The Dissertation Committee for Stanley Reid Seibert

certifies that this is the approved version of the following dissertation:

A Low Energy Measurement of the ^{8}B Solar Neutrino Spectrum at the Sudbury Neutrino Observatory

Committee:

Joshua R. Klein, Supervisor

Sacha E. Kopp

Karol Lang

Duane E. Dicus

Pawan Kumar

A Low Energy Measurement of the ^{8}B Solar Neutrino Spectrum at the Sudbury Neutrino Observatory

by

Stanley Reid Seibert, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2008

This works is dedicated to Tim Horton's, without whose sugary delights I would have surely fallen asleep on shift and destroyed up the experiment.

Acknowledgments

Physics is enormous fun in large part due to the interesting people you get to work with. First and foremost, I would like to thank my advisor, Josh Klein, for making this thesis possible. Josh is the consummate experimental optimist, believing nearly any problem can be overcome with the right combination of clever ideas and hard work. He will happily supply a stumped graduate student with an endless list of crazy and creative suggestions, and entertain new ideas flowing the other direction as well. Much to the annoyance of said graduate student, however, he also refuses to allow people to give up when the right solution turns out to be more difficult than expected. Through it all, he treats his graduate students as real people and colleagues, not units of labor, a trait which is (sadly) less common among advisors than one might hope. Without a doubt, I cannot imagine a better advisor, or a better graduate experience.

Down in the concrete bunker that is ENS 16N, Aubra Anthony, Chris Tunnell, Melissa Jerkins, and Julia Majors made the lab a pleasant and fun place to be. They always offered a sympathetic ear when things were not going well, and had great stories to liven up a dull afternoon. Chris taught me many things, including the proper air hose safety and alternate fillings for piñatas. Most especially, he was a constant reminder never to be complacent about my

knowledge or my way of doing things. If there was a weakness in either, Chris would find it and pry it open like a heat-seeking crowbar.

As part of my extended SNO-family, I have enjoyed working with many interesting and talented people. Huaizhang Deng, Chris Kyba and Noël Gagnon taught me how to debug and repair broken electronics. Gene Beier was generous with advice and resources for our fledgling group at UT and Bill Heintzelman was always available with analysis ideas. I am amazed by the Reverend Doctor Jeff Secrest, who is like a machine, harvesting the work that everyone else has forgotten about. At Oxford, I had many great conversations with Steve Biller, Nick Jelley, and Helen O'Keefe.

In completing the analysis described in this thesis, am indebted to both Monica Dunford and Gabriel Orebi Gann. While Monica and I overlapped on SNO for only two years, her thesis has been my nearly constant companion while performing this analysis. She is by far the hardest working physicist I have ever met, and provided an immense foundation of work to build upon. Thankfully, when she graduated I was not left alone with the futile task of filling her shoes. Gabriel and I were able to each take one shoe and push forward. Through the magic of the Internet and instant messaging, Gabriel has been my virtual officemate for the last 2 years. Both a fantastic colleague and a great friend, Gabriel was always there to carry on the work (with a disturbing amount of cheer) when I couldn't type or think anymore.

I also want to thank my family and friends for always being there, even when I disappeared for months at a time in Sudbury or in my code. My parents, without realizing it, taught me how to be a good experimentalist. Bookkeeping, diagnosing broken medical equipment, and learning how to solder were all activities I was not terribly excited about as a kid, but have

become enormously important skills now. Although the RAT software package does not appear in this thesis, I want to thank my sister, Sheryl Seibert, for producing the disgustingly cute logo for the program.

Most of all, I am grateful to Lydia Pyne for her intellectual and emotional support all these years. She is truly a kindred spirit with a similarly warped view of the world, and she thinks weasels are funny animals. It doesn't get any better than that.

<div align="right">

STANLEY REID SEIBERT

</div>

The University of Texas at Austin

August 2008

A Low Energy Measurement of the ^{8}B Solar Neutrino Spectrum at the Sudbury Neutrino Observatory

Stanley Reid Seibert, Ph.D.

The University of Texas at Austin, 2008

Supervisor: Joshua R. Klein

The Sudbury Neutrino Observatory has previously shown that ^{8}B solar neutrinos undergo flavor transformation between the Sun and the Earth. This work presents a joint analysis of one-third of the pure D$_2$O and salt phase data sets with an electron energy threshold of 3.5 MeV. A measurement of the neutral current interaction rate, sensitive to all neutrino flavors, gives a total ^{8}B flux of $4.79 \pm 0.22\,(\text{stat})\,\pm 0.13(\text{syst}) \times 10^6\,\text{cm}^{-2}\text{s}^{-1}$. This is in agreement with the prediction of the Standard Solar Model.

For the first time we obtain the ν_e survival probability separately for downward-going (day) and upward-going (night) solar neutrinos. No significant distortion is observed day or night for ^{8}B neutrinos with energies greater

than 6 MeV. The lack of distortion, but overall suppression of electron neutrinos relative to the total flux, is consistent with matter-induced neutrino transformation in the Sun and the large mixing angle solution.

Contents

List of Tables

xix

List of Figures

xxiii

xxviii

Chapter 1

Introduction

In this thesis I describe a low energy threshold re-analysis of the data collected at the Sudbury Neutrino Observatory in the first two phases of operation, from 1999 through 2003. With a detector energy threshold of 3.5 MeV, this is the lowest energy threshold ever used by a water Cherenkov neutrino experiment. The increased statistics give us a much better total measurement of the ^{8}B solar neutrino flux and energy spectrum, leading to improved constraints on the solar neutrino mixing parameters.

Chapter 2 briefly reviews the current state of neutrino physics, solar neutrinos, and solar neutrino experimental results, motivating the effort to make an improved solar measurement. Chapter 3 describes the Sudbury Neutrino Observatory (SNO), its unique capabilities among solar neutrino experiments, and its previous results. The overall design of the low energy analysis is described in Chapter 4. Energy reconstruction is a key part of this analysis, so SNO's energy reconstruction method and an improved approach to spatial variation of energy scale is discussed in Chapter 5. Chapter 6 is an overview of some of the verification work done to validate our background model. In

Chapter 7, I introduce the kernel density estimation method of PDF construction, and show how it can be performed much more efficiently than currently available implementations. The kernel estimation method enables a very flexible approach to maximum likelihood fitting and neutrino signal extraction, described in Chapter 8. Finally, in Chapter 9 I present the results of a low-energy ^{8}B neutrino measurement using data from the first two phases of SNO.

Chapter 2

Neutrino Theory and Solar Neutrinos

B^8: *The existence [in the Sun] is doubtful ... This nucleus is not*
very important for astrophysics.
—*Hans Bethe, 1939[1]*

One generation's great discoveries often become the next generation's tools. Such is the progression we are seeing now in neutrino physics. Originally proposed in an act of desperation by Wolfgang Pauli[2] in 1930, it was another 26 years until the neutrino was finally observed[3], and another 68 years before the non-zero mass of the particle was established through neutrino oscillation[4]. Many experiments are planned or underway to finally measure the neutrino masses, and precisely constrain the oscillation parameters.

Even with the remaining mysteries, neutrinos are already being transformed from objects of study into observational tools. Neutrinos are being used to probe regions which would normally be inaccessible to electromag-

netic measurements. Through the measurement of neutrinos, we are beginning to see into the center of the Sun[5], the core of the Earth[6], and distant supernovae[7]. Our understanding of fundamental physics will advance as neutrinos may someday help to explore CP violation, non-standard interactions, leptogenesis, lepton/quark unification in grand unified theories, and possible connections between the cosmological constant and neutrino mass.

Such studies require a model of neutrino physics with which to interpret their results, so it is very important that we carefully examine the experimental data available. One never knows what new physics might be lurking just outside of a previous analysis box.

This chapter contains a brief overview of our current theoretical understanding of neutrinos, as well as a review of solar neutrino experimental results, and prospects for new physics.

2.1 The Extended Standard Model

The Standard Model of particle physics, which came together in the late 1970s, gave structure to the explosion of new particles and interactions discovered in the 20th century. It continued to successfully predict new experimental results for another 20 years until experimental evidence from Super Kamiokande[4] gave significant support to the neutrino oscillation hypothesis. The Standard Model does not predict any lepton flavor transformation as was later observed in solar neutrinos, atmospheric neutrinos, accelerator neutrinos and neutrinos from nuclear reactors. Therefore, the model is clearly incomplete in some way. A straightforward extension of the Standard Model, however, does accurately describe currently observed neutrino phenomena. We will briefly present this

Extended Standard Model[1] in this section.

The Extended Standard Model, like the original, is a Yang-Mills quantum field theory. The theory is characterized by $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge invariance. The matter content of the theory is packed into representations of this composite group. In order to make the Lagrangian invariant under local ("gauge") transformations, we must introduce massless vector bosons in the adjoint representation of this group. The bosons are often written out as

$$g^a, W^b, B, \qquad (2.1)$$

where $a = 1..8$ and $b = 1..3$, corresponding to the number of generators of $SU(3)$ and $SU(2)$, respectively. The strong nuclear force is associated with g^a, and the weak nuclear force and electromagnetic force are a linear transformation of the W^b and B fields.

These vector bosons facilitate interactions between matter, with a coupling strength that depends on the charge of the matter field, which relates to the gauge transformation properties, and universal dimensionless coupling constants. There are three such coupling constants, α_C, α_L and α_Y, corresponding to the three compact subgroups of the Standard Model. The model makes no prediction of the magnitude of these parameters and, so, they must be measured in experiments.

The matter content of the theory is comprised of 3 nearly identical families of fermionic fields. These fields are often written in a chiral representation[2]

---

[1]There is no obvious consensus on what to call this extension to the Standard Model. Many still call it the "Standard Model," but this terminology is somewhat misleading. It suggests that the original Standard Model included neutrino mass as a free parameter, when in fact neutrino mass was excluded by construction[8].

[2]The discussion in this section uses the notation and some results from [9].

$$L_{Li} = \begin{pmatrix} \nu_{Li} \\ E_{Li} \end{pmatrix}, Q_{Li} = \begin{pmatrix} U_{Li} \\ D_{Li} \end{pmatrix}, E_R, \nu_{Ri}, U_{Ri}, D_{Ri}, \qquad (2.2)$$

where i denotes the particle family. The L and R subscripts identify the two-component Weyl fermions as either having left or right-handed chirality. The charged lepton fields, electron, muon and tau, correspond to $E_{Li,Ri}$, and the corresponding electron, muon, and tau neutrinos are $\nu_{Li,Ri}$. The up-type quarks (up, charm, top) are denoted by $U_{Li,Ri}$, and the down-type quarks (down, strange, bottom) are denoted by $D_{Li,Ri}$. Quarks are also arranged into triplets of the $SU(3)_C$ color group, but we have suppressed the color indices above.

Equation 2.2 highlights the $SU(2)_L$ structure of the theory. Left-handed fields are packed into $SU(2)$ doublets, while right-handed fields are singlets, unaffected by $SU(2)$ transformations. The inclusion of ν_{Ri} is where the original Standard Model and the Extended Standard Model differ. Weinberg deliberately excluded the right-handed neutrino, as all available evidence indicated that parity was maximally violated in weak decays, i.e. all neutrinos had left-handed chirality (and antineutrinos, right-handed chirality). In addition, searches for the neutrino mass were consistent with zero mass, and in Section 2.1.1 we find that neutrino mass required the existence of a ν_R field.

The Standard Model contains one more, so far unobserved, field called the Higgs field. The Higgs is a complex scalar field which is a doublet of $SU(2)_L$:

$$\phi = \begin{pmatrix} \phi_0 \\ \phi_- \end{pmatrix}. \qquad (2.3)$$

This field interacts with the fermions through a Yukawa coupling term

$$-\mathcal{L}_{\text{Yukawa}} = Y^d_{ij} \bar{Q}_{Li} \phi D_{Rj} + Y^u_{ij} \bar{Q}_{Li} \tilde{\phi} U_{R_j} + Y^l_{ij} \bar{L}_{Li} \phi E_{Rj} + Y^\nu_{ij} \bar{L}_{Li} \tilde{\phi} \nu_{Rj} + \text{h.c.}. \quad (2.4)$$

These terms preserve the $SU(2)_L$ invariance of the theory, but an effective potential can break $SU(2)_L$ by generating an expectation value for ϕ in the low-energy vacuum

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \quad (2.5)$$

We get mass-like terms in the low energy theory, where

$$m^f_{ij} = Y^f_{ij} \frac{v}{\sqrt{2}}. \quad (2.6)$$

The implications of these mass terms for neutrinos are discussed further in the next section.

2.1.1   The Neutrino Mass Matrix

The right-handed neutrino in the extended Standard Model is a singlet of all gauge groups. Gauge invariance allows mass terms derived from electroweak symmetry breaking as mentioned before

$$-\mathcal{L}_{\mathcal{D}} = m^\nu_{ij} \bar{\nu}_{Ri} \nu_{Lj}, \quad (2.7)$$

but also a new mass term not available to the other fermions,

$$-\mathcal{L}_{\mathcal{M}} = M^\nu_{ij} \bar{\nu}_{Ri} \nu^c_{Rj}. \quad (2.8)$$

This is the *Majorana* mass term. If $M_{ij}^{\nu} \neq 0$, then the neutrino states can be transformed into a basis where each neutrino corresponds to its own anti-particle, the definition of a Majorana fermion. Neutrino oscillation experiments are not sensitive to the difference between the Dirac and Majorana mass, so we will not explore this issue further, and simply consider the Dirac case.

If we have lepton flavor conservation, then m_{ij}^{ν} must be a diagonal matrix. However, if we relax this requirement (motivated by compelling experiments) to just lepton number conservation, then the mass terms can be written in matrix notation:

$$-\mathcal{L}_D = \left(\begin{array}{ccc} \bar{\nu}_{eR} & \bar{\nu}_{\mu R} & \bar{\nu}_{\tau R} \end{array} \right) \left(\begin{array}{ccc} m_{ee} & m_{e\mu} & m_{e\tau} \\ m_{\mu e} & m_{\mu\mu} & m_{\mu\tau} \\ m_{\tau e} & m_{\tau\mu} & m_{\tau\tau} \end{array} \right) \left(\begin{array}{c} \nu_{eL} \\ \nu_{\mu L} \\ \nu_{\tau L} \end{array} \right) + h.c. \qquad (2.9)$$

A spectral decomposition of the matrix reveals the states of definite mass:

$$\left(\begin{array}{ccc} \bar{\nu}_{eR} & \bar{\nu}_{\mu R} & \bar{\nu}_{\tau R} \end{array} \right) \mathbf{U} \left(\begin{array}{ccc} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{array} \right) \mathbf{U}^{-1} \left(\begin{array}{c} \nu_{eL} \\ \nu_{\mu R} \\ \nu_{\tau R} \end{array} \right) + h.c., \qquad (2.10)$$

where $\mathbf{U}$ is the Pontecorvo-Maki-Nakagava-Sakata (PMNS) mixing matrix, a unitary matrix which transforms from the mass eigenstates to the weak, or flavor, eigenstates. Many of the degrees of freedom in $\mathbf{U}$ amount to unobservable phase conventions for the fermion fields[3]. The remaining degrees of freedom

---

[3]The existence of Majorana mass terms limits the number of complex phases that can be canceled in this way, so there can also be two additional *Majorana phases* which are only observable in lepton number violating processes, like double beta decay.

are parameterized as three angles: θ_{12}, θ_{23}, θ_{13}, and complex phase δ. The three-neutrino mixing matrix can be written as the product of three separate two-neutrino mixing matrices.

$$
U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13}e^{i\delta} \\ 0 & 0 & 0 \\ -s_{13}e^{-i\delta} & 0 & c_{13} \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix},
$$

$$(2.11)$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. This factorization turns out to be convenient when the mass splittings $\Delta m_{21}^2 = m_2^2 - m_1^2$ and Δm_{32}^2 are very different. A two neutrino system becomes a very good approximation for many experiments.

