33] G. Leinweber, D. P. Barry, M. J. Trbovich, J. A. Burke, N. J. Drindak, H. D. Knox, R. V. Ballad, R. C. Block, Y. Danon, and L. I. Severnyak, Nuclear Science and Engineering **154**, 261 (2006).

[34] Y. Nakajima, Daya Bay DocDB **9418** (2013).

[35] J. H. Cheng, Z. Wang, L. Lebanowski, G. L. Lin, and S. Chen, Nuclear Instruments and Methods A **827**, 165 (2016).

[36] Y. Malyshkin, Daya Bay DocDB **10578** (2015).

[37] Y. Malyshkin, Daya Bay DocDB **10542** (2015).

[38] J. B. Birks, Proceedings of the Physical Society. Section A **64**, 874 (1951).

[39] P. A. Cerenkov, Dokl. Akad. Nauk. SSSR **2**, 451 (1934).

[40] H. R. Band *et al.*, JINST **7**, P06004 (2012).

[41] G. Barrand *et al.*, Computer Physics Communications **140**, CHEP2000, 45 (2001).

[42] Altair Engineering, Inc., `https://www.altair.com/grid-engine/`.

[43] A. B. Yoo, M. A. Jette, and M. Grondona, in Job Scheduling Strategies for Parallel Processing (2003), pp. 44–60.

[44] VMWare, inc., `https://www.rabbitmq.com/`.

[45] L. Gerhardt, W. Bhimji, S. Canon, M. Fasel, D. Jacobsen, M. Mustafa, J. Porter, and V. Tsulaia, Journal of Physics: Conference Series **898**, 082021 (2017).

[46] W. L. Zhong, Daya Bay DocDB **7738** (2012).

[47] S. Jetter, D. Dwyer, W.-Q. Jiang, D.-W. Liu, Y.-F. Wang, Z.-M. Wang, and L.-J. Wen, Chinese Physics C **36**, 733 (2012).

[48] L. J. Wen, J. Cao, S. Jetter, and L. Zhang, Daya Bay DocDB **7143** (2011).

[49] M. Chadwick, M. Herman, P. Obložinský, *et al.*, Nuclear Data Sheets **112**, Special Issue on ENDF/B-VII.1 Library, 2887 (2011).

[50] KamLAND Collaboration, Physical Review C **81**, 025807 (2010).

[51] J. P. Ochoa, Daya Bay DocDB **8860** (2013).

[52] M. Kramer, Daya Bay DocDB **9823** (2014).

[53] H. Wei and W. Tang, Daya Bay DocDB **11920** (2016).

[54] B. Ma and Q. Wu, Daya Bay DocDB **12296** (2020).

[55] J. P. Ochoa, Daya Bay DocDB **8772** (2013).

[56] Y. Nakajima and J. P. Ochoa, Daya Bay DocDB **8769** (2013).

[57] J. J. Ling, B. Littlejohn, Y. Meng, Y. Nakajima, J. P. Ochoa, X. Qian, K. V. Tsang, E. Worcester, and C. Zhang, Daya Bay DocDB **8810** (2013).

[58] B. Z. Hu, X. P. Ji, and K. Treskov, Daya Bay DocDB **10948** (2016).

[59] W. Gu, G. Cao, X. Chen, X. Ji, G. Li, J. Ling, J. Liu, X. Qian, and W. Wang, Nuclear Instruments and Methods A **833**, 27 (2016).

[60] L. H. Wei, Daya Bay DocDB **11632** (2018).

[61] J. Zhao, Z. Y. Yu, J. L. Liu, X. B. Li, F. H. Zhang, and D. M. Xia, Chinese Physics C **38**, 116201 (2014).

[62] Y. Nakajima, J. P. Ochoa, and P. Tsang, Daya Bay DocDB **8774** (2013).

[63] P. Huber, Physical Review C **84**, 024617 (2011).

[64] T. A. Mueller *et al.*, Physical Review C **83**, 054615 (2011).

[65] C. Lewis, Daya Bay DocDB **8609** (2013).

[66] F. James and M. Roos, Computer Physics Communications **10**, 343 (1975).

[67] J. R. Hu, Daya Bay DocDB **12099** (2020).

[68] Daya Bay Collaboration, Physical Review Letters **121**, 241805 (2018).

[69] Z. Y. Yu, Daya Bay DocDB **9667** (2014).

[70] S. Agostinelli *et al.*, Nuclear Instruments and Methods A **506**, 250 (2003).

[71] J. F. Ziegler, M. Ziegler, and J. Biersack, Nuclear Instruments and Methods B **268**, 1818 (2010).

[72] K. Shibata *et al.*, Journal of Nuclear Science and Technology **48**, 1 (2011).

[73] N. Otuka *et al.*, Nuclear Data Sheets **120**, 272 (2014).

[74] K. V. Tsang and K. Wong, Daya Bay DocDB **9818** (2014).

[75] J. Tuli, Nuclear Instruments and Methods A **369**, 506 (1996).

[76] D. A. Dwyer and T. J. Langford, Physical Review Letters **114**, 012502 (2015).

[77] P. Vogel, G. K. Schenter, F. M. Mann, and R. E. Schenter, Physical Review C **24**, 1543 (1981).

[78] A. C. Hayes, J. L. Friar, G. T. Garvey, G. Jungman, and G. Jonkmans, Physical Review Letters **112**, 202501 (2014).

[79] K. Schreckenbach, H. Faust, F. von Feilitzsch, A. Hahn, K. Hawerkamp, and J. Vuilleumier, Physics Letters B **99**, 251 (1981).

[80] F. von Feilitzsch, A. Hahn, and K. Schreckenbach, Physics Letters B **118**, 162 (1982).

[81] K. Schreckenbach, G. Colvin, W. Gelletly, and F. von Feilitzsch, Physics Letters B **160**, 325 (1985).

[82] A. Hahn, K. Schreckenbach, W. Gelletly, F. von Feilitzsch, G. Colvin, and B. Krusche, Physics Letters B **218**, 365 (1989).

[83] N. Haag, A. Gütlein, M. Hofmann, L. Oberauer, W. Potzel, K. Schreckenbach, and F. M. Wagner, Physical Review Letters **112**, 122501 (2014).

[84] X. B. Ma, Daya Bay DocDB **11687** (2018).

[85] V. I. Kopeikin, L. A. Mikaelyan, and V. V. Sinev, Physics of Atomic Nuclei **67**, 1892 (2004).

[86] T. A. Mueller *et al.*, Physical Review C **83**, 054615 (2011).

[87] P. Vogel and J. F. Beacom, Physical Review D **60**, 053003 (1999).

[88] Particle Data Group, J. Phys. G **37**, 075021 (2010), 2011 partial update for the 2012 edition.

[89] J. C. Hardy and I. S. Towner, AIP Conference Proceedings **481**, 129 (1999).

[90] C. Lewis, B. Littlejohn, M. McFarlane, and D. Webber, Daya Bay DocDB **8608** (2013).

[91] Y. Huang, Daya Bay DocDB **11611** (2018).

[92] S. Jetter, Daya Bay DocDB **9826** (2014).

[93] Y. Nakajima, J. P. Ochoa, P. Tsang, and H. L. H. Wong, Daya Bay DocDB **9999** (2014).

[94] W. Gu, G. Cao, X. Chen, X. Ji, G. Li, J. Ling, J. Liu, X. Qian, and W. Wang, Nuclear Instruments and Methods A **833**, 27 (2016).

[95] J. P. Ochoa, C. J. Lin, Y. Nakajima, K. V. Tsang, H. L. H. Wong, B. Roskovec, and M. Kramer, https://github.com/mjkramer/DybBerkFit.

[96] Daya Bay Collaboration, Physical Review Letters **108**, 171803 (2012).

[97] L. Lebanowski, Daya Bay DocDB **10838** (2016).

[98] Y. Nakajima, Daya Bay DocDB **10956** (2016).

[99] Y. Huang *et al.*, Nuclear Instruments and Methods A **895**, 48 (2018).

[100] C. J. Lin, Y. Nakajima, J. P. Ochoa, and P. Tsang, Daya Bay DocDB **7334** (2012).

[101] Y. Nakajima, Daya Bay DocDB **7536** (2012).

[102] R. Davis, Progress in Particle and Nuclear Physics **32**, 13 (1994).

[103] Z. Maki, M. Nakagawa, and S. Sakata, Progress of Theoretical Physics **28**, 870 (1962).

[104] B. Pontecorvo, Journal of Experimental and Theoretical Physics **6**, 429 (1957).

[105] K. Lande, L. M. Lederman, and W. Chinowsky, Physical Review **105**, 1925 (1957).

[106] K. Hirata *et al.*, Physics Letters B **205**, 416 (1988).

[107] D. Casper *et al.*, Physical Review Letters **66**, 2561 (1991).

[108] Y. Suzuki, Nuclear Physics B - Proceedings Supplements **38**, Neutrino 94, 54 (1995).

[109] V. Gavrin *et al.*, Nuclear Physics B - Proceedings Supplements **91**, Neutrino 2000, 36 (2001).

[110] D. Vignaud, Nuclear Physics B - Proceedings Supplements **60**, 20 (1998).

[111] Super-Kamiokande Collaboration, Physical Review Letters **81**, 1562 (1998).

[112] SNO Collaboration, Physical Review Letters **89**, 011301 (2002).

[113] KamLAND Collaboration, Physical Review Letters **90**, 021802 (2003).

[114] K2K Collaboration, Physical Review D **74**, 072003 (2006).

[115] T2K Collaboration, Nuclear Instruments and Methods A **659**, 106 (2011).

[116] MINOS Collaboration, Physical Review Letters **101**, 131802 (2008).

[117] NOνA Collaboration, Physical Review Letters **123**, 151803 (2019).

[118] OPERA Collaboration, New Journal of Physics **14**, 033017 (2012).

[119] ICARUS Collaboration, Journal of Instrumentation **6**, P07011 (2011).

[120] L. Mikaelyan and V. Sinev, Phys. Atom. Nucl. **62**, 2008 (1999), arXiv:hep-ph/9811228.

[121] Double Chooz Collaboration, Preprint (2006), arXiv:hep-ex/0606025.

[122] RENO Collaboration, Preprint (2010), arXiv:1003.1391 [hep-ex].

[123] Daya Bay Collaboration, Preprint (2007), arXiv:hep-ex/0701029.

[124] H. Kwon, F. Boehm, A. A. Hahn, H. E. Henrikson, J. L. Vuilleumier, J. F. Cavaignac, D. H. Koang, B. Vignon, F. von Feilitzsch, and R. L. Mössbauer, Physical Review D **24**, 1097 (1981).

[125] G. Zacek *et al.*, Physical Review D **34**, 2621 (1986).

[126] A. I. Afonin, A. A. Borovoi, Y. L. Dobrynin, S. N. Ketov, V. I. Kopeikin, L. A. Mikaelyan, M. D. Skorokhvatov, S. V. Tolokonnikov, and A. N. Kheruvimov, Journal of Experimental and Theoretical Physics Letters **41**, 435 (1985).

[127] V. I. Aleshin *et al.*, Instruments and Experimental Techniques **51**, 499 (2008).

[128] M. Abbes *et al.*, Nuclear Instruments and Methods A **374**, 164 (1996).

[129] N. Bauman, H. Gurr, W. R. Kropp, M. Mandelkern, E. Pasierb, L. Price, F. Reines, and H. W. Sobel, Progress in Physics **6**, 30 (1982).

[130] G. Mention, M. Fechner, T. Lasserre, T. A. Mueller, D. Lhuillier, M. Cribier, and A. Letourneau, Physical Review D **83**, 073006 (2011).

[131] Daya Bay and MINOS Collaborations, Physical Review Letters **117**, 151801 (2016).

[132] Chooz Collaboration, European Physical Journal C **27**, 331 (2003).

[133] Palo Verde Collaboration, Physical Review D **64**, 112001 (2001).

[134] RENO Collaboration, Physical Review Letters **108**, 191802 (2012).

[135] Daya Bay Collaboration, Physical Review D **100**, 052004 (2019).

[136] RENO Collaboration, Preprint (2020), arXiv:2010.14989 [hep-ex].

[137] Double Chooz Collaboration, Nature Physics **16**, 558 (2020).

[138] Daya Bay Collaboration, Physical Review Letters **118**, 251801 (2017).

[139] Daya Bay Collaboration, Physical Review Letters **123**, 111801 (2019).

[140] F. Kaether, W. Hampel, G. Heusser, J. Kiko, and T. Kirsten, Physics Letters B **685**, 47 (2010).

[141] SAGE Collaboration, Physical Review C **80**, 015807 (2009).

[142] LSND Collaboration, Physical Review D **64**, 112007 (2001).

[143] MiniBooNE Collaboration, Physical Review Letters **121**, 221801 (2018).

[144] NEOS Collaboration, Journal of Physics: Conference Series **1216**, 012004 (2019).

[145] DANSS Collaboration, Journal of Instrumentation **11**, P11011 (2016).

[146] STEREO Collaboration, Journal of Instrumentation **13**, P07009 (2018).

[147] PROSPECT Collaboration, Nuclear Instruments and Methods A **922**, 287 (2019).

[148] A. P. Serebrov *et al.*, (2020), arXiv:2005.05301 [hep-ex].

[149] SoLid Collaboration, Journal of Instrumentation **16**, P02025 (2021).

[150] NEOS Collaboration, Physical Review Letters **118**, 121802 (2017).

[151] D. Svirida *et al.*, Proceedings of Science **NOW2018**, 066 (2019).

[152] STEREO Collaboration, Physical Review D **102**, 052002 (2020).

[153] PROSPECT Collaboration, Physical Review D **103**, 032001 (2021).

[154] A. P. Serebrov and R. M. Samoilov, Journal of Experimental and Theoretical Physics Letters **112**, 199 (2020).

[155] Y. F. Li, International Journal of Modern Physics: Conference Series **31**, 1460300 (2014).

[156] M. Ghosh, S. Gupta, Z. M. Matthews, P. Sharma, and A. Williams, Proceedings of Science **NuFact2017**, 133 (2018), arXiv:1712.06714 [hep-ph].

[157] M. Shaposhnikov, Progress of Theoretical Physics **122**, 185 (2009).

[158] A. Giuliani, J. J. G. Cadenas, S. Pascoli, E. Previtali, R. Saakyan, K. Schaeffner, and S. Schoenert, Preprint (2020), arXiv:1910.04688 [hep-ex].

[159] T. Li, X. Xia, X. T. Huang, J. H. Zou, W. D. Li, T. Lin, K. Zhang, and Z. Y. Deng, Chinese Physics C **41**, 066201 (2017).

[160] P. Fernandez, PhD thesis (Autonomous University of Madrid, Feb. 2017).

[161] L. Wen, J. Cao, K.-B. Luk, Y. Ma, Y. Wang, and C. Yang, Nuclear Instruments and Methods A **564**, 471 (2006).

[162] O. Tange, *GNU Parallel* (Mar. 2018).

[163] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, Physical Review Letters **13**, 138 (1964).

[164] Palo Verde Collaboration, Physical Review D **62**, 072002 (2000).

[165] KamLAND Collaboration, Nature Geoscience **4**, 647 (2011).

[166] O. Dalager, Poster presented at the Neutrino 2020 conference, Fermilab.

[167] Z. Yu, L. Zhan, L. Wen, and Z. Wang, Daya Bay DocDB **7418** (2012).

[168] X. Qian, W. Wang, and C. Zhang, Daya Bay DocDB **7537** (2012).

[169] X. Qian, Daya Bay DocDB **7946** (2012).

[170] O. Dalager, Daya Bay DocDB **12340** (2021).

[171] X. B. Ma, W. L. Zhong, L. Z. Wang, Y. X. Chen, and J. Cao, Physical Review C **88**, 014605 (2013).

[172] P. Huber and T. Schwetz, Physical Review D **70**, 053011 (2004).

[173] P. Huber and P. Jaffke, Physical Review Letters **116**, 122503 (2016).

Appendix A

Details of minor backgrounds

For the interested reader, we provide here a review of the studies undertaken within the Daya Bay Collaboration to determine the rate and spectra of three minor correlated backgrounds: fast neutrons, AmC backgrounds, and $^{13}\mathrm{C}(\alpha,\mathrm{n})^{16}\mathrm{O}$ reactions. As discussed in Sections A.2–A.4, we use the results of these studies in order to these backgrounds from our sample of IBD candidates. This appendix also contains a discussion of uncorrelated instrumental ("flasher") backgrounds, which are removed during the IBD pre-selection (Section 5.2.1). We begin with the flashers.

A.1 PMT light emission ("flashers")

During detector commissioning, most PMTs were found to occasionally emit varying amount of light due to arcing in their bases. The rate and intensity of this "flashing" would change over time for each PMT. At any given moment, some 5% of the PMTs in each AD will have the tendency to flash brightly enough to trigger the detector [17], in some cases producing as much as 100 MeV of reconstructed energy. Within the delayed energy region of 6-12 MeV, the flasher rate has averaged at around 0.7 Hz for each AD. These "delayed-like" flashers, if included in the analysis, would significantly increase the rate of backgrounds caused by the accidental coincidence of two uncorrelated signals. As discussed in Section 6.2 and Chapter 7, the rate of such "accidentals" is proportional to the rate of delayed-like signals, and this rate (excluding flashers) ranges from around 0.05 Hz at EH3 to 1 Hz at EH1. While the flashers would merely (roughly) double the 1% accidental background in the near halls, in the far hall it would increase this background by an order of magnitude to the 10% level, counter to Daya Bay's goal of perecent-level background contamination.

Fortunately, flashers are easily distinguished from "physical" singles due to their unique pattern of light emission, enabling them to be removed from the analysis with high efficiency while minimally affecting true IBDs. This light pattern is characterized by two "hot spots" on opposite sides of the AD. When a PMT base emits light, much

of the light is absorbed by the black radial shield and the conical magnetic shield. The remainder escapes within a conical profile; some of the photons will strike the flasher's photocathode (resulting in the flashing PMT having the highest charge), and others will primarily illuminate the PMTs across the AD from the flasher, especially the one that lies directly opposite to it. In addition, the time distribution of the PMT hits is broadened for flashers due to the geometry of light propagation across the AD. By taking advantage of these telltale distributions of charges and times, it is possible to achieve excellent discrimination of flashers from physical events.

The flasher identification criteria were developed in a somewhat ad-hoc fashion, by defining quantities that could conceivably serve as discriminators, and then further defining combinations of these quantities, and finally plotting the distributions of these (combined) quantities until a clean separation between flashers and non-flashers was apparent. It is far more important to minimize (IBD) signal inefficiency instead of maximizing flasher rejection, because a small amount of unrejected flashers will simply slightly increase the rate of accidental backgrounds (which can be easily quantified), whereas a signal inefficiency could vary among the ADs and thereby bias the oscillation fit.

Early in the experiment, this prolonged and iterative process eventually gave rise to the *ellipse cut* (based on the charge distribution) and the *PSD cut* (based on the time distribution), which demonstrated excellent performance, and these cuts continue to be used in this analysis. The ellipse cut is based on two quantities, termed f_{max} and f_{quad}. The first, f_{max}, is simply the ratio of Q_{max} (the maximum individual PMT charge across all PMTs) over the total charge Q_{tot}:

$$f_{\text{max}} = \frac{Q_{\text{max}}}{Q_{\text{tot}}}. \tag{A.1}$$

For flasher events, Q_{max} belongs to the flashing PMT itself, and f_{max} is typically higher for flashers than for physics events. However, physical events near the PMTs can exhibit high f_{max}, so this variable alone is insufficient to cleanly discriminate flashers. As such, we also consider f_{quad}, which is based on dividing the AD into four quadrants (Figure A.1): "Quadrant 1" (q1) is the one that is centered on the highest-charge PMT, q3 is the one across from q1, and q2 and q4 are the two "to the side." Then, f_{quad} captures the conical nature of the light emission:

$$f_{\text{quad}} = \frac{Q_{\text{q3}}}{Q_{\text{q2}} + Q_{\text{q4}}}. \tag{A.2}$$

Like f_{max}, f_{quad} alone is not a good discriminator, due to overlap between flashers and physics events. However, their combination

$$f_{\text{ID}} = \log_{10}\left[f_{\text{quad}}^2 + \left(\frac{f_{\text{max}}}{0.45}\right)^2\right] \tag{A.3}$$

turns out to be an excellent discriminator. Indeed, as shown by Figures A.2 and A.3, requiring

$$f_{\mathrm{ID}} < 0 \tag{A.4}$$

reduces the flasher rate to a negligible level, and many analyses have relied on f_{ID} alone to identify flashing 8" PMTs.

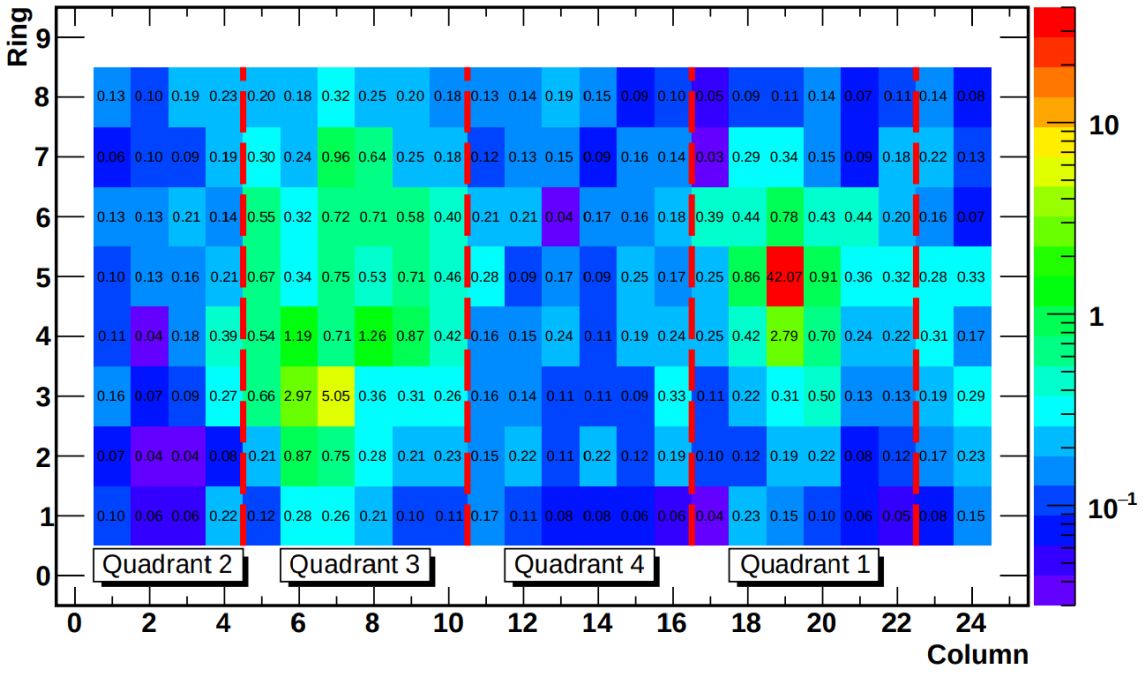

Figure A.1: PMT charge distribution of a flasher candidate, illustrating the division of the AD into four azimuthal quadrants, with Quadrant 1 being centered around the PMT with the highest charge. From [19].

Even further flasher reduction can be achieved by incorporating timing information. To capture the broadening of the time distribution shown by flashers, we use the variable(s) f_{t1} (f_{t2}), defined as the ratio of the number of hits in the first 200 (150) ns of the signal, over the number of hits in the first 400 ns. The discriminator f_{PSD} is then defined as

$$f_{\mathrm{PSD}} = \log_{10}[4 \cdot (1 - f_{\mathrm{t1}})^2 + 1.8 \cdot (1 - f_{\mathrm{t2}})^2]. \tag{A.5}$$

By requiring both $f_{\mathrm{ID}} < 0$ and $f_{\mathrm{PSD}} < 0$, we eliminate virtually all 8" PMT flashers from the analysis.

However, in addition to the 192 8" PMTs, there are six 2" PMTs located at the top and bottom of each AD along the calibration axes, and these can also flash. Such events were easily identified as those in which any 2" PMT saw an extreme amount of charge, with a cut of 100 PE providing essentially perfect separation between 2" PMT flashers and other events.

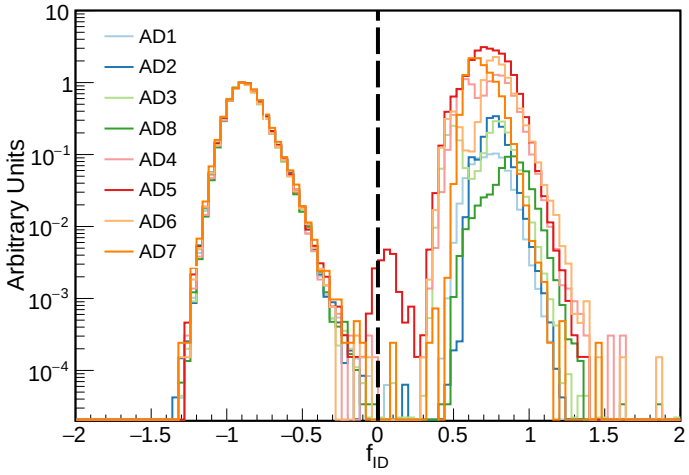

Figure A.2: Distribution of the flasher discriminator f_{ID} for the delayed triggers of IBD candidates. The tagged flashers ($f_{\mathrm{ID}} > 0$) are cleanly separated from physics signals, with the exception of a single flasher in AD5 which leaked into the signal region (leading to a slight increase in the singles rate). In all ADs, there is essentially no signal lost. From [19].

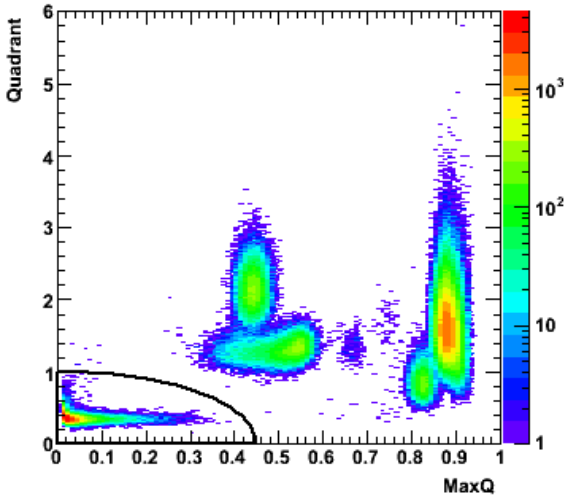

Figure A.3: Two-dimensional distribution of f_{max} ("MaxQ") and f_{quad} ("Quadrant") for physical triggers in EH1-AD2. The black "ellipse" corresponds to $f_{\mathrm{ID}} = 0$; events outside the ellipse are identified as flashers. High flasher rejection and negligible signal inefficiency are apparent. From [48].

A.2 Cosmogenic fast neutrons

A.2.1 Event selection

Two event samples are used in the fast-neutron analysis. The first, which we shall refer to as the "untagged" sample, is obtained using the standard IBD selection, as described in Chapter 5, with the modification that the upper limit on the prompt energy is extended to 300 MeV instead of the usual 12 MeV. In this sample, the prompt spectrum below 12 MeV is essentially the same as the one used in the oscillation fit (i.e., dominated by true IBDs), whereas the high-energy region almost exclusively contains the (recoil protons from) fast-neutron events.

The other, "tagged," sample contains IBD-like events that occur right after a muon that triggers *only* the outer water pool. When a muon passes through the AD or the IWS, the resulting spallation products generate a significant amount of prompt activity in the AD. On the other hand, when a muon triggers only the OWS, without involvement of the IWS or AD, most of the debris is unable to penetrate into the GdLS. Fast neutrons, however, are an exception. The OWS tagging therefore provides a highly pure sample of fast neutrons for analysis.

The tagged sample is obtained by extending the upper cut to 300 MeV (as in the untagged sample) while disabling the standard muon veto. An additional requirement (see Figure A.4) is that the prompt signal be timestamped within (-300, 600) ns of an OWS trigger (defined by NHit $>$ 15, in this case). Furthermore, the delayed signal must occur at least 15 μs after the muon in order to eliminate Michel electrons, and there must be no AD or IWS muons within 600 µs of the OWS muon. This selection results in fairly low statistics (amounting to a few hundred events in the near halls), but the size of the sample is still sufficient to provide strong constraints on the fast-neutron background. This sample is statistically independent from the untagged sample, as all of its events would have been excluded by the (OWS) muon veto in the untagged selection.

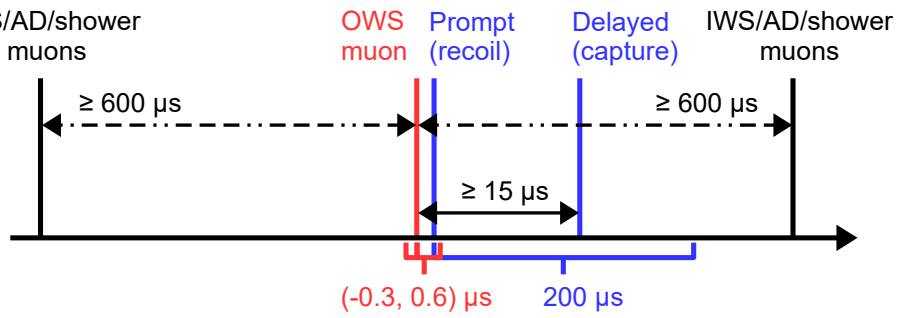

Figure A.4: Diagram illustrating the event selection scheme for the tagged fast-neutron sample.

The prompt spectra from the two samples (scaled to have equal normalization above 12 MeV) are overlaid in Figure A.5, showing excellent agreement in the region between 12 and 150 MeV. Both samples are employed by the scaling method, as described next. Meanwhile, the extrapolation method only makes direct use of the tagged sample; however, the untagged sample still serves a role in validating the form of the fitting function used.

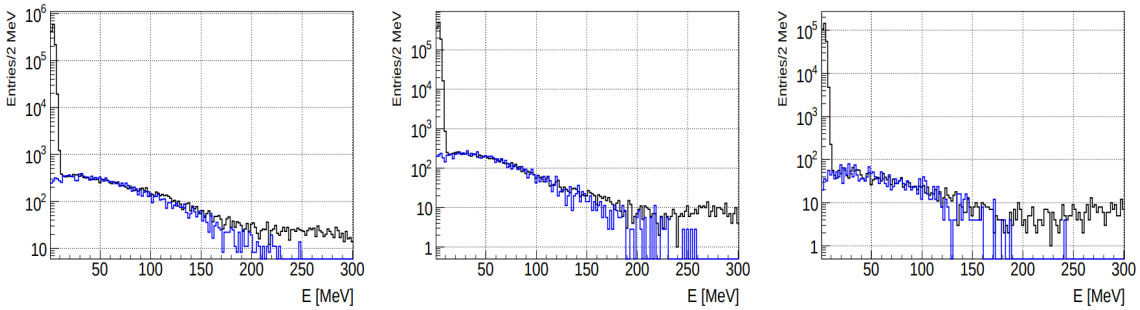

Figure A.5: Prompt spectra of the tagged (blue) and untagged (black) samples. The tagged spectrum has been rescaled to match the normalization of the untagged spectrum above 12 MeV. From [58].

A.2.2 Scaling method

In the scaling method, we assume that the shape of the tagged spectrum is an accurate representation of the shape of the fast-neutron background. In other words, we assume that, for any given fast neutron, there is no energy dependence on the probability of its association with an OWS-only muon. Previous studies within the collaboration have supported the validity of this assumption.

We define the *scaling factor* F as

$$F(E_{\max}) = \frac{N_{\text{untag}}[12, E_{\max}]}{N_{\text{tag}}[12, E_{\max}]}, \tag{A.6}$$

i.e., the ratio of the integral of the two samples between 12 MeV and $E_{\max}$. The dependence on $E_{\max}$ reflects the arbitrary choice of the upper energy limit, which contributes to the uncertainty on the result (primarily due to statistical fluctuations in the small sample of tagged neutrons). This uncertainty is quantified by comparing the results for $E_{\max}$ of 80, 100, 120, and 150 MeV (Figure A.6) [58] , resulting in an estimated systematic uncertainty given by the half-range divided by the mean, as listed in Table A.2. In addition, for a *fixed* $E_{\max}$, there is a purely statistical

uncertainty[1],

$$\sigma_F(E_{\max}) = F \cdot \sqrt{N_{\text{untag}}^{-1}[12, E_{\max}] + N_{\text{tag}}^{-1}[12, E_{\max}]}, \tag{A.7}$$

which also contributes to the final uncertainty.