With only two neutrinos, the mixing matrix has only one free parameter, an angle θ. In solar experiments the two states are ν_e and $\nu_Y \equiv \nu_\mu + \nu_\tau$. Muon and tau neutrinos cannot be distinguished in solar experiments because there is not enough energy to produce the charged lepton that would identify the flavor.

The two neutrino mixing matrix is

$$
U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.
$$

$$(2.12)$$

In the flavor basis, the Hamiltonian for a two neutrino state (ν_e, ν_Y) of energy E, is

$$
\mathbf{H} = \frac{m_1^2 + m_2^2}{4E} + \frac{\Delta m_{21}^2}{4E} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}
$$

$$(2.13)$$

When $\theta \neq 0$, the off-diagonal elements in the Hamiltonian lead to flavor trans-

formation between ν_e and ν_Y. Of particular interest for disappearance experiments is the survival probability,

$$P_{\nu_e \to \nu_e} = < \nu_e | e^{-it\mathbf{H}} | \nu_e > = 1 - \sin^2 2\theta \sin^2 \frac{\Delta m_{21}^2}{4E} t. \qquad (2.14)$$

A survival probability with time as the independent variable is not very useful, but for highly relativistic neutrinos we can convert time to distance from the source using c. This gives the standard two neutrino formula,

$$P_{\nu_e \to \nu_e} = 1 - \sin^2 2\theta \, sin^2 \frac{1.27 \Delta m_{21}^2 L}{E}, \qquad (2.15)$$

where Δm_{21}^2 is in units of eV^2, L is the distance between source and detector in km, and E is the neutrino energy in GeV.

2.1.2 Current Mixing Parameters

As of the 2006 Review of Particle Physics[10], our best measure of the mixing angles and masses are:

$$\sin^2(2\theta_{12}) = 0.86^{+0.03}_{-0.04} \qquad (2.16)$$

$$\sin^2(2\theta_{23}) > 0.92 \; (90\% \text{ C.L.}) \qquad (2.17)$$

$$\sin^2(2\theta_{13}) < 0.19 \; (90\% \text{ C.L.}) \qquad (2.18)$$

$$\Delta m_{21}^2 = (8.0^{-0.4}_{0.3}) \times 10^{-5} \text{eV}^2 \qquad (2.19)$$

$$|\Delta m_{32}^2| = (2.74^{+0.44}_{-0.26}) \times 10^{-3} \text{eV}^2, \qquad (2.20)$$

where the PDG value for $|m_{32}^2|$ has been substituted with the more precise value obtained by MINOS[11].

Combining the individual mixing angles, a 90% C.L. range for the magnitude of the elements of U is obtained by [9],

$$|U_{PMNS}|_{2\sigma} = \begin{pmatrix} 0.80 \to 0.84 & 0.53 \to 0.60 & 0.00 \to 0.17 \\ 0.29 \to 0.52 & 0.51 \to 0.69 & 0.61 \to 0.76 \\ 0.26 \to 0.50 & 0.46 \to 0.66 & 0.64 \to 0.79 \end{pmatrix}. \qquad (2.21)$$

The PMNS matrix is very different from the quark mixing matrix, as is discussed further in Section 2.5.1.

2.2 Neutrino Propagation in Matter

The addition of matter to the system brings a whole new range of phenomena to neutrino oscillations. In matter, the neutrino propagates through a sea of electrons. The coherent forward scattering contribution to the propagator acts like an effective potential that is different for electron neutrinos compared to muon and tau neutrinos. Electron neutrinos can interact with the electrons through both charged current and neutral current channels, whereas the muon and tau neutrinos can only interact via the neutral current. The idea was first proposed by Wolfenstein[12] and later applied to the Sun by Mikheyev and Smirnov[13], so it is now known as the *MSW effect*.

The effective potential due to matter adds to the Hamiltonian in Equation 2.13 to give

$$\mathbf{H} = \frac{\Delta m_{21}^2}{4E} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix} + \begin{pmatrix} \sqrt{2}G_F n_e & 0 \\ 0 & 0 \end{pmatrix}, \qquad (2.22)$$

where G_F is the Fermi coupling constant, and n_e is the number density of electrons in the medium[4]. While not obvious in this form, the Hamiltonian can be rewritten in terms of a new set of mixing parameters,

$$\mathbf{H} = \frac{\Delta \tilde{m}_{21}^2}{4E} \begin{pmatrix} -\cos 2\tilde{\theta} & \sin 2\tilde{\theta} \\ \sin 2\tilde{\theta} & \cos 2\tilde{\theta} \end{pmatrix}, \tag{2.23}$$

where,

$$A \equiv 2\sqrt{2}EG_F n_e \tag{2.24}$$

$$\cos 2\tilde{\theta} = \frac{-A/\Delta m_{21}^2 + \cos 2\theta}{\sqrt{(A/\Delta m_{21}^2 - \cos 2\theta)^2 + \sin^2 2\theta}} \tag{2.25}$$

$$\tilde{m}_1^2 = \frac{A}{2} - \frac{1}{2}\sqrt{(A/\Delta m_{21}^2 - \cos 2\theta)^2 + (\Delta m_{21}^2)^2 \sin^2 2\theta} \tag{2.26}$$

$$\tilde{m}_2^2 = \frac{A}{2} + \frac{1}{2}\sqrt{(A/\Delta m_{21}^2 - \cos 2\theta)^2 + (\Delta m_{21}^2)^2 \sin^2 2\theta}. \tag{2.27}$$

Maximal mixing is obtained when $A/\Delta m_{21}^2 = \cos 2\theta$. A neutrino starting from a very dense region, propagating into the vacuum (solar neutrinos, for example) will pass through this region of maximal mixing. If the vacuum mixing angle θ is small enough, then the two matter eigenstates will have very similar masses, leading to a high probability of a level crossing between the two.

In the case of high energy neutrinos and if the mixing angle is large, passage from a high density region to the vacuum in the adiabatic limit produces a pure mass state with the heavier of the two masses. To see this, consider

---

[4]The diagonal portion of this Hamiltonian has already been subtracted out, as it contributes just to an overall phase of the neutrino state.

the case where $\Delta m_{21}^2/4E \ll \sqrt{2}G_F n_e$. The Hamiltonian reduces to:

$$\mathbf{H} = \begin{pmatrix} \sqrt{2}G_F n_e & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.28}$$

A neutrino produced in the pure ν_e state is already the more massive of the two eigenstates of the matter Hamiltonian. If n_e slowly approaches zero as the neutrino propagates, then the ν_e will adiabatically transform into the ν_2 eigenstate of the vacuum Hamiltonian, Equation 2.13. This turns out to be a reasonable model for much of the spectrum of ^{8}B solar neutrinos, which is discussed in more detail in the next section.

2.3 Solar Neutrinos

Nuclear reactions, including fission, fusion and nuclear beta decay, produce neutrinos and anti-neutrinos. One of the great triumphs of nuclear and astrophysics was the discovery that the primary power generation mechanism in stars is nuclear fusion of hydrogen. For the neutrino physicist, this is very exciting because it means there is a large, powerful source of neutrinos in our backyard, with an enormous baseline between the neutrino source and terrestrial-based detectors. This makes the Sun-Earth system a large "neutrino interferometer" of sorts, which can be very sensitive to new physics.

2.3.1 Fusion Paths

Two major fusion sequences have been identified in stars: the *pp chain* and the *CNO cycle*[1]. Both convert hydrogen to helium in stars, with the pp
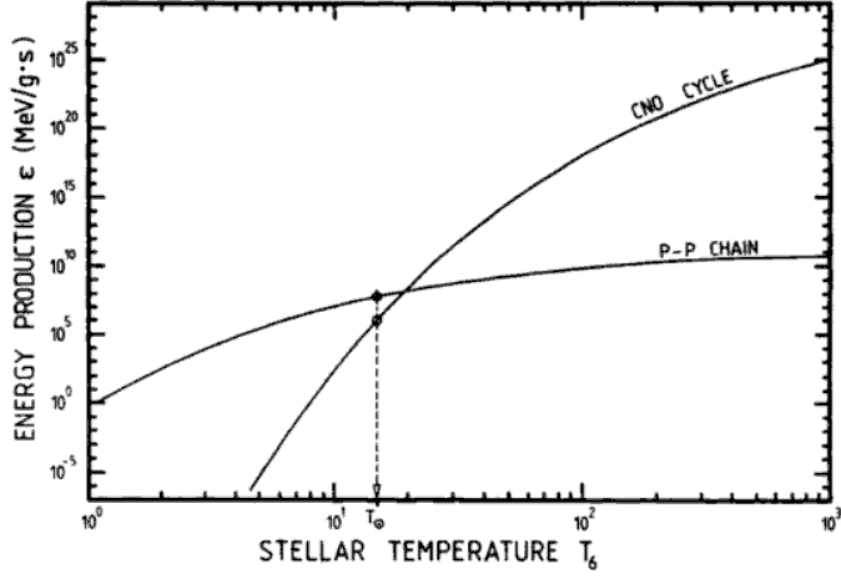
Figure 2.1: Energy output from pp chain and CNO cycle as a function of stellar core temperature. Figure from [14].

chain dominating in stars with lower core temperatures, such as the Sun. Figure 2.1 shows the energy contribution of the two processes as a function of temperature. Both sequences ultimately facilitate the reaction:

$$4p + 2e^- \rightarrow^4 \text{He} + 2\nu_e (26.7 MeV). \qquad (2.29)$$
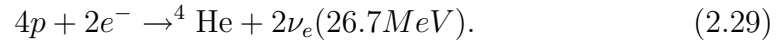
The pp chain is a multi-path reaction, shown in Figure 2.2, which produces neutrinos at five of the stages. These are[15]:

- pp neutrinos: $p + p \rightarrow d + e^+ + \nu_e$ ($E_\nu < 0.42\,\text{MeV}$)

- pep neutrinos: $p + e^- + p \rightarrow d + \nu_e$ ($E_\nu = 1.44\,\text{MeV}$)

- hep neutrinos: $^3\text{He} + p \rightarrow^4 \text{He} + e^+ + \nu_e$ ($E_\nu < 18.77\,\text{MeV}$)

14
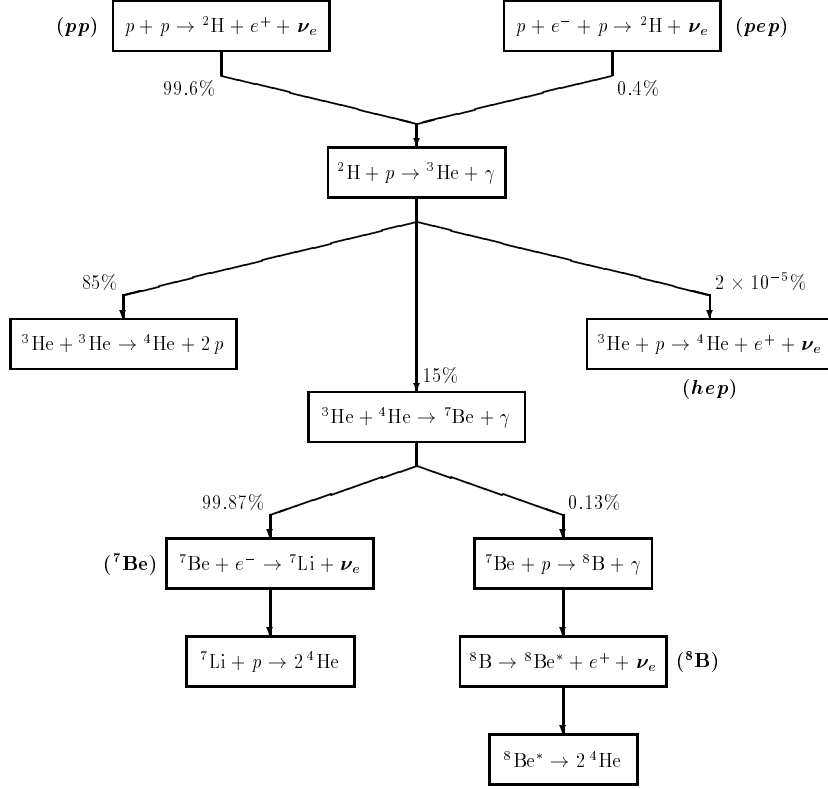
Figure 2.2: Fusion sequence of the pp-chain in the Sun. Figure from [16].

- ^{7}Be neutrinos: $^{7}\text{Be} + e^{-} \rightarrow^{7} \text{Li} + \nu_{e}$ ($E_{\nu} = 0.384, 0.862\,\text{MeV}$)

- ^{8}B neutrinos: $^{8}\text{B} \rightarrow^{8} \text{Be}^{*} + e^{+} + \nu_{e}$ ($E_{\nu} < 15\,\text{MeV}$)

It is important to note that the Sun is a pure ν_{e} source[5]. If muon or tau neutrinos are observed coming from the Sun, then it is a clear sign of flavor transformation.

The rates of all these reactions can be calculated theoretically for the Sun given the chemical composition, surface temperature, luminosity, mass,

---

[5]The much less active CNO cycle, not discussed here, also produces only electron neutrinos.

and many other input parameters. These models have been refined and tested so much over the years that the current model is now called the Standard Solar Model (SSM), in analogy to the Standard Model of particle physics. To distinguish improved models from previous versions, they are often identified by the author initials and the year. As of this writing, the latest model is BS2005[17].

Figure 2.3 shows the energy spectrum and total fluxes for the different neutrino reactions in the Sun. The pp neutrinos have the highest overall flux, but the lowest energy. Of particular interest to SNO, the ^{8}B neutrino spectrum extends to much higher neutrino energies, but with a greater flux than the hep neutrinos. Additionally, ^{8}B neutrinos are produced much closer to the center of the Sun, as shown in Figure 2.4, providing an observational window to the conditions in the stellar core. BS2005 predicts a total ^{8}B neutrino flux of $(5.69 \pm 0.91) \times 10^6 \mathrm{cm}^{-2}\mathrm{s}^{-1}$.

2.3.2 Vacuum-Matter Transition

With large mixing, the MSW effect predicts a transition from "matter-dominated" to "vacuum-dominated" oscillation as a function of neutrino energy[19], where these two terms refer to the relative importance of the vacuum and matter Hamiltonians in Equation 2.22. In the matter-dominated regime, we observe a pure ν_2 state coming out of the Sun. The survival probability in this limit,

$$P_{\nu_e \to \nu_e} = \sin^2 \theta_{12}, \tag{2.30}$$

is independent of energy. In the vacuum-dominated regime, neutrinos produced over a wide range of locations in the Sun are all at different phases in
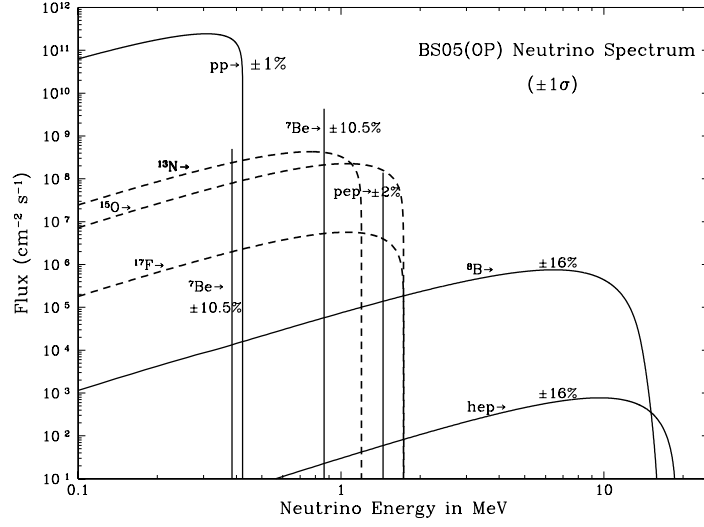
Figure 2.3: Neutrino energy spectrum, as predicted by the BS2005 model. Figure from [17].
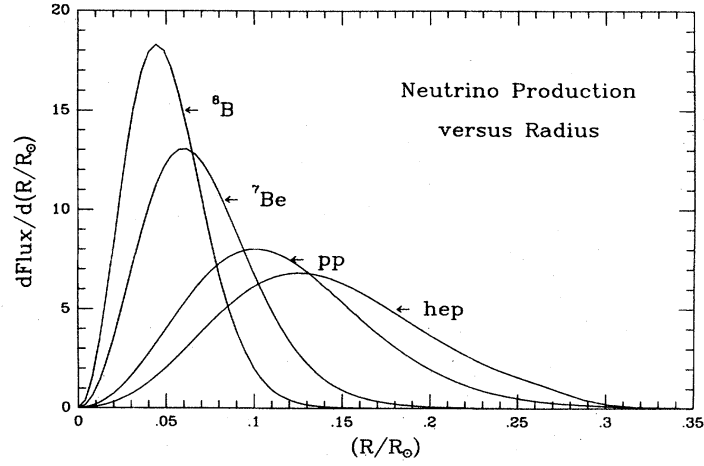
Figure 2.4: Production of neutrinos as a function of radius in the Sun. Figure from [18].
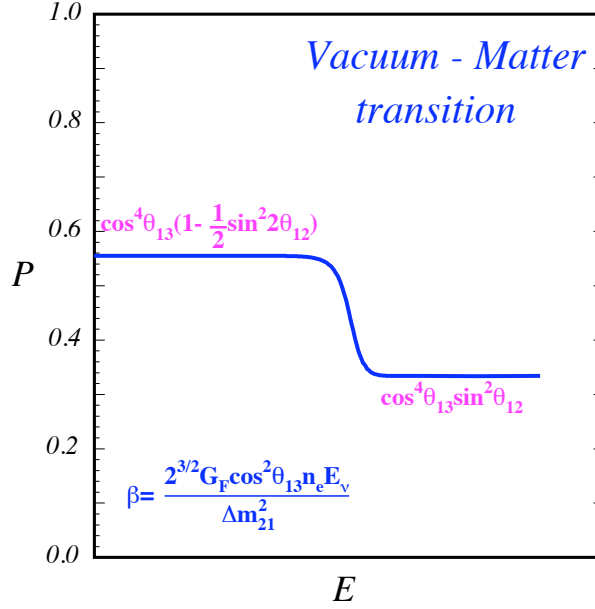
Figure 2.5: Transition from vacuum to matter-dominated oscillations in the Sun as a function of neutrino energy. β sets the energy of the transition point. Figure from [19].

their oscillation when they reach Earth. The survival probability in this case becomes the average of Equation 2.15,

$$P_{\nu_e \to \nu_e} = 1 - \frac{1}{2}\sin^2 2\theta_{12}, \tag{2.31}$$

also independent of energy. In between the two limits, there is an energy dependent survival probability. Figure 2.5 shows the transition region schematically.

2.3.3 Day/Night Effect

Terrestrial detectors of solar neutrinos are constantly in motion due to the Earth's rotation. During the day, the incoming neutrinos pass through at most
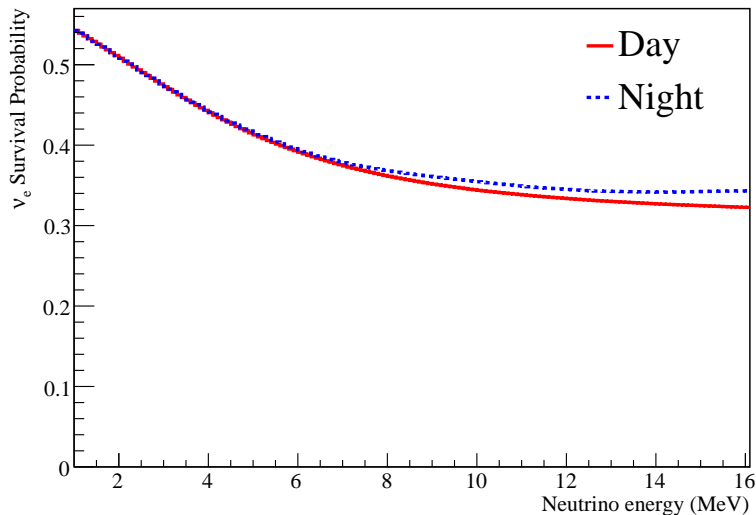
Figure 2.6: Comparison of day and night survival probabilities for $\Delta m_{21}^2 = 8.0 \times 10^{-5} \text{eV}^2$ and $\sin^2 2\theta_{12} = 0.86$. Probability tables generated by O. Simard[20].

a few kilometers of rock to reach the target, whereas at night, neutrinos go through thousands of kilometers of rock. This leads to a day-night asymmetry of the survival probability of electron neutrinos. Figure 2.6 shows the day and night survival probability for the current best fit solar neutrino mixing parameters. The slight increase in the survival probability at night indicates a regeneration of ν_1 in the Earth.

2.4 Experimental Results

The amount of independent evidence for neutrino flavor transformation under the three neutrino oscillation model is immense. Oscillations have been observed for neutrinos over vastly different energy scales and detector baselines.

19

In this section we will focus on solar neutrino experiments and KamLAND as most relevant to SNO. The SNO experiment itself is discussed in detail in Chapter 3, so experimental results from SNO are deferred to Section 3.8.

2.4.1 Homestake Chlorine Experiment

Arguably, the first evidence for neutrino flavor transformation was found in the solar sector, although it was not initially recognized as such. The first solar neutrinos were observed by Davis[21] with a tank of 615 tons of C_2Cl_4 installed in the Homestake mine in South Dakota. With 4200 m.w.e. of shielding from cosmic rays, the Davis experiment could search for the rare process

$$^{37}Cl + \nu_e \rightarrow \, ^{37}Ar + e^-. \tag{2.32}$$

The 814 keV energy threshold makes this reaction sensitive to ^{7}Be, pep, ^{8}B, and hep neutrinos, with ^{8}B comprising the bulk of the detected neutrinos. Assuming that all of the detected neutrinos were from ^{8}B, the group was able to set a limit, using 108 data collection runs spread over a 24 year period (see Figure 2.7), of

$$\Phi_{^8B} < (2.25 \pm 0.21) \times 10^6 cm^{-2} s^{-1}, \tag{2.33}$$

which is less than half the prediction of the Standard Solar Model. The initial version of this measurement, combined with an early version of the SSM in 1968 started what later became known as the *solar neutrino problem.*
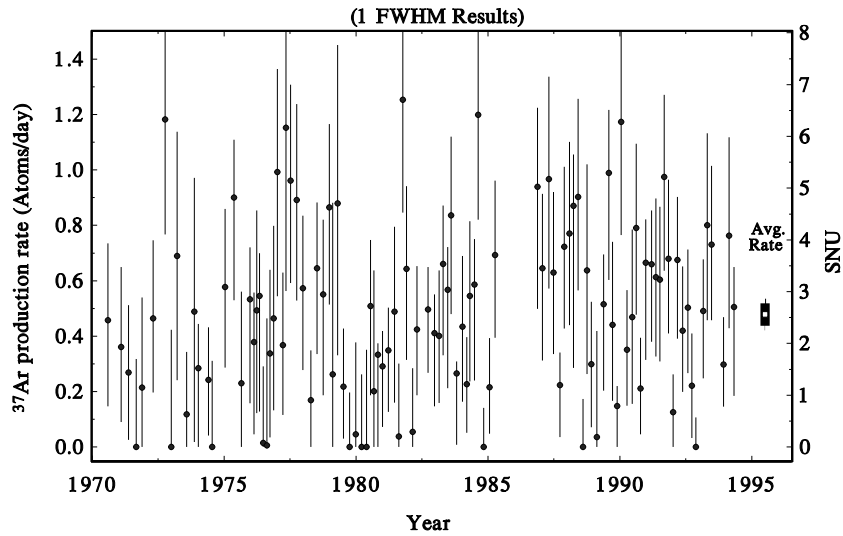
Figure 2.7: Solar neutrino interactions observed in the 24 year running time of the Homestake experiment, in units of 10^{-36} captures per target atom per second (1 SNU). Figure from [21].

2.4.2 Gallium Experiments

Later radiochemical observations of solar neutrinos were made using gallium by the GALLEX[22], GNO[23] and SAGE[24] experiments. Gallium has the advantage of a much lower energy threshold of 233 keV for electron neutrino absorption via the reaction

$$^{71}\text{Ga} + \nu_e \rightarrow\ ^{71}\text{Ge} + e^-. \tag{2.34}$$

The germanium is extracted from the gallium target periodically and observed to decay through electron capture back to gallium. Neutrino capture on gallium has a low enough threshold to be sensitive to all kinds of solar neutrinos, including the pp neutrinos, which have the lowest endpoint energy. As the

21

pp neutrinos are produced in the first stage of the solar fusion cycle, their flux is very directly connected to the total power output of the Sun. These experiments reported values of

$$77.5 \pm 6.2(\text{stat})^{+4.3}_{-4.7}(\text{syst}) \text{ SNU} \qquad (\text{GALLEX})$$

$$62.9 \pm 5.4(\text{stat}) \pm 2.5(\text{syst}) \text{ SNU} \qquad (\text{GNO})$$

$$69.1^{+4.3}_{-4.2} \text{ SNU} \qquad (\text{SAGE})$$

where SNU is the *solar neutrino unit* (the preferred unit for radiochemical experiments), defined to be 10^{-36} captures per target atom per second. The Standard Solar Model[25] predicted 128 ± 8 SNU, clearly in disagreement with the experiments.

2.4.3   Super-Kamiokande

Super-Kamiokande[26] is a 50 kiloton water Cherenkov detector located under 2700 mwe of rock in Kamioka, Japan. During its initial run, the detector contained 11,146 inward facing 50 cm diameter Hamamatsu photomultiplier tubes (PMTs). This array of PMTs observe Cherenkov light produced by relativistic charged particles moving in the inner volume. An accident in 2001 destroyed a number of PMTs, which necessitated redistributing the remaining PMTs and protecting them with acrylic shields[27].

Although Super-Kamiokande's first major result was regarding atmospheric neutrinos, the detector also has substantial sensitivity to solar neutrinos. Incoming neutrinos can elastically scatter electrons in the water target:

$$\nu + e^- \rightarrow \nu + e^- \tag{2.35}$$

With electron neutrinos, this reaction can proceed through both W and Z exchange, whereas for mu and tau neutrinos, only the Z channel is possible. This gives the elastic scattering interaction some sensitivity to all flavors of neutrinos in the proportion $\phi_{\nu_e} + \frac{1}{6}(\phi_{\nu_\mu} + \phi_{\nu_\tau})$. In the presence of neutrino oscillations, a measurement of either the electron neutrino flux or the total (all flavors) flux is not possible with Super-Kamiokande alone.

The elastic scattering interaction is especially interesting for solar neutrinos as the outgoing electron direction is strongly correlated with the incoming neutrino direction. The Cherenkov cone from a relativistic particle points along the particle's velocity vector. As a real time detector (in contrast to the radiochemical experiments) can reconstruct the direction of motion in each event from the image of the cone projected onto the PMT array. Figure 2.8 shows the plot of the cosine of the angle between the event direction and the vector pointing from the Sun to the Earth at the time of the event. A clear forward peak is visible, indicating that the neutrinos observed by Super-Kamiokande are indeed solar in origin.

Intrinsic backgrounds limited the solar analysis in the second run (SK-II) to a reconstructed electron energy threshold of 7 MeV. This makes Super-Kamiokande only sensitive to ^{8}B and hep neutrinos, which have high enough energy to scatter electrons above this threshold. In practice, the ^{8}B neutrinos dominate, as the hep neutrinos are 3 orders of magnitude lower in flux. The
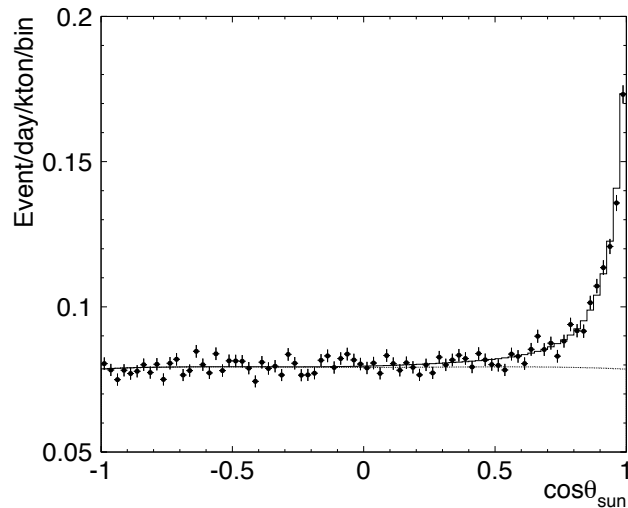
Figure 2.8: Cosine of event direction relative to incoming vector from the Sun. Figure from [27].

solar neutrino flux observed by SK-II was

$$\phi_{ES} = 2.38 \pm 0.05\text{stat}^{+0.16}_{-0.15} \times 10^6 \text{ cm}^{-2}\text{sec}^{-1}, \tag{2.36}$$

far below the $(5.69 \pm 0.91) \times 10^6 \text{cm}^{-2}\text{s}^{-1}$ prediction of BS2005.