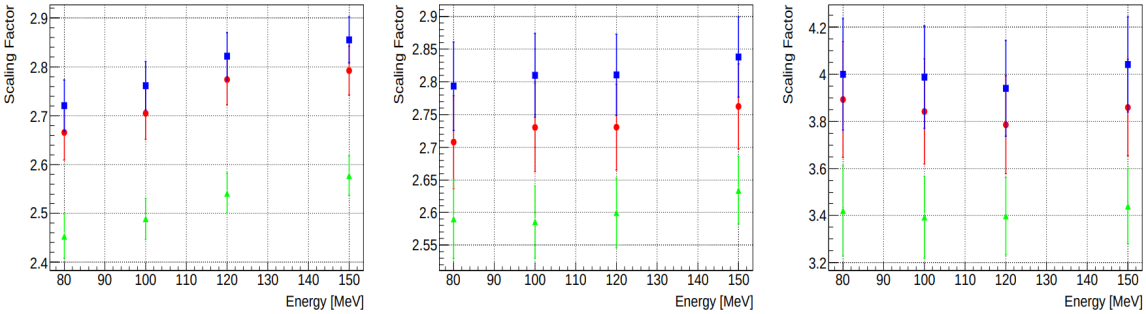

Figure A.6: Comparison of the scaling factor $F(E_{\max})$ obtained for different values of $E_{\max}$. The three colors correspond to different IBD selection criteria, with the blue squares representing the criteria used in this analysis. From [58].

For the standard IBD selection, the number of fast neutrons within the fiducial energy range of [0.7, 12] MeV is determined simply as

$$N_{\text{FN}} = F \cdot N_{\text{tag}}[0.7, 12]. \tag{A.8}$$

Its statistical uncertainty, in turn, is

$$\sigma_{\text{FN}} = \sqrt{N_{\text{tag}}^2[0.7, 12] \cdot \sigma_F^2 + F^2 \cdot N_{\text{tag}}[0.7, 12]}. \tag{A.9}$$

Finally, the normalized daily fast-neutron rate is

$$R_{\text{FN}} = \frac{N_{\text{FN}}}{T_{\text{DAQ}} \cdot \epsilon_\mu \cdot \epsilon_m}, \tag{A.10}$$

where T_{DAQ}, ϵ_μ, and ϵ_m are the DAQ livetime, muon veto efficiency, and multiplicity cut efficiency for the untagged sample. The resulting R_{FN}, for the four different values of $E_{\max}$, are plotted in Figure A.7.

A.2.3 Extrapolation method

Previous simulation studies have found that the fast neutron spectrum can be accurately described by the PDF

$$f(E; E_0, a) = A \cdot \left(\frac{E}{E_0}\right)^{-a - E/E_0}, \tag{A.11}$$

[1] As noted previously, the tagged and untagged samples are statistically independent.

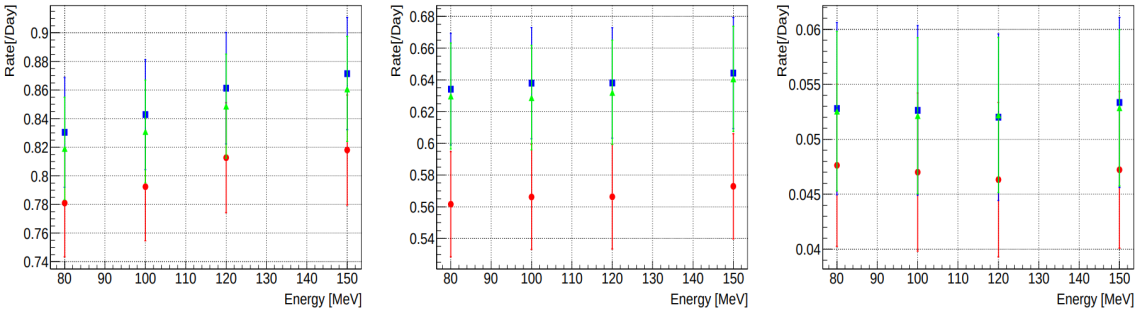

Figure A.7: Daily fast-neutron rate from the scaling method, as a function of $E_{\max}$. The blue squares correspond to the IBD selection criteria used in this analysis. From [58].

where E_0 and a are shape parameters, and $A(E_0, a)$ normalizes the PDF. When a measured fast-neutron spectrum is then fit to the form

$$S(E; N, E_0, a) = N \cdot f(E; E_0, a), \tag{A.12}$$

the best-fit value of N represents the number of events "under the curve" from 0 to ∞ MeV.

In turn, if we have a (hypothetical) pure fast-neutron spectrum containing N_{FN} events in the fiducial region of [0.7, 12] MeV, then the full spectrum ([0, ∞] MeV) should contain an event count equal to

$$N = \frac{N_{\mathrm{FN}}}{\int_{0.7}^{12} f(E; E_0, a)\, dE}. \tag{A.13}$$

Here, the denominator represents the fiducial region's fraction of the total spectrum. Substituting this form of N into Equation A.12, and allowing the floated parameter to be N_{FN} instead of N, we obtain the form

$$S(E; N_{\mathrm{FN}}, E_0, a) = \frac{N_{\mathrm{FN}}}{\int_{0.7}^{12} E'^{-a-E'/E_0}\, dE'} E^{-a-E/E_0}. \tag{A.14}$$

When Equation A.14 is then fit to the measured fast-neutron spectrum, the best-fit N_{FN} indicates the number of events in the fiducial region. The key to the extrapolation method is that this fit is performed outside the fiducial region, where the fast-neutron sample is uncontaminated.

The validity and robustness of Equation A.14 was verified by fitting it to the OWS-tagged samples from each hall (Figure A.8) [58]. Four fitting ranges were used, all starting at 0.7 MeV, and ending at 80, 100, 120, and 150 MeV. All fits produced satisfactory goodness-of-fit and consistent values of E_0 (Figure A.9). Disabling of

the a parameter was found to introduce negligible differences. As summarized in Table A.1, we define the nominal value of E_0 to be the average from the eight fits in each hall (four fitting ranges, with and without a). This value is then inserted into Equation A.11 (with $a = 0$) to generate the spectral shape of the fast-neutron background. Normalizing this shape to the predicted fast-neutron rate then gives the prompt spectrum that we subtract from that of the IBD candidates.

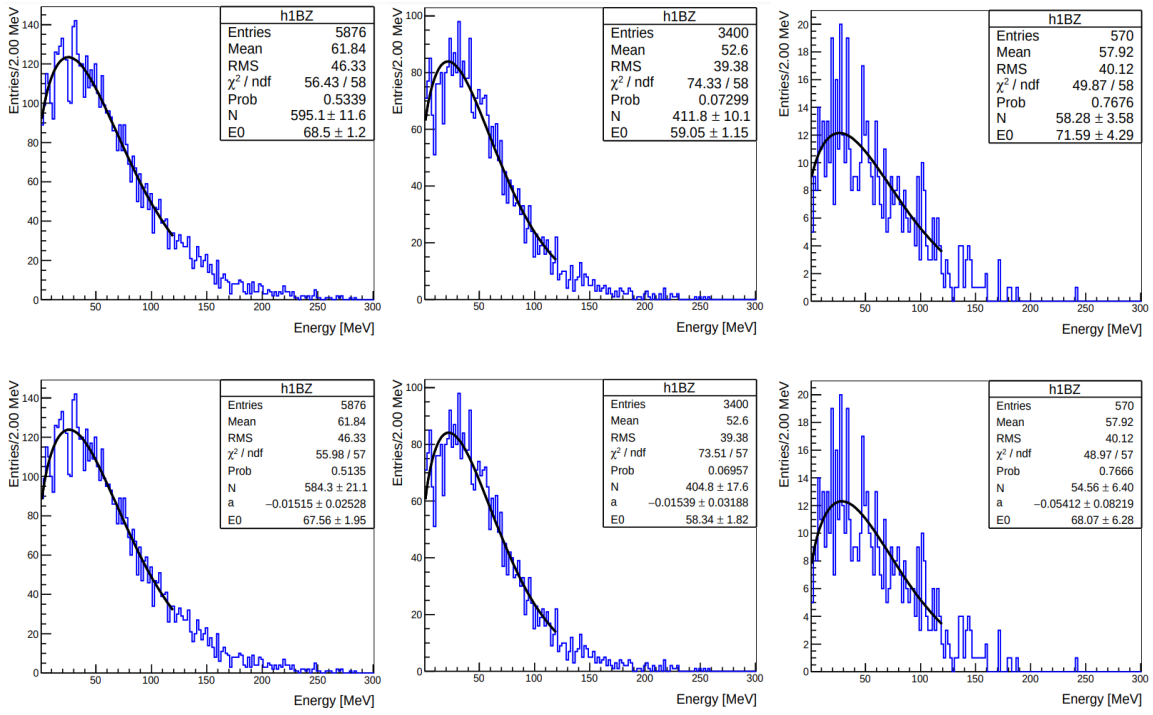

Figure A.8: Fits of the OWS-tagged fast-neutron spectra to Equation A.14 for EH1 (left), EH2 (center), and EH3 (right). From [58].

Hall	E_0 (MeV)
EH1	68.68
EH2	59.16
EH3	67.91

Table A.1: Fast-neutron shape parameters E_0 (inserted into Equation A.11) used for subtracting this background in each hall [58].

Finally, the fit was performed on the untagged sample (Figure A.10), using the same four upper limits as before, but with the lower limit set to 12 MeV. As with the scaling method, each value of N_{FN} was converted to a livetime- and efficiency-normalized daily rate according to Equation A.10 (Figure A.11). The spread between

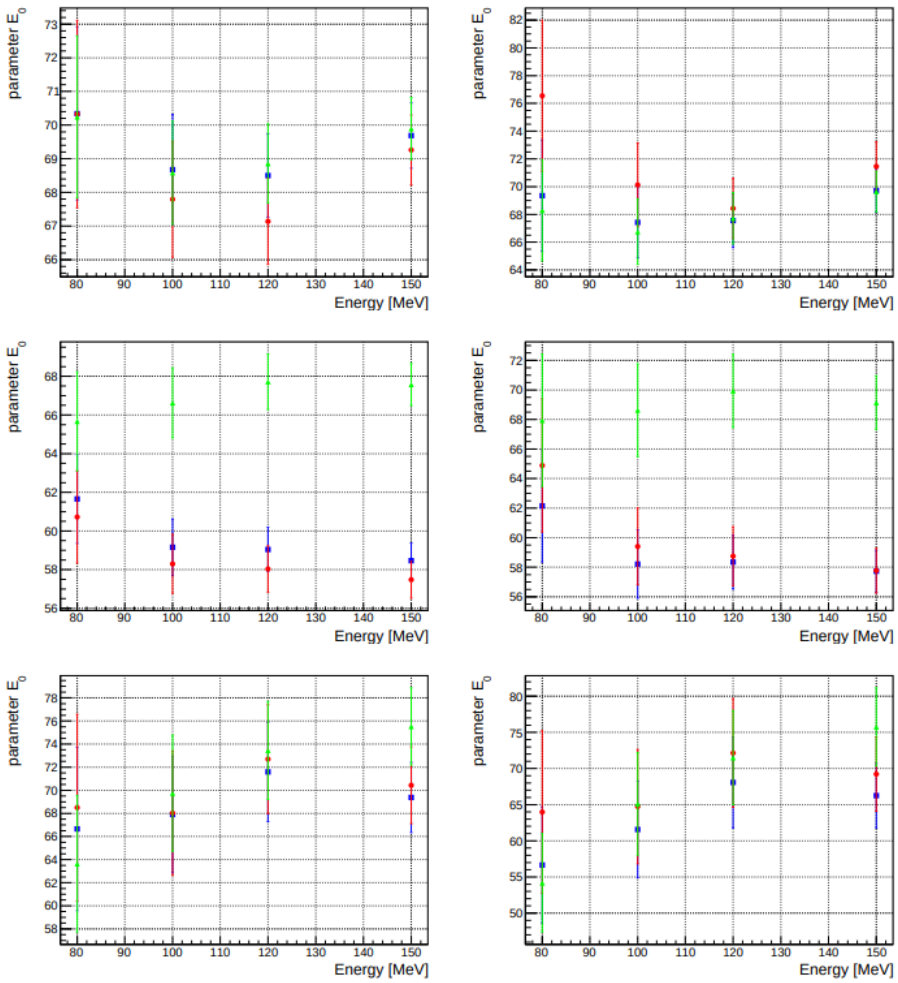

Figure A.9: Values of the shape parameter E_0 obtained from fitting the OWS-tagged fast-neutron samples to Equation A.14. The parameter a was fixed to zero for the plots on the left, while it was allowed to float for the plots on the right. The blue squares correspond to the IBD selection criteria used in this analysis; the average of the eight values for each hall, as reproduced in Table A.1, is used when subtracting the fast-neutron background in our analysis. From [58].

the resulting four values was incorporated into the total uncertainty, as described in the next section.

A.2.4 Final result and total uncertainty

In total, eight consistent estimates of R_{FN} were obtained for each hall, derived from four scaling ranges in the scaling method, and four fitting ranges in the extrapolation

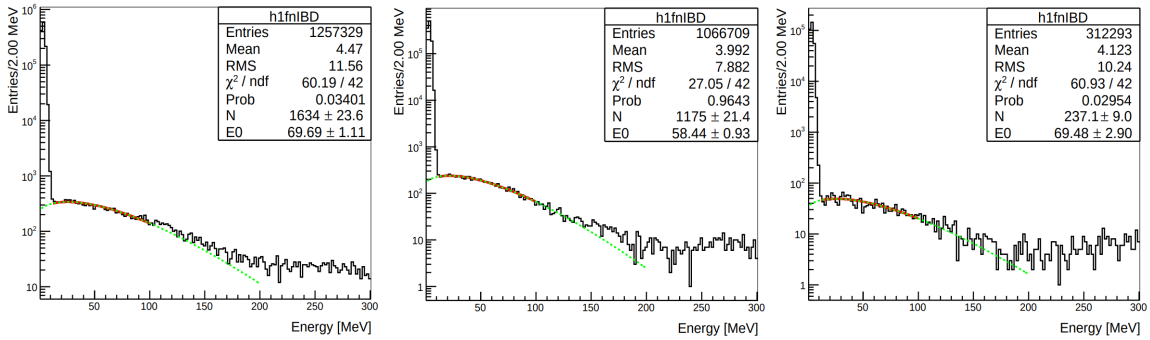

Figure A.10: Fits of the untagged fast-neutron spectra to Equation A.14 for EH1 (left), EH2 (center), and EH3 (right). From [58].

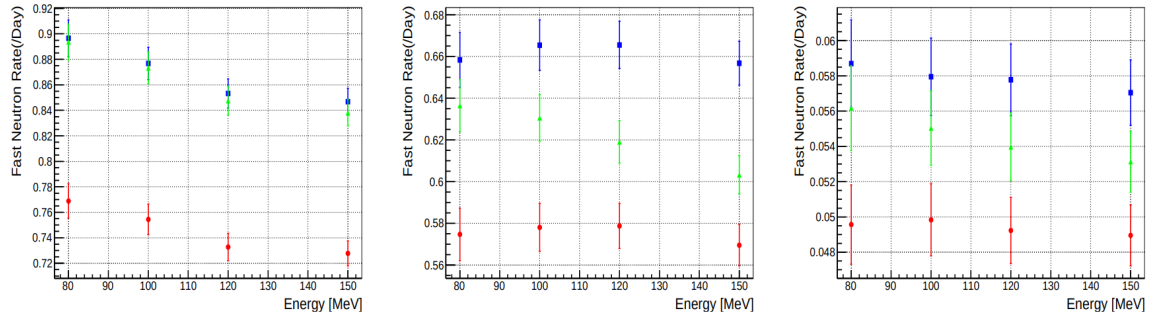

Figure A.11: Daily fast-neutron rates from the extrapolation method, as a function of the upper limit of the fit range. The blue squares correspond to the IBD selection criteria used in this analysis. From [58].

method [58]. As in the official Daya Bay analysis, we arbitrarily choose the 12-100 MeV scaling method as the source of the daily fast neutron rates in this analysis.

Six uncertainties were added in quadrature to obtain the total [58]:

- Statistical: The statistical uncertainty on the scaling factor F, described by Equation A.9.
- Scaling range: The uncertainty from the choice of scaling range, which was determined from the difference between the highest and lowest fast-neutron rates across the four ranges used.
- Fit range: The analogue for the choice of the fitting range.
- Fit result: The statistical uncertainty in the fitted value of N_{fid}.
- Bin width: From the dependence of the fit results on the choice of binning, which was determined by varying the bin widths and repeating the fits.

- Methods: From the difference in results between the scaling and extrapolation methods.

These uncertainties are summarized in Table A.2, and the final results given in Table A.3.

Uncertainty (%)	EH1	EH2	EH3
Statistical	4.6	5.5	14.7
Scaling range	2.4	0.8	1.2
Fit range	2.9	0.6	0.8
Fit result	1.5	1.8	4.0
Bin width	0.4	0.4	2.0
Methods	7.7	7.7	7.7
Total	9.8	9.7	17.2

Table A.2: Uncertainty budget for the fast-neutron rates [58].

EH1	EH2	EH3
0.843 ± 0.083	0.638 ± 0.062	0.053 ± 0.009

Table A.3: Final estimated fast-neutron rates (per AD per day) [58].

A.3 AmC source

The first experimental suggestion of this background came from an observed excess of neutron-like (i.e. 6–12 MeV) *uncorrelated* events in the top half of the detector (Figure A.12). Subsequent MC simulations showed that these uncorrelated events (from neutron capture in the stainless steel or the GdLS overflow tank) are associated with a correlated background. Further MC studies were then used to characterize the relationship between the rate of uncorrelated events and the rate of the correlated background.

A high-activity AmC source (HAS) was used to benchmark the MC. The HAS produced ∼59 neutrons/s and was enclosed in a nearly solid cylinder of stainless steel, maximizing the rate of neutron captures and inelastic scatters. It was placed on the lid of EH3-AD1 (labeled AD4) in the summer of 2012, and data was collected for ten days.

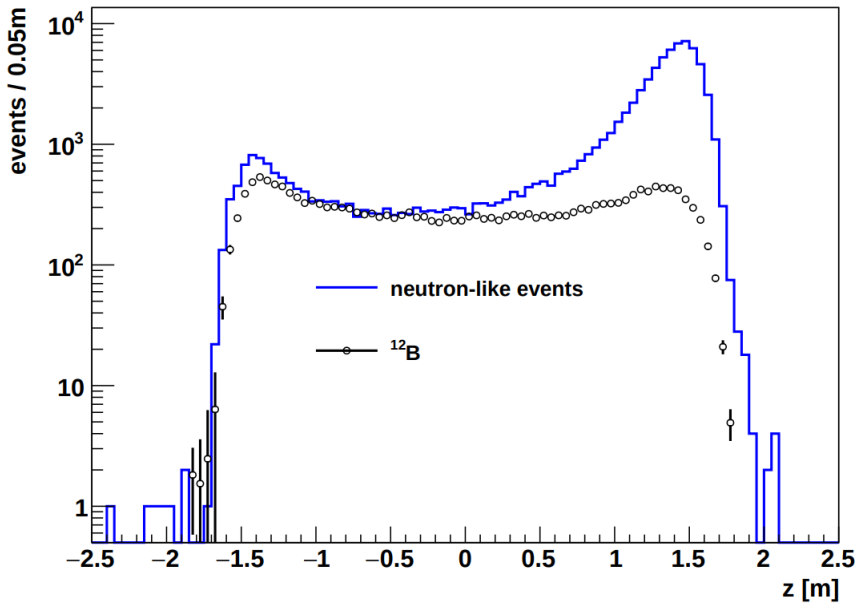

Figure A.12: Vertical distribution of uncorrelated neutron-like events, illustrating an excess in the upper half of the AD due to the AmC source in the ACUs. From [59].

The evaluation of this AmC background is detailed in [59]. Here we briefly summarize the results, which are based on the following fundamental relationship:

$$R_{\text{corr}} = R_{\text{uncorr}} \times \xi \tag{A.15}$$

Here, R_{uncorr} is the rate of uncorrelated neutron-like events produced by the AmC source (or HAS), as measured directly by taking the difference in the number of isolated neutron-like events between the top and bottom halves of each AD. Meanwhile, ξ is the ratio of correlated to uncorrelated events, as determined from a combination of simulations and HAS data. All of the complexity lies in the determination of ξ, for which two independent estimates were made, one using HAS data, and the other using MC simulations.

To measure ξ from HAS data, the number of uncorrelated HAS-induced neutron-like events was first determined by subtracting the neutron-like samples between AD4 (with the HAS) and the adjacent AD5, which observed ∼50,000 and ∼4,000 events, respectively (Figure A.13). Next, the number of *correlated* HAS events was measured by taking the spectrum of IBD candidates in AD4, subtracting the accidental background, and then subtracting the (background-subtracted) IBD sample measured by AD5 (Figure A.14).[2] The delayed spectrum from this sample, consisting of the

[2]This procedure doesn't account for other correlated backgrounds in AD4, such as ^{9}Li and fast neutrons, but their rates of < 0.2/d are insignificant compared to the 63 correlated events per day produced by the HAS.

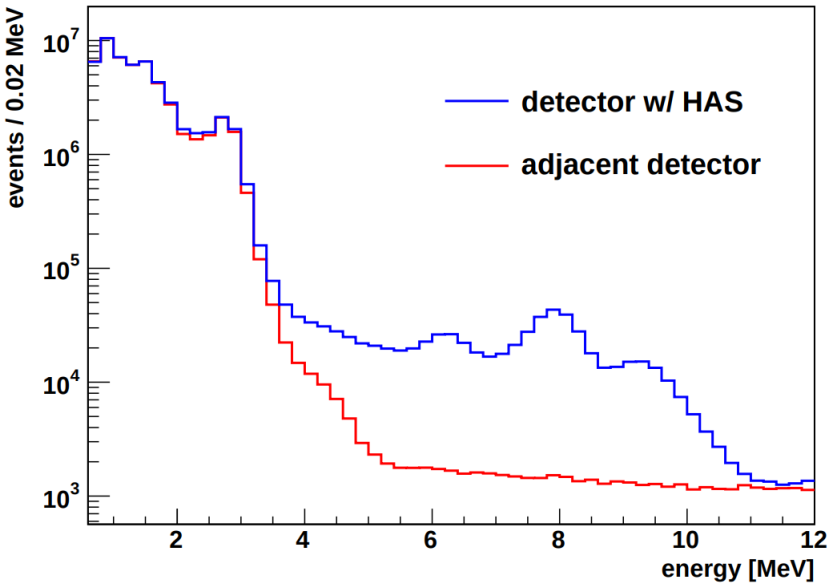

Figure A.13: Spectra of uncorrelated events (singles) in the HAS-deployed AD and an adjacent AD, showing a clear excess of neutron-like events from the HAS. From [59].

sum of neutron-capture peaks from Fe, Ni, Cr, and Mn, demonstrated excellent agreement with the prediction of the MC, providing further validation of the MC's predictions (Figure A.15). Relating the uncorrelated and correlated rates gave a value of ξ, *for the HAS*, of $(1.5 \pm 0.3) \times 10^{-3}$. The Geant4 MC, on the other hand, returned a ξ of $(1.2 \pm 0.1) \times 10^{-3}$ for the HAS. The 25% difference was then assigned as the uncertainty (and bias) of the MC. With the addition in quadrature of the 20% statistical uncertainty in the data, a total uncertainty of 30% was assigned to ξ.

Compared to the HAS, the ordinary (low-intensity) AmC source (LAS) is expected to have a lower ξ, since it lies farther from the AD and has a lower density of surrounding stainless steel. For the LAS, the MC predicted a ξ of 0.9×10^{-3}. Based on the MC/data comparison for the HAS, this value was scaled up by 25% to 1.125×10^{-3}, with an uncertainty of 30%.

In addition to the rates, the prompt spectrum of the AmC background also required determination. Excellent agreement in the HAS prompt spectra was found between the data and MC (Figure A.16). Furthermore, similar agreement was found between the MC HAS and MC LAS prompt spectra (Figure A.17), in spite of the differences in geometry and materials between the HAS and the LAS. As such, any one of these spectra could have been chosen as a reference. The choice was to use the measured HAS spectrum, which was fit to an exponential function,

$$f(E) = p_0 \times e^{-E/p_1}. \tag{A.16}$$

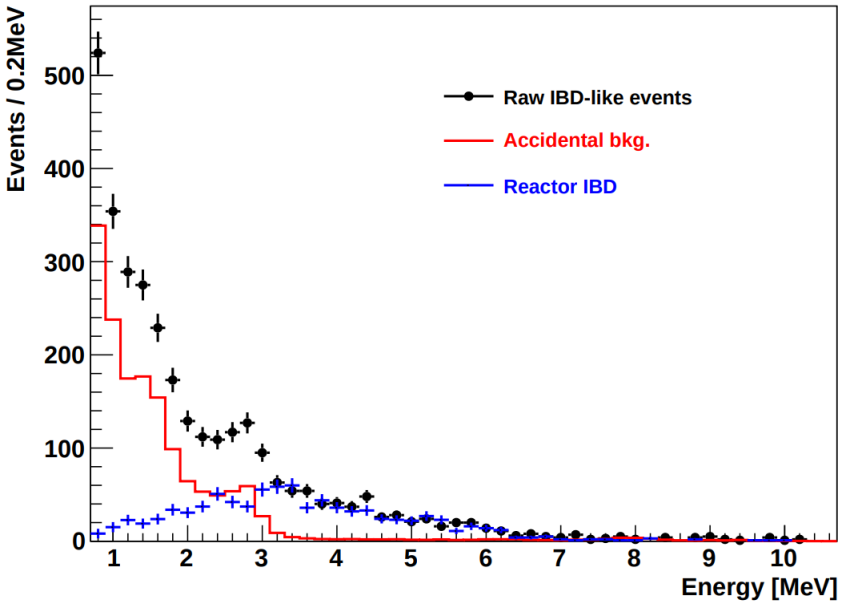

Figure A.14: Prompt spectrum of IBD-like events in the HAS-deployed detector, along with the spectrum of accidentals (determined from the singles spectrum) and the (background-subtracted) reactor IBD spectrum from an adjacent AD. From [59].

The fit gave $p_1 = 0.783\,\text{MeV}$ with a 10% statistical uncertainty, compared to 0.794 MeV and 0.830 MeV for the LAS and HAS MC samples, respectively. This 5% spread, in combination with the 10% statisical uncertainty, gave a conservative total uncertainty of 15% on p_1 (essentially a shape uncertainty). Meanwhile, p_0 was fixed by the normalization condition $\int f(E)\,dE = \xi$, giving (for $\xi = 1.125{\times}10^{-3}$) $p_0 = 3.606 \times 10^{-3}\,/\text{MeV}$. Conservatively combining the 30% uncertainty on ξ with the 15% uncertainty on p_1 gave a total uncertainty of 45% on the AmC background. Given the identical design of the ADs, identical behavior was assumed with respect to the AmC background, and no attempt was made to calculate AD-specific quantities.

After the determination of ξ, the prompt spectrum (i.e. p_1), and the uncertainty, evaluation of each AD's AmC background then amounted to the simple task of measuring R_{uncorr} (after correcting for the muon veto efficiency) and multiplying it by ξ according to Equation A.15, resulting in the final values used in this analysis. It should be noted that the ACU-B and ACU-C AmC sources were removed from the EH3 ADs in 2012, during installation of EH2-AD2 and EH3-AD4.[3] This significantly reduced the AmC background at the far site from 0.3% to 0.1% of the IBD rate. Furthermore, over the first two years of data, a 50% decline was observed in the rate

[3]In principle, the effective value of ξ could vary between the three-sources and one-source scenarios, but this subtlety is not discussed in [59]. Presumably, any such effects fall within the 45% uncertainty.

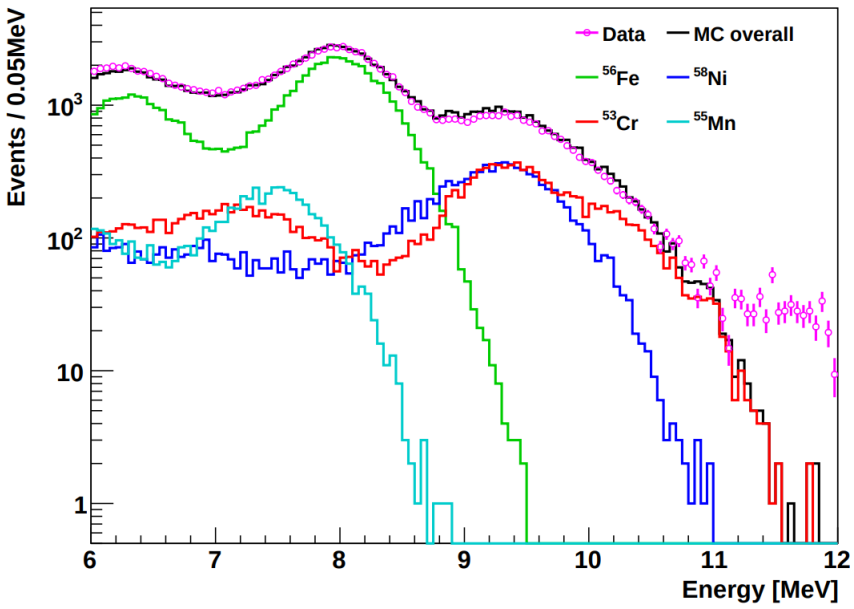

Figure A.15: Comparison of the AmC delayed spectrum between data and the MC, showing the latter's contributions from the individual neutron capture peaks. From [59].

from each AmC source, in all three halls, likely due to accumulation of scintillator into the source enclosures from the weekly calibration. This led to an ultimate background rate of only 0.03% and 0.05%[4], near and far, respectively (although the mean rate over the entire data sample is higher, due to the fact that earlier rates were higher—0.05% and 0.3%, near and far, respectively.)

For the P17B data set used in this analysis, the AmC background rates were re-estimated in [60] using updated measurements of the uncorrelated event rates from the AmC sources. The values are listed in Table A.4.

	EH1		EH2		EH3			
Period	AD1	AD2	AD1	AD2	AD1	AD2	AD3	AD4
6AD	0.29 ± 0.13	0.27 ± 0.12	0.30 ± 0.14		0.24 ± 0.11	0.23 ± 0.10	0.23 ± 0.10	
8AD	0.15 ± 0.07	0.15 ± 0.07	0.12 ± 0.06	0.14 ± 0.06	0.04 ± 0.02	0.03 ± 0.01	0.03 ± 0.02	0.04 ± 0.02
7AD		0.11 ± 0.05	0.09 ± 0.04	0.08 ± 0.04	0.02 ± 0.01	0.02 ± 0.01	0.03 ± 0.01	0.02 ± 0.01

Table A.4: AmC background rates for the P17B data set [60].

[4]Within the single-digit precision of these percentages, the same value is obtained regardless of whether the denominator is chosen to be the signal rate or signal+background.

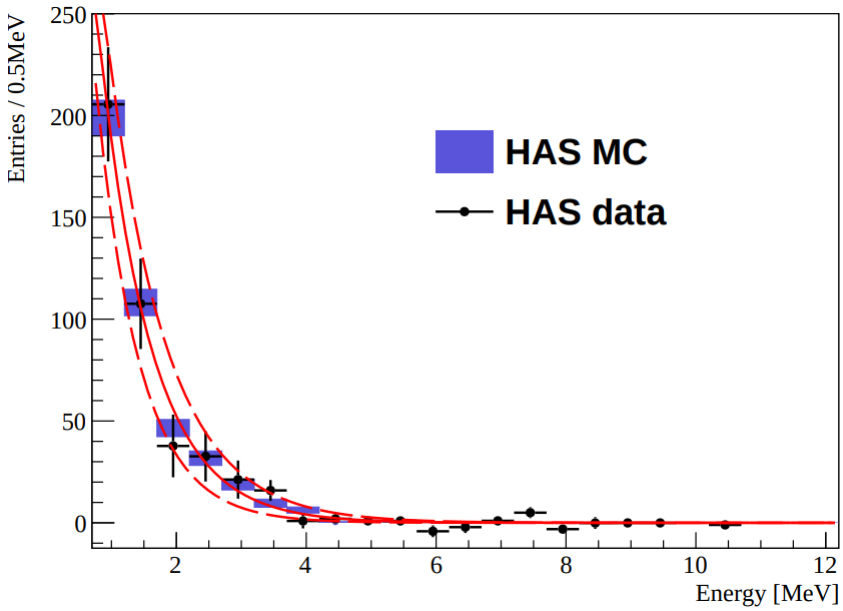

Figure A.16: Comparison of HAS prompt spectra from data and MC, showing the results of fitting the data to Equation A.16. From [59].