2.4.4 Borexino

The Borexino experiment[28][29] is a liquid scintillator detector designed to observe solar neutrinos in real time, but at energies lower than is possible in water Cherenkov detectors. This makes it possible to observe ^{7}Be neutrinos directly for the first time. The inner target volume is a sphere composed of 300 tons of pseudocumene with 1.5 g/L of PPO mixed in as a fluor. Encased in a thin nylon vessel, the scintillator is suspended in a larger sphere of buffer
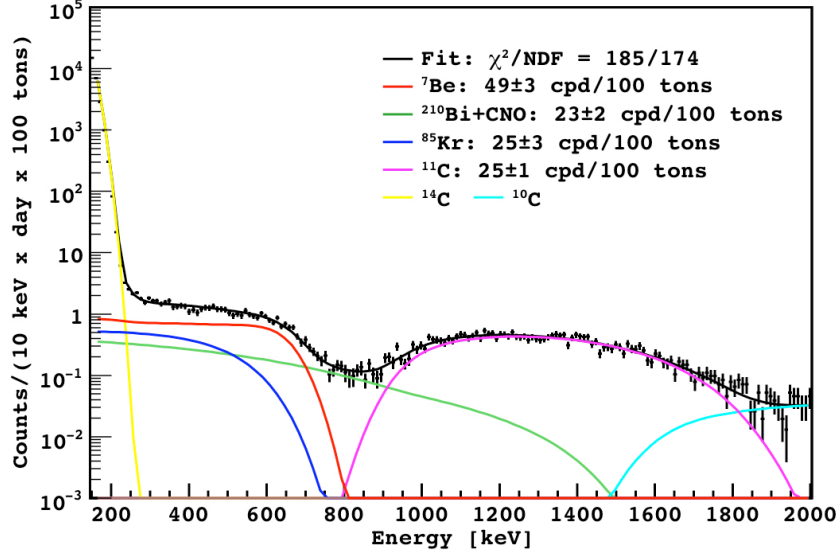
Figure 2.9: Energy spectrum of reconstructed events in Borexino after α/β cuts. Figure from [29] pre-print.

liquid (more pseudocumene with a scintillation quencher) surrounded by 2212 8" PMTs in a spherical, inward looking configuration. Of these, 1828 also have light concentrators to increase the effective photocathode coverage.

Similar to Super-Kamiokande, Borexino also observes solar neutrinos through the elastic scattering of electrons by incoming neutrinos. Unlike Cherenkov light, the scintillator does not provide any directional information, so background rejection becomes much more difficult. All detector materials must be of high radiopurity, and coincidence cuts are used to eliminate radon daughters. Figure 2.9 shows the energy spectrum observed by Borexino's latest 192 day result. The interaction rate of ^{7}Be neutrinos observed is $49 \pm 3(\text{stat}) \pm 4(\text{syst})$ counts per day per 100 tons, compared to a predicted unoscillated rate of 74 ± 4 counts per day per 100 tons.

2.4.5 KamLAND

KamLAND[30] is another liquid scintillator detector, but designed for a completely different purpose. Rather than search for solar neutrinos, KamLAND detects *antineutrinos* produced by nuclear reactors in Japan and neighboring areas, with an average baseline of 180 km. The L/E for reactor antineutrinos over this baseline gives KamLAND sensitivity to the same range of mixing parameters observed in the solar experiments. This provides a valuable cross-check, and complementary measurements to the solar results.

The detector is located in the original Kamiokande detector hall, and is comprised of one kiloton of 80% dodecane, 20% pseudocumene, and 1.52 g/L of PPO. Antineutrinos interact with hydrogen in the detector through inverse beta decay

$$p + \nu_e \rightarrow n + e^+, \tag{2.37}$$

the very same reaction used to first discover neutrinos. The prompt positron annihilation followed by a delayed neutron capture provide a coincidence signal with high background rejection. Both signals are observed in the scintillator by a combination of 1,325 17" PMTs and 554 20" PMTs. The outgoing positron energy is strongly correlated with the incoming neutrino energy[6]. In the latest high precision results from KamLAND[31], they can now observe the L/E oscillation pattern, as shown in Figure 2.10. Fitting the energy dependence of the distortion gives KamLAND excellent precision for measuring Δm_{12}^2, but not as much sensitivity to θ_{12}, as that requires observing the minimum P_{ee} in one oscillation (See Equation 2.15). On the other hand, high energy solar experiments, like SNO, most directly observe $\sin^2 2\theta_{12}$. Together, KamLAND

---

[6]This is analogous to the CC interaction in SNO. See Section 3.1.
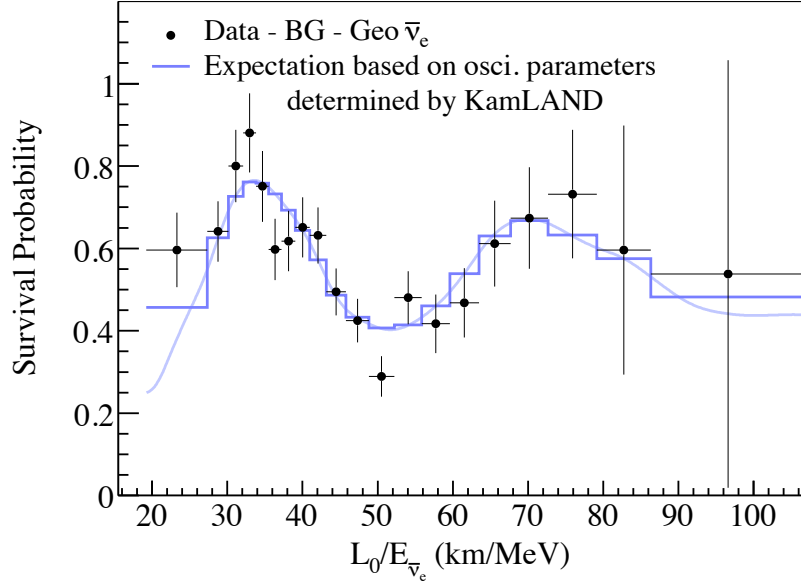
Figure 2.10: L/E distribution for events observed by the KamLAND detector. Figure from [31].

and solar experiments provide nearly orthogonal constraints on the mixing parameters.

2.4.6 Experimental Summary

All of the solar neutrino results can be combined to produce a best estimate for the solar mixing parameters. Figure 2.11 shows the global solar-only (including SNO) best fit contours as well as the solar+KamLAND contours. The only allowed region left is called the *large mixing angle* (LMA) region, to distinguish it from previously allowed regions where the solar neutrino deficit was explained by a small mixing angle (SMA) or vacuum oscillation.
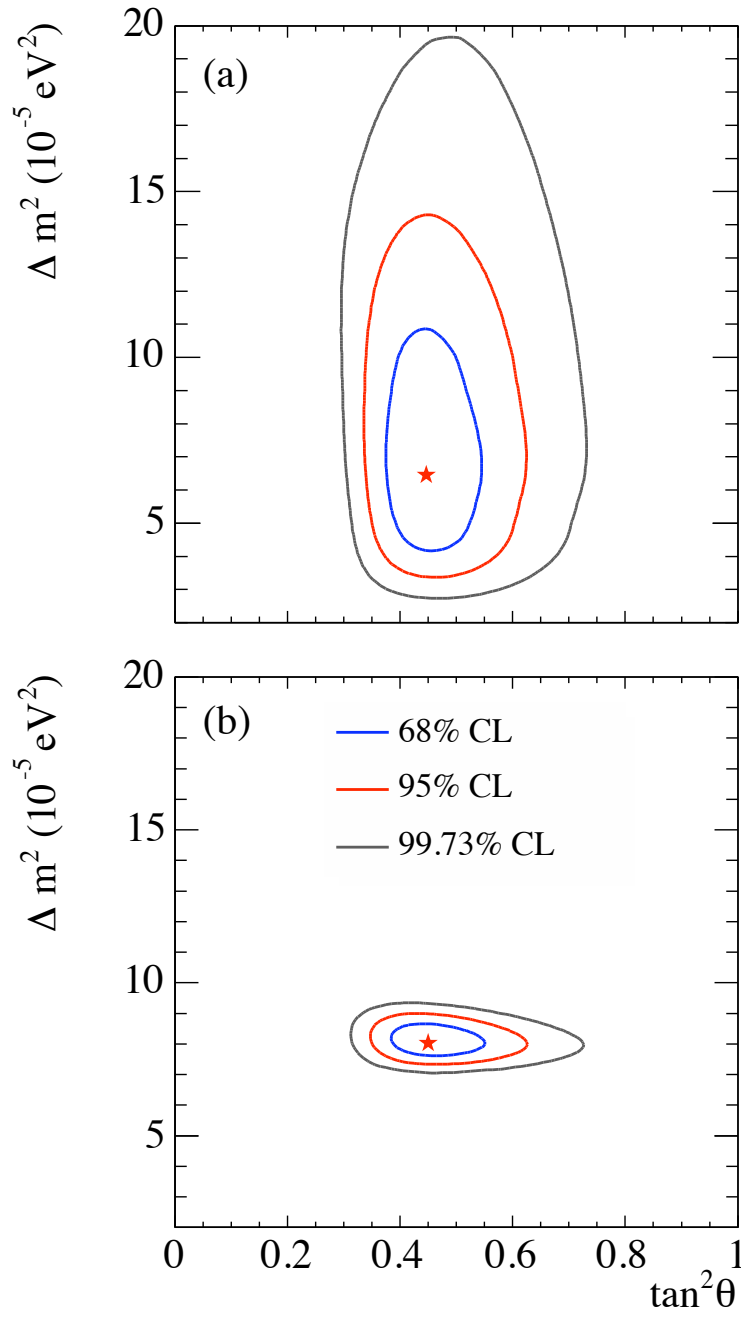
Figure 2.11: Allowed mixing parameters for Δm_{21}^2 and θ_{12} with only solar neutrino experiments (left) and solar+KamLAND (right). Figure from [32].

2.5 Prospects for New Physics

As fundamental parameters in the Standard Model, the neutrino mixing angles and masses are intrinsically useful to know, if only to better prepare neutrino experiments to be future probes in astrophysics and geophysics. But more importantly, improved measures of the neutrino oscillation parameters also have something to tell us about fundamental new physics. Just a few of these ideas are mentioned in the following sections.

2.5.1 Tri-Bimaximal Mixing and Lepton-Quark Unification

One of the great peculiarities of the PMNS neutrino mixing matrix is how very different it is from the analog in the quark sector, the CKM matrix[10]:

$$|V_{CKM}| = \begin{pmatrix} 0.97383^{+0.00024}_{-0.00023} & 0.2272^{+0.0010}_{-0.0010} & (3.96^{+0.09}_{-0.09}) \times 10^{-3} \\ 0.2271^{+0.0010}_{-0.0010} & 0.97296^{+0.00024}_{-0.00024} & (42.21^{+0.10}_{-0.80}) \times 10^{-3} \\ (8.14^{+0.32}_{-0.64}) \times 10^{-3} & (41.61^{+0.12}_{-0.78}) \times 10^{-3} & 0.999100^{+0.000034}_{-0.000004} \end{pmatrix}.$$

(2.38)

Quark flavor number is only weakly broken, whereas in the neutrino sector, the mixing is maximal for one angle, and near maximal for the other. The second mass state, ν_2, even has "flavor democracy," with equal parts ν_e, ν_μ and ν_τ.

The extremely non-diagonal form of the PMNS matrix has been observed[33] to be very nearly the product of a 45° rotation matrix (the angle between x or y axis and the vector (1,1)) and a $\approx$ 35° rotation matrix (the angle between the x, y or z axis and the vector (1,1,1)). This idealized matrix is called the

tri-bimaximal mixing matrix and has the form

$$|U_{TBM}| = \begin{pmatrix} \sqrt{\frac{2}{3}} & \sqrt{\frac{1}{3}} & 0 \\ \sqrt{\frac{1}{6}} & \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{6}} & \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{2}} \end{pmatrix} \approx \begin{pmatrix} 0.816 & 0.577 & 0 \\ 0.408 & 0.577 & 0.707 \\ 0.408 & 0.577 & 0.707 \end{pmatrix} \qquad (2.39)$$

The tri-bimaximal values are separately consistent with the 90% C.L. measured values shown in Equation 2.21.

At first glance, the tri-bimaximal mixing form might seem like idle numerology. However, the specific form may give some clue about the fundamental physics which gives mass to neutrinos. The tri-bimaximal matrix has been found to have relations to group theory [34][35], which may also indicate why quarks have so little mixing. Certainly this question will have to be addressed by any grand unified theories which postulate quark-lepton unification at high energy. It has even been shown[36][37] that if the CKM and the PMNS matrices are equal at the GUT energy scale, renormalization effects in the MSSM can cause the the two matrices to diverge from each other at low energy, under the assumption that neutrinos have Majorana masses and are quasi-degenerate with $m_{1,2,3} > 0.1$ eV. The CKM matrix can be used in this framework to estimate the value of θ_{13}, which [37] predicts to be $\theta_{13} \approx 3.5° - 10°$. Clearly, more experimental input will help test and refine these ideas.

2.5.2 Spectral Distortion

There is an entire class of new physics that can be discovered by looking for the vacuum-matter transition in the ν_e survival probability, show in Figure 2.5. The exact location of the cross-over point depends upon the relative

magnitudes of the terms in the Hamiltonian shown in Equation 2.22. Any physical interaction which creates an effective potential for neutrinos will have an associated term in the Hamiltonian and distort the ν_e spectrum.

One such idea is that of non-standard interactions (NSI), like flavor-changing neutral currents. Miranda, et. al.[38] considered the case of a Fermi-like interaction between electron neutrinos and the down quark. This leads to a Hamiltonian term of the form

$$\mathbf{H}_{NSI} = \sqrt{2} G_F N_d \begin{pmatrix} 0 & \epsilon \\ \epsilon & \epsilon' \end{pmatrix}, \tag{2.40}$$

where ϵ and ϵ' parameterize the strength of NSI. Using both solar and KamLAND data as a constraint, they find two other allowed solar neutrino mixing parameters in the presence of non-standard interactions. Figure 2.12 shows these additional solutions, called LMA-D and LMA-0, in contrast to the standard solution, LMA-I. These different solutions can be distinguished by their effect on the survival probability, shown in Figure 2.13.

Even more exotic physics, like mass-varying neutrinos[39], or a scale-invariant field ("unparticle") coupled to the Standard Model[40][41] can also be detected in a similar way, through their distortion of the ν_e spectrum beyond the three-flavor neutrino oscillation prediction.

2.5.3   CPT Invariance

We obtain excellent constraints on Δm_{21}^2 from KamLAND and θ_{12} from SNO and Super-Kamiokande. If we assume that the mass matrix of neutrinos and antineutrinos are identical, then we can combine these two measurements for
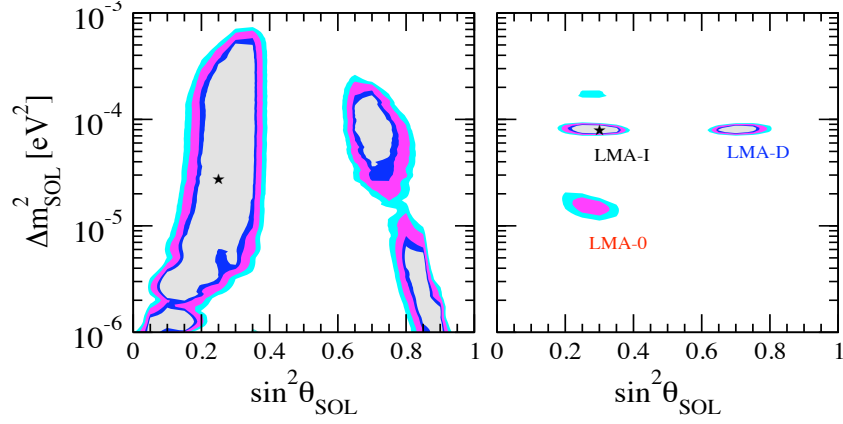
Figure 2.12: Allowed mixing parameters in the presence of non-standard interactions for solar-only (left) and solar+KamLAND (right). LMA-I represents the standard LMA region, while LMA-D and LMA-0 are new solutions. Figure from [38].
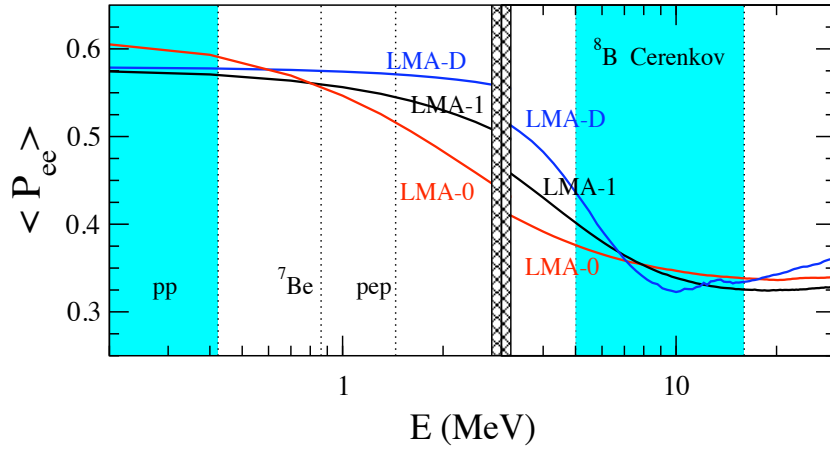
Figure 2.13: Predicted survival probabilities for different NSI solutions. The left side is the result of a fit to low energy solar experiments and the right is a fit to ^{8}B solar neutrino experiments. Figure from [38].

the best estimate of the solar mixing parameters. However, as these two experiments improve, the possibility arises to test whether the two-neutrino mixing of each experiment separately is consistent with the other.

De Gouvêa and Peña-Garay[42] performed such an analysis and set a limit of

$$|\Delta m^2 - \Delta \bar{m}^2| < 1.1 \times 10^{-4} \, \text{eV}^2 \text{ 95\% C.L.,} \qquad (2.41)$$

where the bar refers to antineutrinos. While pessimistic that additional Kam-LAND data would help this limit, they do suggested that additional solar measurements could improve the result.

2.6 Conclusions

The neutrino oscillation model has been fantastically successful in explaining neutrino experiments across many orders of magnitude in baseline and energy. The "solar neutrino problem" has become a "solar neutrino solution," which opens up the possibility of using the Sun to explore new physics through precision measurement. In the previous section, we briefly mentioned new physics that could be investigated through improving the uncertainty on the mixing parameters and looking for the vacuum-matter transition in the ν_e spectrum. The next two chapters will show how SNO is uniquely situated to assist in both of these goals.

Chapter 3

The Sudbury Neutrino
Observatory

After many decades of difficult measurements, the chlorine and gallium so-
lar neutrino experiments showed a significant deficit of neutrinos from the
Sun relative to the predictions of the Standard Solar Model, constrained by
independent, non-neutrino solar observations. Either the solar model was in-
accurate, or electron neutrinos were somehow disappearing on their way to
Earth. The Sudbury Neutrino Observatory was built to resolve this question.
In this chapter, we will describe the design principles, construction, and ca-
pabilities of SNO, as well as review previous results. Most of the information
in this chapter is summarized from [43] and [44], although special emphasis is
given to details which are relevant for the low energy analysis.

3.1 Principles of Operation

H. Chen conceived of SNO in 1985[45] as a water Cherenkov detector, similar to Kamiokande, but using heavy water (D_2O) instead of light water. By replacing the hydrogen in the water molecule with deuterium, three interactions with solar neutrinos are possible:

- Elastic scattering (ES): $\nu_x + e^- \rightarrow \nu_x + e^-$

- Charged current (CC): $\nu_e + d \rightarrow p + p + e^- \; (-1.44\,\mathrm{MeV})$

- Neutral current (NC): $\nu_x + d \rightarrow p + n + \nu_x \; (-2.2\,\mathrm{MeV})$

The elastic scattering interaction involves the bound electrons in the water molecule, and so is the only one of the three interactions in common with light water Cherenkov detectors. The second two interactions are named to correspond with the currents carried by the charged W and the neutral Z, the gauge bosons which mediate the process.

Water Cherenkov detectors observe visible light produced by relativistic particles traveling faster than the speed of light in the medium through which they are passing. The Cherenkov radiation comes out in a forward cone with angle[46]

$$\cos \theta_c = \frac{1}{\beta n(\omega)} \tag{3.1}$$

relative to the the direction of travel, where $n(\omega)$ is the index of refraction in the medium as a function of angular frequency. Over a frequency range where the index of refraction is approximately constant, the intensity (in units of energy per unit frequency) of Cherenkov radiation[44] follows the approximate

formula

$$\frac{dI}{d\omega} = \frac{\omega e^2 L \sin^2 \theta_c}{c^2}. \tag{3.2}$$

Transforming to units of wavelength, the distribution in terms of number of photons per unit wavelength is

$$\frac{dN}{d\lambda} \propto \frac{1}{\lambda^2}. \tag{3.3}$$

As a result, Cherenkov light is biased toward very short wavelengths, a feature which is important in Section 3.3.

In the case of the ES and CC reactions, we observe the outgoing electron when it produces Cherenkov light in the target volume. The NC reaction does not produce a relativistic charged particle, and therefore generates no Cherenkov light directly. It does, however, produce a free neutron which can be detected in a variety of ways.

SNO was operated in three different configurations, distinguished by their neutron capture mechanism:

- Phase I, "D_2O phase": The target volume was composed of pure heavy water, with signature: $n +^2 H \rightarrow^3 H + \gamma$ (6.25 MeV).

- Phase II, "salt phase": The heavy water was doped with 0.2% NaCl (by mass), which produces a gamma cascade: $n +^{37} Cl \rightarrow^{38} Cl + N \times \gamma$ ($\sum E_\gamma = 8.6$ MeV).

- Phase III, "NCD phase": Thirty-six proportional counters filled with ^{3}He were installed in the heavy water, providing neutron capture signature: $n +^3 He \rightarrow p + T$(764keV), where T is a "triton," or fully ionized ^{3}H.
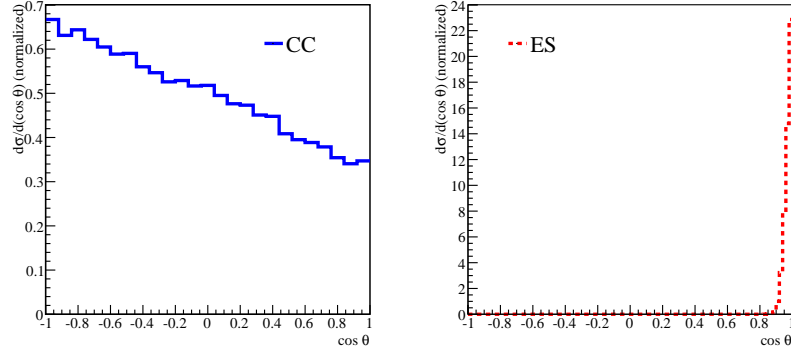
Figure 3.1: Normalized differential cross section as a function of outgoing electron angle for charged-current and elastic scattering interactions. Angle is relative to incoming neutrino direction.

More information on this configuration of the detector in the third phase can be found in [47].

The differential cross-sections for each of these processes give them different, and complementary sensitivities. Elastic scattering is very sensitive to the incoming direction of the neutrino, as shown in Figure 3.1. It also has sensitivity to all neutrino flavors, although the cross section is 6 times larger for electron neutrinos compared to μ and τ neutrinos. Overall, though, the elastic scattering cross section is the smallest of the three, making it the most statistically limited neutrino signature in SNO.

The charged current process is sensitive only solar neutrinos of the electron flavor, as ^{8}B neutrinos are not energetic enough to produce a muon in the center-of-mass frame of the interaction. The charged current interaction provides some information about the direction of the incoming neutrino, but strongly correlates the neutrino energy and the outgoing electron energy. Figure 3.2 shows the differential cross section $d\sigma/dE_e$ for the charged current and

37
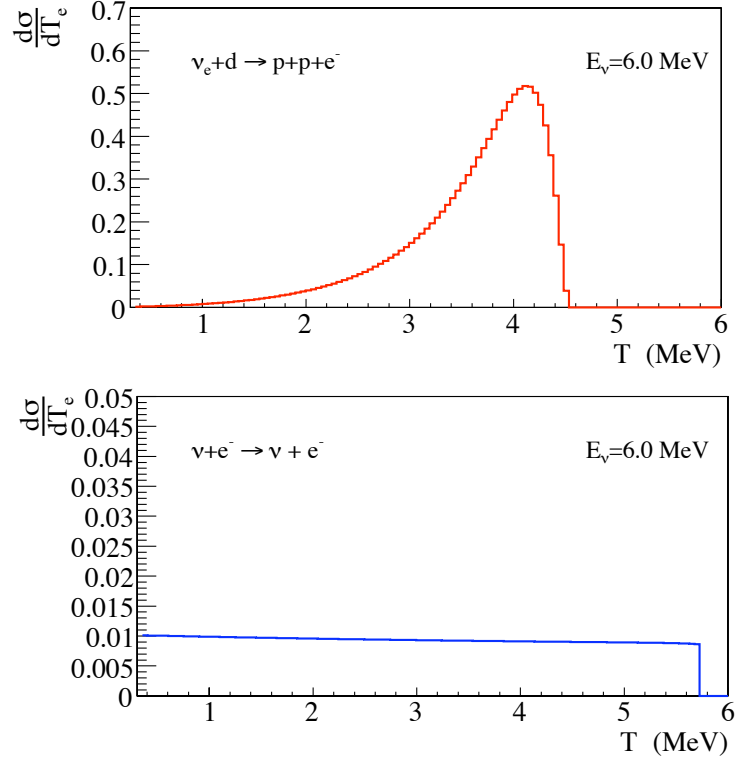
Figure 3.2: Normalized differential cross section as a function of outgoing electron energy for charged-current and elastic scattering interactions. Figure provided by J. Klein.

elastic scattering processes.

The neutral current interaction is equally sensitive to all flavors of neutrino. This serves to normalize the spectrum, and also is directly comparable to the Standard Solar Model prediction regardless of flavor oscillations.

3.2 Detector Construction

The SNO detector is situated in the Creighton nickel mine operated by Vale Inco, Ltd., near Sudbury, Ontario, Canada. Figure 3.3 shows the overall detector design. The inner target volume is approximately 1 kiloton of heavy water, where 99.917% of the hydrogen is the ^{2}H isotope. To protect the (very expensive) heavy water from contamination with light water, the inner volume is encased in a solid acrylic vessel ("AV"), with inner diameter of 12 meters, and a wall thickness of 5.5 cm. Access to the inner volume is provided through a long acrylic chimney (called the "neck") which has an inner diameter of 1.46 meters. The neck and the acrylic vessel are bonded into a single, continuous container from which heavy water cannot escape, except through circulation piping which exits the neck and takes heavy water to a purification plant. Calibration sources are deployed through an airlock at the top of the neck.

The acrylic vessel is suspended in a rock cavity filled with ultrapure light water. As heavy water is more dense than light water at a given temperature (hence the name), the acrylic vessel has a net negative buoyancy. Vectran ropes loop through grooved plates bonded to the equator of the vessel to provide support. Surrounding the acrylic vessel is a stainless steel PMT support ("PSUP") structure to which 9547 photomultiplier tubes (PMTs) are mounted 8.4 meters from the center of the D_2O volume. Most PMTs are inward looking, aimed at the acrylic vessel, but 91 of them are outward looking ("OWL") PMTs which are used to detect cosmic-ray muons which could potentially generate backgrounds that leak into the inner region.

Both the acrylic vessel and the PSUP hang from deck at the top of the cavity, where the data acquisition electronics and calibration source manip-
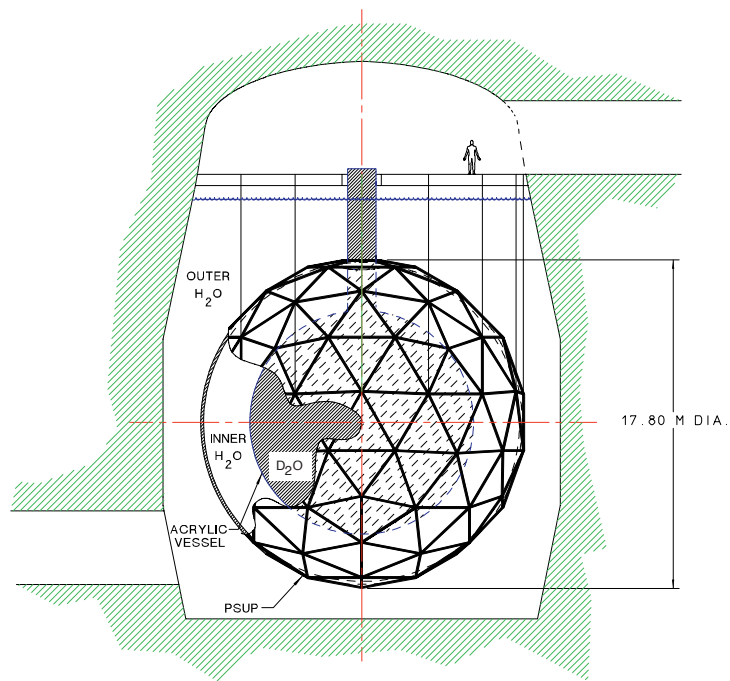
Figure 3.3: Diagram of SNO detector. Figure from [43].

ulator are located. After the completion of construction, the only personnel access was to the deck area of the detector.