A.4 $^{13}C(\alpha, n)^{16}O$

Based on the chemical composition of the scintillator and the known cross sections of (α, n) reactions, it was determined that $^{13}C(\alpha, n)^{16}O$ is the only such reaction to occur in the ADs at any significant rate. Meanwhile, there are three natural decay chains that can lead to alpha-particle activity in the AD: The uranium, thorium, and actinium chains, which begin, respectively, with the long-lived isotopes ^{238}U, ^{232}Th, and ^{235}U[5]. Given that U, Th, and Ac all have similar chemical properties to Gd, a small amount of contamination is difficult to avoid during the Gd-doping process. In addition to these three decay chains, additional alpha particles come from the decay of ^{210}Po, a moderately stable ($t_{1/2}$ = 138 d) daughter of ^{222}Rn (which itself comes from the uranium chain). ^{210}Po was deposited on detector surfaces by ^{222}Rn during detector construction, and is essentially the only significant alpha-particle emitter outside the GdLS region.

Quantifying the $^{13}C(\alpha, n)^{16}O$ background consists of two parallel tasks. One task is to determine the level of alpha activity produced by the three decay chains and by ^{210}Po. The other task is to determine, for the set of alpha particles produced by a given chain (see Figure A.18), the probability and prompt spectrum of the $^{13}C(\alpha, n)^{16}O$

[5] In practice, the rate of the thorium chain is determined by the concentration of the shorter-lived ^{228}Th ($t_{1/2}$ = 1.9 yr), and likewise, for the actinium chain by ^{227}Ac ($t_{1/2}$ = 21.8 yr).

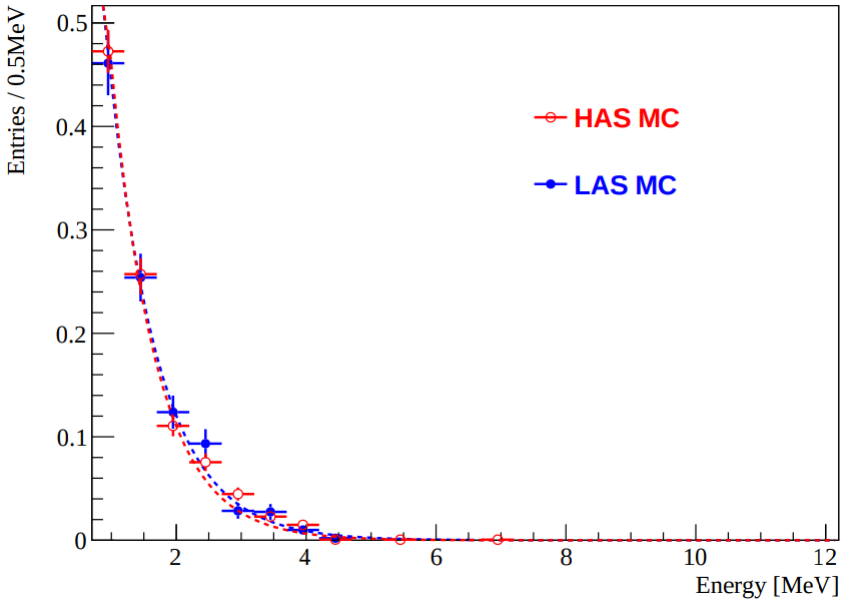

Figure A.17: Comparison of HAS and LAS prompt spectra from the MC, showing the results of fitting the HAS spectrum to Equation A.16. From [59].

events. These two pieces of knowledge can then be combined to yield a predicted rate and spectrum for the $^{13}\mathrm{C}(\alpha, \mathrm{n})^{16}\mathrm{O}$ background.

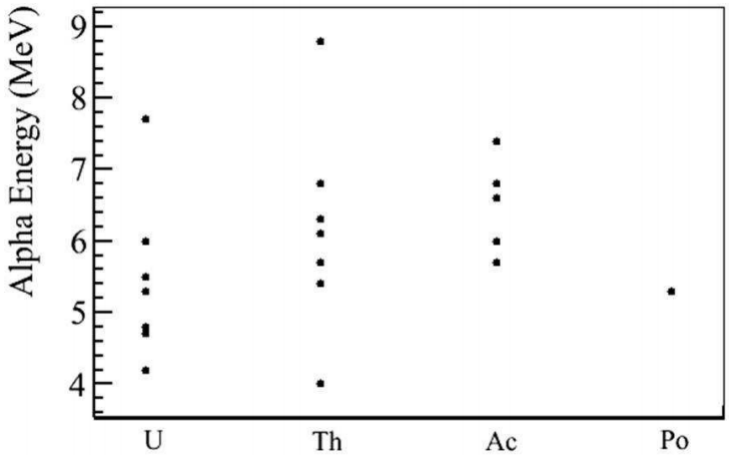

Figure A.18: Distribution of α particle energies from decays of ^{238}U, ^{232}Th, ^{227}Ac, and ^{210}Po. From [61].

The three chains all share a fortuitous property that enables a straightforward estimation of their rates. Namely, they each contain a rapid α-α or β-α cascade whose

time correlation and energy distribution allow for clean extraction from the data (Figure A.19). For the uranium, thorium, and actinium chains, these cascades are, respectively, ^{214}Bi $\rightarrow$ ^{214}Po $\rightarrow$ ^{210}Pb, ^{212}Bi $\rightarrow$ ^{212}Po $\rightarrow$ ^{208}Pb, and ^{219}Rn $\rightarrow$ ^{215}Po $\rightarrow$ ^{211}Pb, with Po half-lives of 164.3 µs, 0.3 µs, and 1.781 ms, respectively.

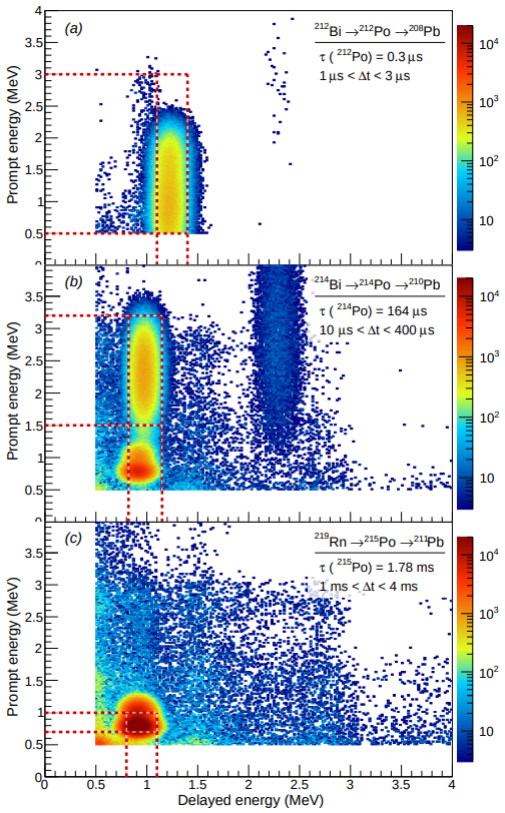

Figure A.19: Illustration of the prompt and delayed energy distributions for the Bi-Po cascades used in determining the rates of the ^{238}U, ^{228}Th, and ^{227}Ac decay chains. From [19].

To extract these events, time coincidence windows of [10, 400] µs, [1, 3] µs, and [1, 4] ms were used, respectively [19].[6] Accidentals (most significant for the actinium chain's Po cascade) were subtracted via the usual procedure, and for the uranium chain's Po cascade, contamination with nH IBDs was not an issue given that the (quenched) delayed energy of the alpha particle for these events is around 1-1.5 MeV,

[6] It is curious that, for the case of ^{212}Po (i.e. ^{232}Th), the time window used is significantly larger, relative to the half-life, compared to the other two isotopes. As shown in Figure A.19, the ^{212}Po sample is the "cleanest" (in terms of accidental backgrounds), enabling the use of a (perhaps excessively) wide window in order to maximize the statistics. Since Th is the most active chain, maximal statistics are desirable.

significantly below the 2.2-MeV gamma ray from nH capture. For the thorium chain, the prompt spectrum had to be extrapolated below 0.5 MeV in order to determine the total rate; otherwise, there were no major complications. Under the assumption that each chain is in equilibrium[7], the rate of the polonium cascade gives the rate of the entire chain.

At the point in time when Daya Bay began taking data, this procedure determined average (across ADs) rates of 0.009, 0.2, and 0.02 Bq for U, Th, and Ac, respectively [60]. Since the U chain is initiated by ^{238}U in the AD, its rate is essentially constant, given the ^{238}U half-life of 4.5 Gyr. On the other hand, the Th and Ac rates do decrease over time, since the parent half-lives (i.e. those of ^{228}Th and ^{227}Ac) are 1.9 and 21.8 yr, respectively. To determine the average Th and Ac rates for the P17B dataset used in this analysis, the Bi-Po cascade selection was repeated, and AD-averaged rates of 0.136 and 0.0178 Bq for ^{228}Th and ^{227}Ac, respectively, were obtained [60].

For ^{210}Po, a single decay instead of a chain, time correlations could not be exploited. Instead, the 5.3 MeV alpha particle produced by this isotope was, after quenching, visible as a peak around 0.5 MeV in the singles spectrum. Fitting this peak gave rates ranging from 4–10 Hz for each AD [69]. Although the ^{210}Po half-life is only 138 days, its parent, ^{210}Pb (which determines the ^{210}Po rate decrease over time), has a 22.3-year half-life. Within the precision of the measurement procedure, the decrease of the ^{210}Po rate is effectively unobservable.

For a given decay chain (or ^{210}Po decay), the set of emitted alpha particles is known (Figure A.18). For each of these alpha particles, in turn, simulations (using Geant4 [70] and SRIM [71]) can be used to determine the rate and prompt spectrum of the $^{13}\mathrm{C}(\alpha,\mathrm{n})^{16}\mathrm{O}$ events. At each step in the simulation, the alpha particle loses some energy and travels some distance according to its dE/dx profile in the LS. With some probability (i.e. cross section), during this step the alpha particle may be captured, producing one of the excited states of ^{17}O. If this happens, the ^{17}O will emit a neutron, whose energy depends on both the initial excited state of ^{17}O and the final (excited or ground) state of ^{16}O. The neutron produces prompt energy through proton recoils and, if it is sufficiently energetic, may scatter inelastically on ^{12}C to produce a $\sim$5 MeV gamma ray. Additional prompt energy will come from de-excitation gamma rays if the ^{16}O had been produced in an excited state.

The simulation calculates the (very low) probability of an (α,n) reaction occurring for a single cascade of alpha particles in a given chain (Table A.5). The simulation also produces, for the (α,n) reactions from a given chain, the 2D PDF of (a) the amount of energy deposited in the scintillator and (b) the kinetic energy of the emitted neutron (illustrated in Figure A.20). Finally, the prompt-energy spectrum (Figure A.21) is obtained by sampling this PDF and performing MC simulations of the detector's response to the alpha particle, the neutron's recoil protons, and any gamma rays from

[7] Up to ^{228}Th and ^{227}Ac for the thorium and actinium chains, rather than all the way up to ^{232}Th and ^{235}U, as noted previously.

^{12}C inelastic scattering and ^{16}O* de-excitation.

Chain	N_{ground}	N_{excited}	N_{total}	Uncertainty
^{210}Po	5.26×10^{-8}	4.90×10^{-9}	5.75×10^{-8}	7.2%
^{238}U	4.34×10^{-7}	2.96×10^{-7}	7.30×10^{-7}	16.9%
^{228}Th	4.49×10^{-7}	4.92×10^{-7}	9.41×10^{-7}	27.7%
^{227}Ac	4.72×10^{-7}	6.18×10^{-7}	1.09×10^{-6}	25.9%

Table A.5: Neutron yield (i.e., number of (α, n) events) per decay chain. N_{ground} and N_{excited} refer to the number of events that leave ^{16}O in the ground and excited states, respectively. N_{total} is their sum, whose uncertainty is given in the final column. From [61].

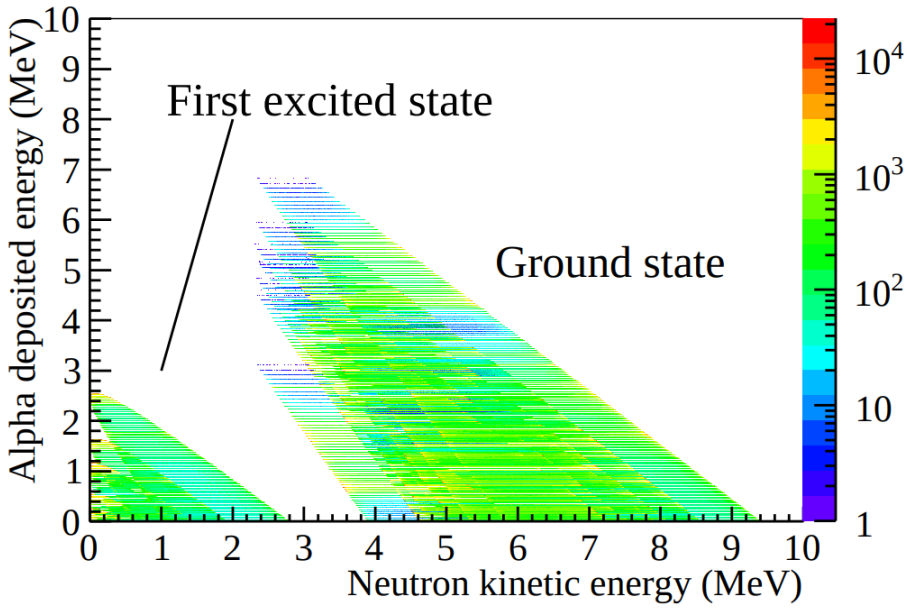

Figure A.20: Probability distribution function of alpha-particle energy deposition and neutron kinetic energy for (α, n) reactions from the ^{228}Th chain. From [61].

Uncertainties in the $^{13}\text{C}(\alpha, \text{n})^{16}\text{O}$ prediction arise from a number of sources. The uncertainty coming from the (α, n) cross section was estimated by repeating the MC procedure using two different cross-section tables, JENDL [72] and EXFOR [73] (Figure A.22). This suggested an uncertainty ranging from 6.6% (for ^{210}Po) up to 27.5% (for ^{232}Ac) [61]; a nominal uncertainty of 20% was thus assigned. Meanwhile, there was negligible effect on the predicted neutron reconstructed energy spectrum from switching between the cross-section tables and from changing the assumed

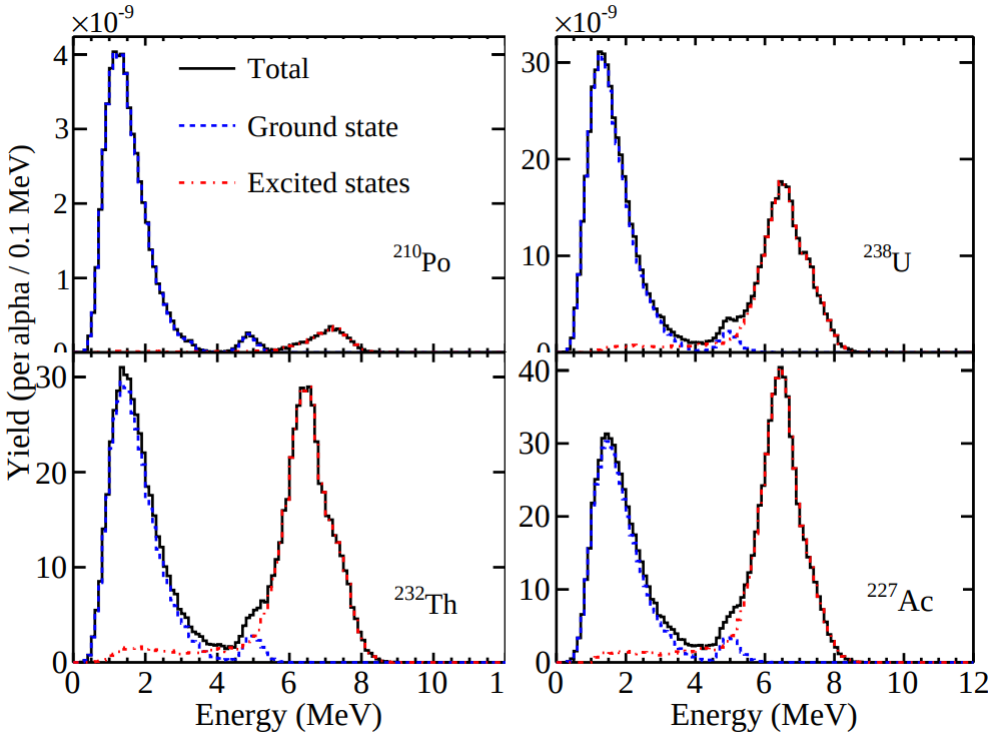

Figure A.21: Simulated prompt-energy spectra for (α, n) reactions produced by the four decay chains. From [19].

neutron angular distribution (Figure A.23). Additional uncertainty could come from the fundamentals of the MC simulation, i.e., the dE/dx table and the numerical integration of discrete steps. This was evaluated by comparing the results of Geant4 [70] and SRIM [71] (Figure A.24), which differed, overall, at a negligible level of less than a percent. Finally, the assumption of decay-chain equilibrium, and the efficiency of the cascade selection, both could introduce additional uncertainty, leading to the conservative assignment of an overall 30% uncertainty[8] in the measurement of the rates of the decay chains. This 30% uncertainty is combined with the 20% uncertainty assigned to the neutron yield calculation; conservatively adding the two uncertainties then gives a total uncertainty of 50% on the $^{13}\mathrm{C}(\alpha, \mathrm{n})^{16}\mathrm{O}$ rate estimation.

For the P17B data set used in this analysis, the $^{13}\mathrm{C}(\alpha, \mathrm{n})^{16}\mathrm{O}$ background rates were re-estimated using updated measurements of the rates of alpha activity [60]. The values are listed in Table A.6.

[8]Presumably validated by varying the Bi-Po cascade selection criteria, although this is not explicitly stated in the references.

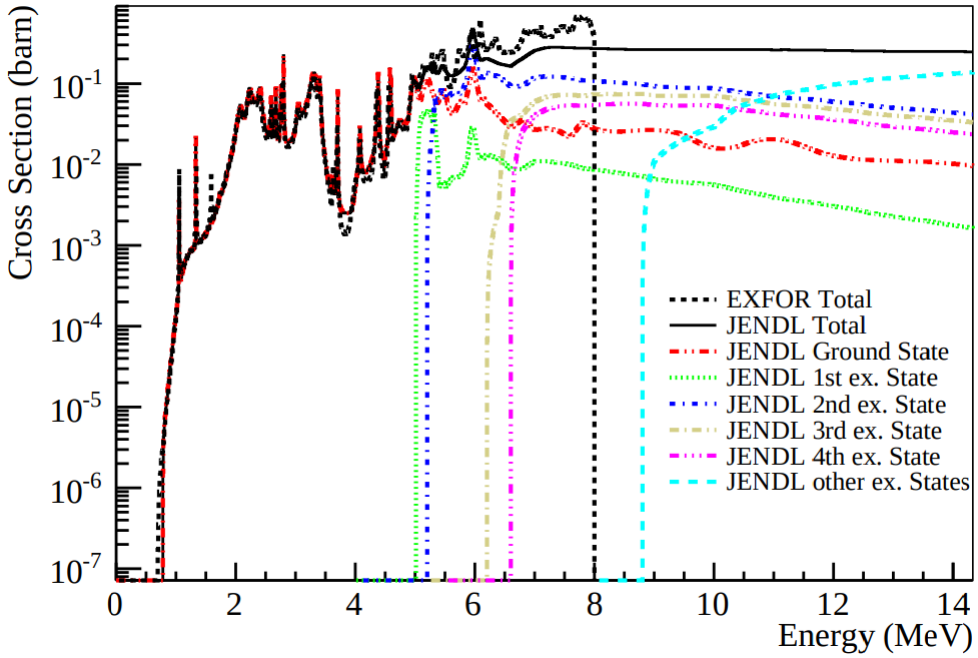

Figure A.22: Cross section of the $^{13}C(\alpha, n)^{16}O$ reaction as reported by the JENDL and EXFOR tables. From [61].

	EH1		EH2		EH3			
Period	AD1	AD2	AD1	AD2	AD1	AD2	AD3	AD4
6AD	0.09 ± 0.04	0.07 ± 0.04	0.05 ± 0.02		0.05 ± 0.02	0.04 ± 0.02	0.04 ± 0.02	
8AD	0.08 ± 0.04	0.06 ± 0.03	0.04 ± 0.02	0.06 ± 0.03	0.04 ± 0.02	0.04 ± 0.02	0.03 ± 0.02	0.04 ± 0.02
7AD		0.05 ± 0.03	0.03 ± 0.02	0.06 ± 0.03	0.03 ± 0.02	0.03 ± 0.02	0.03 ± 0.01	0.03 ± 0.02

Table A.6: $^{13}C(\alpha, n)^{16}O$ background rates for the P17B data set [60].

A.4.1 Efficiencies

The exact efficiency of these cuts (i.e., the rejection factor for flashers), is unimportant, as long as it is high enough.[9] Any residual flashers will automatically be counted in the singles rate, and thus so will their contribution to the accidental background rate.[10] As shown in Figure A.2, the rejection factor (and hence the purity of the IBD sample, with respect to flashers) is very nearly 100% in all ADs, with the minor exception of AD5, where the small contribution of unrejected flashers will, in any case, be removed during the subtraction of the accidental background spectrum.

[9] According to [17], the efficiency is greater than 99.99%.

[10] There is a small second-order correction due to the fact that an accidental cannot be formed by two flashers in the same PMT, given that it takes on the order of a second for a PMT to "recharge" after flashing. However, this correction is negligible given the extremely low rate of residual flashers.

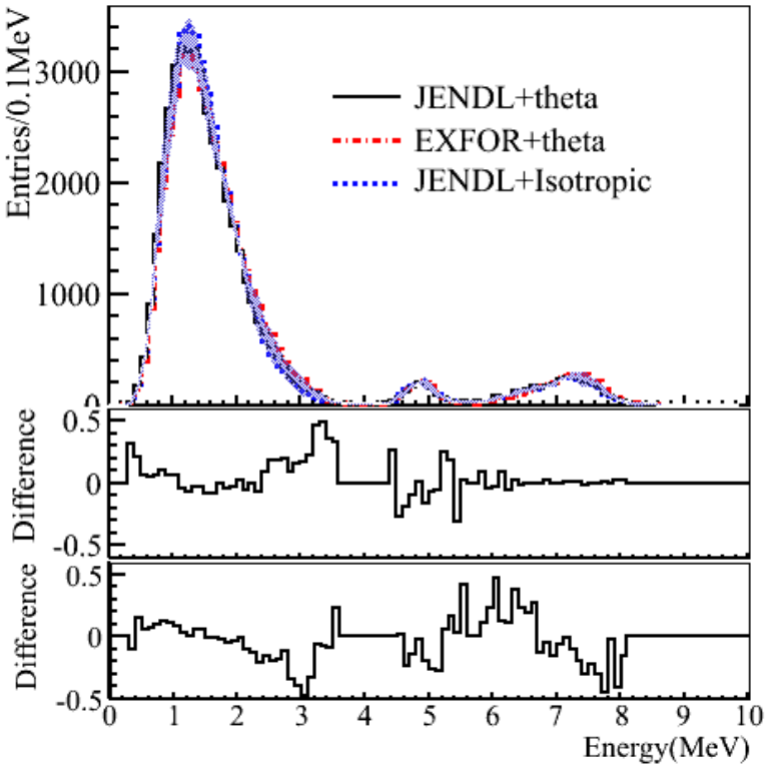

Figure A.23: Predicted reconstructed energy of neutrons for different cross-section tables (JENDL and EXFOR) and different neutron angular distributions ("theta", $d\sigma/d\Omega \sim 1/\sin\theta$, and "isotropic", $d\sigma/d\Omega \sim 1$). The differences are negligible. From [61]. (The particular decay chain is not specified.)

Compared to the efficiency, it *is* important to study the signal inefficiency, i.e. the probability of improperly rejecting an IBD prompt or delayed trigger. If this inefficiency differs significantly among the ADs, it could bias the oscillation result. The inefficiency was estimated [74] using a ∼600-day sample of IBD-like events (without any flasher cuts). For each of the three flasher discriminators (2", f_{ID}, and f_{PSD}), a 2D histogram was constructed of its values for the prompt and delayed events. These histograms indicated that essentially all rejected IBDs had a flasher-like prompt (but not delayed) trigger. Accordingly, 1D histograms were then constructed of the discriminators of the prompt triggers for all pairs that lacked a flasher-like delayed trigger. Near the region of overlap between the IBD and flasher distributions, the true IBD distribution was fit (to both exponential and Gaussian functions) and extrapolated in order to estimate the fraction within the rejection region (Figure A.25). Systematic uncertainties were determined by varying the fit range and function. These results were then cross-checked by exploiting the different prompt-delayed time correlations

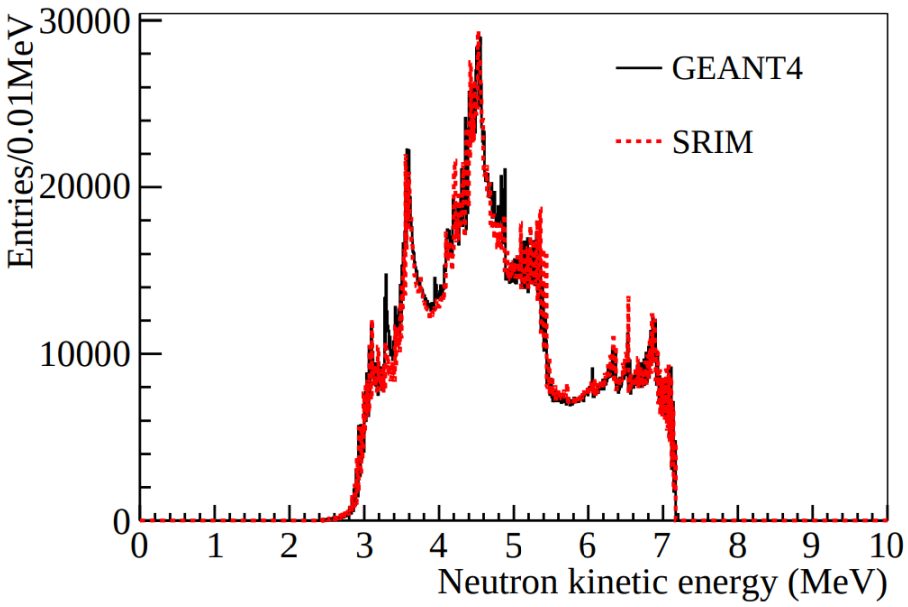

Figure A.24: Neutron kinetic energy as predicted by Geant4 and SRIM. From [61]. (The particular decay chain is not specified.)

of the two event classes, giving an independent breakdown of the total distribution into its IBD and flasher components (Figure A.26). The total inefficiency was found to be $0.039\% \pm 0.006\%$, as summarized in Table A.7. AD-to-AD variations lay within this uncertainty band.

Cut	Inefficiency (%)
2" PMT	0.000 ± 0.000
f_{ID}	0.023 ± 0.004
f_{PSD}	0.016 ± 0.002
Total	0.039 ± 0.006

Table A.7: Inefficiencies of the flasher cuts [74]. For true IBDs, the distributions of the three discriminator were assumed to be uncorrelated, so that the total inefficiency was determined as the sum of the three. The systematic uncertainties were conservatively assumed to be fully correlated, and were thus added linearly rather than in quadrature.

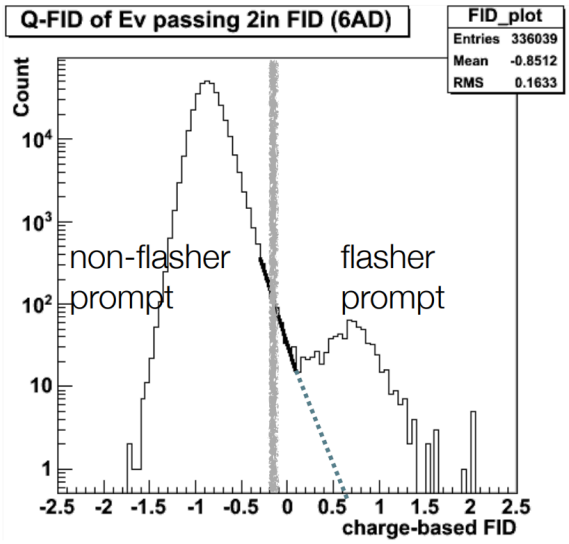

Figure A.25: The prompt f_{ID} distribution of IBD-like events for which the delayed trigger is *not* flasher-like (as determined by f_{ID}). The true IBD component is fit and extrapolated in order to determine the inefficiency of the cut. From [74].

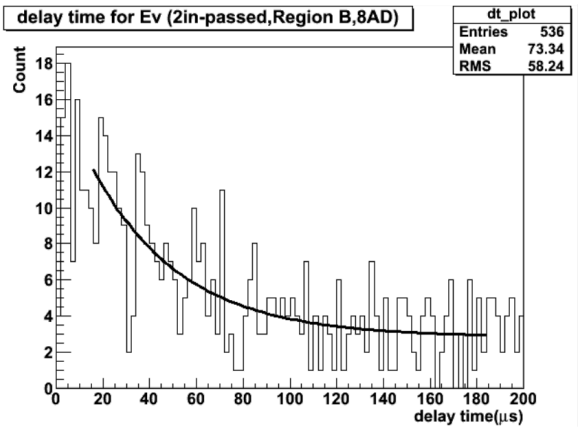

Figure A.26: The prompt-delayed time difference of IBD-like events for which the delayed trigger is *not* flasher-like (as determined by f_{ID}). The true IBD component follows an exponential distribution, while the flasher component is flat. A fit is performed in order to quantify the two components. From [74].

Appendix B

Reactor antineutrino prediction

In principle, a reactor-based near/far experiment can measure the oscillation parameters without any recourse to reactor physics: Simply measure the spectrum of the $\bar{\nu}_e$s at the near site, "unoscillate" it back to the reactor, and then oscillate it out to the far site. The oscillation parameters are then those that give the best fit to the far-site data. No information is needed on the reactor power, fission fractions, or theoretical antineutrino spectra.

In practice, the situation is complicated by the fact that Daya Bay features three reactor complexes and two near sites. Instead of two baselines, there are effectively nine, and the reactors can differ from each other in terms of instantaneous power and burnup. These factors must be accounted for when predicting the far-site $\bar{\nu}_e$ spectra from the near ones.

In the simplest approach, one can just use the known operating power and baseline information to split each near $\bar{\nu}_e$ spectrum into components from each reactor, and then unoscillate/oscillate each component separately to the far site. This can be done separately for each near site, and the two far-site predictions can then be averaged. This is sufficient for getting a good measurement of θ_{13}, thanks to the cancellation of absolute efficiency uncertainties and the lack of need for reactor modeling.

However, measuring Δm^2 involves examining the detailed shapes of the near and far spectra. In the simple approach just described, it is assumed that all reactors are producing the same spectral shape (and ratio of $\bar{\nu}_e$ flux to reactor power). This is not necessarily accurate, as the reactors may be in different stages of their fuel cycles. Therefore, to get the most out of a rate/shape analysis, we must perform a full prediction of the spectrum from each reactor separately. This will, naturally, introduce uncertainties from the imperfect nature of the predictions. As long as the predictions are applied consistently, their uncertainties will be largely (but not completely) canceled in the near/far ratios, so this primary benefit of a near/far measurement is not lost.

This chapter describes the reactor prediction, beginning with the question of predicting the antineutrino spectrum from a single fission of a given isotope, and

then proceeding to discuss how these are combined, with the aid of information on generated thermal power and burnup (i.e., fission fractions), to form a full spectrum prediction.

B.1 Spectrum prediction

In a conventional pressurized water reactor, each megawatt of thermal power corresponds to about 2×10^{20} fissions per second, with each fission producing an average of about six antineutrinos from the beta decays of the fission fragments (or their daughters) [63]. Virtually all of the power originates from the fission of four isotopes: ^{235}U, ^{239}Pu, ^{241}Pu, and ^{238}U. In a fresh fuel assembly of low-enriched (3–5% ^{235}U) uranium, initially some 92% of fissions will be of ^{235}U, while the remaining 8% will be fast-neutron fissions of ^{238}U. This ^{238}U fraction remains nearly constant throughout a fuel cycle. At the same time, some of the ^{238}U will undergo neutron capture and subsequent beta decay to ^{239}Pu, whose fission rate rapidly reaches 10-20% of the total, eventually catching up to the (decreasing) ^{235}U rate by the end of the cycle (∼450 days). A fraction of ^{239}Pu will produce ^{241}Pu after a pair of neutron captures, and by the end of the cycle this isotope will contribute a fission fraction comparable to that of ^{238}U. This evolution of the fission fractions is illustrated in Figure B.1.

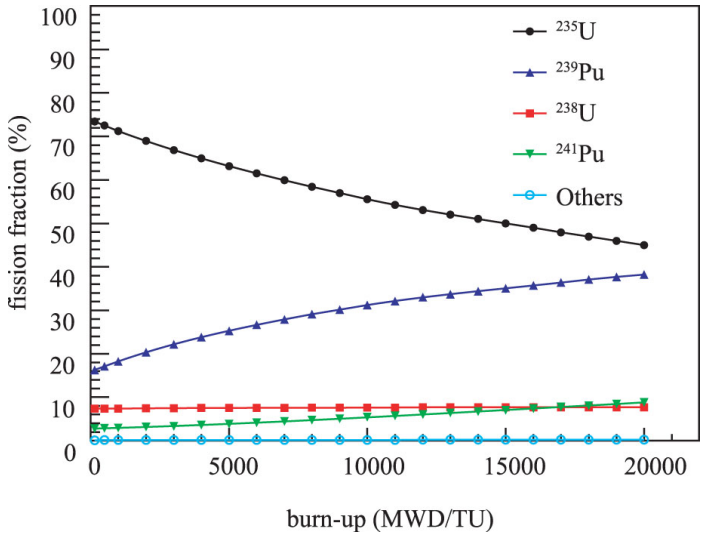

Figure B.1: Fission fractions over time in a typical pressurized water reactor. From [25].

Clearly, all four isotopes contribute a non-negligible fraction of the total thermal power, and since their antineutrino spectra are fairly similar, they all contribute significantly to the flux. Therefore, it is imperative that each isotope's spectrum be predicted accurately. This can be done either *ab initio*, by summing theoretical

spectra based on the individual beta-decay branches of all fission products listed in nuclear databases, or via the *conversion method*, in which the total beta-ray spectrum is measured and then converted into an antineutrino spectrum. The latter method is generally preferred when measurements are available, as nuclear databases are known to be incomplete. In what follows, we discuss the existing predictions produced using both methods, and determine an optimal set to use in the oscillation analysis.

B.1.1 *Ab initio* method

In a beta decay, conservation of energy implies a one-to-one correspondence between electron energy E_e and antineutrino energy E_ν.[1] For a single beta-decay branch of endpoint energy E_0,

$$E_\nu = E_0 - E_e. \tag{B.1}$$

Hence, if we know the spectrum of every beta-decay branch of every fission product, we can invert them all and sum them up to derive the total antineutrino spectrum. Unfortunately, this procedure is hampered by two significant issues. First of all, nuclear databases contain the beta-decay *endpoints*, not the spectra, and there are theoretical uncertainties involved in calculating the spectra. Secondly, existing databases [72, 75] are known to be missing some 10% of beta-decay branches, and errors exist in the listed branching ratios for the known branches.

The theoretical difficulties arise because, to properly model the spectrum for a single beta-decay branch, it is not enough to merely know the endpoint energy. One must also know the *type* of decay: Is it a Fermi decay, with antiparallel electron and antineutrino spins? Or a Gamow-Teller (GT) decay, with parallel spins? Or a mixture? Does the lepton system carry any orbital angular momentum, making it a *forbidden* (as opposed to an *allowed*) decay? If the decay is forbidden, does the nucleus undergo the maximum possible change in angular momentum ΔJ (a *unique* decay), or not (a *non-unique* decay)? The type of decay directly affects the shape of the spectrum.

Traditionally, *ab initio* calculations have assumed that all beta-decay branches are of the allowed type. Unfortunately, nuclear databases indicate that some ∼25% of fission product beta-decay branches are forbidden. Even if we generously assume that the databases correctly list the type of each decay, there remains a problem: While it is possible to calculate a general shape correction for *unique* forbidden decays, the correction for a *non-unique* decay depends on the exact combination of nuclear matrix elements involved in the decay, and this information is largely unknown.

Furthermore, the idealized Fermi model of beta decay ignores a number of subtle effects that add further corrections to the spectrum. Most of these are well-understood, including the effects of the finite size of the nucleus, charge screening, and radiative

[1] Ignoring nuclear recoil effects, which introduce a negligible smearing of $\mathcal{O}(E_0/(Am_p)) \sim 10^{-4}$, where A is the mass number of the nucleus, and m_p is the proton mass [63].

corrections. These corrections can be applied with minimal uncertainty. However, there is an additional effect known as *weak magnetism* (WM): Essentially, the total weak current of the nucleus contains a contribution from the spatial distribution of the vector current, and this factor depends on the specific (and typically unknown) details of each nucleus's structure. In *ab initio* calculations, numerous assumptions and simplifications are applied in order to produce a tractable model for the WM correction, which is then fit to measured observables. While this approach is better than nothing, it still results in one of the largest components of the uncertainty produced by the *ab initio* method.

Finally, *ab initio* calculations are hampered by the fact that nuclear databases suffer from missing and incorrect information. This problem is particularly acute for the rarest and/or most unstable isotopes, some of which are entirely missing from the databases. Due to their short livetime, these isotopes are expected to possess high-energy beta-decay branches, well above the inverse beta-decay threshold. Among known daughter isotopes, ∼6% lack *any* tabulated data on beta-decay branches [76]. In [64], discussed later, their hybrid *ab initio*/conversion procedure suggests that nuclear databases fail to account for some 10% of the measured beta-decay spectrum. Even among isotopes for which data exists, this data may be biased by the *pandemonium effect*, in which low-E_0 beta-decay branches are undercounted relative to the high-E_0 ones for the same isotope, due to the fact that the deexcitation of the daughter nucleus involves low-energy transitions between closely spaced energy levels, and these gamma-rays often evade measurement.

Altogether, while the *ab initio* method is attractive and elegant in principle, in practice it suffers from deficiencies in the underlying nuclear databases, as well as a poor understanding of the weak magnetism effect. For this reason it has traditionally been rejected in favor of the conversion method, described next. Recent *ab initio* calculations still predict a flux that is some 5–10% lower than that obtained from the conversion method, but the other systematic uncertainties are nevertheless comparable between the two methods. The overall preference for the conversion method has one exception, however: ^{238}U. Since it is only fissioned by fast neutrons, measuring the total beta-ray spectrum of its fission daughters is difficult, and data only became available in the last few years. Given that ^{238}U only contributes some 10% of the total fission rate, even a conservative 10% error on the *ab initio* result will only result in an uncertainty of 1% on the total antineutrino spectrum, which is acceptable.

B.1.1.1 ^{238}U calculations

In the 1980s, Vogel carried out a prediction of the ^{238}U antineutrino spectrum using the nuclear data available at the time [77]. Although this was a very careful analysis, certain approximations were made in order to keep the calculation tractable; for instance, the finite size and weak magnetism effects were parameterized by a single

energy-dependent correction applied to the spectrum as a whole, rather than being treated branch-by-branch.

In 2011, the group of Mueller *et al.* revisited the problem using the latest available nuclear data, aggregated and curated from multiple sources [64]. This time, all higher-order corrections were applied on each branch individually. With these two improvements, along with an increase in the calculated IBD cross section due to a change in the measured lifetime of the neutron, a 9.8% increase was found in the predicted ^{238}U contribution to the IBD rate, relative to Vogel's result. To this day, the Mueller spectrum remains the state-of-the-art among ^{238}U *ab initio* predictions, and it is the one that we use in this oscillation analysis.

B.1.2 Conversion method

In contrast to the *ab initio* method, which depends on thousands of measurements of the individual fission daughters and beta-decay branches, the conversion procedure relies on just one measurement: The total beta-ray spectrum from fissions of a given isotope. For a single beta-decay branch, the measured electron spectrum can be directly converted to an antineutrino spectrum via Equation B.1:

$$E_\nu = E_0 - E_e.$$

In this case, E_0 can be immediately inferred from the endpoint of the measured spectrum. On the other hand, the *total* beta-ray spectrum is the sum of many beta-decay branches with different endpoints, and these endpoints cannot be determined from the total spectrum alone. If one were to use the endpoints listed in nuclear databases, the resulting antineutrino spectrum would be incomplete, given the lack of data for a significant fraction of branches. The traditional solution to this problem is to fit the total spectrum with a series of fictional *virtual* branches, which are then inverted separately and summed to give the total antineutrino prediction. Typically, one starts at the endpoint of the total spectrum, positing a virtual branch (an allowed decay, generally[2]) of the same endpoint. This virtual branch is normalized by fitting it to the end of the total spectrum, and then subtracted out. The procedure is repeated at the new endpoint of the subtracted total spectrum, until one finally ends up with a few dozen virtual branches that together fit the entire spectrum.

While this approach avoids the uncertainties caused by the incompleteness of the nuclear databases, it gains uncertainty from the arbitrary nature of the virtual branch technique. Fortunately, this uncertainty can be characterized fairly well by varying the

[2]To first order, the shape of the beta-decay spectrum is unaffected by the distinction between Fermi and GT decays. However, higher-order corrections, such as those for weak magnetism, are sensitive to this detail. According to [63], the reactor antineutrino spectrum is dominated by GT decays, so this type is assumed for the virtual branches. Furthermore, [63] argues that the shape corrections for forbidden decays are small due to reasons of symmetry, so that the virtual branches can be assumed to be allowed. This argument is challenged in [78].

procedure and observing the changes in the result. It should be noted that, in spite of their relative independence, both approaches suffer from some of the same theoretical uncertainties involved in inverting single beta-decay branches (whether real or virtual), particularly from weak magnetism. In the end, however, the 3–5% uncertainty of the conversion method is a worthwhile tradeoff in order to avoid the 5–10% bias imposed on the *ab initio* method by the deficiencies in the available nuclear data.

B.1.2.1 ILL beta-ray spectra measurements

For ^{235}U, ^{239}Pu, and ^{241}Pu, all conversion predictions make use of the same measurements of the total beta-ray spectra. These were taken at ILL in 1980s by Schrekenbach *et al.* [79–82]. Thin foils of each isotope were subjected to a thermal neutron flux from the ILL research reactor, and a small, extremely pure sample of beta-decay electrons escaped through a narrow vacuum pipe for measurement by a high-resolution magnetic spectrometer, BILL. The normalization, that is, the total number of fissions N_f, was determined by the neutron flux ϕ_n, the fission cross-section σ_f, the number of fissile atoms n_f, and the irradiation time t_f:

$$N_f = \phi_n \cdot \sigma_f \cdot n_f \cdot t_f. \tag{B.2}$$

In turn, the neutron flux was measured by irradiating calibration targets made from materials with well-understood neutron capture reactions. The counts N_{cal} of these reactions were determined by measuring the associated internal conversion lines and beta-ray spectra. Using the known cross-section σ_{cal} of a given neutron capture reaction, ϕ_n can be calculated by rearranging the analogue of Equation B.2:

$$\phi_n = \frac{N_{\text{cal}}}{\sigma_{\text{cal}} n_{\text{cal}} t_{\text{cal}}} \tag{B.3}$$

Although the original ILL conversion results (i.e. $\bar{\nu}_e$ spectra, as described next) have been refined by later authors, the ILL measurements remain the standard for the total spectra of beta-rays from fission of ^{235}U, ^{239}Pu, and ^{241}Pu.

B.1.2.2 Schreckenbach ILL conversion

After these beta-ray spectra were measured, Schreckenbach *et al.* proceeded to convert them into antineutrino spectra [79–82]. Their method was a "pure" conversion, based only on virtual branches with no input from nuclear databases. As with Vogel's ^{238}U *ab initio* prediction, certain corrections (finite size, weak magnetism) were applied in a simplified manner to the total $\bar{\nu}_e$ spectrum. Until 2011, the Schreckenbach conversion was considered canonical.

B.1.2.3 Mueller ILL conversion

The situation changed in 2011 due to the aforementioned Mueller group [64], who took the attitude that, while the nuclear databases are indeed somewhat incomplete, the data they *do* include is still a precious constraint that deserves to be considered in the conversion procedure. Accordingly, they began with an *ab initio* calculation for all four isotopes. For ^{238}U, as described above, they had to stop at that point. But for the other three isotopes, they proceeded to subtract the *ab initio* spectra (for the electron, not the antineutrino) from the ILL measurements, leaving a $\sim$10% residual, which was then fit with virtual branches. In a further departure from the previous ILL conversion procedure, the Mueller procedure employed a more accurate, branch-by-branch correction for weak magnetism and other higher-order effects. Although the quoted systematic uncertainties on the spectra were not substantially different from ILL's, there was a significant $\sim$3% increase in the total predicted IBD rate, which was later corroborated by the work of Huber [63], as described next.

B.1.2.4 Huber ILL conversion

Shortly after the publication of the Mueller prediction, Huber undertook an independent calculation of the antinuetrino spectra from ^{235}U, ^{239}Pu, and ^{241}Pu [63]. Unlike Mueller *et al.*, he avoided using nuclear databases, instead converting each beta-ray spectrum using only virtual beta-decay branches, as in the original ILL results. However, [63] included careful studies of the variance introduced by the details of the conversion procedure; based on these studies, the procedure was tuned to minimize any introduced bias. As in [64], a careful branch-by-branch treatment was applied to the WM and other corrections. With these improvements, the result was largely consistent with [64]. In this oscillation analysis, we use the Huber predictions for these three isotopes. Previous studies have shown negligible differences from using the Mueller results from [64] instead.

B.1.2.5 FRM II U238 measurement and conversion

It is worth noting that in 2013, the ^{238}U total beta-ray spectrum was finally measured by N. Haag *et al.* at the FRM II neutron source in Germany [83]. They exposed foils of natural uranium to both thermal and fast neutrons, measuring the beta-ray spectra with a gamma-ray suppressing electron telescope, whose efficiency was accurately determined by comparing their ^{235}U data to the results from ILL. The beta-ray spectrum was then converted to an antineutrino spectrum using the virtual branch technique. Although it is worth exploring the possibility of using this latest result, in practice the effects are unlikely to be substantial, given the relatively minor fission rate of ^{238}U. In keeping with the official Daya Bay results published to-date, our ^{238}U prediction uses the pure *ab initio* Mueller calculation from [64].

B.1.3 Off-equilibrium correction

In an operating reactor, thousands of fission-derived unstable isotopes are continually decaying and being regenerated (from parent isotopes or directly from fission). The number of nuclei N_d of a "daughter" isotope d changes at a rate that depends on the number of parent nuclei N_p for all "parent" isotopes p (or fission), the rate constants R_{pd} for the production of d, and the rate constant R_d for the decay of d:

$$\frac{dN_d}{dt} = \sum_p R_{pd} N_p - R_d N_d. \tag{B.4}$$

The physical solution to this differential equation is an exponential function in which dN_d/dt asymptotically goes to zero. Thus, over time, a reactor will tend toward *equilibrium*[3], in which each isotope is held at a constant concentration, decaying at the same rate at which it is generated. The longer an isotope's lifetime (or those of its ancestors), the longer it takes to reach equilibrium. In practice, most of the isotopes in a reactor reach equilibrium early in the fuel cycle. The longer-lived isotopes, however, introduce a small time-dependence to the antineutrino spectrum predicted for each fission fuel.

At ILL, the target foils were irradiated for about a day. However, reactor fuel typically lives in the core for more than a year. As such, the ILL beta-ray spectra do not properly account for isotopes that take more than a day to reach equilibrium. Around 10% of the fission products meet this criterion, highlighting the importance of correcting for this issue. As longer-lived isotopes are the ones affected, the spectral distortion is restricted to low energies, up to about 4 MeV. Any correction must obviously be time-dependent, spanning the range from initial irradiation up to the end of a fuel cycle.

Mueller *et al.* provide correction factors for ^{235}U, ^{239}Pu, and ^{241}Pu at five energy values from 2 to 4 MeV, calculated at irradiation times of 100 d, 10^7 s (∼115 d), 300 d, and 450 d [64]. The ILL reference spectra provide an anchor at 36 h (except for ^{235}U, which was measured at 12 h; a 36 h correction is thus provided for the isotope). These corrections, at 450 d, are at most 2% (5% for ^{235}U). They also account for another, subdominant difference between the reactor environment and the ILL apparatus, namely, the presence of epithermal and fast neutrons, which alter the distribution of fission products. The reactor simulation code MURE was used to determine the correction, which was validated against the FISACT code.

In this oscillation analysis, we use Mueller's corrections verbatim, collectively referring to them as the "off-equilibrium correction". Interpolation in time is performed between the provided points, and the appropriate value of the irradiation timem is determined from data provided by the power company (discussed in Section B.2). In the Daya Bay cores, one third of the fuel is replaced during each refueling, so each

[3]Ignoring any (slow) variation in the relative fission fractions.

core contains three fuel batches at different levels of burnup. Accordingly, a weighted sum of off-equilibrium corrections is applied. Conveniently, the batched fueling tends to wash out the differences between cores as well as the overall time dependence of the correction.

For our purposes, we assign a 30% uncertainty to the Mueller correction factors. The correction increases the predicted flux by some 0.5%, so the ensuing uncertainty on the absolute rate is around 0.2%.

B.1.4 Spent nuclear fuel

During refueling, spent nuclear fuel (SNF) is moved to water-filled storage pools close to the core. Long lived fission products in the SNF will continue to decay, producing an additional low-energy antineutrino flux whose spectrum can be calculated from nuclear data, assuming that the fuel's irradiation history (and time since removal) is known. The P17B-period SNF flux (Figure B.2) was predicted using data on the history of each batch of stored SNF [84]. In total, SNF is predicted to account for around 0.3% of the IBD rate, largely below 3.5 MeV (given the relatively long lifetimes of the decaying isotopes), with an assigned uncertainty of 100%.

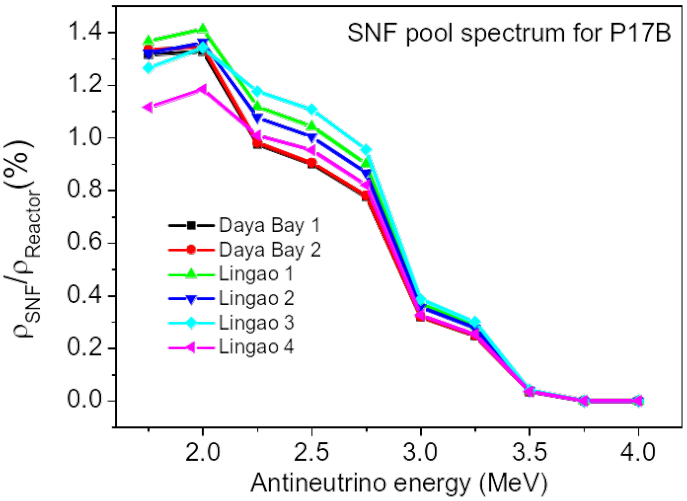

Figure B.2: The spectrum of antineutrinos from each pool of spent nuclear fuel, relative to the spectrum of antineutrinos from the corresponding core. From [84].

B.2 Power, burnup, and fission fractions

In order to combine the various ingredients above into a final reactor antineutrino prediction, we must know the generated thermal power of each core, its fission fractions, and its refueling history. This information is provided by the power company in a number of forms.

Each core's average daily thermal power is provided as a fraction of the full nominal power (2895 MW). This information is crucial for calculating the fission fractions (discussed below), but the Daya Bay detectors do not always produce a full 24 hours of "good" data each day. As such, the collaboration periodically supplies the power company with a "good hour" list, from which the company produces a livetime-averaged daily power. Combined with the daily livetime, this provides a normalization for the daily predicted spectrum at each detector.

The fission fractions of a core can be expressed as a function of *burnup* (Figure B.1). Burnup is defined as the total energy extracted from each unit mass of fuel, typically given in units of MWd/(Metric Ton Uranium). The power company has performed simulations to determine the fission fractions versus burnup of each core[4], and this information is provided to the collaboration in intervals of ∼1000 MWd/MTU. The company also provides the burnup of each core at two-week intervals; using the daily generated thermal power to interpolate between the provided points, the daily burnup, and hence fission fractions, can be calculated.

B.3 Final AD spectrum

Putting it all together, the predicted IBD spectrum at a given detector, from a single core, is

$$S_{\text{ibd}}(E,t) = \frac{N_p(t)}{4\pi L^2}\,\epsilon(E,t)\,\sigma(E)\left(\sum_k\left(\frac{W(t)\,f_k(t)}{\sum_k f_k(t)\,e_k}S_k(E)\,c_k^{\text{ne}}(E,t)\right) + S_{\text{snf}}(E,t)\right). \tag{B.5}$$

Here, E is the $\bar{\nu}$ energy, not the IBD prompt energy; $N_p(t)$ is the number of target protons; L is the baseline; $\epsilon(E,t)$ is the detection efficiency; $\sigma(E)$ is the IBD cross section; $W(t)$ is the thermal power; $f_k(t)$ is the fission fraction of isotope k, e_k is its energy per fission (Table B.1), $S_k(E)$ is its per-fission spectrum, and $c_{\text{ne},k}(E,t)$ is its non-equilibrium correction; and finally, $S_{\text{snf}}(E,t)$ is the spectrum from spent fuel. The full prediction at each detector is obtained by adding the contributions from each core.

[4]Data is provided for a full fuel cycle; for cycles that haven't yet run their course, nominal projections of thermal power are used for the "future" fractions.

	^{235}U	^{238}U	^{239}Pu	^{241}Pu
Energy (MeV)	201.92 ± 0.46	205.52 ± 0.96	209.99 ± 0.60	213.60 ± 0.65

Table B.1: Energy per fission of the four main fuels in a pressurized water reactor [85].

B.4 Covariance matrix

The uncertainties of the reactor antineutrino spectra exhibit significant correlations between energy bins, isotopes, and reactor cores. As such, it would be inappropriate to simply assign an independent error bar to each energy bin of the predicted spectra. Instead, a full covariance matrix is needed in order to account for the correlations. Such a covariance matrix was generated by taking into account the following uncertainties [65]:

- The uncertainties in the ^{235}U, ^{239}Pu, and ^{241}Pu spectra resulting from the statistics of the ILL measurements and from the biases of the conversion procedure [63]. These are uncorrelated between energy bins and isotopes, but correlated between cores.
- The uncertainties in the ^{235}U, ^{239}Pu, and ^{241}Pu spectra resulting from nuclear physics (Coulomb and weak magnetism corrections) [63]. These are completely correlated between energy bins, isotopes, and cores.
- The uncertainties in the ^{238}U spectrum, primarily due to missing information in the nuclear databases [64]. These are conservatively treated as correlated between energy bins and cores.
- The uncertainties of the energies per fission (Table B.1). The errors are assumed to be Gaussian, and uncorrelated between isotopes, but correlated between cores.
- The uncertainties in the fission fractions determined from data provided by the power company. Based on comparisons of simulations and measurements of spent fuel rods, an uncertainty of 5% is assigned, correlated between cores.
- The uncertainties in the generated thermal power reported by the power company. Monthly measurements of power were taken using an offline heat-balance system with an 0.48% uncertainty. The daily power measurements for each core, in turn, were taken using an online system calibrated to the offline one, with an 0.1% calibration error. As such, the reactor power uncertainty is assigned to be 0.1% uncorrelated and 0.5% correlated between cores.
- The 30% uncertainties in the non-equilibrium correction factors from [64]. These are correlated between energy bins and cores, but uncorrelated between isotopes.

- The uncertainties in the SNF spectra. An uncertainty of 100%, correlated between spent fuel pools and energy bins, is assigned, as well as an additional 100% uncertainty correlated between energy bins but not between pools (to account for the uncertainty in the quantity of spent fuel in each pool).

A large sample of M simulated total reactor antineutrino spectra was simulated[5], with random fluctuations applied according to the uncertainties listed above [65]. The covariance matrix was then calculated according to

$$V_{ab} = \frac{1}{M}\sum_{n=1}^{M}(S_a^n - S_a^0)(S_b^n - S_b^0), \tag{B.6}$$

where the indices a and b run over the energy bins (each of 50 keV) and cores, the index n runs over the simulated spectra, S^n is the nth simulated spectrum, and S^0 is the nominal spectrum. The covariance matrix V_{ab} is then used in this analysis's toy Monte Carlo in order to fluctuate the toy reactor $\bar{\nu}_e$ spectra used in generating the signal systematic covariance matrix, as discussed in Section C.2.5.

B.5 Implementation details

This section describes some of the more technical implementation details of the reactor prediction. By convention, we define these spectra in terms of true neutrino energy, and omit any weighting by the IBD cross section. The spectra are divided into 220 bins, 50 keV wide, ranging from 1.85 to 12.8 MeV. To produce this prediction, a number of basic inputs are required:

- Weekly average thermal power and fission fractions for each reactor, as reported by the Reactor Working Group using data provided by the power company. The fission fractions are determined from simulations, as described in Appendix B.
- Weekly detector livetimes, in order to properly weight the spectra for each week. Given that we only use data where all three halls were operating (and passing data quality requirements) simultaneously, there is only a single value for each week, not three values.
- Nominal fission spectra, in 10 keV increments of E_ν from 1.8 to 13 MeV, and in units of $\bar{\nu}_e$ MeV^{-1} fission^{-1}. For ^{235}U, ^{239}Pu, and ^{241}Pu, Huber's spectra are used, whereas the French spectrum is used for ^{238}U. We rebin these into 50 keV bins by sampling the midpoints.

[5]The exact value of M is unspecified in [65] or in the associated code, but presumably it was large enough for convergence of the covariance matrix.

- Nominal corrections to account for the ILL measurements having been made out of equilibrium. These consist of five correction per isotope (none for ^{238}U) defined at evenly spaced E_ν points from 2 to 4 MeV (see Table B.2).
- A nominal spectrum for spent nuclear fuel (SNF).
- The IBD cross section as a function of E_ν. At this stage, this is only used in calculating the normalization of the SNF contribution, as described below.

B.5.1 Nominal spectra

For each core c, the first step is to sum over each week w and calculate the time-averaged nominal flux $F_{ci}^{\text{nom}}(E)$ from isotope i at energy E. By convention, this quantity is specified in units of 10^{18} $\bar{\nu}_e$ MeV^{-1} s^{-1}, and calculated as:

$$F_{ci}^{\text{nom}}(E) = \frac{S_i(E)}{N \sum_w T_w} \sum_w T_w f_{wci} R_{wc},$$

where $S_i(E)$ is the theoretical spectrum, in $\bar{\nu}_e$ MeV^{-1} fission^{-1},
$N \equiv 10^{18}$ is a normalization factor,
T_w is the weekly total detector livetime (i.e. weekly weighting factor), in days,
f_{wci} is the weekly average fission fraction of isotope i, and
R_{wc} is the weekly average fission rate, in s^{-1}.

In turn, the weekly fission rate R_{wc} is

$$R_{wc} = \frac{\overline{P} P_{wc}}{q \overline{E}_{wc}},$$

where $\overline{P}$ is the nominal core power, 2895 MW,
P_{wc} is the actual core power, as a fraction of $\overline{P}$,
q is 1.602×10^{-19} J/eV,
$\overline{E}_{wc} \equiv \sum_i f_{wci} E_i$ is the weekly average energy per fission, in MeV, and
E_i is the average energy per fission of isotope i, in MeV.

Thus, from data files containing T_w, P_{wc}, and f_{wci}, along with static definitions of $S_i(E)$ and E_i, the livetime-weighted average nominal flux emitted by each core is calculated.

B.5.2 Corrected spectra

B.5.2.1 Non-equilibrium correction

As discussed in Appendix B, the nominal spectra are derived from measurements taken with foils irradiated for a few dozen hours. Since, in such experiments, longer-lived fission fractions are not given enough time to reach their equilibrium concentrations, the measured spectra deviate slightly from those emitted by long-running nuclear reactors. Table B.2 shows the percentage corrections to the spectra of the three fissile isotopes, which were tabulated by Lewis [65] based on [86]. At energies between the five tabulated points, the corrections are linearly interpolated. Above 4.0 MeV, no correction is applied. Below 2.0 MeV, the corrections at 2.0 and 2.5 MeV are linearly extrapolated. This procedure results in a continuous, piecewise linear correction function, $C_i^{\mathrm{ne}}(E)$ for each isotope i. For U-238, the function is identically zero. Applying the correction and summing over isotopes then gives the intermediate result $F_c^{\mathrm{ne}}(E)$,

$$F_c^{\mathrm{ne}}(E) = \sum_i \Big(1 + C_i^{\mathrm{ne}}(E)\Big)\, F_{ci}^{\mathrm{nom}}(E).$$

E (MeV)	2.0	2.5	3.0	3.5	4.0
U-235	5.7	4.4	1.5	0.7	0.1
Pu-239	2.1	1.7	0.5	0.0	0.0
Pu-241	1.9	1.5	0.5	0.0	0.0

Table B.2: Non-equilibrium corrections to antineutrino spectra, in percentage terms. The corrections are linearly interpolated when applied at intermediate energies. No correction is defined for U-238.

B.5.2.2 Spent nuclear fuel

An additional term must be added to the reactor prediction due to the presence of spent nuclear fuel in storage pools near the reactor cores. Based on studies described in [65], it was determined that a fraction R^{snf} of 0.3% of the *total, cross-section weighted* antineutrino flux must come from spent nuclear fuel. The uncertainty on this percentage is unspecified (FIXME?), but the oscillation analysis is largely insensitive even to large errors in the reactor prediction, so the impact of the SNF uncertainty is essentially negligible. A core-dependent SNF spectrum $S_c^{\mathrm{snf}}(E)$ (also described in [65]), with arbitrary normalization, is added to $F^{\mathrm{ne}}(E)$ (after non-equilibrium correction), subject to the aforementioned constraint on the measured rate. This

constraint determines the normalization factor A_c^{snf},

$$A_c^{\mathrm{snf}} = \frac{\frac{1}{6}\sum_{c'} \int F_{c'}^{\mathrm{ne}}(E)\,\sigma(E)\,dE}{\int S_c^{\mathrm{snf}}(E)\,\sigma(E)\,dE}\; R^{\mathrm{snf}},$$

where $\sigma(E)$ is the IBD cross section and the integrals are understood to represent sums over binned spectra. The factor of 1/6 is included because R^{snf} is defined with respect to the *total* flux, while A_c^{snf} is specific to the core c. Essentially, this procedure divides the total integrated SNF flux evenly among the six cores, while still allowing for a different shape in each core. In the current implementation, a single shape is used for all six cores.

At this point, the final, fully-corrected reactor prediction can be written simply as

$$F_c(E) = F_c^{\mathrm{ne}}(E) + A_c^{\mathrm{snf}}\, S_c^{\mathrm{snf}}(E) \tag{B.7}$$

It is this $F_c(E)$ that is fed into the fitter (for extrapolation) and the toy MC. The spectra are specified in 220 bins (50-keV wide) of E_ν from 1.85 to 12.8 MeV, in units of 10^{18} $\bar{\nu}_e$ MeV^{-1} s^{-1}, with one such spectrum per core per data period (6AD, 8AD, or 7AD). The livetime weighting uses the average weekly livetime across the three halls, rather than treating each hall individually; this is valid, since the data sample only includes periods where all three halls were running.

Appendix C

Fitting details

C.1 Overview

In order to extract neutrino oscillation parameters, Daya Bay data is compared to the predictions associated with different parameter values, and the extracted parameters are then those that give the best fit to the data. Given knowledge of the reactor $\bar{\nu}_e$ flux, detector response, and expected backgrounds, it is conceptually straightforward to generate a set of predictions. However, calculating the goodness of fit (while properly accounting for systematic and statistical uncertainties), and then assigning errors to the extracted parameters, is where the procedure becomes more subtle and complex. In Daya Bay, separate analysis groups have historically employed two different approaches, theoretically equivalent but implemented very differently. These are the method of pull terms (a.k.a. nuisance parameters), and the covariance matrix approach. In this analysis, we use the latter, but both will be briefly described in this introductory section. In the rest of this chapter, we detail the fitting machinery developed at LBNL [56, 62].

C.1.1 Method of pull terms

In the method using pull terms, the fitter is "smart" in the sense that it has knowledge of the various underlying models (reactor $\bar{\nu}_e$ flux, detector response, backgrounds, etc.) and knows how their predictions will change under changes of the systematic uncertainties. In this approach, each systematic is represented by a *pull term* (or *nuisance parameter*), which is in turn assigned an uncertainty of its own. An example of such a pull term might be the relative energy scale of a given AD. Each pull term is assigned a nominal value, corresponding to our best estimate given available knowledge. Then, during the fit, not only are the oscillation parameters varied, but so are the pull terms, and the predictions are transformed accordingly. The total χ^2

then takes a form similar to

$$\chi^2 = \sum_i \frac{(x_i - \bar{x}_i)^2}{\sigma_i^2} + \sum_j \frac{(\eta_j - \bar{\eta}_j)^2}{\varsigma_j^2}, \tag{C.1}$$

where x_i are the measured data (e.g., AD spectra), $\bar{x}_i$ are the predictions (which vary as we scan the oscillation parameters and pull terms), σ_i are the *statistical* uncertainties on the data, η_j are the pull terms, $\bar{\eta}_j$ are their nominal values, and ς_j are the uncertainties on the pulls.