3.3   Photon Detection and Optics

Photons are produced in Cherenkov radiation more toward shorter wavelengths. In an optically transparent medium like water, that translates into blue and soft ultraviolet light. Light created in the D_2O volume of the detector must pass through at least 5.5 cm (more if it the path is not exactly radial) of acrylic to reach the PMTs. Commercial acrylic is typically UV-absorbing to prevent degradation when exposed to sunlight and fluorescent lights. Polycast produced for SNO a special batch of UV-transmitting acrylic which has an attenuation length of 5.5 cm for 320 nm light, and improves at longer wavelengths.

Once photons exit the acrylic vessel, they travel through the light water and reach the PMTs and support structure. Photons are detected by the 9456 inward-facing 20 cm (8") Hamamatsu R1408 PMTs, an example of which is shown in Figure 3.4. By themselves, these PMTs would provide 31% coverage, but each PMT is also mounted in a 27 cm diameter reflective light concentrator, boosting the effective coverage to 54% (accounting for reflectivity of concentrators). The concentrators are designed in the shape of a Winston cone to provide optimal collection efficiency for light generated inside the D_2O volume[48]. The detection efficiency as a function of angle of the PMT/Concentrator unit is shown in Figure 3.5. The drop in the angular response at high angles is called the *concentrator cutoff* and helps to make the detector less sensitive to Cherenkov light produced at high radius, outside the
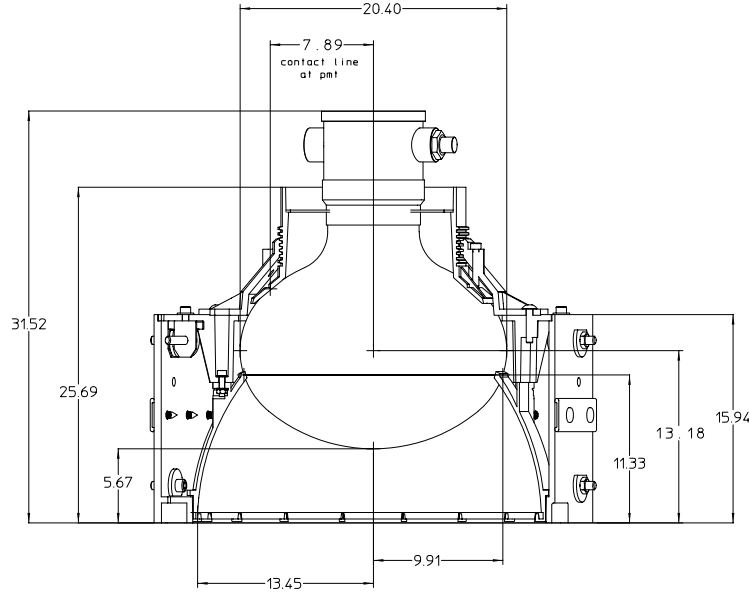
Figure 3.4: Photomultiplier tube with light concentrator mounted to front. Figure from [43].

D$_2$O volume of interest.

The R1408 contains 9 dynode stages with a charge amplification of more than 10^7. Single-photoelectrons can be easily observed in this PMT. However, estimating the number of photoelectrons from a charge integral is difficult due to the very broad charge spectrum of single photoelectrons. Figure 3.6 shows the charge distribution, which has a long, high energy tail. In the energy range of interest for ^{8}B neutrinos, most PMTs are likely to only generate one photoelectron[1] at most, so the charge resolution is not a major concern. For most reconstruction tasks, the time of the PMT hit is the primary observable of interest, although several background cuts rely on charge to identify abnormal events. The assumption of 1 hit equal 1 photoelectron also influences the

---

[1] As a general rule of thumb, SNO detects 7–8 photoelectrons per MeV of kinetic energy for an electron in the D$_2$O volume.
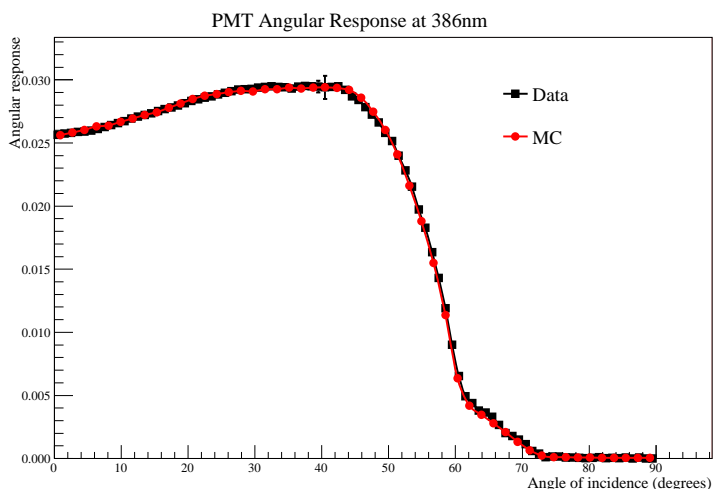
Figure 3.5: Angular response of PMT-Concentrator unit. Angle of incidence is defined relative to the central axis of the PMT. Figure provided by G. Orebi Gann.

design of the trigger system, discussed further in Section 3.5.

Stray magnetic fields affect the trajectory of photoelectrons in the PMT, reducing their overall efficiency depending on the PMT orientation relative to the field. Since the SNO PMTs are aligned radially on the surface of a sphere, there is no orientation for the detector to avoid this problem. Instead, large have been installed into the walls of the SNO cavity, allowing the vertical component of the Earth's magnetic field to be canceled out inside the detector. This increases the overall efficiency of the array from 82% to 97.5% of the zero-field efficiency.
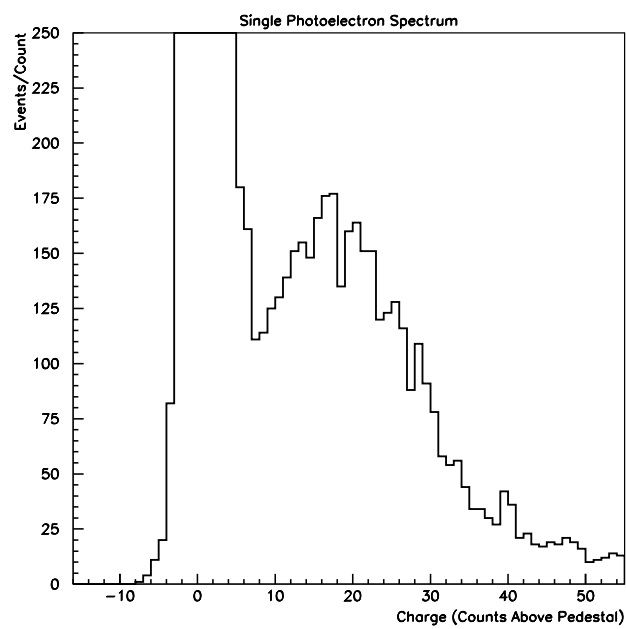
Figure 3.6: Single photoelectron charge distribution for the Hamamatsu R1408 PMT. Figure from [43].

3.4 Control of Backgrounds

Above all else, unwanted neutrons are the primary background concern for SNO, as they can easily be confused with neutrons produced in the NC neutrino reaction. The first line of defense against unwanted neutrons is SNO's actual location. Situated under 6100 mwe of rock shielding, only 65 muons enter the detector per day. Spallation neutrons and short-lived isotopes produced by such muons are easily removed by cutting all data for 20 seconds after the initial muon. The outer light water volume provides additional shielding from neutrons coming from the rock wall of the cavity.

The remaining backgrounds are intrinsic to the detector, due to uranium, radon (also the uranium chain), and thorium contamination. Strategies for managing this depend on the type of material in question. Of the solids, the PMT glass and the acrylic vessel are the biggest concern. In the case of the PMTs, a special run of glass was manufactured by Schott Glaswerke for the PMT envelopes which contained less than 40 parts-per-billion uranium and thorium. Acrylic, being a chemically simple hydrocarbon, is much easier to make clean, and bulk contamination levels of 1.1 parts-per-trillion were achieved.

The liquids in the detector are susceptible to continuous contamination from radon present in the mine air, and uranium/thorium leeching from solids immersed in the detector. As a result, both the heavy and light water periodically is circulated out of the detector, filtered, de-ionized, de-gassed to remove radon, then re-gassed with pure nitrogen before being reintroduced to the detector. In order to improve the radio-purity of the light water region, a nearly leak-proof plastic barrier was installed on the back side of the PMT support

structure, dividing the light water into two volumes: 1.7 kilotons between the acrylic vessel and the PMTs, and 5.7 kilotons between the PMTs and the walls of the cavity. The water inside the PMT support structure can be made much cleaner, which helps reduce backgrounds in the region where they can cause the greatest trouble.

Along with the purification process, the water is also assayed to determine the amount of uranium and thorium present in the water. Two different processes are used to do this, called the HTiO[49] and the MnOx[50] assay techniques. The HTiO process uses HTiO deposited on filtration media to adsorb ^{228}Th, ^{224}Ra, ^{226}Ra, and ^{212}Pb from the water, which are then later stripped from the filter with acid, concentrated in a small volume of liquid, and then observed for radioactive alpha and beta decays. The MnOx technique uses columns of acrylic beads covered with MnOx (the "x" indicating variable amounts of oxygen atoms in the molecule) through which the water flows, depositing radium in the process. The results from both techniques are compared and combined to produce an estimate of the radioactivity in the water in terms of ^{238}U and ^{232}Th-equivalent grams per gram of D_2O (or H_2O). In the case of ^{238}U, the actual source of radioactivity in the decay chain is mostly from radon diffusing into the detector rather than uranium dissolved in the water. Tables 3.1 and 3.2 list the radioactivity measurements of the D_2O and H_2O volumes obtains with these techniques.

3.5   Electronics

The SNO data acquisition (DAQ) and triggering electronics are organized into a hierarchy of crates, interface cards, and individual channels. Each PMT

Type	D$_2$O Phase (g/gD$_2$O)	Salt Phase (g/gD$_2$O)
^{238}U	$1.01^{+0.34}-0.20 \times 10^{-14}$	Upper limit: 2.0×10^{-14} Lower limit: $1.14 \pm 0.46 \times 10^{-16}$
^{232}Th	$2.09 \pm 0.21(\text{stat}) \, ^{+0.96}_{-0.91}(\text{syst}) \times 10^{-15}$	$1.76 \pm 0.44(\text{stat}) \, ^{+0.70}_{-0.94}(\text{syst}) \times 10^{-15}$

Table 3.1: Concentration of ^{238}U and ^{232}Th in the D$_2$O volume as measured by radioassays in both the D$_2$O and salt phases. A leak in the assay system in the salt phase contaminated the samples with small amounts of radon, so only upper and lower limits on ^{238}U are available. Table from [51].

Type	D$_2$O Phase (g/gH$_2$O)	Salt Phase (g/gH$_2$O)
^{238}U	$29.5 \pm 5.1 \times 10^{-14}$	$20.6 \pm 5.0 \times 10^{-14}$
^{232}Th	$8.1^{+2.7}_{-2.3} \times 10^{-14}$	$5.2 \pm 1.6 \times 10^{-14}$

Table 3.2: Concentration of ^{238}U and ^{232}Th in the H$_2$O volume as measured by radioassays in both the D$_2$O and salt phases. Table from [51].

is connected to an electronics channel via RG59 coaxial cable which supplies high voltage and carries the signal pulse back up to the DAQ hardware on the deck. Cables from all PMTs are approximately the same length to ensure that pulses from simultaneously hit PMTs arrive at the front-end at the same time.

At the highest level, 9728 (of which only 9547 are used) channels are bundled into 19 groups of 512 channels each. A group of 512 channels is handled by a *crate*. A crate provides high voltage to the PMTs through 16 PMT Interface Cards (PMTICs), each assigned to 32 PMTs. PMT pulses are separated from the high voltage at the PMTIC through an AC coupling to the low-voltage front-end electronics in the crate.

The front-end of the crate also contains 16 motherboards, which connect through a backplane to the high-voltage PMTICs. A motherboard has mounted on it 4 daughterboards, each of which provide channel discrimina-

tors, integrators and analog charge memory for 8 channels. An incoming PMT pulse is first received by a custom ASIC which determines when the voltage crosses a programmed threshold. Once the channel threshold is crossed, several tasks are initiated. First, another custom ASIC begins integrating of high and low gain (with ratio 16:1) copies of the incoming channel pulse. The high gain signal is also integrated over two intervals, 35 ns and 100 ns. At the same time, another ASIC starts a linear voltage ramp, called the time-to-amplitude converter (TAC). Finally, several different signals are sent from the channel into the trigger system.

Events in SNO a triggered by a fast, asynchronous analog summing network, as shown in Figure 3.7. When a channel discriminator fires, it generates four different trigger pulses:

- 100 ns square pulse

- 20 ns square pulse

- Hi-gain shaped copy of PMT pulse

- Low-gain shaped copy of PMT pulse

These four pulses are independently summed for 8 channels on each daughterboard, then pairs of daughterboards are summed on each motherboard, and finally motherboards are summed at the crate level. Trigger signals from the inward looking PMTs, and the outward looking PMTs are summed separately. Summed trigger signals from each crate are passed via coaxial cable to a central timing rack on the deck.