Fitting is complete when the fitter has found the values of the oscillation parameters *and pull terms* that minimize the total χ^2. The 1σ errors on the oscillation parameters are then based on the amount of variation required to increase the reduced[1] χ^2 by 1 unit (or 2.3 for a 2D fit,[2] etc.), while minimizing over the pull terms at every step. Correlations between energy bins are handled implicitly; the information is encoded in how the predictions in different bins are shifted together when the pull terms are varied.

C.1.2 Covariance matrix approach

As an alternative to using pull terms, uncertainties and correlations can be encoded in a single covariance matrix generated using Monte Carlo techniques. In this approach, the fitter is relatively "dumb": It knows only how to generate a prediction using a *nominal* model (of, again, reactors, backgrounds, detectors, etc.) and how to vary the prediction for different values of the oscillation parameters. It has no idea how the prediction will transform under varying assumptions with respect to systematic uncertainties. (This knowledge belongs to the Monte Carlo.) The fitter's job is simply to take the measurements x_i, the predictions $\bar{x}_i$ (which vary according to the oscillation parameters), and the covariance matrix V_{ij}, and then to find the oscillation parameters which minimize the χ^2,

$$\chi^2 = (x_i - \bar{x}_i) V_{ij}^{-1} (x_j - \bar{x}_j). \tag{C.2}$$

Here we note that this is a linearized model, and thus, χ^2 will precisely follow the standard χ^2 distribution only if the x_i are all Gaussian-distributed. Any non-Gaussianity could not only distort the interpretation of χ^2 as a measure of goodness-of-fit, but could also bias any confidence intervals calculated from this model. However, given the ample statistics in each energy bin and the large number of independent systematics, Gaussianity is a reasonable assumption. It is further justified, in this analysis, by the fact that we obtain a χ^2 (per degree of freedom) of approximately unity (Section 8.5), indicating a high goodness-of-fit. This implies that the differences between the x_i and the $\bar{x}_i$ are generally small at the best-fit point, minimizing the effects of any non-Gaussian tails in the x_i distributions.

[1] We use the term "reduced χ^2" to refer to the χ^2 divided by the number of degrees of freedom.

[2] Based on the fact that $(1/2\pi) \int_{\theta=0}^{2\pi} \int_{r=0}^{r_0} e^{-r^2/2}\, r\, dr\, d\theta = 0.68$ when $r_0 = 2.3$.

A major advantage of the covariance matrix approach is that it greatly reduces the dimensionality of the minimization—instead of minimizing over potentially dozens of pulls (as well as the oscillation parameters), only the two oscillation parameters must be considered. The computational cost and risk of falling into false minima are thus avoided. A disadvantage is the need to possibly take extra steps in order to ensure an invertible covariance matrix (see Section C.3.3), and the need to carefully treat correlations, which are implicitly accounted for when using pull terms.

In the LBNL fitter, the full NuWa-based Monte Carlo is not used for generating the covariance matrix, due to its complexity and computational cost. Instead, a "toy" MC, described in Section C.2, was developed for this purpose.

C.1.3 Relative or absolute?

In addition to the choice between using pull terms or a covariance matrix (or some hybrid of the two), there is also the orthogonal option of performing a relative or an absolute fit. In a relative fit, the near-site data is used to generate predictions for the far sites, which are then compared to the far-site data. Whereas in an absolute fit, predictions are generated for both the near and the far sites, for comparison to both the near and far observations. The main advantage of performing a relative fit is that the absolute detection efficiency (both its nominal value and its uncertainty) can be ignored completely, since it has no bearing on the process of near-to-far extrapolation. On the other hand, a relative fit requires dealing with multiple correlated predictions and/or the merging of data between ADs, and implementing it in the context of a pull-term fitter is not as straightforward as doing an absolute fit. In any case, both approaches should ultimately give the same results. The LBNL fitter takes the relative approach, as discussed further in Section C.3.

C.2 Toy Monte Carlo

The primary purpose of the LBNL toy MC [56] is to enable the computation of the covariance matrix V_{ij} in Equation C.2. It does so by generating a large collection of simulated "toy" experiments, with each one represented by the prompt spectrum measured by each AD. The covariance matrix can then be derived from the variations within the collection, as detailed in Section C.2.5. Each toy experiment may include fluctuations due to statistics and/or a chosen set of systematics.[3] As detailed later, a given systematic is incorporated by, first, describing it (analogously to a pull term) by a parameter (such as the detection efficiency), and then fluctuating the parameter; the value of the parameter is then used by the toy MC's model for generating the

[3]In practice, statistical fluctuations are disabled, since the statistical covariance matrix is calculated analytically as described in Section C.3.2.

prompt spectra. This basic functionality enables the production of three essential inputs used by the fitter:

- The covariance matrices for signal and background systematics, used in calculating the χ^2 during the fit. As will be explained, the signal covariance matrix is rescaled according to the signal size, and then added to the background covariance matrix to produce the full systematic matrix, which is finally added to the statistical covariance matrix (generated on the fly by the fitter).
- The "super histograms": The nominal (i.e. non-fluctuated) predicted cross section-weighted antineutrino spectrum produced by each core. This is used when breaking down the near-site spectra into reactor contributions for extrapolation to the far site.
- The conversion matrix between prompt energy E_{rec} and true antineutrino energy E_ν; this is also used in the extrapolation.

In addition, the toy MC provides a method of validating the fitter, since toys can be generated for any chosen values of θ_{13} and Δm^2_{ee}, thus enabling the testing of the fitter's ability to recover the same values.

C.2.1 Binning

In the fitter and toy MC, two (somewhat arbitrary) options are available for the reconstructed energy binning:

1. "LBNL" binning: One bin from 0.7 to 1 MeV, 35 bins of 0.2 MeV from 1 to 8 MeV, and one bin from 8 to 12 MeV, for a total of 37 bins.
2. "BCW" binning: One bin from 0.7 to 1.3 MeV, 24 bins of 0.25 MeV from 1.3 to 7.3 MeV, and one bin from 7.3 to 12 MeV, for a total of 26 bins.

The spectrum below 1.3 MeV is particularly susceptible to distortion from the IAV effect. In turn, if the IAV effect (Section C.2.3.2) is imperfectly modeled, this can lead to a $\sim 1\sigma$ bias in the extraction of Δm^2_{ee}, as was indeed observed during comparisons of the results between the independent analysis groups [19]. The use of a coarser binning at low energies thus removes any sensitivity to this distortion. For this reason, we employ the BCW binning in this analysis. A comparison of the oscillation fit between the two binnings can be seen in Figure 9.16.

C.2.2 IBD spectrum at each AD

The toy MC chain begins with the prediction of the livetime-averaged antineutrino spectrum $F_c(E_\nu)$ produced (per unit time) at each reactor core c, as specified by Equation B.7.[4] This is then used to calculate the IBD spectrum at each AD i as

$$R_i(E_\nu) = T_i\, N_i\, \epsilon_i\, \sigma(E_\nu) \sum_c F_c(E_\nu) \frac{1}{4\pi L_{ci}^2} P_{\text{osc}}(E_\nu, L_{ci}) \tag{C.3}$$

where T_i is the livetime, N_i is the number of target protons, and L_{ci} is the baseline.

The detection efficiency ϵ_i includes the calculated efficiencies of the muon veto and the multiplicity cut (both of which are considered to have negligible uncertainty), as well as an additional factor which accounts for all of the remaining efficiency components. This factor is identical for all ADs in the nominal case, but when allowing fluctuations (as when generating toy samples for the construction of the covariance matrix), the efficiency is assigned an AD-to-AD uncorrelated uncertainty of $0.11\% \oplus 0.072\%$ (Section D.1), with the latter component fully correlated with the variation in the energy scale of the AD.

The cross section $\sigma(E_\nu)$ is calculated by performing a 4π numerical integration of the differential cross section (to first order in the inverse nucleon mass $1/M$) [87]. Letting θ be the angle between the antineutrino and the positron in the lab frame (with the proton initially at rest),

$$\frac{d\sigma}{d\cos\theta} = \frac{\sigma_0}{2}\left[(f^2+3g^2) + (f^2-g^2)v_e^{(1)}\cos\theta\right] E_e^{(1)} p_e^{(1)} - \frac{\sigma_0}{2}\left[\frac{\Gamma}{M}\right] E_e^{(0)} p_e^{(0)}, \tag{C.4}$$

where

$$\begin{aligned}
\Gamma = {} & 2(f+f_2)g\left[(2E_e^{(0)}+\Delta)(1-v_e^{(0)}\cos\theta) - \frac{m_e^2}{E_e^{(0)}}\right] \\
& + (f^2+g^2)\left[\Delta(1+v_e^{(0)}\cos\theta) + \frac{m_e^2}{E_e^{(0)}}\right] \\
& + (f^2+3g^2)\left[(E_e^{(0)}+\Delta)(1-\frac{1}{v_e^{(0)}}\cos\theta) - \Delta\right] \\
& + (f^2-g^2)\left[(E_e^{(0)}+\Delta)(1-\frac{1}{v_e^{(0)}}\cos\theta) - \Delta\right] v_e^{(0)}\cos\theta,
\end{aligned} \tag{C.5}$$

[4] The technical details of this prediction are described in Section B.5, and the underlying theory is discussed in Appendix B.

and

$$\sigma_0 = \frac{G_F^2 \cos^2\theta_C}{\pi}(1+\Delta_{\text{inner}}^R), \tag{C.6}$$

$$E_e^{(1)} = E_e^{(0)}\left[1 - \frac{E_\nu}{M}(1 - v_e^{(0)}\cos\theta)\right] - \frac{y^2}{M}, \tag{C.7}$$

$$E_e^{(0)} = E_\nu - \Delta, \tag{C.8}$$

$$v_e^{(i)} = \frac{p_e^{(i)}}{E_e^{(i)}}, \tag{C.9}$$

$$\Delta = M_n - M_p, \tag{C.10}$$

$$M = \frac{M_n + M_p}{2}, \tag{C.11}$$

$$y^2 = \frac{\Delta^2 - m_e^2}{2}. \tag{C.12}$$

The various constants that enter this cross section are taken from [88] (unless noted otherwise) and summarized in Table C.1.[5]

Symbol	Description	Value
G_F	Fermi coupling constant	1.16637×10^{-5} $(\hbar c)^3$ GeV^{-2}
$\hbar c$	Planck constant times speed of light	197.3269631 MeV fm
f	Vector coupling g_V/g_V	1
g	Axial coupling g_A/g_V	1.2701
f_2	Anomalous isovector magnetic moment $\mu_p - \mu_n$	3.70589
$\cos\theta_C$	Cabbibo angle	0.974
Δ_{inner}^R	Inner radiative correction [89]	0.024
M_p	Proton mass	938.272013 MeV
M_n	Neutron mass	939.565345 MeV
m_e	Electron mass	0.51099891 MeV

Table C.1: Constants used in the evaluation of Equation C.4 [88].

Returning to Equation C.3, P_{osc} is calculated according to Equation 1.17, which of course depends on the oscillation parameters.[6] Table C.2 summarizes the nominal

[5] It is often stated (e.g., in [25]) that the cross section is calculated using the neutron lifetime τ_n from a particular edition of the PDG tables. However, τ_n can only be directly used when calculating σ to zeroth-order in $1/M$ [87]. In the first-order calculation Equation C.4, τ_n does not appear explicitly. Nevertheless, other constants, such as g, are connected to τ_n and are thus modified by updated measurements of the latter.

[6] By default, the toy MC assumes the normal hierarchy when using Equation 1.18 to convert Δm^2_{ee} into Δm^2_{32} and Δm^2_{31} for insertion into Equation 1.17.

oscillation parameters used in generating the covariance and response matrices. Other values of the oscillation parameters can be used for benchmarking the fitter, e.g., verifying that it recovers the parameters used by the toy MC.

Parameter	Value
$\sin^2 2\theta_{13}$	0.0850
Δm^2_{ee}	2.5010×10^{-3} eV^2
$\sin^2 2\theta_{12}$	0.851 ± 0.021
Δm^2_{21}	$(7.53 \pm 0.18) \times 10^{-5}$ eV^2

Table C.2: Nominal oscillation parameters used by the toy Monte Carlo. For the solar parameters ($\sin^2 2\theta_{12}$ and Δm^2_{21}), the indicated errors are used for generating random fluctuations about the nominal values.

C.2.3 Detector response

Once the toy MC has determined each AD's IBD spectrum (in terms of the antineutrino energy), the next step is to convert it into a prompt-energy spectrum. This entails four steps:

1. Converting antineutrino energy into positron energy
2. Accounting for the "loss" of visible energy incurred when positrons deposit some of their energy in the acrylic wall of the IAV
3. Converting scintillator-deposited energy into mean visible energy, according to the absolute energy scale and the nonlinearity model
4. Smearing the visible energy according to the resolution of the AD

C.2.3.1 Positron energy

As with the IBD cross section discussed previously, the positron energy is calculated to first order in the inverse nucleon mass scale $1/M$. Equation C.7 gives the first-order positron energy $E_e^{(1)}$ as a function of the lab-frame antineutrino-positron angle θ. Meanwhile, the mean value of $\cos\theta$, weighted by the differential cross section Equation C.4, is [87]

$$\langle \cos\theta \rangle = -0.034\, v_e^{(0)} + 2.4 \frac{E_\nu}{M}. \tag{C.13}$$

Since $E_e^{(1)}$ is a linear function of $\cos\theta$, the mean positron energy is then simply obtained by inserting $\langle\cos\theta\rangle$ into Equation C.7 in place of $\cos\theta$:

$$\langle E_e^{(1)}\rangle = E_e^{(0)}\left\{1 - \frac{E_\nu}{M}\left[1 - v_e^{(0)}\left(-0.034\, v_e^{(0)} + 2.4\frac{E_\nu}{M}\right)\right]\right\} - \frac{y^2}{M}, \tag{C.14}$$

where $E_e^{(0)}$, M, etc. were defined in Section C.2.2.

Later, in Section C.2.3.4, we show that it is safe to neglect the energy spread caused by the angular distribution, as it is negligible in comparison to the dominant factors in the energy resolution (primarily photon statistics, as well as detector nonuniformity and noise). Likewise, there is no need to extend the calculation to higher order, given that the neutron carries away only $\mathcal{O}(10\text{ keV})$ of kinetic energy.

C.2.3.2 IAV effect

For IBDs that occur near the edge of the inner acrylic vessel, some of the positron's kinetic energy may be deposited in the acrylic, rendering that energy invisible.[7] In order to model this effect, it was simulated in the full Daya Bay MC [90], producing a matrix $\mathbf{M}_{ij}^{\text{IAV}}$ which converts "true" positron energy E_e^{true} into "LS-deposited" positron energy E_e^{LS},

$$E_{e,i}^{\text{LS}} = \mathbf{M}_{ij}^{\text{IAV}} E_{e,j}^{\text{true}} \tag{C.15}$$

where i and j are bin indices, and $\mathbf{M}^{\text{IAV}}$ is subject to the normalization condition $\sum_i \mathbf{M}_{ij}^{\text{IAV}} = 1$ for all true energy bins j (i.e., the conversion preserves the total antineutrino count, as it should). Each axis of $\mathbf{M}^{\text{IAV}}$ contains 240 bins of width 50 keV ranging from 0 to 12 MeV.

The uncertainty of the IAV wall thickness is assigned a conservative 4%, which is assumed to translate to a 4% uncertainty on the elements of $\mathbf{M}^{\text{IAV}}$. As implemented, the matrix is fluctuated by applying an independent 4% Gaussian variation to each off-diagonal element, and then setting the diagonal elements so as to restore the normalization condition.

C.2.3.3 Positron to mean reconstructed energy

As discussed in Section 4.2.4, the reconstructed energy is affected by nonlinearity in the scintillator (quenching, Cherenkov radiation) and in the electronics. Our analysis makes use of the "unified" nonlinearity model described in [91]. The unified nonlinearity model takes the form of a nominal curve (tabulating the ratio of reconstructed to positron energy, as a function of positron energy), along with four pull curves ([92], p10) that express the uncertainty in the model. The nominal curve and the ±68% CL bands were shown in Figure 4.4. All curves (nominal and pulls) were generating

[7] Some of the energy from the annihilation gammas can also disappear in this way. The Daya Bay MC accounts for this.

using the full Daya Bay MC, as controlled by five parameters (absolute energy scale, Birks/Cherenkov constants, electronics parameters). The curve that best fit a collection of source/^{12}B data was designated the nominal curve. Meanwhile, among 250 random curves within 68% CL, the pull curves were taken as the four that best spanned the remaining 246. The four pulls thus account for the correlations inherent in the shape uncertainty, without requiring the use of a full covariance matrix during analysis. Figure C.1 shows the nominal curve along with the four pulls. The model parameters for these five curves were never published, and cannot easily be recovered from a fit (since the Cherenkov nonlinearity function f_c was tabulated numerically and was also never published). Instead, the author and other Daya Bay analyzers were simply provided with tabulated values of the five curves for use in analysis.

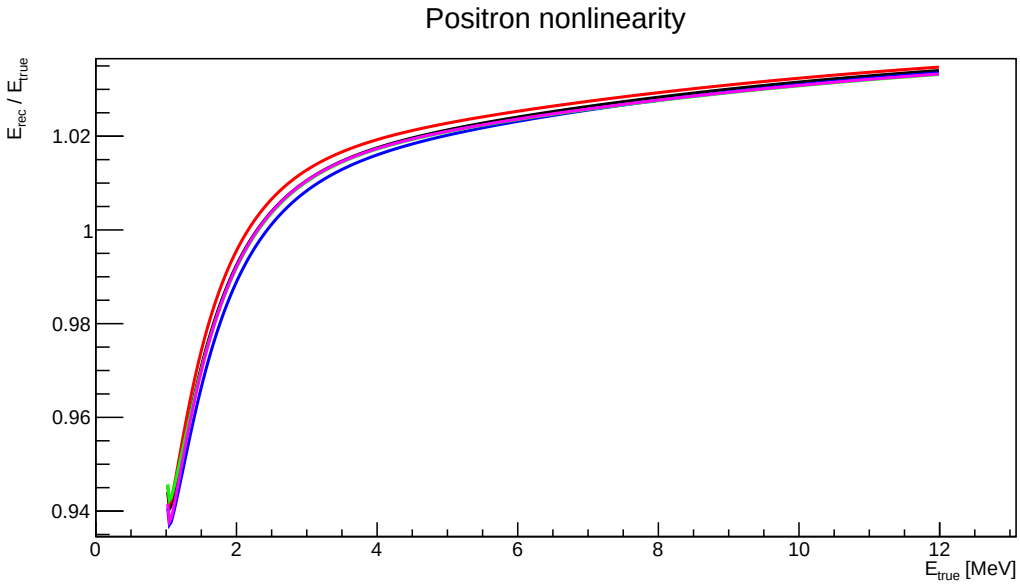

Figure C.1: The nominal curve (black) from the unified nonlinearity model, along with the four pull curves (red, blue, magenta, green). The green curve only differs noticeably from the nominal curve at the lowest energies.

Based on the unified model, a nonlinearity curve is generated according to

$$f_{\text{ran}}(E_e) = f_{\text{nom}}(E_e) + \sum_p^4 a_p[f_p(E_e) - f_{\text{nom}}(E_e)], \tag{C.16}$$

where E_e is the positron energy, and the a_p are standard Gaussian random variables. (We will continue using a to denote such variables.) The same curve is used in all detectors. However, where the ADs *can* differ is in their overall energy scales. The relative energy scale uncertainty has been estimated to be 0.2% [93]. Accordingly, for

each toy, and each AD d, a scaling factor k_d is calculated as

$$k_d = 1 + a_d \cdot 0.002. \tag{C.17}$$

The final relationship between the positron and reconstructed energy, for AD d, is then

$$E_{\text{rec}} = F_d(E_e) = k_d \cdot f_{\text{ran}}(E_e) \cdot E_e. \tag{C.18}$$

Using this relationship, the number of events $N_{\text{rec},i}$ in the ith bin of reconstructed energy (centered at $E_{\text{rec},i}$) is calculated as

$$N_{\text{rec},i}(E_{\text{rec},i}) = N_{e,j} \left(\frac{dF}{dE_e}\Big|_{E_{e,j}} \right)^{-1} \frac{\Delta_{\text{rec}}}{\Delta_{\text{e}}} \tag{C.19}$$

where the index j is such that

$$F(E_{e,j}) = E_{\text{rec},i}, \tag{C.20}$$

and the Δ are the bin widths.

C.2.3.4 Detector resolution

As described in [19], the Daya Bay detector resolution has been characterized as taking the form

$$\frac{\sigma(E_{\text{rec}})}{E_{\text{rec}}} = \sqrt{(1.6\%)^2 + \frac{(8.1\%)^2}{E_{\text{rec}}} + \frac{(2.6\%)^2}{E_{\text{rec}}^2}}, \tag{C.21}$$

where the three terms correspond to detector nonuniformity, photoelectron statistics, and noise, respectively. In Figure C.2 we plot this function, along with its uncertainty bands (discussed later). The coefficients were determined by performing a fit to the gamma-ray peaks from various calibration sources and neutron captures, as well as to the alpha-particle peaks from radioactivity [19]. This model does not account for the intrinsic energy spread of the IBD positron due to its angular distribution. We proceed to justify this omission, first heuristically and then quantitatively.

From Equation C.7, it can be seen that the first-order angular dependence of the positron energy takes the form

$$\delta E = E_e^{(0)} \frac{E_\nu}{M} v_e^{(0)} \cos\theta. \tag{C.22}$$

We aim to determine the worst-case energy spread, so we may conservatively let $v_e^{(0)} = 1$. Now, for the highest antineutrino energy observed at Daya Bay, around 9 MeV, we have $E_\nu/M \approx 1\%$. Then, since $\cos\theta$ ranges from -1 to 1, we have

$$\frac{\max(\delta E) - \min(\delta E)}{E_e^{(0)}} \approx 2\% \left(\frac{E_\nu}{9\,\text{MeV}} \right). \tag{C.23}$$

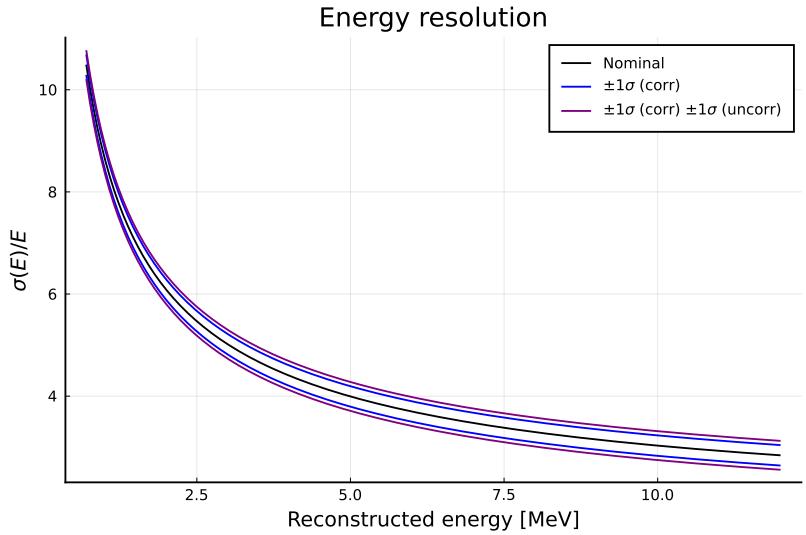

Figure C.2: The energy resolution model (Equation C.21) and its uncertainty bands (AD-correlated and total).

That is, the *total spread* of the positron energies is, in the worst case ($E_\nu = 9\,\text{MeV}$), 2%, and gets linearly smaller for lower antineutrino energies. Translating this 2% total spread into a standard deviation, for comparison with the resolution model, requires knowledge of the shape of the energy distribution. As we show later, however, this distribution is reasonably flat, so for the sake of this qualitative argument, we will assume a uniform energy distribution. For a uniform distribution of total spread S, the standard deviation σ is given by[8]

$$\sigma = \frac{S}{2\sqrt{3}}. \tag{C.24}$$

For the positron energy distribution, this gives

$$\begin{aligned} \sigma &\approx \frac{2\%}{2\sqrt{3}} \left(\frac{E_\nu}{9\,\text{MeV}} \right) \\ &= 0.6\% \left(\frac{E_\nu}{9\,\text{MeV}} \right). \end{aligned} \tag{C.25}$$

The question now is whether this kinematic energy spread is significant in comparison to the modeled energy resolution in Equation C.21. From Figure C.2 it is clear that we need only consider the highest-energy case, since at lower energies, the fractional kinematic spread gets smaller while the modeled resolution gets larger. According to

[8] In what follows, σ (without a derivative symbol) will always refer to a standard deviation, while cross-sections will always be treated in differential form, $d\sigma$, thus avoiding any ambiguity.

Figure C.2, the modeled resolution levels off at around 4%. Adding the 0.6% kinematic spread in quadrature then increases the total resolution to 4.05%. Thus, heuristically, the effect is negligible.

To address the issue more rigorously, we can directly calculate the standard deviation of the positron energy by using the differential cross-section (Equation C.4) and the relation between the first-order energy $E_e^{(1)}$ and the scattering angle (Equation C.7). In Figure C.3, we normalize $d\sigma/d\cos\theta$ to obtain the probability distribution function (PDF) of $\cos\theta$, plotted for various values of the antineutrino energy E_ν. The PDFs are relatively flat, as was claimed earlier, with a modest bias toward backscattering. In order to visualize the energy spread that arises from the angular distribution, in Figure C.4 we plot the *positron energy deficit* $E_\nu - E_e^{(1)}$ (i.e., the amount of energy transferred to the hadronic system) as a function of $\cos\theta$ for various values of E_ν. All of the curves approximately intersect in the forward-scattering limit, where the neutron kinetic energy is minimal and the deficit is approximately the neutron-proton mass difference of 1.29 MeV. As expected, the deficit increases for harder scattering angles and higher E_ν as more kinetic energy is transferred to the neutron. For the case of $E_\nu = 9\,\text{MeV}$, the difference between the maximum and minimum $E_e^{(1)}$ is 0.15 MeV. Relative to $E_e^{(1)}$ ($\approx E_\nu - 1.3 = 7.7\,\text{MeV}$), the size of this spread is 1.9%, consistent with the 2% we estimated earlier.

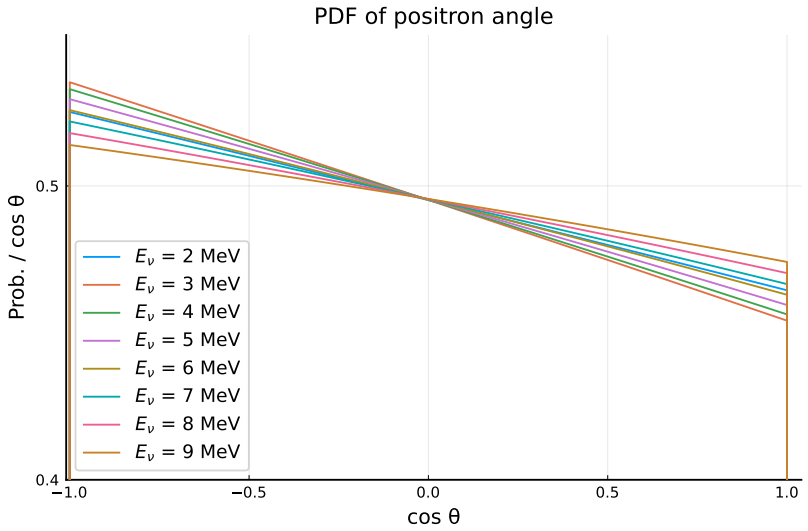

Figure C.3: Probability distribution functions of the scattering angle for IBD positrons.

To go further and directly calculate the standard deviation of the positron energy, we require the PDF of the latter, which can be obtained by normalizing the energy-

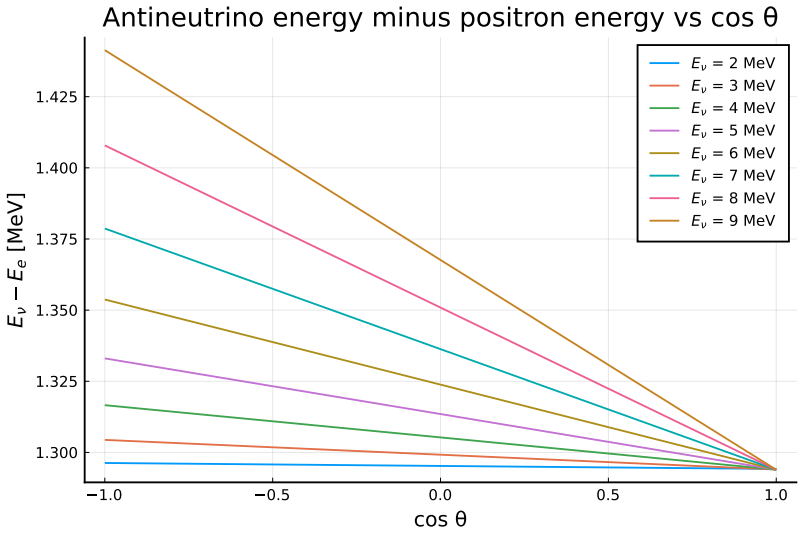

Figure C.4: Dependence of the positron energy deficit (i.e. the amount of energy transferred to the hadronic system) on the scattering angle.

differential cross-section

$$\frac{d\sigma}{dE_e^{(1)}} = \frac{d\sigma}{d\cos\theta}\left(\frac{dE_e^{(1)}}{d\cos\theta}\right)^{-1}, \tag{C.26}$$

where the derivative $dE_e^{(1)}/d\cos\theta$ is evaluated at the angle θ' corresponding to energy $E_e^{(1)}$. Operationally, we determine θ' by numerically inverting Equation C.7. In Figure C.5 we plot the PDFs (for various values of E_ν) as a function of the *percent deviation of the positron energy from the mean* (where the mean is individually computed for each value of E_ν). As was claimed earlier, these distributions are relatively flat. As expected, they are broader for higher values of E_ν, and there is a slight bias toward lower energies (i.e. backscattering).

At this point, all that remains is to take the standard deviations σ of these energy PDFs. The resulting values are in percentage terms, which can be directly compared to the modeled resolution (Equation C.21 and Figure C.2) in order to determine the significance of this energy spread. In fig Figure C.6, we plot this σ as a function of E_ν. The results are in excellent agreement with our previous estimate: In the worst case (at 9 MeV), σ is only about 0.6%. As was argued earlier, this is insignificant relative to the energy resolution model.[9]

In the official Daya Bay publications (e.g., [19]), the uncertainty of the resolution has been declared to be negligible. However, we apply the conservative uncertainties

[9]Since some of the reconstructed energy comes from the annihilation electron, which is not subject to this spread, this analysis overestimates the size of the effect, further highlighting its insignificance.

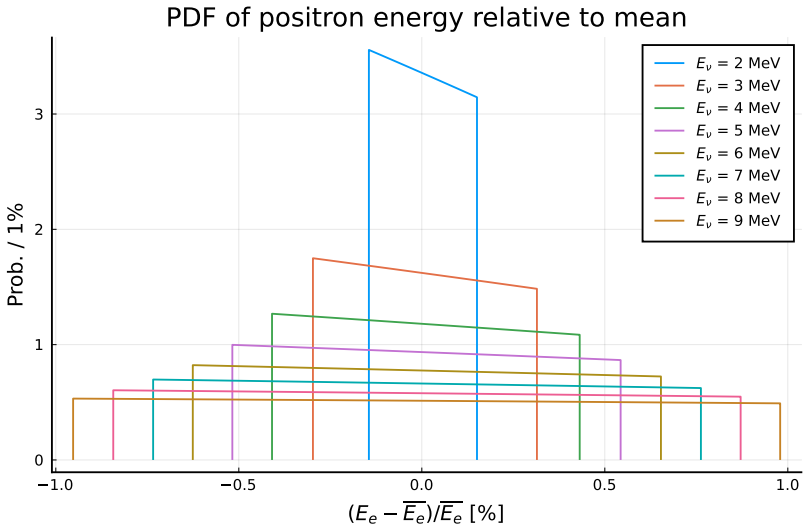

Figure C.5: Probability distribution functions of the positron energy (expressed as percentage deviations from the mean energy for fixed E_ν).

originally employed by the LBNL analysis: 0.2% correlated and 0.2% uncorrelated. The correlated component is due to the statistical uncertainty of the (old) resolution fits (with limited statistics), while the uncorrelated part is attributed to the observed 0.2% AD-to-AD differences in the reconstructed energy [56]. Resolution fluctuations are applied as a constant shift to the fractional uncertainty. As such, for detector d, σ/E_{rec} is fluctuated according to

$$\frac{\sigma_{\text{ran},d}(E_{\text{rec}})}{E_{\text{rec}}} = \frac{\sigma(E_{\text{rec}})}{E_{\text{rec}}} + (0.2\% \times a_{\text{corr}}) + (0.2\% \times a_d). \tag{C.27}$$

Application of the detector resolution begins with the "source" E_{rec} histogram produced according to Equation C.19. An identically-shaped (but empty) "destination" histogram is then constructed, and for each destination bin centered at $E_{\text{rec},i}$, we loop over the "source" bins that span $E_{\text{rec}i} \pm 8\,\sigma(E_{\text{rec},i})$,[10] and increase the count in the destination bin by the contents of the source bin, times the appropriate Gaussian factor. This produces the final, smeared, IBD spectrum.

C.2.4 Backgrounds

The toy MC's treatment of backgrounds is based on the discussion in Chapter 6. Here we briefly review the determination of the rate and shape (and their respective uncertainties) of each background. Fluctuations may be applied differently for different

[10] 8 standard deviations was chosen in order to ensure "almost all" relevant events would be included. In practice, 4 or 5 standard deviations would have been more than sufficient.

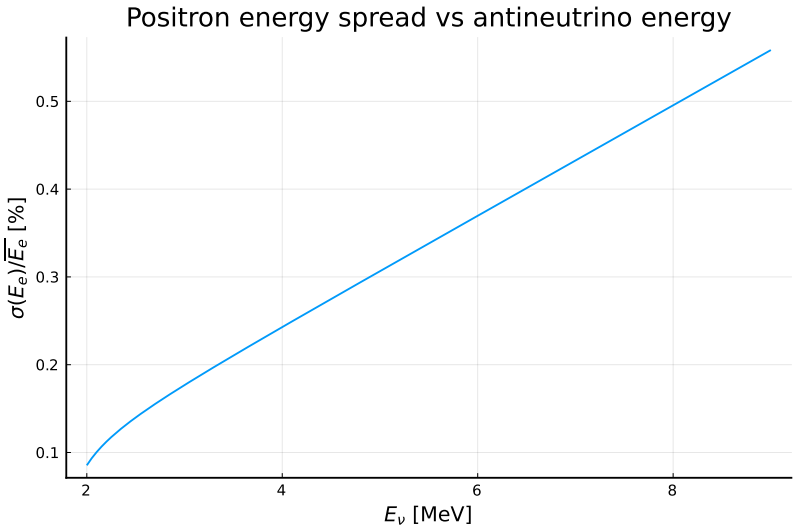

Figure C.6: Standard deviation of the positron energy (due to its angular distribution) as a function of antineutrino energy.

backgrounds; for instance, for some backgrounds each AD is fluctuated independently, while for others the fluctuations are correlated. Such details will also be covered in what follows. In all cases, the rate and shape uncertainties are decoupled; that is, shape fluctuations are implemented by distorting the spectrum and then renormalizing it to its previous integral, while rate fluctuations involve uniformly scaling all bins by the same factor.

C.2.4.1 Accidentals

The accidental-background rate is calculated from data per Section 7.7, and the shape is likewise extracted from data as described in Section 5.3 and Section 7.2. The uncertainty assessment is described in Section 7.7.2 (statistical) and Section 7.7.3 (systematic); in total, a practically negligible (yet conservative) 2% total uncertainty is assigned to the rate, the same as for the accidentals rate in data.

Rate fluctuations are applied independently for each AD. No shape uncertainty is assigned, given the substantial statistics of the background sample and the fact that the uncertainty in each energy bin is more than covered by the overall rate uncertainty.

C.2.4.2 ^{9}Li/^{8}He

The calculation of the ^{9}Li/^{8}He rate is described in Section 6.3, as is the prediction of the spectrum, which assumes a 5.5% proportion of ^{8}He. Given the limited statistics of the ^{9}Li sample, it is not possible to measure statistically significant differences in the ^{9}Li rate within a given hall. Accordingly, all ADs in a hall are treated identically.

Rate fluctuations are thus applied to each *hall* independently. In order to enable fluctuation of the shape, a sample of 250 random ^{9}Li/^{8}He spectra were produced under 100% variations of the alpha-particle and neutron quenching factors. (The ^{8}He proportion was not varied during this process, since its impact is small compared to that of the quenching factors.) For each fluctuated toy sample, a spectrum is chosen at random from this set. The same random spectrum is used among all ADs.

C.2.4.3 Fast neutrons

The fast neutron rate and shape were estimated using a number of techniques, as described in Section A.2. As with ^{9}Li/^{8}He, rate fluctuations are applied on a per-hall basis. The nominal shapes take the form of Equation A.11, with the parameter E_0 specified for each hall, and a fixed to be zero.[11] based on fits to data samples enriched in fast neutrons. For shape fluctuations, the key consideration is the fact that most of the shape uncertainty lies at low energies. Accordingly, for each toy sample, we first generate a Gaussian random variable k:

$$k = \text{Gaus}(0, 1), \tag{C.28}$$

and then an empirical distortion factor is generated:

$$y(E) = k \cdot E^{-0.1} + c, \tag{C.29}$$

where c is determined such that that $y = 1$ (i.e. no distortion) at 12 MeV:

$$c = 1 - k \cdot E^{-0.1}. \tag{C.30}$$

When $k = \pm 1$, the size of this distortion is 25% at 0.7 MeV. The distorted shape is obtained by multiplying the nominal spectrum by $y(E)$. The same distorted spectrum is used for all ADs.

C.2.4.4 AmC

As described in Section A.3, the AmC rates and a nominal measured spectrum from the high-activity AmC source (HAS) were obtained from studies within the collaboration [94]. The ensuing uncertainty was deemed to be dominated by potential biases in the procedure itself (i.e. MC simulations and measurement of a high-activity source), rather than by statistics or AD-to-AD variations. As such, both rate and shape fluctuations are applied consistently among all ADs. The nominal spectrum was obtained by fitting the HAS spectrum to an exponential according to Equation A.16. Shape fluctuations are then implemented by varying the exponential's slope by ±0.15%, based on the recommendation of studies internal to the collaboration.

[11] As noted in Section A.2.3, it has been shown that there is no significant effect from allowing a nonzero a when fitting Equation A.11 to the fast-neutron spectrum [58].

C.2.4.5 $^{13}C(\alpha, n)^{16}O$

The rate and its uncertainty for the $^{13}C(\alpha, n)^{16}O$ background are described in Section A.4. Rate fluctuations are applied independently among all ADs. No shape fluctuations are applied, since this background is very small and the rate uncertainty is very conservative.

C.2.5 Outputs

C.2.5.1 "Super histograms"

The so-called *super histograms* S_c are essentially the cross section-weighted antineutrino spectra produced by each core, with some arbitrary (but consistent) normalization:

$$S_c(E_\nu) \propto \sigma(E_\nu)\, F_c(E_\nu). \tag{C.31}$$

The calculation of the cross section $\sigma(E_\nu)$ is described in Section C.2.2, and the core spectrum $F_c(E_\nu)$ is given by Equation B.7. As an implementation detail, these histograms are calculated in the toy MC by applying Equation C.3 for a single core, with all of the AD-specific quantities (baseline, etc.) set to unity (or an arbitrary constant), and θ_{13} set to zero (likewise for θ_{12}). The super histograms are used to calculate the fraction of each AD's spectrum that is attributable to each core, as needed when extrapolating near-site measurements to the far site. The normalization is unimportant, since only the ratios matter in this calculation.

C.2.5.2 Reverse response matrix

In the fitter (Section C.3), we require a "reverse response" matrix $\mathbf{M}$ to convert the prompt energy to the antineutrino energy for extrapolation to the far site. This matrix is generated as a two-dimensional ROOT histogram, in which the x axis represents antineutrino energy in 240 bins from 0 to 12 MeV, and the y axis represents prompt energy (in either the LBNL or BCW binning).[12]

For the near-to-far extrapolation, we require the ability to convert from E_{rec} to E_ν. However, the toy MC can only go in the opposite direction, from E_ν to E_{rec}. We could thus easily produce a "forward" response matrix $\mathbf{M_{fwd}}$ by looping over E_ν bins and populating their columns with the corresponding (normalized) E_{rec} spectra. The result could then be trivially used for converting E_ν to E_{rec}:

$$\mathbf{S_{rec}} = \mathbf{M_{fwd}}\mathbf{S}_\nu \tag{C.32}$$

[12] In our mathematical treatment here, we "flip" the axes in accordance with the normal convention for matrices: For an element $\mathbf{M}_{ij}$, the "vertical" index i refers to a particular antineutrino energy (the "output"), and the "horizontal" index j refers to a particular reconstructed energy (the "input"). It is merely a historical implementation detail that the ROOT histogram uses the opposite convention.

where the $\mathbf{S}$ are the spectra.[13] Unfortunately, this matrix would not be directly invertible, greatly complicating the reverse transformation that we need. Meanwhile, use of the transpose, as in

$$\mathbf{S}_\nu = \mathbf{M}^{\mathbf{T}}_{\mathbf{fwd}}\mathbf{S}_{\mathbf{rec}}, \tag{C.33}$$

would only be valid if the antineutrino spectrum were flat, which is obviously not the case.

The solution to these difficulties is to reweight the E_{rec} spectrum in each E_ν column of $\mathbf{M}_{\mathbf{fwd}}$ according to the expected shape (in E_ν) of the IBD spectrum. Each *row* of this $\mathbf{M}^{\mathbf{reweight}}_{\mathbf{fwd}}$ (corresponding to an E_{rec} bin) can then be normalized to give an E_ν spectrum for each E_{rec} bin. Under the assumption that the measured spectrum reasonably matches this shape, it is then valid to use the (transposed) result for converting an E_{rec} bin to an E_ν distribution.

Accordingly, to construct the reverse response matrix, the toy MC first generates a nominal antineutrino energy spectrum $\mathbf{S}^{\mathbf{nom}}_\nu$ (for the arbitrary case of AD1). Backgrounds are not included, since the fitter subtracts them before converting E_{rec} to E_ν. The toy MC then loops over the 240 antineutrino energy bins. For each antineutrino energy bin i (centered about $E_{\nu,i}$), it produces the corresponding PDF $f(E_{\text{rec}}; E_{\nu,i})$ of reconstructed energy (assuming a flat distribution of antineutrino energies within bin i), then scales it by the value of $\mathbf{S}^{\mathbf{nom}}_{\nu,i}$; the result is then assigned to the ith row of the matrix. Finally, each column of the matrix is normalized to unity:

$$\mathbf{M}^{\mathbf{un\text{-}norm}}_{ij} = \mathbf{S}^{\mathbf{nom}}_{\nu,i}\, f(E_{\text{rec},j}; E_{\nu,i}), \tag{C.34}$$

$$\mathbf{M}_{ij} = \frac{\mathbf{M}^{\mathbf{un\text{-}norm}}_{ij}}{\sum_{j'=0}^{240} \mathbf{M}^{\mathbf{un\text{-}norm}}_{ij'}}. \tag{C.35}$$

C.2.5.3 Covariance matrices

The covariance matrix, which encodes the scale of fluctuations between the far-site data and the prediction from the near sites (including correlations between ADs and energy bins), can be decomposed as the sum of three components, corresponding to signal (antineutrino) systematics, background systematics, and statistics. The statistical covariance matrix is calculated analytically by the fitter, as described in Section C.3.2. The two matrices for the systematics are generated by the toy MC, as detailed here.

Although the toy MC is capable of simultaneously varying the signal and background systematics, the two categories are treated separately due to differences in their scaling behaviors: The signal uncertainties scale with the size of the signal

[13] In practice, the fitter does not use this method to convert E_ν to E_{rec} at the end of the extrapolation. Instead, a separate E_ν spectrum is extrapolated for each E_{rec} bin, and the result is then integrated back into the original E_{rec} bin. This method is perhaps not quite as rigorous, but there is no evidence of any resulting bias as long as the binning is sufficiently fine, as is the case for both the LBNL and BCW binnings.

(which depends on the oscillation parameters), while the background uncertainties are constant.

To generate the covariance matrix for the signal systematics, a sample of $M = 1{,}000$ toy experiments is generated, using the nominal values of $\sin^2 2\theta_{13}$ and Δm^2_{ee} from Table C.2, subject to the following fluctuations:

- Solar oscillation parameters (fluctuated according to the uncertainties listed in Table C.2)
- Reactor power (0.5%, core-to-core uncorrelated)
- Fission fractions (uncorrelated), isotope $\bar{\nu}$ spectra, non-equilibrium corrections, and spent fuel contributions (all correlated), as encapsulated by the covariance matrix from [65] (see Section B.4)
- IAV thickness (4%, AD-to-AD uncorrelated)
- Nonlinearity model, fluctuated according to Equation C.16 (correlated)
- Relative energy scale (0.2%, uncorrelated)
- Energy resolution (0.2%, correlated $\oplus$ 0.2%, uncorrelated). The two parts are treated as *absolute* shifts to the (relative) resolution, and combined additively: $\sigma/E = \sigma_{\text{nominal}}/E + \text{Gaus}_1(0, 0.002) + \text{Gaus}_2(0, 0.002)$.

From this sample of M toy experiments, a "normalized" signal covariance matrix is constructed according to

$$(V_{\text{sig}}^{\text{norm}})_{ij} = \frac{1}{M} \sum_{t}^{M} \frac{(F_i^{\text{obs},t} - F_i^{\text{pred},t})(F_j^{\text{obs},t} - F_j^{\text{pred},t})}{F_i^{\text{pred},t} \cdot F_j^{\text{pred},t}} \tag{C.36}$$

Here, F^{obs} is the observed far-site data (after background subtraction), and F^{pred} is the predicted far-site data, as given by Equation C.46, based on the near-site observations. The denominator carries out the normalization, enabling the matrix to be rescaled according to the size of the signal in data, as described next. The indices i and j can potentially[14] span (a) far ADs, (b) the near site(s)/AD(s) used for the prediction, (c) energy bins, and (d) data periods with different detector configurations. When the fitter calculates the χ^2 at a given set of oscillation parameters, it rescales $(V_{\text{sis}}^{\text{norm}})_{ij}$ according to the predicted signal at the far site:

$$(V_{\text{sig}})_{ij} = (V_{\text{sig}}^{\text{norm}})_{ij} \cdot F_i^{\text{pred}}(\sin^2 2\theta_{13}, \Delta m^2_{ee}) \cdot F_j^{\text{pred}}(\sin^2 2\theta_{13}, \Delta m^2_{ee}). \tag{C.37}$$

[14] Depending on how data is combined among ADs, as described in Section C.3.3.

This uniform scaling procedure does not account for second-order variations in the "shape" of the covariance matrix as the oscillation parameters vary, but this simplification was found to have a negligible affect on the fit, so long as the assumed nominal parameters are reasonable.

For the background systematics, another set of $M = 1{,}000$ toy experiments is generated (again with nominal oscillation parameters), subject to the rate and shape fluctuations described in Section C.2.4. From this sample the covariance matrix is calculated as

$$(V_{\text{bkg}})_{ij} = \frac{1}{M}\sum_{t}^{M}(F_i^{\text{obs},t} - F_i^{\text{pred},t})(F_j^{\text{obs},t} - F_j^{\text{pred},t}) \tag{C.38}$$

Note that this matrix is not "normalized" (i.e., there is no denominator, unlike in Equation C.36), and thus there is no need to rescale it (i.e., multiply it by F^{pred}). This difference in treatment between the signal and background matrices is due to the fact that the backgrounds (and hence their uncertainties) do not vary as a function of the oscillation parameters, while the converse is true for the signal.

C.2.6 Handling of multiple data periods

Some additional complexity arises from the fact that Daya Bay has operated under three different detector configurations: First with 6 ADs (missing EH2-AD2 and EH3-AD4), then with all 8, and finally with 7 (after the repurposing of EH1-AD1 for the JUNO experiment R&D). In principle, the data from the three periods could simply be merged together. However, as explained in Section C.3.3, obtaining an invertible covariance matrix requires combining data from ADs in the same sites, and this procedure is the simplest to implement when the (active) ADs have the same livetime, since they can be weighted equally. For this reason, the LBNL toy MC and fitter treat the three data periods as separate experiments (thus increasing the dimensionality of the prediction/observation vectors and the covariance matrix). However, some systematic fluctuations must be correlated among the periods, given that the overall experimental setup was the same (aside from some missing ADs). These correlations are summarized in Table C.3.

C.3 Fitter

Fundamentally, the task of the fitter [62] is simply to find the oscillation parameters that best fit the observed data, where the quality of the fit is given by

$$\chi^2 = \sum_{i,j}(F_i^{\text{obs}} - F_i^{\text{pred}})V_{ij}^{-1}(F_j^{\text{obs}} - F_j^{\text{pred}}), \tag{C.39}$$

Fluctuated systematic	Correlation between: Detector	Period
Relative efficiency/response		✓
Solar parameters, nonlinearity	✓	✓
Correlated background shapes	✓	✓
Accidentals rates		
Fast-neutron and ^{9}Li/^{8}He rates	✓(same hall)	✓
AmC rates		✓
^{13}C$(\alpha, \mathrm{n})^{16}$O rates		

Table C.3: Correlations of systematics across detectors and periods. Reproduced from [93].

with the total covariance matrix V defined as

$$V = V_{\mathrm{sig}} + V_{\mathrm{bkg}} + V_{\mathrm{stat}}. \tag{C.40}$$

V_{sig} and V_{bkg} were described in Section C.2.5, while V_{stat} is described below in Section C.3.2. The indices i and j can run over detectors (or halls), energy bins, and data periods (6, 8, 7AD), depending on how the data (and covariance matrix) is merged across detectors, as described later. The main distinguishing characteristic of the LBNL fitter is that it takes a *relative* approach: Rather than simultaneously fitting the near-site and far-site data directly, the near-site data is used to generate a far-site prediction, which is then compared to the data. The chief advantage of such an approach is that the absolute detection efficiency, shared among all ADs, cancels automatically, and therefore the covariance matrix can be generated without considering fluctuations in the absolute efficiency. Indeed, we could even ignore reactor uncertainties that are correlated among all cores, but because the reactor covariance matrix was designed to be useful for both relative and absolute fitters, they are accounted for anyway, unnecessarily but harmlessly.

The bulk of the fitter's complexity lies in the determination of the far-site prediction F^{pred} based on the near-site data, as well as in the aforementioned merging of data and the corresponding reduction of the covariance matrix.[15] In what follows we detail the steps taken by the fitter to ultimately determine the χ^2 for a given set of oscillation parameters. Given the ability to calculate this χ^2, finding the best fit can be done using standard minimization techniques, and a map of the χ^2 across parameter space can be used to generate two-dimensional contours at any desired confidence level.

[15]This merging is necessary in order to obtain an invertible covariance matrix, given the high degree of correlation in the full covariance matrix.

C.3.1 Far-site prediction from near sites

As was mentioned, the far-site data is predicted based on the measurements at the near sites. If Daya Bay were measuring antineutrinos from a single reactor, then this prediction could take place without any reference to reactor power and fission fraction data, resulting in a truly pure relative measurement. In reality, however, each near AD samples the flux from multiple reactors at different locations, which must each be treated individually when extrapolating to the far site. As such, reactor operational information must be used to decompose each near-site measurement into the components from each core. This process of near-to-far extrapolation, as implemented, essentially treats each bin of E_{rec} as a separate sub-experiment. For a single near AD, a single far AD, and a single E_{rec} bin, the extrapolation proceeds as follows:

1. The reverse response matrix $\mathbf{M}$ from Equation C.35 is used to generate a distribution of E_ν, by taking the row corresponding to the E_{rec} bin being extrapolated.
2. For each bin of E_ν, the contribution from each core is determined via Equation C.41 below, based on the predicted flux (according to power, fission fractions, and isotope spectra), the baselines, and the oscillation parameters.
3. For a single core's contribution to a single E_ν bin, the extrapolation factor of Equation C.44 is applied to obtain the predicted contribution at the far AD.
4. This is repeated for all cores, over all bins of E_ν, to generate a predicted E_ν spectrum at the far AD *for the single bin of* E_{rec} *in question.*
5. The E_ν spectrum is integrated back into the original E_{rec} bin.

This process is illustrated by Figure C.7, and further details are given in the next section. After repeating this procedure for each E_{rec} bin, a predicted prompt spectrum is thereby obtained at the far AD. Given that there are four near ADs and four far ADs (when all 8 ADs are operating), this results in a total of 16 predictions for each E_{rec} bin. The 6AD and 7AD periods are essentially considered separate experiments (whose correlations due to systematics are encoded in the covariance matrix, as described in Section C.2.6), with 9 and 12 predictions for each E_{rec} bin, respectively.

C.3.1.1 Mathematical details

For a near AD i and a given antineutrino energy E_ν, the fraction of the flux coming from core k is given by the flux factor

$$f_{ik}(E_\nu) = \frac{F_k(E_\nu) \times P^{\text{osc}}_{ik}(E_\nu) \times 1/L^2_{ik}}{\sum_c F_c(E_\nu) \times P^{\text{osc}}_{ic}(E_\nu) \times 1/L^2_{ic}}, \tag{C.41}$$

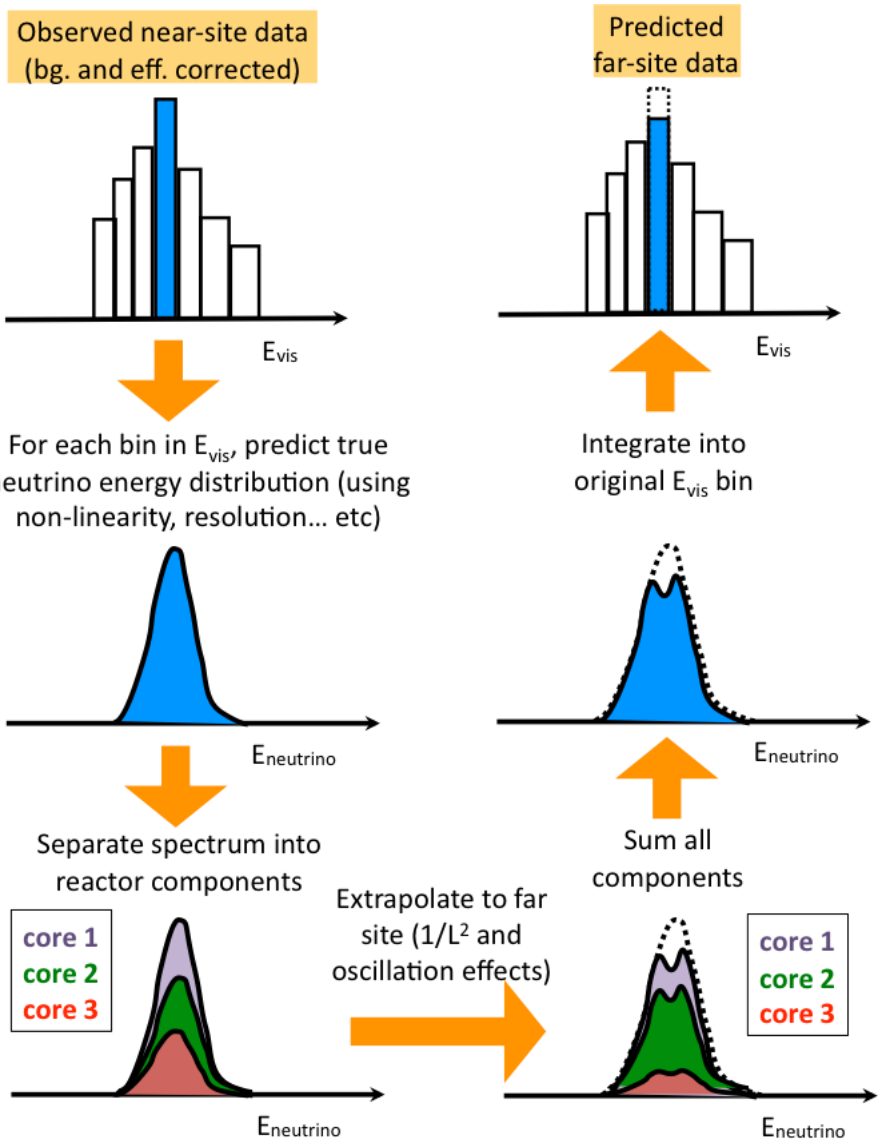

Figure C.7: Conceptual illustration of the procedure for extrapolating a near AD measurement to a far AD. In the bottom-right panel, the dashed curve represents the no-oscillation prediction. From [62].

where $F_c(E_\nu)$ is the antineutrino flux from core c, given by Equation B.7, and L_{ic} is the baseline from core c to AD i.

The observation at near AD i, attributable to core k, in true-energy bin E_ν, for the extrapolation of reconstructed-energy bin E_{rec}, is then simply

$$N_{ik}(E_\nu; E_{\text{rec}}) = f_{ik}(E_\nu) \times N_i(E_\nu; E_{\text{rec}}) \tag{C.42}$$

where $N_i(E_\nu; E_{\rm rec})$ is the background-subtracted count in the E_ν bin:

$$N_i(E_\nu; E_{\rm rec}) = \mathbf{M}(E_{\rm rec}, E_\nu) \times N_i(E_{\rm rec}), \tag{C.43}$$

and $N_i(E_{\rm rec})$ is the (background-subtracted) number of observed events in reconstructed-energy bin $E_{\rm rec}$. For clarity, here we express the reverse response matrix $\mathbf{M}$ (from Equation C.35) as a function of $E_{\rm rec}$ and E_ν rather than of bin indices.

In order to take near AD i's observation of core k and calculate the predicted observation at far AD j, we use the extrapolation factor

$$e_{ij,k}(E_\nu) = \frac{P^{\rm osc}_{jk}(E_\nu) \times 1/L^2_{jk}}{P^{\rm osc}_{ik}(E_\nu) \times 1/L^2_{ik}}, \tag{C.44}$$

which essentially removes the effects of baseline and oscillation from the near observation, and then applies them for the case of the far detector.

We then have the predicted flux contribution at the far AD:

$$F^{\rm pred}_{ij,k}(E_\nu; E_{\rm rec}) = e_{ij,k}(E_\nu) \times N_{ik}(E_\nu) \tag{C.45}$$

Then to get the final prediction at far AD j from near AD i, we sum over all cores k and integrate all E_ν bins into the original $E_{\rm rec}$ bin:

$$F^{\rm pred}_{ij}(E_{\rm rec}) = \sum_k \sum_{E_\nu} F_{ij,k}(E_\nu; E_{\rm rec}) \tag{C.46}$$

C.3.2 Statistical covariance matrix

The near-to-far extrapolation produces multiple predictions for each far AD, one per near AD. Likewise, each near AD furnishes multiple predictions, one per far AD. Accordingly, statistical fluctuations in the near and far AD data will produce correlated fluctuations (across multiple near/far pairs) in the deviation between prediction and measurement. These correlations are encoded in the statistical covariance matrix, of which we describe the calculation here.

Technically speaking, fluctuations in the near-site data will produce fluctuations in $F^{\rm pred}$, while fluctuations in the far-site data will produce fluctuations in $F^{\rm obs}$. However, for the sake of simplicity, we choose to "transfer" the uncertainty on $F^{\rm obs}$ over to $F^{\rm pred}$, essentially leaving $F^{\rm obs}$ without any error bar. This simplifies the conceptual description of the covariance matrix (in the context of calculating χ^2): It encodes fluctuations in $F^{\rm pred}$ around the "fixed" $F^{\rm obs}$. Of course, in reality, both quantities will fluctuate, but since it is only the *difference* between the two that enters the χ^2 calculation, we are free to divide the uncertainty between them as we please.

Given this convention, the statistical uncertainty on $F^{\rm pred}$ that arises from the near-site fluctuations is

$$\sigma_{\rm near}(F^{\rm pred}) = \frac{F^{\rm pred}}{N^{\rm obs}} \sqrt{N^{\rm obs} + N^{\rm bkg}}, \tag{C.47}$$

where N^{obs} is the background-subtracted observation at the near AD, and N^{bkg} is likewise the predicted near AD background. In other words, this is the statistical uncertainty of the (background-subtracted) near measurement, scaled by the expected far-to-near ratio.

Meanwhile, the "transferred" statistical uncertainty on F^{pred} from far site fluctuations is simply

$$\sigma_{\text{far}}(F^{\text{pred}}) = \sqrt{F^{\text{pred}} + F^{\text{bkg}}}, \tag{C.48}$$

where F^{bkg} is the predicted background at the far AD. The total uncertainty on F^{pred} is then the sum of σ_{near} and σ_{far}, in quadrature.

As was noted, statistical correlations exist between pairs of predictions that share a near or far AD. It can be shown that, for two predictions (in the same E_{rec} bin and data period) that arise from the same near AD, the covariance is the product of the two values of σ_{near}. Similarly, for two predictions at the same far AD, the covariance is the product of the σ_{far} values. And, of course, there is no covariance between different energy bins, data periods, or predictions that don't share a common AD. Based on the discussion thus far, it is then a simple matter to fill in the statistical covariance matrix. The tedious details depend on the way that the data is merged between ADs and on the technical definition of the indices i and j; the code can be seen at [95, ShapeFit/Predictor.C, Calculate(NearSite)StatError]. Unlike the systematic components of the covariance matrix (which are fixed throughout the fit procedure, aside from rescaling), the statistical component is recalculated at each point in the parameter space.

C.3.3 Combination of data

As was noted, the extrapolation process produces multiple predictions, many of which are highly correlated by virtue of sharing a near or far AD. For the basic case of a single 8AD data period, we are essentially turning four observations (one from each far AD) into 16 variables (one per prediction). Due to this correlation, the full covariance matrix is non-invertible. In order to obtain an invertible matrix, the dimensionality must be reduced down to the level of the original input data (or smaller). This is done by combining data between ADs.

There are multiple ways this combination could be carried out. Due to the limited statistics at the far site, it is reasonable in all cases to combine the four far ADs. As for the near ADs, they could be kept distinct, resulting in four predictions, matching the underlying dimensionality of the data. Or, the four near ADs could be fully combined, producing a single prediction. As another alternative, near ADs could only be combined within the same near site, giving two predictions. The fitter supports all three of these "combination modes", and they all give similar results. In practice, we use the last of these modes, with two predictions, since there should be no significant

loss of information from combining data at a given near site, but there could be from combining EH1 and EH2.[16]

By convention, the far ADs are added, while the ADs of each near hall are averaged. Ultimately this convention is irrelevant as long as it is consistently applied between the observations, the predictions, and the covariance matrix. For the 8AD case, we then have the combined far-site observation

$$F^{\text{obs, comb}}(E_{\text{rec}}) = \sum_{b=1}^{4} F_b^{\text{obs}}(E_{\text{rec}}), \tag{C.49}$$

where the index b runs over the far ADs. Likewise, we have the two near-site predictions, where the index i now runs over the two near sites:

$$F_i^{\text{pred, comb}}(E_{\text{rec}}) = \frac{1}{n_i} \sum_{a=1}^{n_i} \sum_{b=1}^{4} F_{ab}^{\text{pred}}. \tag{C.50}$$

Here, the index a runs over the n_i ADs in the near site i. Finally, the combined covariance matrix is given by

$$V_{ij}^{\text{comb}} = \frac{1}{n_i} \frac{1}{n_j} \sum_{i',j'} (V_{\text{sig}} + V_{\text{bkg}} + V_{\text{stat}})_{i'j'}, \tag{C.51}$$

where the indices i' and j' run over all predictions from the ith near site to all far ADs.

C.3.3.1 Handling of multiple data periods

As noted in Section C.2.6, data is *not* combined between the 6AD, 8AD, and 7AD periods. As such, the dimensions of the (full) prediction vector, observation vector, and covariance matrix are enlarged by a factor of $1+6/8+7/8 = 21/8$ (with respect to the case of a single 8AD period). In generating the full covariance matrix, fluctuations are correlated (or not) across periods according to Table C.3. For the combined vectors and matrix, the dimensionality is increased by a factor of 3. The above discussion generalizes to this scenario in a straightforward fashion.

[16] Consider an extreme case: Two distant nuclear reactors, each with a nearby experimental hall. Treating the two halls separately, one can measure the antineutrino flux from each reactor independently. When data from the two halls is combined, a degeneracy is introduced: There is a continuum of solutions, in terms of the two reactor fluxes, that can produce the combined antineutrino rate. This illustrates how combining the near halls can produce a loss of information, in the statistical sense.