The timing rack contains two trigger-related boards:
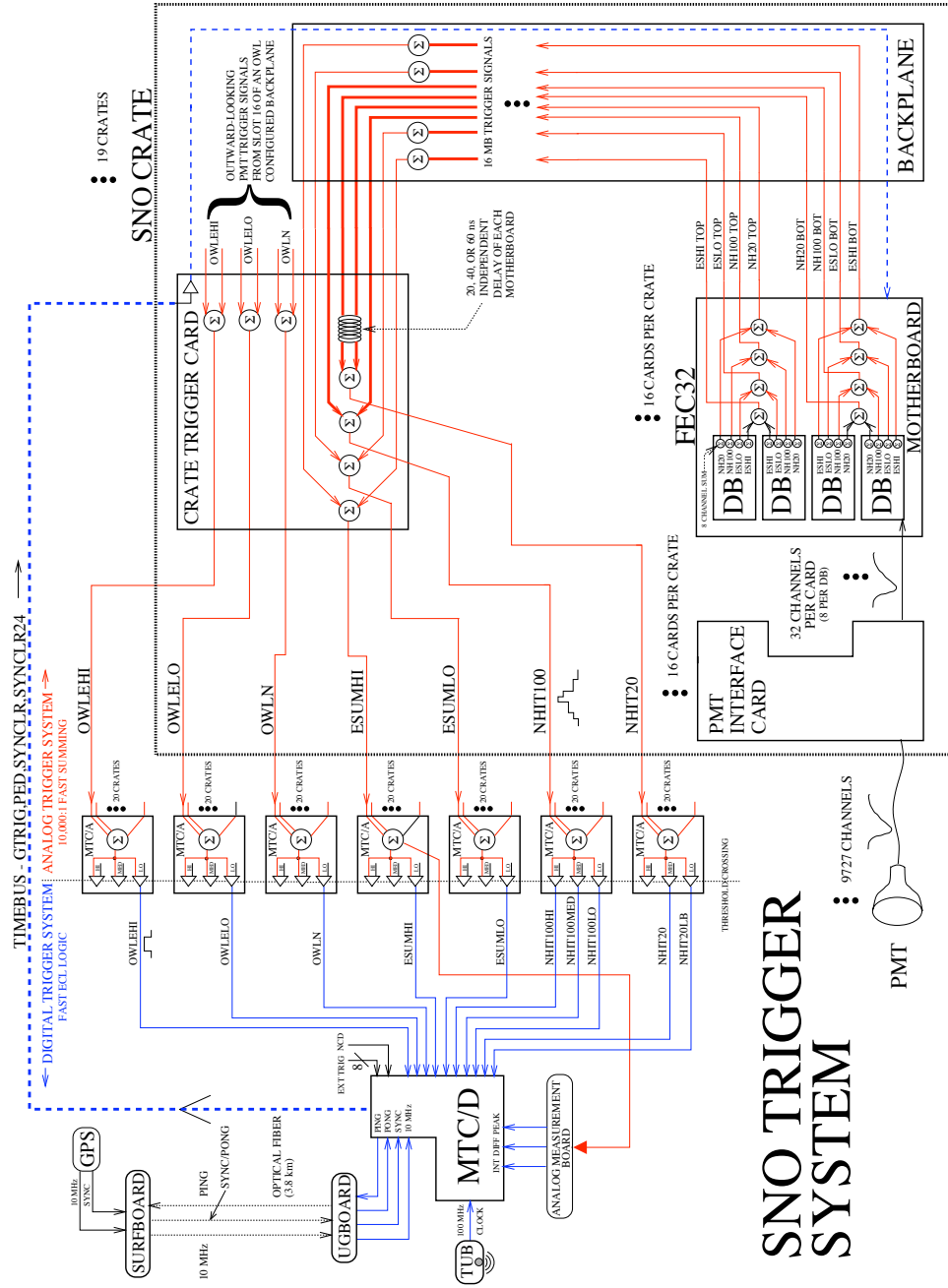
Figure 3.7: Diagram of the SNO trigger system. Figure from [52].

- Master Trigger Card, Analog (MTC/A): The MTC/A performs the final sum of trigger signals from all 19 crates and compares with the programmed trigger threshold. When the trigger threshold is crossed, the MTC/A sends a digital signal to the MTC/D. There is separate MTC/A for each kind of trigger.

- Master Trigger Card, Digital (MTC/D): When triggers from the MTC/A cards arrive, the MTC/D checks if the trigger is enabled, and if so, issues a global trigger signal to all crates synchronous with the next tick of an internal 50 MHz clock. The current time of the event as indicated by the 50 MHz clock counter is recorded, as is a 10 MHz clock counter that is driven via a fiber from GPS unit on surface. Global triggers are recorded in on-board memory.

The summed pulses are used to construct several different kinds of event triggers:

- NHIT100: Trigger if 16 or more PMTs are hit in a 100 ns coincidence window.

- NHIT20: Trigger if 16 or more PMTs are hit in a 20 ns coincidence window.

- OWLN: NHIT100 trigger for outward-looking PMTs, which are summed separately to look for cosmic ray muons.

- ESUMHI: Trigger if the analog sum of PMT pulses exceeds a charge threshold.

- ESUMLO: Same as ESUMHI, but with lower gain and therefore a wider dynamic range.

- OWLEHI: ESUMHI trigger for outward looking PMTs.

- OWLELO: ESUMLO trigger for outward looking PMTs.

- Pulsed Global Trigger: An automatic global trigger generated at 5 Hz. By sampling the detector at regular intervals not coincident with actual Cherenkov events, we can determine the PMT noise rate, which fluctuates with time and temperature.

- Software Trigger: A software-initiated trigger to mark the start and stop of a run with a special event for precise livetime calculation.

- External Trigger/External Asynchronous Trigger: Calibration sources can force a global trigger. The asynchronous version of the trigger bypasses the 50 MHz clock, and generates a global trigger immediately. This is used in PMT timing calibrations to avoid the 20 ns jitter that the clocked global trigger would produce.

The primary trigger used to select neutrino events is the NHIT100 trigger.

When the global trigger arrives at each crate, the channel integrators on the daughterboards transfer their three accumulated pulse integrals and the TAC (all in the form of charges on a capacitor) to a 16-cell deep analog memory buffer. Charge buffering for each channel is required because the four analog-to-digital converters (ADC) which measure the charge are located on the motherboard and shared by all 32 channels. A hit channel can be prepared for another hit with very little deadtime. As the ADCs read out the hits stored

in the analog memory, they store the digitized charges into 4 MB of RAM on the motherboard, along with a header which allows the hit to be matched up with the global trigger that initiated it. This gives the electronics the ability to buffer up to a million hits in the event of a nearby supernova.

Read-out of the motherboards is performed by a Motorola 68040 processor (called the "eCPU") installed in a VME crate. The eCPU continuously polls the crates for hits over a custom bus, accumulating the information in a local buffer. At the same time it also copies global trigger records from the MTC/D. The eCPU buffer is read out by a Sun UltraSparc 1/170 workstation, which does the final reassembly of global trigger records and associated PMT hits. These rebuilt events are then written to disk for offline analysis.

3.6 Calibration System

Frequent calibration of the SNO detector is critical for the interpretation of the data. As we will see in Section 3.7, the simulation depends heavily on detector calibration to produce accurate results. The calibration program can be divided into two parts: electronics calibration and source calibration.

3.6.1 Electronics Calibration

The calibration of the electronics takes place weekly to minimize the effect of drift in the response of the charge integrators and ADCs. The front-end cards contain diagnostic circuitry which can inject artificial pulses into the channel discriminators. A *pedestal* run is performed periodically in which the discriminators are fired with no incoming pulse, and the charge recorded and digitized. This provides a definition of zero charge on each channel. In a

second run, called the TAC-slope run, the channel discriminators are fired, and followed by the issue of a global trigger with various fixed delays. With the lookup table generated by this procedure, a recorded ADC value from a TAC can then be converted to a time in nanoseconds for the channel hit.

3.6.2 Source Calibrations

During operation, a variety of radioactive sources can be deployed to track detector performance and detector properties over time. The two primary workhorses of the SNO calibration program are

- ^{16}N[53]: Gaseous ^{16}N made with a deuterium-tritium neutron generator on deck is pumped down an umbilical line to a stainless steel decay chamber, where it undergoes beta decay to an excited state of ^{16}O, which promptly relaxes, emitting a 6.13 MeV gamma ray. The initial beta is observed by a scintillator and small PMT in the source, providing a tag to separate ^{16}N events from random backgrounds. This source's primary purpose is to provide the overall calibration of the detector's light collection efficiency, and systematic uncertainties in energy reconstruction as a function of position and time. It is also useful for evaluation of the fraction of neutrino events removed by background cuts ("cut sacrifice"), and isotropy of Cherenkov light production.

- Laserball[54]: A dye laser of adjustable wavelength is used to send monochromatic light down a fiber to a diffuser ball. Like the ^{16}N source, the laserball system provides an event tag, although the tag is normally used to trigger the detector directly. The laserball creates a short isotropic light pulse from a fixed, well-defined location. It is used to

calibrate the relative PMT efficiencies (whereas the ^{16}N sets the global efficiency) and PMT timing. By moving the source off-center to sample different photon paths, attenuation lengths in the optical media and the PMT angular response can be simultaneously extracted as well.

Two additional sources generate high energy events that are useful for measuring energy systematics:

- ^{8}Li[55]: The D-T generator in conjunction with a ^{11}B target produces ^{8}Li which is transported very quickly in a helium carrier gas through an umbilical line to a decay chamber lowered into the detector. The beta decay of ^{8}Li has an endpoint of 14 MeV, and the chamber walls are thin enough to allow the beta to escape. Scintillation of the helium carrier gas from alphas coincident with the beta also provides an event tag.

- pT[56]: Protons collide with ^{3}H to produce ^{4}He and a 19.8 MeV gamma ray. The source also produces enormous amounts of neutrons as a by-product, which activate sodium into a problematic radioactive isotope, ^{24}Na. For this reason, the source was not deployed in the salt phase.

Finally, there were a number of neutron and low energy background sources:

- ^{252}Cf: A spontaneous fission isotope which produces gammas and one or more neutrons.

- AmBe: Another neutron source in which an alpha emitted by ^{241}Am interacts with beryllium to eject a free neutron.

- ^{238}U/^{232}Th[57]: These sources are radioactive isotopes encased in acrylic which produce either uranium or thorium chain decays. Of interest are

the 3.3 MeV beta decay of ^{214}Bi in the uranium chain, and the 2.6 MeV gamma decay of ^{208}Tl in the thorium chain. As ^{238}U and ^{232}Th are very long lived isotopes and therefore have low decay rates, the acrylic sources are often composed of different, shorter-lived isotopes, which still feed into either the uranium or thorium chain above the decays of interest.

- Radon Spike[58]: In the salt phase, radon-enriched D_2O was deliberately injected into the detector and mixed to provide a distributed source of uranium chain decays, in particular the . The radon was then removed (what was left after several 4 day half-lives) through the normal purification process. With no source container to absorb Cherenkov photons, the spike is free of the "shadowing" problems which make interpreting other sources more difficult. A second spike was performed in the H_2O region as well.

- ^{24}Na Spike: Very similar to the Rn spike, except a sample of neutron activated salt dissolved in D_2O was introduced into the detector. The ^{24}Na decay provides a 2.6 MeV gamma source.

All sources are deployed using a computer-controlled manipulator system which can place the source in a wide range of positions in the detector, shown in Figure 3.8. The manipulator can be operated in three configurations: single-axis, $x - z$ and $y - z$ mode. In single-axis mode, the source is suspended by a single rope, and therefore can only be moved along the z-axis of the detector. In $x - z$ and $y - z$ mode, additional ropes are attached to the source which can pull it off axis, either in the $x - z$ plane or the $y - z$ plane. The source cannot be moved to an arbitrary (x, y, z) coordinate off these planes, though.
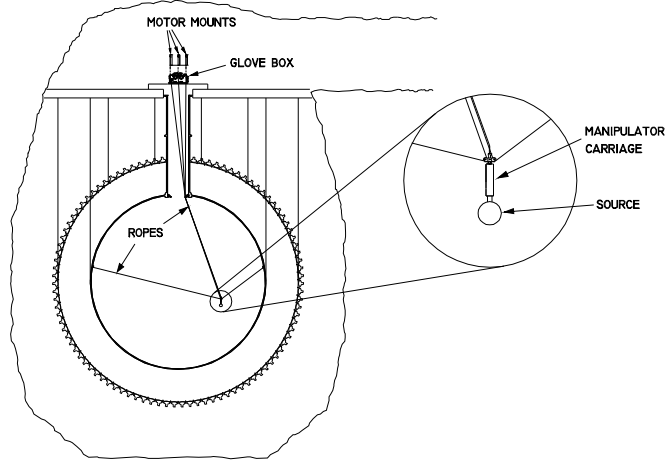
Figure 3.8: SNO calibration source manipulator. Figure from [43].

3.7 Simulation

The SNO detector simulation is nearly as important to the experiment as the detector itself. It is a critical tool for PDF generation (see Chapter 4.2.5) and precise energy reconstruction (see Chapter 5). The simulation program, *SNO M*onte Carlo and *An*alysis (SNOMAN), is the product of more than 15 years of development work by the collaboration. SNOMAN is a FORTRAN application which is used by the experiment both to simulate the response of the detector to various kinds of events, but also for the first stage of data processing and reconstruction.

The Monte Carlo component is a full photon-tracking simulation which starts with seed particles, such as electrons, gammas, neutrons and muons, and propagates them through the detector. Electron and gamma physics are calculated using code from EGS4[59] which has been embedded into the sim-

ulation framework. EGS4 computes energy loss due to ionization, Compton scattering of electrons by gammas, pair production, and electron multiple scattering. The production of Cherenkov light has been added to the simulation by seeding photon vertices along each segment of the electron track according to Equations 3.1 and 3.2. Neutron propagation and capture are calculated with MCNP[60]. Higher energy particles like muons and electrons above 2 GeV are simulated with LEPTO[61], and hadrons with FLUKA[62] and GCALOR[63], all from the CERNLIB software library.

Particle tracking through the detector geometry is handled by custom SNOMAN code which reads the dimensions, materials and optical properties of each element of the SNO detector from a database on startup. Calibration source geometries are also included when simulating runs in which they were deployed. The particle tracking code is also used when propagating optical photons through the detector, with reflection and refraction happening at media boundaries, and Rayleigh scattering and absorption in the bulk.

Once photons cross into the PMT concentrator region, one of two simulations can be used. There is a phenomenological simulation of PMT response derived from laserball calibrations (called "greydisk"), and there is also a full 3D photon tracking model which traces the photon through all parts of the PMT/Concentrator unit. In the default configuration, the Monte Carlo uses the 3D PMT model, but for speed, the energy estimator described in Chapter 5 uses the greydisk model as an approximation.

If the photon produces a photoelectron in the PMT, a full simulation of the data acquisition hardware integrates the pulse, and simulates the operation of the trigger system. At the end, a data structure is generated for the event which is very similar to that produced by the real detector. The Monte Carlo

event differs only from real data in that the quantization error of the ADCs and the charge pedestals are not simulated. As a result, the simulation and the real data are only comparable after the electronics calibration has been applied to the real data, converting charges back into hit times in nanoseconds.

After this point, processing for data or Monte Carlo is identical. A user-supplied command file pipelines the events through a series of analysis tasks which can include position reconstruction, energy estimation, application of cuts, and finally generation of output files in either HBOOK or ROOT format.