C.3.4 χ^2 calculation

The data combination procedure results in an invertible covariance matrix, and thus the final χ^2 can be calculated as

$$\chi^2 = \sum_{i,j} (F^{\text{obs, comb}} - F_i^{\text{pred, comb}}) V_{ij}^{-1, \text{comb}} (F^{\text{obs, comb}} - F_j^{\text{pred, comb}}), \tag{C.52}$$

where i and j run over the EH1 and EH2 predictions for each E_{rec} bin and data period. Here, χ^2, F^{pred} and V are implicitly functions of the oscillation parameters.[17] To find the best fit oscillation parameters, ROOT's MINUIT package is used to minimize the χ^2 over the 2D parameter space.

C.3.5 Generation of contours

To obtain the uncertainty on the oscillation parameters, a 2D grid of χ^2 values is generated in parameter space. Based on the shape of the χ^2 distribution for two degrees of freedom, the 1σ, 2σ, and 3σ contours are then determined according to the condition $\chi^2 = 2.30$, 6.18, and 11.83, respectively. The individual error bar on $\sin^2 2\theta_{13}$ is finally given by the distance along the $\sin^2 2\theta_{13}$ axis from the best fit point to the 1σ contour, and similarly for Δm^2_{ee}.

[17] Again, the systematic components of V are only rescaled, but not fully recomputed, for reasons of computational tractability.

Appendix D

Miscellaneous details

D.1 Uncorrelated detection efficiency uncertainty

In order to calculate proper errors on the oscillation parameters (particularly $\sin^2 2\theta_{13}$), it is necessary to account for possible AD-to-AD variations in the IBD detection efficiency, as any such variations could potentially bias the measured rate deficit at the far site. As with the other systematics, this one is treated by the fitter's toy MC (Section C.2): For each random set of toy prompt spectra, a random value is generated for each AD's efficiency, and its prompt spectrum is scaled accordingly. The covariance matrix generated from these prompt spectra then encodes the effects of the relative efficiency uncertainty. Here we briefly describe the determination of this uncertainty and its implementation in the toy MC.

The detection efficiency can be decomposed into various components, as detailed in Table D.1. With the exception of the (dominant) first two rows (the delayed energy cut efficiency and the Gd capture fraction), the measurements of these components (and their AD-to-AD uncertainties) have remained unchanged since they were first described in [17] and compiled in [96].

Meanwhile, the first two rows of Table D.1 were updated in 2016 by Lebanowski in [97], taking advantage of the increase in statistics since 2012. For the delayed energy cut, the uncorrelated efficiency uncertainty was evaluated by loosening the delayed cut to 3.6 MeV and comparing, between ADs, the percentage of delayed energies lying in the [6, 12] MeV region. And for the nGd-capture fraction, the prompt-delayed time difference for each AD was fit to an exponential in order to determine each AD's average capture time, which in turn gave the capture fraction. The spread of these capture fractions then gave the uncorrelated uncertainty.

An additional complication arises due to the correlation between the energy scale and the detection efficiency (via the delayed energy cut efficiency). When, in the toy MC, the energy scale is fluctuated, the detection efficiency must reflect this fluctuation, in addition to *independent* fluctuations of the fraction of the efficiency

Efficiency	Uncorrelated unc. (%)
Delayed energy	0.075
Gd capture fraction	0.1
Target protons	0.03
Flasher cut	0.01
Multiplicity cut	0.01
Spill-in	0.02
Capture time	0.01
Prompt energy	0.01
Total	0.132

Table D.1: Uncorrelated detection efficiency uncertainties [97]. Uncertainties were added in quadrature to obtain the total.

that is *not* correlated with the energy scale. The task is then to decompose the efficiency uncertainty into two parts, one correlated and one uncorrelated with the energy scale.

Nakajima carried out this decomposition in [98]. His reasoning was as follows. First, the total efficiency uncertainty σ_{tot} (0.132%) can be broken down into two parts: The 0.075% uncertainty due to the delayed energy cut (which we denote σ_{E_d}), and the remainder (σ_{other}):

$$\sigma_{\text{tot}} = 0.132\% = \sqrt{\sigma_{E_d}^2 + \sigma_{\text{other}}^2}, \tag{D.1}$$

where (Table D.1)

$$\sigma_{E_d} = 0.075\%. \tag{D.2}$$

In turn, σ_{E_d} is *partially* correlated to the energy scale. According to [97], 91.4% of the AD-to-AD variance in the delayed energy cut efficiency is due to variance of the energy scale (Table D.2). Thus,

$$\sigma_{\text{corr}} = \sqrt{0.914 \times \sigma_{E_d}^2} = 0.072\%. \tag{D.3}$$

We can then subtract (in quadrature) σ_{corr} from σ_{tot} to obtain the part of the detection efficiency uncertainty that is uncorrelated with the energy scale:

$$\sigma_{\text{uncorr}} = \sqrt{\sigma_{\text{tot}}^2 - \sigma_{\text{corr}}^2} = 0.11\%. \tag{D.4}$$

In summary,

$$\begin{aligned} \sigma_{\text{tot}} &= \sqrt{\sigma_{\text{corr}}^2 + \sigma_{\text{uncorr}}^2} \\ &= \sqrt{(0.072\%)^2 + (0.11\%)^2}. \end{aligned} \tag{D.5}$$

Component	Fraction (%)
Energy scale	91.4
OAV thickness	7.8
Nonuniformity	0.8

Table D.2: Decomposition of the variance of the delayed energy cut efficiency [97].

As described in [19, Sec. III B 5 b], the AD-to-AD variation of the energy scale, δ_E, is ∼0.20%. It is then assumed that this 0.20% variation leads to the observed 0.072% energy-scale-correlated detection efficiency variation δ_{ϵ_d}:

$$0.072\% = \frac{\delta_{\epsilon_d}}{\delta_E} \times 0.20\%, \tag{D.6}$$

or

$$\frac{\delta_{\epsilon_d}}{\delta_E} = \frac{0.072}{0.20} = 0.36. \tag{D.7}$$

Therefore,

$$\delta_{\epsilon_d} = 0.36 \times \delta_E. \tag{D.8}$$

In the toy MC, after a fractional energy scale fluctuation δ_E is generated via $\text{Gaus}(0, 0.0020)$, the nominal[1] detection efficiency is first multiplied by $1 + 0.36 \times \delta_E$. The result is then multiplied by an additional factor of $1 + \text{Gaus}(0, 0.0011)$ to account for the detection efficiency variance that is uncorrelated with the energy scale. That is,

$$\epsilon_d = \epsilon_{d,\text{nom}} \cdot (1 + 0.36 \times \text{Gaus}(0, 0.0020)) \cdot (1 + \text{Gaus}(0, 0.0011)), \tag{D.9}$$

where the first Gaussian random variable is the same one used in fluctuating the energy scale. This gives the final detection efficiency used by the toy MC. The prompt spectrum is uniformly scaled by this factor in generating the toy sample.

D.2 Energy nonlinearity model

The AD's response on the type of particle in question. By construction, AdSimple's E_{rec} will report the correct deposited energy for the ∼8 MeV gamma-ray peak produced by nGd capture. However, due to the various sources of nonlinearity (from quenching, Cherenkov emission, and the electronics, as summarized in Section 4.2.4 and detailed below), E_{rec} will report a different value for a *positron* that deposits 8 MeV. As described below, this effect must be considered when constructing the conversion

[1] And arbitrary, since we are performing a relative measurement.

between E_{rec} and E_{dep} for positrons.[2]) All in all, the ratio of E_{rec} to E_{dep}varies by some 10% across the range of energies used in the analysis (0.7–12 MeV). If this nonlinearity is not corrected for,[3] then the resulting distortion in the IBD positron spectrum will potentially bias the extraction of Δm^2_{ee}.

The nonlinearity correction thus converts E_{rec} into E_{dep}, which can then be trivially converted to the true neutrino energy, E_ν, using Equation 2.4:

$$E_\nu \simeq E_{\text{dep}} + 0.80\,\text{MeV} \tag{D.10}$$

. In what follows, the phrase "positron energy" refers to E_{dep}, which can be trivially related to the kinetic energy E_{kin} and the relativistic energy E_{rel} via

$$E_{\text{dep}} = E_{\text{kin}} + 2m_e = E_{\text{rel}} + m_e. \tag{D.11}$$

The scintillator (quenching + Cherenkov) and electronics nonlinearities are both on the scale of 10%, and work in opposite directions: The scintillator response is suppressed for low-energy events, while the electronics show a reduced response for high-energy events. Nevertheless, they do not cancel each other very effectively, so each must be treated separately and carefully. We begin with the scintillator. In terms of physical processes, the scintillator nonlinearity determines E_{vis} from E_{dep}, and then the electronics nonlinearity determines E_{rec} from E_{vis}. In analysis, this logic is reversed in order to determine E_{dep} given E_{rec}[4].

In detailing the nonlinearity correction, we begin by discussing the AD's response to electrons, rather than positrons. For organic scintillators such as the Daya Bay LS, quenching is quantitatively well described by *Birks' law* [38],

$$\frac{dQ}{dx} \propto \frac{\frac{dE}{dx}}{1 + k_B \frac{dE}{dx}} \tag{D.12}$$

where Q is the amount of emitted light, dE/dx is the linear density of *ionization* energy deposition, and the scintillator-specific value k_B is known as Birks' constant.

If the dE/dx profile is known across a particle's range, then it can be used to integrate Equation D.12. Such a "semi-empirical" analytic approach was used to compute the shape of the quenching curve for electrons in the Daya Bay LS [29]. This

[2]Given that, for the purpose of oscillation physics, the IBD positron spectrum is the only one whose shape carries any significance, there is no need in this analysis to apply separate corrections for purported gamma ray, electron, or alpha particle events. Indeed, the delayed (i.e. nGd gamma ray) spectrum does not undergo any nonlinearity correction here.

[3]The nonlinearity correction is not actually performed during reconstruction. Instead, the correction is applied by the fitter (Chapter 8), which takes AdSimple's E_{rec} and converts it into E_ν. However, as this is merely an implementation detail, we discuss the conversion in this chapter, along with the other preceding steps of the energy reconstruction.

[4]Recall that any effects from geometric nonuniformity have already been removed in the computation of E_{rec}.

relationship was encoded in a function denoted $f_q(E_{\mathrm{dep},e^-}, k_B)$, where k_B remained to be determined from data:

$$f_q(E_{\mathrm{dep},e^-}, k_B) = \int_0^{E_{\mathrm{dep},e^-}} \frac{\frac{dE}{dx}}{1 + k_B \frac{dE}{dx}}. \tag{D.13}$$

Plots of f_q for different values of k_B are shown in Figure D.1.

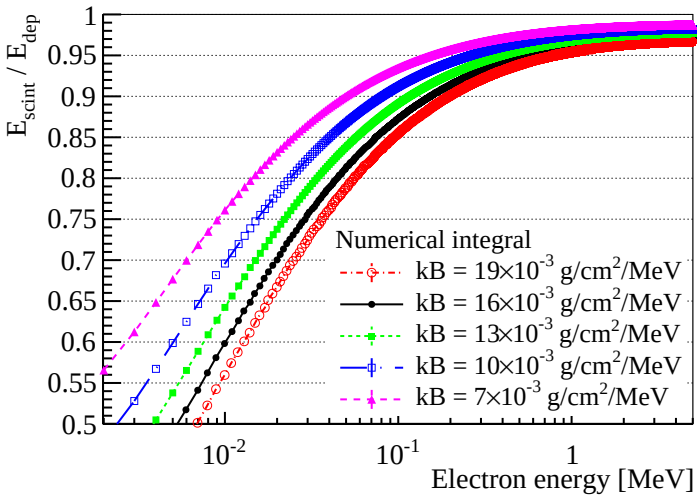

Figure D.1: Dependence of the scintillation energy on the deposited energy of electrons, for various values of Birks' constant k_B. From [29].

Although Cherenkov emission is the sole source of light in the Daya Bay water pools, it plays a sub-percent role in the ADs. The energy dependence of Cherenkov emission was determined from Geant4 simulations (and independently confirmed by an analytic calculation), giving the tabulated function $f_c(E_{\mathrm{dep},e^-})$ (Figure D.2), which was arbitrarily assigned a (k_B-dependent) normalization such that $f_c(1\,\mathrm{MeV}) = f_q(1\,\mathrm{MeV}, k_B)$.

Altogether, the combined effects on electrons of quenching and Cherenkov emission are therefore described by the relation

$$\frac{E_{\mathrm{vis},e^-}}{E_{\mathrm{dep},e^-}} = \beta_{\mathrm{vis}}[f_q(E_{\mathrm{dep},e^-}, k_B) + k_c f_c(E_{\mathrm{dep},e^-})], \tag{D.14}$$

where f_q is given by Equation D.13, and f_c (which, being tabulated from simulation data, cannot be expressed as an equation) is, again, shown in Figure D.2. Here, we have introduced the constant k_c, which describes the ratio of the amount of Cherenkov emission to scintillation. β_{vis} is an arbitrary normalization. The parameters k_B and k_c are determined from data, as will be described shortly.

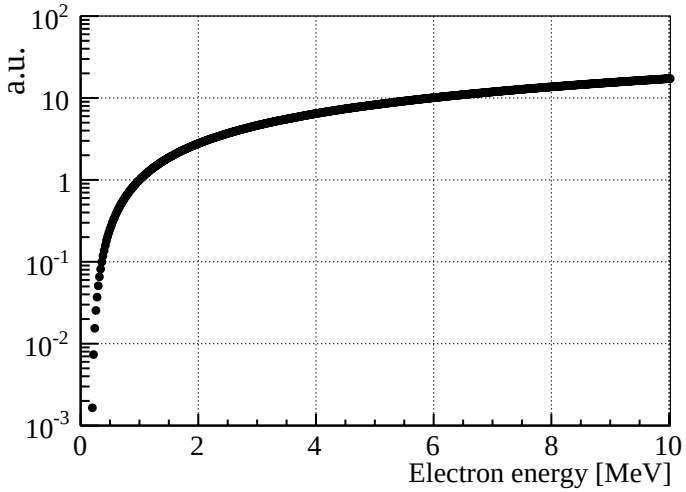

Figure D.2: Energy dependence of the Cherenkov contribution to light emission by electrons in the LS (here, arbitrarily normalized at 1 MeV), from [29]. The actual normalization of this function was determined by a fit to measurements of the nonlinearity, as described in [29].

Thus far we have only discussed the response of the scintillator to electrons. The end goal, however, is to characterize the response to *positrons,* since that is what will allow the measurement of the antineutrino spectrum. Before the positron annihilates, it effectively ionizes the medium in the same manner as an electron would. Following ionization, we measure the response of the scintillator to two 511 keV gamma rays. Accordingly,

$$E_{\text{vis},e^+}(E_{\text{kin}}) = E_{\text{vis},e^-}(E_{\text{kin}}) + 2 \times E_{\text{vis},\gamma}(0.511\,\text{MeV}). \tag{D.15}$$

Gamma rays, in turn, do not themselves ionize, but they do produce and scatter electrons and positrons. The total response to gamma rays is, accordingly, rather complex, given the need to account for annihilation and pair production *ad nauseam.* As such, Geant4 simulation were used to determine the response of the scintillator to gamma rays, as a function of k_B and k_c. The (ionization) energy deposited by each simulated electron and positron (Figure D.3) was converted into visible energy according to Equation D.14, and the sum gave the visible energy from each gamma ray. With the response to gamma rays thus determined, it could be plugged into Equation D.15 to give the response to positrons.

In addition to the nonlinear light *emission* of the scintillator, there is also nonlinear light *measurement* caused by the design of the electronics. As was described in Section 4.2.1.1, the charge (i.e. light) in each channel is determined by taking the *first* hit within the nominal time window. And as was described in Section 3.3.1, the first hit will tend to accurately measure the light observed by the PMT, *provided*

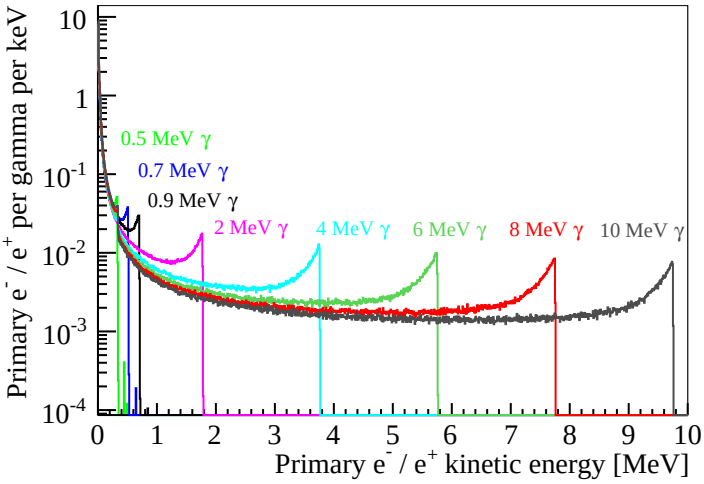

Figure D.3: Distribution of kinetic energies of electrons produced by gamma rays of a variety of initial energies. From [29].

that all of the photons arrived within 100 ns of each other. However, some 5% of the light is from the slow component, and will therefore arrive too late to be captured by the hit that corresponds to the fast/medium light. Thus, high-energy events, in which every channel "sees" some fast/medium light, will fail to record any of the slow light. Conversely, in a lower-energy event, some of the slow light *will* be recorded, since there is not enough fast/medium light to hit every channel. This means that the electronics response is slightly suppressed for high-energy events and slightly enhanced for low-energy ones.

Based on a combination of measurements and simulations, this behavior was found to be adequately described by the relation

$$\frac{E_{\text{rec}}}{E_{\text{vis}}} = \beta_{\text{rec}} \left[1 + \alpha \left(-\frac{E_{\text{vis}}}{\tau} \right) \right], \tag{D.16}$$

where α and τ are, like k_B and k_c, constrained by measurements, and β_{rec} is, like β_{vis}, merely an arbitrary normalization. The product $\beta_{\text{vis}}\beta_{\text{rec}}$, by convention, is chosen to ensure that $E_{\text{rec}} = E_{\text{dep}}$ for 8 MeV electrons.

In order to determine the four parameters k_B, k_c, α, and τ of the nonlinearity model, a fit was performed to a dataset consisting of the peaks from twelve gamma rays (from deployed sources and natural radioactivity) as well as the electron spectrum from decays of cosmogenic ^{12}B [29], as shown in Figures D.4 and D.5. The measured values of E_{rec} were compared to those predicted by simulation (for a given set of the parameters), and the best-fit parameters were determined as those that minimized the χ^2 between measurement and prediction. As no significant differences in nonlinearity were observed between ADs, a single nonlinearity model is used for all eight. The

best-fit parameters were found to be $k_B = 15 \times 10^{-3}$ cm MeV^{-1}, $k_c = 0.5\%$, $\alpha = 0.078$, and $\tau = 2.55$ MeV. However, these values depend on the assumed shapes f_q and f_c, which in turn depend on the configuration of Geant4. Accordingly, our analysis does not make direct use of these four parameters; instead we use the digitized nonlinearity curve published in [29], as shown in Figure 4.4, which includes a 68% uncertainty band (derived from the χ^2 fit), corresponding to a precision of about 1%.

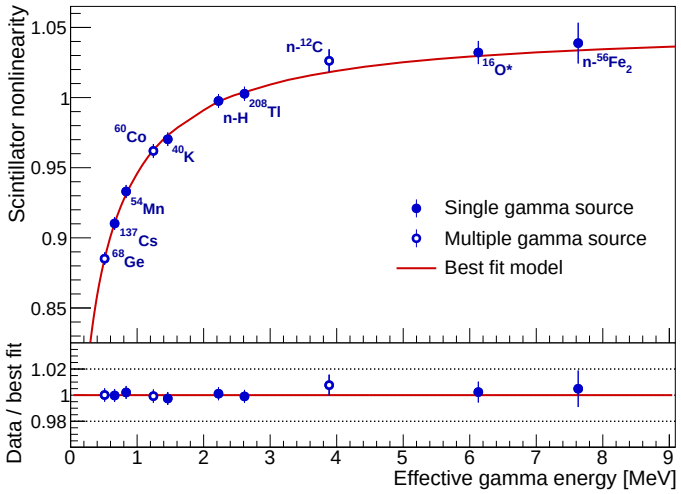

Figure D.4: Measured LS nonlinearity for various energies of gamma-rays. From [29].

The scintillator nonlinearity was cross-checked using the 53-MeV Michel electron from muon decay as well as the $\beta + \gamma$ spectra from ^{212}Bi, ^{214}Bi, and ^{208}Tl decays, while the electronics nonlinearity was validated using data from a fast ADC system that directly recorded the PMT waveforms in one AD [99]. Further validation involved verifying that the best-fit model remained stable under removal of any single calibration point. The results of these studies were consistent with the uncertainty band of the nonlinearity model.

D.3 Vertex reconstruction

The AdSimple vertex reconstruction proceeds in multiple stages [100, 101], as shown in Figure D.6. First, an initial vertex is determined by taking a simple *center of charge* (COC) using the coordinates of the PMTs. As this method suffers from significant biases (largely toward the center of the AD), a correction is then applied, based on interpolating a map of the mean bias (as a function of COC position) determined from a sample of Monte Carlo events, giving the *Monte Carlo Corrected COC* (MCC-COC). However, even with the MC correction, this vertex still suffers from biases, particularly

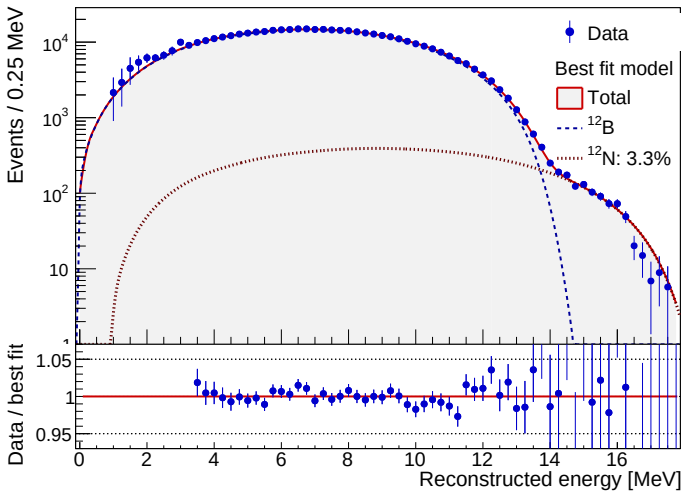

Figure D.5: The reconstructed energy spectrum from ^{12}B (with a minor contribution from ^{12}N). Also shown is the prediction obtained from the best-fit nonlinearity model (equations Equation D.14 × Equation D.16). From [29].

at large z (Figure D.7). In order to reduce such effects, and to also improve the resolution of the position reconstruction, a final vertex is computed by matching the distribution of PMT charges to a library of MC templates. We now discuss these steps in further detail.

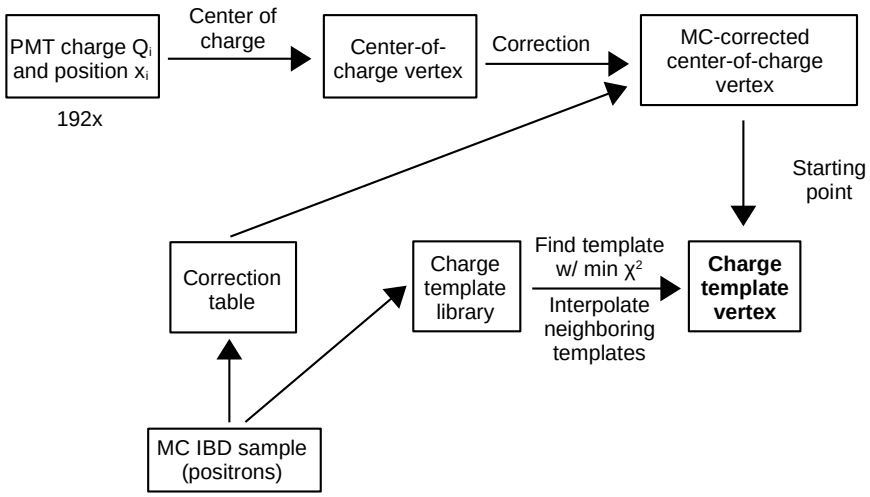

Figure D.6: Flowchart of the steps involved in the AdSimple vertex reconstruction.

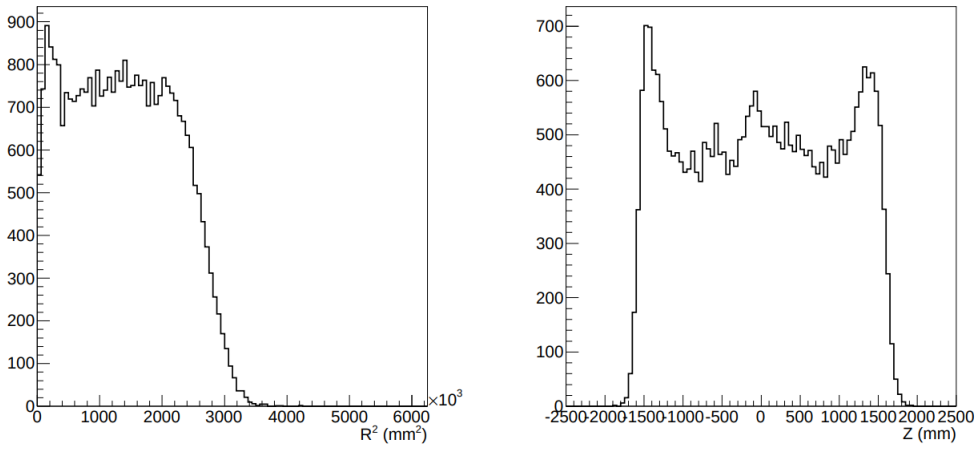

Figure D.7: Biases in the MCC-COC vertex distributions for positrons associated with IBD neutrons. The true distribution of events is uniform in the GdLS. From [100].

The COC vertex is calculated trivially,

$$x_{\mathrm{COC}} = \frac{\sum_i^{\mathrm{PMTs}} Q_i \vec{x_i}}{\sum_i^{\mathrm{PMTs}} Q_i}, \tag{D.17}$$

where $\vec{x_i}$ is the position of the ith PMT (in Cartesian coordinates, with the origin at the AD's center), and Q_i is the corresponding observed charge in photoelectrons (without any correction for electronics nonlinearity). To this vertex, the correction from MC is applied next.

The MC sample consists of a large number of uniformly-distributed IBD events, generated using the ILL-Vogel flux model without oscillation (see Table D.3). A cut is applied on the true interaction to eliminate events that occur outside the OAV. The positrons (i.e. prompt triggers) from these events are then used to generate a correction table.

For each MC event, the COC vertex is calculated, and two corrections, a "radial" and a "vertical" one, are computed (using cylindrical coordinates, again with the origin at the AD's center):

$$\Delta r = \frac{\vec{r}_{\mathrm{true}} \cdot \vec{r}_{\mathrm{COC}}}{|\vec{r}_{\mathrm{COC}}| - r_{\mathrm{COC}}}, \tag{D.18}$$

$$\Delta z = z_{\mathrm{true}} - z_{\mathrm{COC}}, \tag{D.19}$$

Events are divided into 20 bins for $0\,\mathrm{m} < r_{\mathrm{COC}} < 2\,\mathrm{m}$ and 40 bins for $-2\,\mathrm{m} < z_{\mathrm{COC}} < 2\,\mathrm{m}$. For each bin, the mean Δz is computed, while the mean Δr is computed over (and assigned to) all z_{COC} bins for a given r_{COC}, since there is very little z dependence on

Isotope	Fraction
^{235}U	0.60
^{238}U	0.05
^{239}Pu	0.30
^{241}Pu	0.05

Table D.3: Nominal fission fractions used when employing the ILL-Vogel model to generate simulated IBD events for constructing the MC correction table for the MCC-COC.

Δr. The Δr and Δz correction tables are "pre-interpolated" with a spline function[5] to give a $\times$10-finer grid, which is then stored for use by AdSimple. During reconstruction, linear interpolation is applied to this fine-binned grid to look up the corrections for arbitrary (r_{COC}, z_{COC}). Application of the correction is trivial: $r \mapsto r + \Delta r$ and $z \mapsto z + \Delta z$. This gives the MCC-COC vertex.

The final stage of the AdSimple vertex reconstruction relies on a library of 9,600 charge templates (i.e. distributions of charge across the PMTs), with each template corresponding to a voxel of the detector. These voxels (i.e. bins) are defined as the product of 20 bins in $0\,\mathrm{m}^2 < r^2 < 2^2\ \mathrm{m}^2$, 20 bins in $-2\,\mathrm{m} < z < 2\,\mathrm{m}$, and 24 bins in $0 < \phi < 2\pi$. The charge template for each voxel is taken from the mean of the charge distributions (each normalized by total charge) of all MC events whose true position lay within the voxel. The MC sample, in turn, is of the same nature as the one used for the COC correction: IBD positrons lying within the OAV. Taking advantage of the azimuthal symmetry of the detector response, each event was rotated in angular steps of $\pi/12$ to produce 23 "clones", which were added to the sample, thereby providing a "free" boost in statistics without the need to generate additional MC events.[6]

During reconstruction, the charge templates are compared to the event's (normalized) charge distribution (again without correcting for electronics NL) using the "χ^2"

[5] Unfortunately, it was not possible to find a record of the exact spline function used. Presumably, it was a quadratic or cubic function. In any case, since the MCC-COC is ultimately only used as a "pre-fit" starting point for the template-based reconstruction, its precise details are not extremely important.

[6] In principle, the detector response is not completely azimuthally symmetric, due to the Earth's magnetic field. However, such effects are largely canceled by the conical magnetic shields around each PMT. Furthermore, the MC does not take this factor into account; thus, azimuthal symmetry is a valid assumption for the MC sample.

(more precisely, the log-likelihood)

$$\begin{aligned}\chi^2 &= \sum_i^{\text{PMTs}} \left[-2\ln\frac{P(N_i^{\text{obs}}, N_i^{\text{exp}}(r^2,z,\phi))}{P(N_i^{\text{obs}}, N_i^{\text{obs}})}\right] \\ &= 2\sum_i^{\text{PMTs}} \left[N_i^{\text{exp}} - N_i^{\text{obs}} + N_i^{\text{obs}}\ln\left(\frac{N_i^{\text{obs}}}{N_i^{\text{exp}}}\right)\right],\end{aligned}$$

where $P(n,\nu) = \nu^n e^{-n}/(n!)$ is the Poisson probability of observing n events when ν are expected. The MCC-COC vertex is used as a starting point to search the grid for a local minimum (which is presumably also a global minimum). Having thus located the bin with the lowest χ^2, the reconstruction proceeds by quadratically interpolating the χ^2 values at the neighboring grid points. This interpolation is performed independently for the three coordinates r^2, z, and ϕ, each time using two neighbors in the appropriate direction. Letting s to denote any of the three coordinates, with s_1 being the value of s at the χ^2-minimizing grid point, and s_2 and s_3 being those of the neighbors, the value of s at the interpolated minimum is

$$s_{\text{min}} = \frac{(s_1^2 - s_2^2)(\chi_3^2 - \chi_1^2) - (s_1^2 - s_3^2)(\chi_2^2 - \chi_1^2)}{2(s_1 - s_2)(\chi_3^2 - \chi_1^2) - 2(s_1 - s_3)(\chi_2^2 - \chi_1^2)}. \tag{D.20}$$

Calculated in this manner, r_{min}, z_{min}, and ϕ_{min} give the final AdSimple reconstructed vertex. This vertex provides a significantly improved resolution compared to the MCC-COC vertex, of ∼7 cm in r (Figure D.8) and ∼9 cm in z (Figure D.9), compared to 11.5 cm (Figure D.10) and 17 cm (Figure D.11), respectively, for the MCC-COC.[7] (These values are for IBD positrons.) There are also no significant biases within the OAV region; although some high-frequency "wiggles" remain (as a consequence of the finite grid spacing), these are insignificant in the context of performing physics analysis, as shown, for instance, by the consistency of physics results between AdSimple and AdScaled (which lacks such structures).

Note that the reconstructed vertex is not directly used in selecting IBD candidates in this analysis (however, see Section 9.5). Nonetheless, it is still used indirectly in calculating the reconstructed energy (via the nonuniformity correction). In addition, some studies of background rates and spectra also depended on the reconstructed vertex. Finally, there is an independent oscillation analysis (not covered in this thesis), based on selecting IBDs where the neutron is captured on hydrogen, and this one typically relies on the use of the prompt-delayed distance in order to reduce the background of accidental coincidences.

[7] As can be seen, these distributions are non-Gaussian, so the quoted resolutions (from Gaussian fits) should be taken with a grain of salt. The improvement over the MCC-COC remains significant.

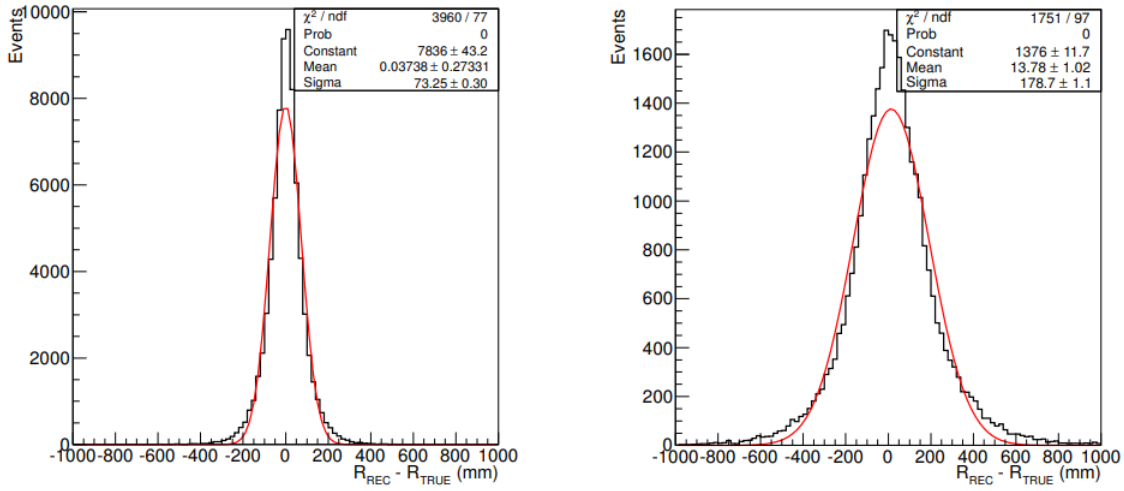

Figure D.8: Distributions of residuals of the radial coordinate for the AdSimple charge template vertex reconstruction. IBD positrons are shown on the left, while IBD nGd captures are shown on the right. From [101].

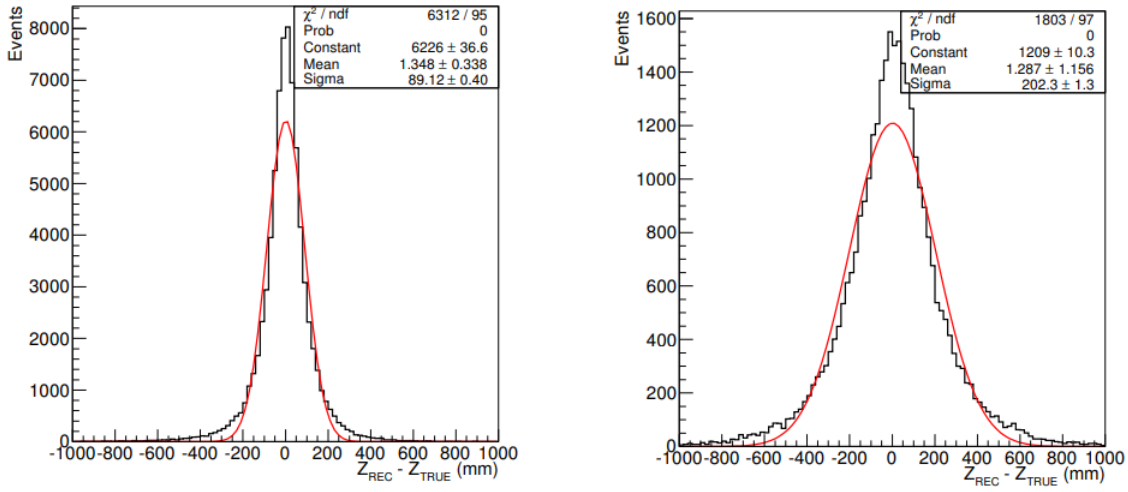

Figure D.9: Distributions of residuals of the vertical coordinate for the AdSimple charge template vertex reconstruction. IBD positrons are shown on the left, while IBD nGd captures are shown on the right. From [101].

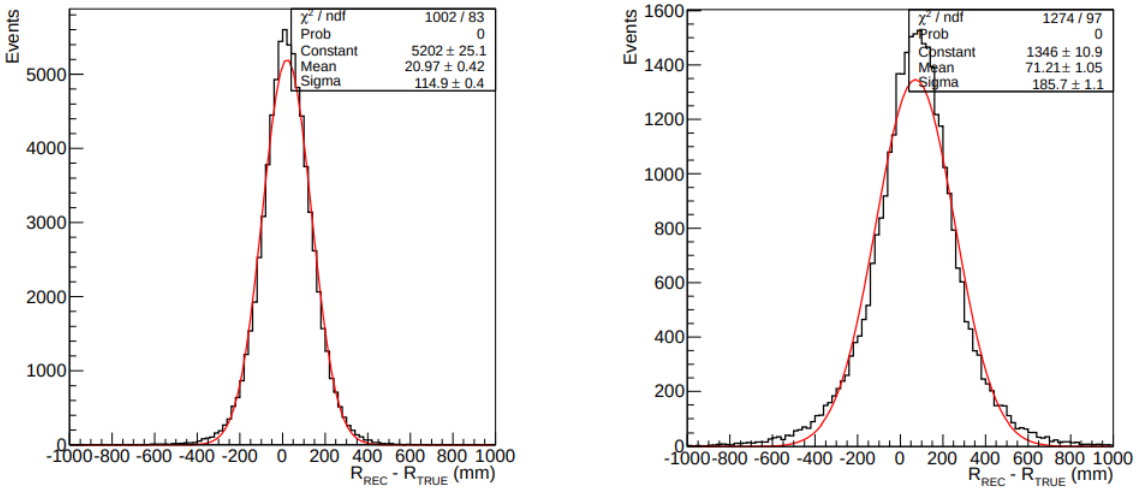

Figure D.10: Distributions of residuals of the radial coordinate for the AdSimple MC-corrected center of charge vertex. IBD positrons are shown on the left, while IBD nGd captures are shown on the right. From [101].

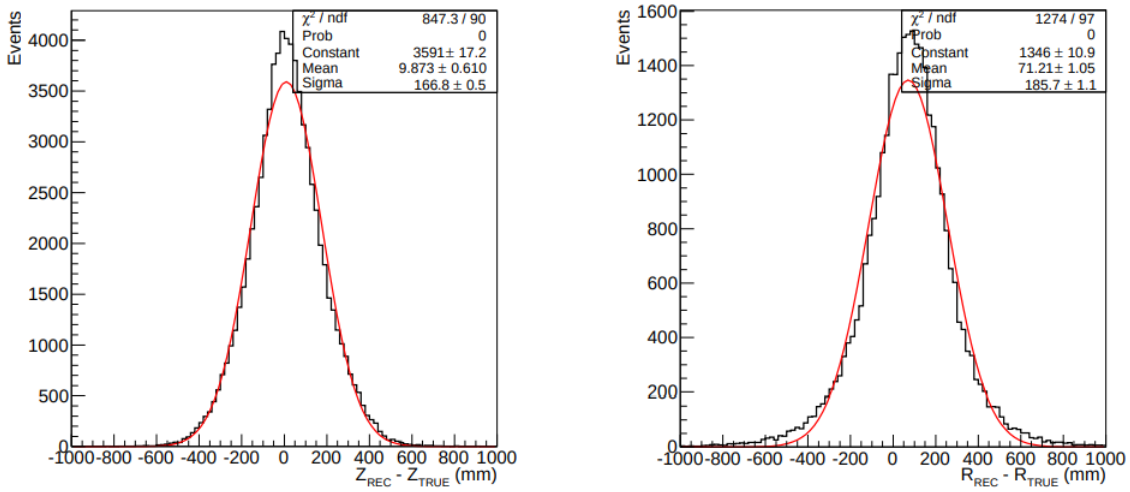

Figure D.11: Distributions of residuals of the vertical coordinate for the AdSimple MC-corrected center of charge vertex. IBD positrons are shown on the left, while IBD nGd captures are shown on the right. From [101].

Appendix E

Looking back, looking forward

E.1 History of neutrino oscillations

In the late 1960s, deep in the Homestake mine located in South Dakota, Ray Davis filled a large tank with tetrachloroethylene, a common dry-cleaning agent, and waited as solar neutrinos interacted with chlorine-37 atoms via the reaction

$$\nu_e + {}^{37}\mathrm{Cl} \longrightarrow {}^{37}\mathrm{Ar} + e^- . \tag{E.1}$$

Every few weeks, from 1970 to 1994, helium was bubbled through the liquid to extract the few dozen argon atoms that were produced in the time since the previous extraction. Each extraction was then placed in a gas-filled proportional chamber, where the ^{37}Ar decays (roughly one per week) were counted over the course of a year or so (long enough to count effectively all of the atoms, given the 35-day half-life of ^{37}Ar). By the early 1970s, the measurements were clearly indicating a reaction rate that was one-third of the prediction derived from the Standard Solar Model (SSM) [102]. This "solar neutrino problem" was interpreted to mean that either the SSM or the experiment was in error, and neutrinos remained, according to the wisdom of the day, massless.

Despite the prevailing belief in masslessness, the possibility of massive neutrinos was considered as early as 1962, soon after the discovery that neutrinos come in separate electron and muon flavors. That year, in the "MNS" paper by Maki, Nakagawa, and Sakata [103], the authors considered the possibility of neutrino flavor mixing, possibly from a nonzero mass. Neutrino mixing had been considered previously by Pontecorvo in 1957 [104], but in the form of neutrino-antineutrino mixing, in analogy with the then-recently discovered phenomenon of neutral kaon mixing [105]. Pontecorvo revisited the subject in 1967, building upon the MNS formalism to describe how oscillations could occur in traveling neutrinos, going so far as to suggest that solar neutrinos could oscillate (well before the first experimental hints of such by the Davis experiment). Still, it would take decades of further observations to prove that these four theorists were correct.

The late 1980s delivered additional anomalies, when the Kamiokande [106] and IMB [107] experiments observed a deficit in the number of atmospheric charged-current ν_μ events relative to expectations. In addition, Kamiokande also confirmed the solar neutrino problem [108], which was yet again confirmed around 1992 by the gallium-based SAGE [109] and GALLEX [110] radiochemical experiments. In 1996, the massive Super-Kamiokande (SK) water Cherenkov detector came online, and in 1998 the SK collaboration published measurements of the zenith angle dependence of the atmospheric neutrino deficit [111]. This geometric dependence was consistent with mass-induced flavor oscillations. Other models, such as neutrino decay or decoherence, were disfavored by the SK data, which was of sufficient quality to provide initial values of θ_{23} and Δm^2_{32} (defined in Section 1.2). Super-Kamiokande's compelling evidence in favor of nonzero neutrino mass would soon receive confirmation from other experiments.

Such confirmation arrived dramatically in 2002, thanks to the Sudbury Neutrino Observatory (SNO) [112]. Owing to its use of heavy water as the target material, SNO had unprecedented sensitivity to neutral current (NC) interactions, which are undergone by all three neutrino flavors. This NC sensitivity stood in addition to SNO's customary sensitivity to charged current (CC) interactions, which (at the relevant energy scale of $\sim$10 MeV) only provide detection of electron neutrinos. As such, SNO was capable of independently measuring both the total and the electron neutrino fluxes. The total flux was in excellent agreement with the SSM, demonstrating that the "missing" neutrinos underlying the solar neutrino problem were, in fact, merely hiding in the form of muon and tau neutrinos. SNO thus succeeded both in confirming the existence of neutrino mass and in resolving the solar neutrino problem, once and for all.

At this point, the existence of neutrino oscillations was no longer a question, but a fact. The precision era had begun, and the Japan-based reactor neutrino experiment KamLAND was one effort that had gotten a head start. KamLAND was initially proposed to search for potential oscillation in 1994, when the solar measurements were still murky, but by 1997 there was enough evidence from the results of Davis, SAGE, and GALLEX to suggest that KamLAND might be able to measure θ_{12} and Δm^2_{21}, if indeed such parameters were responsible for the solar neutrino deficit (and assuming that θ_{12} was sufficiently large, which remained an open question). Around the time of SNO's announcement, KamLAND published results on the disappearance of reactor antineutrinos over long ($\sim$100 km) baselines [113]. KamLAND succeeded in pinning down the value of θ_{12}, in favor of the large mixing angle solution, and furthermore provided what remains the most precise measurement of Δm^2_{21} ever performed.

Meanwhile, the atmospheric results of SK and others on Δm^2_{32} had set off a flurry of successful long-baseline accelerator experiments, such as K2K [114], T2K [115], MINOS [116], NOνA [117], OPERA [118], and ICARUS [119], optimized for this mass splitting and designed to narrow down its value and that of θ_{23}. As the 2010s approached, however, one mixing angle remained elusive: θ_{13}.

The disappearance of electron (anti)neutrinos is controlled by θ_{12} and Δm^2_{21} at longer baselines (such as those employed in solar neutrino experiments and KamLAND) and by θ_{13} and Δm^2_{31} at shorter baselines. As luck would have it, at the energies of reactor antineutrinos, the latter oscillation length is ∼1 km, which is a sufficiently short distance that a relatively modest target mass of ∼10 tons will provide ample statistics from a typical commercial power reactor. By adopting such baselines and target masses, the Chooz experiment in France, as well as the Palo Verde experiment in the USA, set the stage for the reactor-based study of neutrino oscillations. Unfortunately, for Chooz and Palo Verde, sensitivity to oscillations was limited, in large part due to the dependence of a single-detector configuration on modeling the absolute reactor antineutrino flux.[1] These two experiments were unable to discover evidence of oscillation, instead setting upper limits of roughly $\sin^2 2\theta < 0.17$ (for a mass-squared splitting of $|\Delta m^2_{32}| \approx 2.5 \times 10^{-3}\,\mathrm{eV}^2$). Importantly, however, this result excluded $\nu_\mu \to \nu_e$ oscillations as the driver of the atmospheric ν_μ disappearance seen by experiments such as Super-K.

In order to mitigate the uncertainty arising from modeling of the absolute antineutrino flux, the subsequent generation of reactor experiments were designed using identical detectors at different baselines [120]. This would enable the near detectors to measure the antineutrino flux while the far detectors measure any oscillation. Uncertainties on the absolute flux and detection efficiency thus largely cancel in the far/near ratio. The Double Chooz [121], RENO [122], and Daya Bay [123] experiments embarked on this effort in parallel.

E.2 History of reactor neutrino experiments

Although a number of reactor-based experiments were briefly described in E.1, the history of the field merits further discussion, given that this thesis is based on one such experiment. Nuclear reactors have played an important role in experimental neutrino physics since the very beginning. Indeed, the Savannah River reactor provided the antineutrinos that led to the first direct confirmation of the particle's existence by the 1956 Cowan-Reines experiment. Since then, reactors have continued to provide key insights into the nature of the neutrino.

Essentially all reactor neutrino experiments are based on detecting charged-current inverse beta decay (IBD) interactions in a volume of liquid scintillator, observed by photomultiplier tubes. The details of this measurement principle are discussed in Chapter 2. Among its advantages are (a) the fact that the double-pulse signature allows efficient background rejection without the need to evade cosmogenic backgrounds by going deep underground (especially important at km and shorter baselines, where a deep overburden would create challenging logistics), (b) the threshold energy is lower compared to a water Cherenkov detector, and (c) the materials and technology

[1] Rapid deterioration of the gadolinium-doped scintillator was another obstacle.

are inexpensive in comparison to advanced designs such as noble liquid time projection chambers. These advantages have remained pertinent from the Cowan-Reines experiment up to the present day.

Efforts involving reactor neutrinos gained steam in the 1970s, when interest arose in going beyond merely using nuclear reactors to qualitatively confirm the neutrino's existence and interactions. One of the new goals was to quantitatively measure the flux and spectrum of electron antineutrinos produced by nuclear reactors. These efforts bore fruit in the early-to-mid 1980s when a half-dozen experiments[2] published such measurements [124–129]. These early detectors were relatively small, with target masses of a few hundred kg (versus Daya Bay's 80 t at the far site); hence, short baselines (of 10–100 m) were necessary in order to obtain sufficient statistics. The fluxes measured by these experiments were in good agreement with the predictions of the ILL-Vogel model [77, 80–82], which was developed around the same time. Much later, as discussed in Appendix B, these measurements would provide evidence of the so-called reactor antineutrino anomaly [130] associated with the Huber-Mueller reevaluations of the predicted flux [63, 64], which lie $\sim 5\%$ higher than the measured fluxes. In addition, the data from one of these experiments (Bugey-3) would later be combined with Daya Bay's in order to set limits on light sterile neutrino mixing [131].

The 1990s brought the Chooz [132] and Palo Verde [133] experiments (mentioned previously in the context of neutrino oscillations), which employed larger, $\mathcal{O}(10\text{ t})$ detectors so as to acquire useful statistics at longer (km-scale) baselines. Although the search for oscillations was their primary goal, these two experiments were also able to measure the $\bar{\nu}_e$ flux at these longer baselines, comparable to the average baseline for the Daya Bay far site. As with the short-baseline experiments, Chooz and Palo Verde would later indicate a $\sim 5\%$ deficit with respect to the Huber-Mueller flux predictions.

The θ_{13} sector is not the only one in which reactor antineutrinos can play a useful role. The KamLAND experiment used reactors to confirm and complement the solar neutrino results on θ_{12} and Δm^2_{21}, significantly improving the precision on the latter. Located a kilometer underground in Japan's Kamioka mine, in the cavern formerly occupied by the Kamiokande-II detector, KamLAND employed a massive, transparent balloon filled with 1000 t of liquid scintillator [113]. Antineutrinos arrived from some 50 reactors scattered throughout Japan, at a flux-averaged baseline of 180 km. This L/E was well-suited for measuring oscillations driven by Δm^2_{21}, and KamLAND's measurement of this splitting will remain unchallenged for the foreseeable future.

Returning to the topic of θ_{13}, in 1998, Mikaelyan and Sinev [120] rigorously showed that a two-detector (near and far) reactor experiment could overcome the absolute flux uncertainty and allow for precision measurement of $\sin^2 2\theta_{13}$. Although beam experiments also carry the potential to measure θ_{13} by observing $\nu_\mu \to \nu_e$ oscillation, they are susceptible to degeneracies between it and other parameters, such as the mass hierarchy, δ_{CP}, and the other mixing angles. Reactors thus remained a favorable

[2] ILL, Gosgen, Rovno, Krasnoyarsk, Bugey, and Savannah River.

means to a pure θ_{13} measurement, despite the fact that disappearance channels are generally more challenging than appearance channels when measuring a small mixing angle.

Accordingly, in the mid-2000s, the Double Chooz [121], RENO [122], and Daya Bay [123] experiments were proposed. All three shared the same basic design: Near and far detectors at a $\mathcal{O}(1$ km) (far) baseline, using gadolinium-doped liquid scintillator to improve detection efficiency and background rejection. In April 2011, Double Chooz began taking far site data, followed in August by RENO's full configuration, and in December by Daya Bay using a six-detector configuration. In March 2012, with just 55 days of data, Daya Bay announced a 5.2σ discovery of nonzero θ_{13} [96], which was quickly confirmed by RENO in the following weeks [134]. Since then, these experiments have continued to publish increasingly precise measurements of θ_{13} and Δm^2_{ee}, as well as measurements of the absolute flux (confirming the anomaly) [135–137], the spectrum (discovering an as-yet unexplained excess of events around 5 MeV, relative to predictions) [135–137], and, at least in the case of Daya Bay, the time evolution of the flux and spectrum [138], the decomposed fuel isotope components of the flux and spectrum [139], sterile neutrino limits [131], decoherence limits [11], and numerous other valuable results, demonstrating the major scientific utility of reactor neutrino experiments.

Recent years have seen a resurgence of activity in short-baseline (∼5–25 m) reactor experiments, motivated both by the need for precise measurement of the reactor antineutrino flux/spectrum as well as by the prospect of an eV-scale sterile neutrino (as suggested, controversially, by the reactor antineutrino anomaly [130], the SAGE/GALLEX anomalies [140, 141], and the anomalous ν_e results of the LSND and MiniBooNE accelerator experiments [142, 143]). These ton-scale experiements[3] complement each other with different reactor types (commercial or research), baselines/mobility, scintillator materials (liquid or plastic), neutron capture isotopes (Gd or ^{6}Li), and levels of segmentation (3D, 2D, or none). With the exception of SoLiD, all of these experiments have already published valuable results, including new limits on sterile neutrino mixing around 1 eV^2 [150–153], strong rejection of sterile neutrinos as the explanation of the reactor antineutrino anomaly [152, 153], confirmation of the 5 MeV bump [150], and precise measurement of nearly-pure antineutrino spectra from ^{235}U [153]. Intriguingly, NEUTRINO-4 has claimed to observe 4.6σ evidence for sterile neutrino mixing at a Δm^2 of 7.25 eV^2 [154], but time will tell whether this is borne out by other experiments.

Finally, looking to the future, the 20 kt JUNO detector [155] (in southern China) will be by far the largest reactor neutrino detector ever constructed. The driving goal of the JUNO experiment is the determination of the mass hierarchy (see Section E.3.2) without susceptibility to the degeneracies and correlations that complicate the hierarchy

[3] Among them, NEOS [144], DANSS [145], STEREO [146], PROSPECT [147], NEUTRINO-4 [148], and SoLiD [149]

determination in accelerator experiments [156]. JUNO will observe $\bar{\nu}_e$ interactions using more efficient PMTs and significantly larger photocathode coverage in order to achieve sub-3% energy resolution. At this resolution, the two hierarchies produce measurable differences in the high-frequency oscillations due to θ_{13} of the $\bar{\nu}_e$ spectrum. If JUNO is successful, it could determine the mass hierarchy years before other (beam-based) experiments are expected to do so, cementing yet another historical achievement for the humble reactor antineutrino.

E.3 Relevance of θ_{13} to future research

Quantitatively, the value of θ_{13} is broadly useful in the context of beyond-the-Standard-Model (BSM) model building. More concretely, however, θ_{13} is important in the context of unraveling two extremely significant properties of neutrinos: The value of δ_{CP}, and the mass hierarchy (MH). In this context, the importance of θ_{13} arises simply from the fact that δ_{CP} and the MH both influence oscillation probabilities via higher-order terms that are controlled by e.g. $\sin^2\theta_{13}$ or $\sin^2 2\theta_{13}$; thus, a larger θ_{13} (up to a certain point) implies a greater experimental sensitivity to these subtle effects. In turn, δ_{CP} and the MH are pivotal in relation to two of the biggest questions in particle physics: The origin of the the matter-antimatter asymmetry of the Universe, and the Majorana nature (or lack thereof) of the neutrino, respectively.

E.3.1 Baryon asymmetry and δ_{CP}

Most physical processes produce, in aggregate, equal amounts of matter and antimatter. And yet, our present universe contains a clear excess of matter. The origin of this excess thus demands our investigation. Although the specific mechanisms at play remain unknown, together they must fulfill the *Sakharov conditions.* As most of the normal matter in the Universe is baryonic, explaining the matter-antimatter asymmetry largely boils down to explaining *baryogenesis,* the generation of excess baryons over antibaryons. The three Sakharov conditions, in the context of baryogensis, are:

1. Existence of processes that violate baryon number symmetry B; otherwise, every baryon would be created in concert with an antibaryon
2. Existence of processes that violate C and CP symmetry; otherwise the C and CP-conjugates of B-violating interactions would cancel out the asymmetry
3. Lack of thermal equilibrium (at some point in history); otherwise, every process would be counterbalanced by its CPT-conjugate

Within the SM, an asymmetry can theoretically be produced via the mechanism of electroweak baryogenesis. However, the only source of CP violation in the SM

is the complex phase in the CKM matrix, and it is known that this effect alone is not large enough to account for the observed matter-antimatter asymmetry [157]. If the neutrino δ_{CP} is found to deviate significantly from zero (or π), then this would constitute a brand new source of CP violation, and if the effect is large enough, it could fully explain our matter-dominated universe (when harnessed by a BSM mechanism such as leptogenesis). Thus, δ_{CP}, in a sense, could play a critical role in explaining why the universe is not a formless sea of photons. Conversely, if it is found to be small or consistent with zero, then the mystery of baryogenesis will be deepened even further. For these reasons, the measurement of δ_{CP} is a chief driving goal of some of the most ambitious neutrino oscillation experiments underway. This effort would be far more difficult if Nature had chosen a much smaller θ_{13}.

E.3.2 Majorana neutrinos and the mass hierarchy

As is illustrated in Figure 1.1, the ordering of the neutrino masses is not fully known. Determining the mass hierarchy is an important effort in its own right. The MH affects cosmological models, kinematic measurements of the neutrino mass, the measurement of δ_{CP}, and BSM model building, to name a few examples of its importance. It even relates to the question of the very nature of the neutrino mass, as we show here.

In the SM, the neutrino is a simple two-component left-handed *Weyl* spinor. Once mass is added, however, the picture necessarily becomes more complicated. Massive spinors can be classified as either of Majorana or Dirac type. A Majorana spinor is, by definition, one that is invariant under the operation of charge conjugation (i.e., it is identical to its antiparticle). A Dirac spinor, on the other hand, is not self-conjugate, and contains twice as many components.

For a Majorana spinor, the Hamiltonian can contain a term (the Majorana mass) that contributes to the particle's mass and creates particle-particle pairs. Obviously, this is impossible if the particle carries any sort of conserved charge, as is the case for every particle of the SM, except (after electroweak symmetry breaking) the neutrino. Thus, the neutrino could prove to be the first known fundamental Majorana fermion, and if this is so, it would also provide the first example (via the Majorana mass) of lepton number violation.

The possible Majorana nature of the neutrino is closely related to the question of the smallness of the neutrino mass scale. Taking a step back, the simplest model of neutrino mass (which furnishes Dirac neutrinos) can be constructed in analogy with the masses of the other fermions: We introduce a sterile ν_{R} and construct a Yukawa coupling between it, the Higgs field, and the active ν_{R}. The lack of a Majorana mass term means that ν_{R} and ν_{L} can be combined into a single Dirac spinor. Although this construction works, it suffers from the fact that the dimensionless coupling constant must be extremely small, of order m_ν/m_{EW}.

An alternative (and generally more favored) model of neutrino mass is described by the *seesaw mechanism*. In this case, sterile right-handed neutrino fields ν_{R} are

introduced, self-coupled via Majorana mass term near the GUT scale. (Such Majorana masses are not allowed for the fundamental ν_{L} fields as they carry weak hypercharge.) In addition, Dirac mass terms are introduced, linking ν_{R} and ν_{L} at the electroweak scale. The resulting mass matrix contains three small and three large eigenvalues, with the small ones being of the order $m_{\mathrm{EW}}^2/m_{\mathrm{GUT}}$. The corresponding eigenstates behave as Majorana particles: They are self-conjugate, and they have a Majorana mass term (leading to L violation). This model thus naturally generates small neutrino masses. Furthermore, the heavy sterile neutrinos are a fairly generic prediction of GUT and other BSM theories. Finally, even more generic higher dimensional operators can also be shown to generate Majorana mass terms in the effective low-energy theory. For these reasons, the theory community regards Majorana neutrinos as a likely fact of nature. The experimental challenge is then to find experimental evidence of the Majorana neutrino.

The hypothetical process of neutrinoless double beta decay ($0\nu\beta\beta$) holds the key to potentially observing evidence of Majorana neutrinos. In ordinary double beta decay, an isotope emits two electrons and two antineutrinos. A Majorana mass would allow the two antineutrinos to "annihilate", leading to a small peak at the endpoint of the double beta spectrum. If this peak is observed, it would constitute definitive evidence of the Majorana nature of the neutrino, with profound implications for BSM model building.

The rate of $0\nu\beta\beta$ is proportional to the square of the *effective Majorana mass* of the electron neutrino,

$$m_{\beta\beta} \equiv \left| \sum m_i U_{ei}^2 \right|, \tag{E.2}$$

in which the unknown Majorana phases of the U_{ei} can interfere with each other. It turns out that the mass hierarchy has important implications when it comes to the range of possible values of $m_{\beta\beta}$. For a given hierarchy, the three neutrino masses are uniquely determined by the mass m_l of the lightest one, since the values of Δm^2 are known. Then, for a given m_l (and hierarchy), the possible range of $m_{\beta\beta}$ is determined by varying the three Majorana phases. For m_l above ~ 0.1 eV, the two hierarchies give largely the same $m_{\beta\beta}$ ranges. However, at lower m_l the two hierarchies differ considerably. In the case of the inverted hierarchy, $m_{\beta\beta}$ remains above ~ 0.01 eV for all m_l. On the other hand, under the normal hierarchy, the Majorana phases can cancel completely when m_l is between ~ 2 and 8 meV (making $0\nu\beta\beta$ unobservable). Below 2 meV, $m_{\beta\beta}$ is bounded from below, but its range lies an order of magnitude below that for the IH case. See Figure E.1.

The upcoming generation of $0\nu\beta\beta$ experiments seek to probe the entire range of $m_{\beta\beta}$ values allowed by the inverted hierarchy. Therefore, *if it is established that the mass ordering is inverted,* then a null result from these $0\nu\beta\beta$ experiments would demonstrate conclusively that *the neutrino is not a Majorana particle.*[4] Meanwhile,

[4] At least if there are no sterile neutrinos.

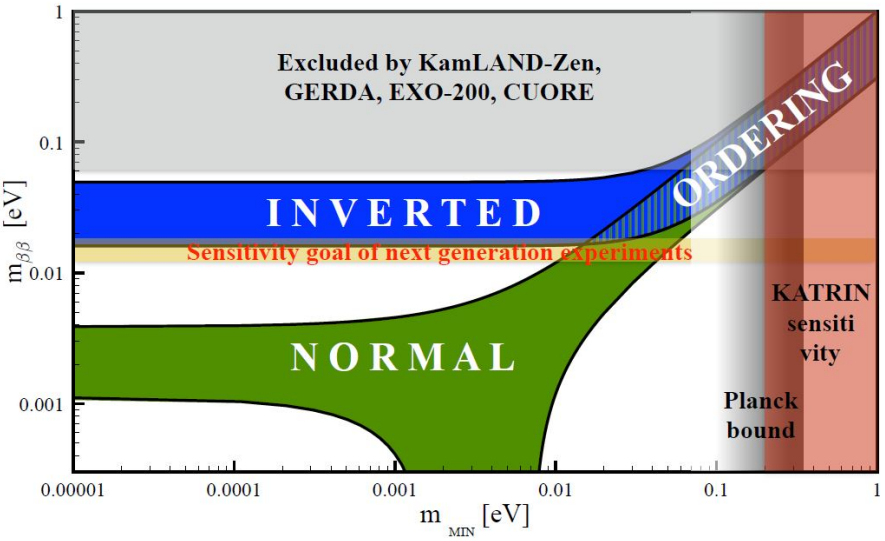

Figure E.1: The dependence of the effective Majorana mass $|m_{\beta\beta}|$ on the minimum neutrino mass m_{MIN}. From [158].

if the normal hierarchy is confirmed, then it becomes impossible to disprove the Majorana nature of the neutrino, since it may be hiding beneath a very small value of $m_{\beta\beta}$. (Of course, proving, as opposed to disproving, the Majorana nature remains possible via observation of $0\nu\beta\beta$.) The mass hierarchy thus is extremely relevant to the task of determining the nature of neutrino mass, and, in turn, the large value of θ_{13} makes it possible to determine the MH via oscillation experiments. As shown by Equation E.2, θ_{13} also plays an important role in the relation between $m_{\beta\beta}$ and the physical masses m_i.

Furthermore, as was mentioned earlier, the determination of the mass hierarchy would carry significant implications for the measurement of δ_{CP}. Many experiments in the pipeline, particularly those that take advantage of matter effects, suffer from a degeneracy between δ_{CP} and the mass hierarchy, forcing them to quote two, possibly very different, confidence intervals for δ_{CP}. The result is a considerably reduced overall sensitivity to leptonic CP violation. Knowledge of the hierarchy would eliminate this issue. With the knowledge that θ_{13} is large, measuring the mass hierarchy is an easier prospect in general, but in particular it allows for experiments, such as JUNO, that aim to efficiently determine the MH independently of δ_{CP}. This accelerates the timeframe of MH determination, while also providing improved sensitivity to long-term δ_{CP}(+MH) experiments such as DUNE and Hyper-Kamiokande. θ_{13}, via its connection to the mass hierarchy, thus also provides an indirect boost to the δ_{CP} measurement effort, in addition to its direct influence on the oscillation probability.