3.7.1   Calibration of Simulation Parameters

SNOMAN is only as accurate as the inputs which are provided. Much of the detector configuration is set *a priori*. That is, parameters like dimensions, materials, and locations of detector elements are input based on design documents. Benchtop measurements of PMT charge and timing distributions, and photocathode efficiencies are also included. Other parameters are measured using the calibration sources and input to the simulation to provide realistic looking Monte Carlo events.

Run-level Parameters

A run in SNO ranges from 30 minutes to 4 days, with most runs lasting 7 hours. At the start of each run, the detector state is recorded. After the run ends, additional information is computed and loaded into the SNOMAN database. These data include

- Start and stop time of the run

- Trigger thresholds and list of enabled triggers

- Channel discriminator thresholds

- Channel status (online, offline)

- Average PMT noise rate

- Source type and location (if deployed)

All of these values are used by SNOMAN when deciding whether PMTs can register hits, how many noise hits to add to the event, and whether to enable a particular source geometry.

Phase-level Parameters

On longer timescales, some parameters are measured only a few times per phase. These include PMT angular response, and attenuation lengths in D_2O, acrylic, and H_2O. The very last parameter to be set is the *global collection efficiency*, which is a final "fudge-factor" of sorts. The global collection efficiency scales all of the PMT photon detection efficiencies in order to make the mean number of hit PMTs observed from the ^{16}N source match the number generated by the simulation. Any additional discrepancies between data and Monte Carlo are either corrected outside the simulation (if there is significant evidence from calibration sources) or taken as systematic uncertainties in the simulation itself.

3.8 Previous SNO Results

SNO has published several observations of ^8B neutrinos, in addition to papers setting limits on periodicities in the ^8B flux[64], hep neutrinos[65], the

diffuse supernova background[65], nucleon decay[66], and antineutrinos from the Sun[67]. In this section, we review previous ^{8}B results from the first two phases of SNO[2].

3.8.1   D$_2$O Phase

Three papers were published based on the D$_2$O phase dataset. The first paper[68] reported the reconstructed electron energy (T_{eff}) spectrum of CC events above a threshold of 6.75 MeV (kinetic) for 241 live days. The signal extraction was performed using a maximum likelihood method in the event observables energy, position, and direction. Combined with Super-Kamiokande's ES measurement, SNO could estimate the total flux of neutrinos, of all flavors, from the Sun, to be $5.44 \pm 0.99 \times 10^6 \, \text{cm}^{-2}\text{s}^{-1}$. This is in good agreement with the BS2005 SSM estimate of $5.69 \pm 0.91 \times 10^6 \, \text{cm}^{-2}\text{s}^{-1}$.

The next paper[69] expanded the data set to 306 live days and lowered the energy threshold to 5 MeV to have sensitivity to neutrons from the NC reaction, allowing SNO to measure the solar neutrino flux without reference to Super Kamiokande. Figure 3.9 shows how the measurements of CC, ES and NC by SNO are all self-consistent, and lead to a slightly more precise determination of the total solar flux to be $5.09^{+0.44}_{-0.43}(\text{stat})\,^{+0.46}_{-0.43}(\text{syst}) \times 10^6 \, \text{cm}^{-2}\text{s}^{-1}$.

The final D$_2$O phase paper[70] was an analysis of the energy spectrum, separating day and night events. Being sensitive to the regeneration of ν_1 as neutrinos passes through the Earth at night allows SNO to set some additional constraints on neutrino mixing parameters. A ν_e asymmetry of

---

[2]Results from the third phase of SNO are under review as of this writing.
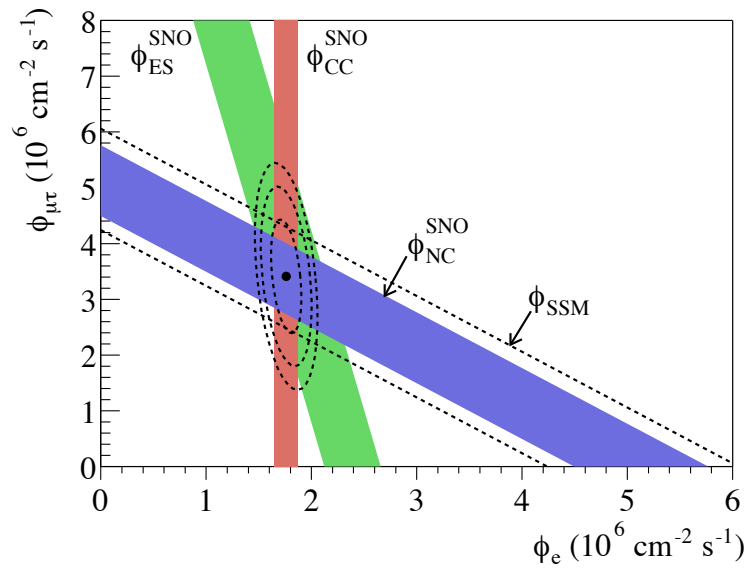
Figure 3.9: Measured D$_2$O-phase ν_e and $\nu_\mu + \nu_\tau$ fluxes at SNO using the CC, NC and ES reactions. Each reaction is a different linear combination of these two neutrino sources, and all three are self-consistent, as shown by the intersection of the bands. Figure from [69].
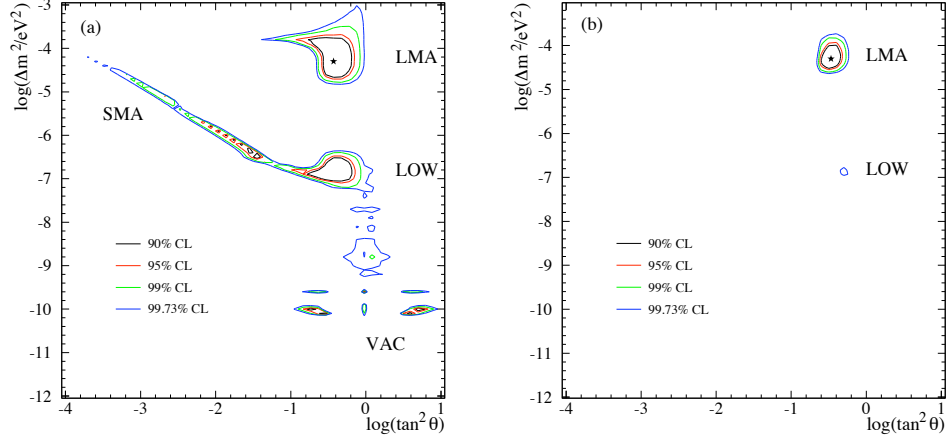
Figure 3.10: Allowed neutrino mixing parameters for SNO-only (left) and SNO+solar experiments (right) based on the D$_2$O-phase results. Star shows the best fit location. Figure from [70].

$7.0\% \pm 4.9\%(\text{stat})\,^{+1.3}_{-1.2}\%(\text{syst})$, where asymmetry is defined to be

$$\mathcal{A} = 2\frac{\phi_{\text{night}} - \phi_{\text{day}}}{\phi_{\text{night}} + \phi_{\text{day}}}. \tag{3.4}$$

This allowed meaningful SNO-only contours to be made in the Δm^2, θ_{12} space for the first time. Figure 3.10 shows the SNO-only contours, and the SNO+other solar experiments. At this stage, SNO alone could not resolve whether solar neutrino mixing was in the large mixing (LMA), small mixing (SMA), low Δm^2 (LOW), or vacuum oscillation-dominated (VAC) regime. The best fit point was in the LMA region, and when other experiments were included, all that remained at the 99.73% confidence level were the LMA and LOW solutions.

3.8.2 Salt Phase

When 2 tons of NaCl were added to the D_2O in the salt phase, the detector sensitivity to neutrons improved by a factor of three, dramatically improving the event statistics in the NC channel. The first salt phase paper[71] reported the results from 254 live days, whereas the second paper[32] used the complete data set of 391 live days. The second paper also performed a day-night analysis as had been done in the D_2O phase. Both papers used an electron energy threshold of 5.5 MeV (kinetic).

The increased livetime and improved neutron statistics dramatically improved SNO's estimate of the total ^{8}B neutrino flux to $4.94^{+0.21}_{-0.21}(\text{stat})\,^{+0.38}_{-0.34}(\text{syst})\times 10^6\,\text{cm}^{-2}\text{s}^{-1}$. Additionally, SNO could now rule out with 95% confidence the vacuum and SMA regions of the mixing parameter space, as shown in Figure 3.11. The combined SNO+solar and SNO+solar+KamLAND contours both reduced the parameter space to a single region, the LMA solution, as shown in Figure 2.11.

The CC spectrum in the salt phase showed (Figure 3.12) no significant distortion compared to the expected ^{8}B energy spectrum. With an energy threshold of 5.5 MeV, SNO still cannot see the predicted MSW transition from matter to vacuum oscillations.

3.9   Summary

The SNO detector has unique sensitivity to all flavors of neutrinos through the NC reaction, and to the energy spectrum of electron neutrinos through the CC reaction. Previous analyses have been very successful in solving the
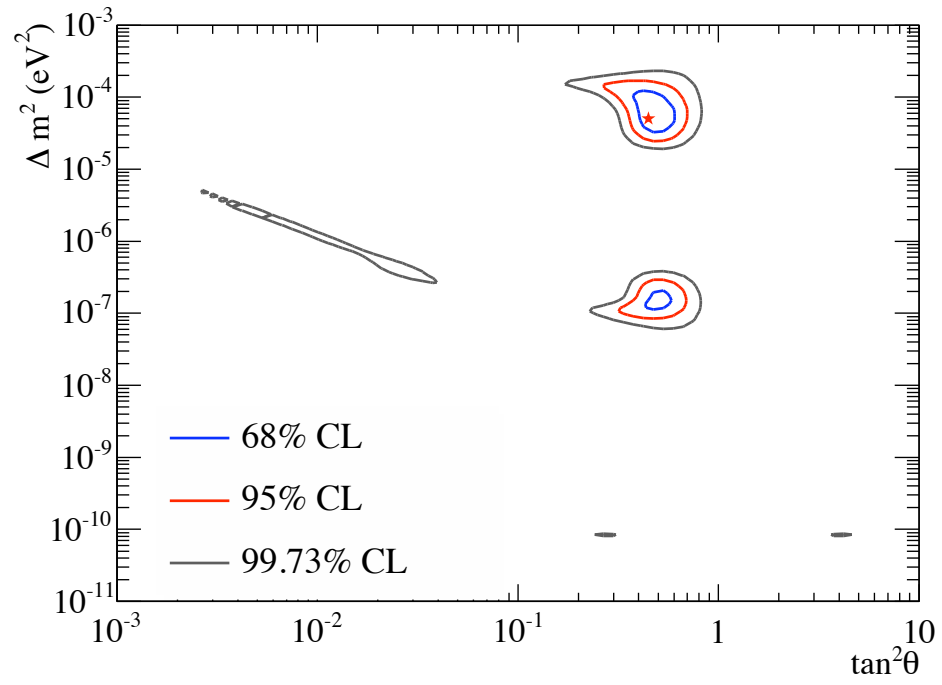
Figure 3.11: Allowed neutrino mixing parameters for D_2O and salt phases of SNO together. Figure from [32].
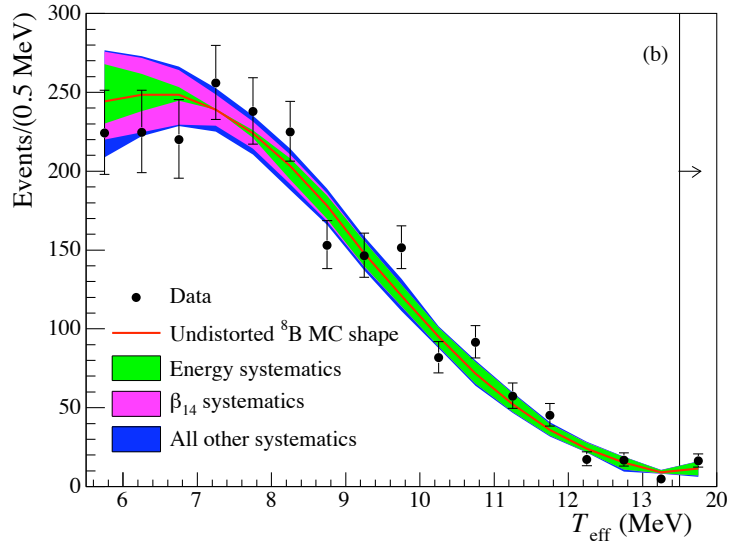
Figure 3.12: Reconstructed electron energy spectrum of CC events with the undistorted ^{8}B prediction overlaid. Systematic uncertainties are shown as bands on the prediction, while statistical uncertainties are shown as error bars on the data. Figure from [32].

solar neutrino problem first observed by Davis and Bahcall. However, at energy threshold of 5 and 5.5 MeV, the MSW transition has not yet been observed.

More can be done, though. Careful attention to background reduction during construction and operation of the experiment, and the low trigger threshold used during operation means that even more neutrino events are lurking in the SNO data set. We need only go lower in energy to find them. The next chapter explains how to achieve this.

Chapter 4

Designing a Low Threshold ^{8}B Neutrino Analysis

SNO's previous results have firmly established that the solution to the solar neutrino problem is in fact new physics! Neutrinos change flavor between the production point in the Sun and the detection point on Earth. In concert with the other solar neutrino experiments and KamLAND, SNO has shown that the neutrino oscillation model is a good fit for the solar data, with mixing parameters in the large mixing angle region of phase space.

After such success, one can reasonably ask: Now what? What can be done in the final years of SNO to best use the data we have collected?

The only place left to go is down; down in energy, that is. Previous SNO publications have reported results with a reconstructed kinetic energy (T_{eff}) threshold of 5 MeV in the D$_2$O phase and 5.5 MeV in the salt phase. With the additional detector knowledge and computing power gained in the time since those publications, we have the ability now to go much, much lower. In

	CC	NC
D_2O phase	22%	74%
Salt phase	35%	68%
D_2O + Salt phase	30%	69%

Table 4.1: Projected improvement in CC and NC event statistics by lowering the energy threshold from 5.0 and 5.5 MeV in the D_2O and salt phase, respectively, to 3.5 MeV in both phases. The combined D_2O+salt improvement weights the phases by their different livetimes and neutron capture efficiencies.

this chapter, we will discuss the goals and methods required for a low energy threshold analysis of the SNO data, as well as lay out the organization for the rest of the dissertation.

4.1 Goals

What does a low energy threshold analysis have to offer? Firstly, lowering the energy threshold dramatically increases the number of neutrino events in the data set. Figure 4.1 shows the reconstructed energy spectrum of simulated NC events with the previous energy threshold marked. More than half of the NC events were discarded before in order to remove backgrounds also present at lower energy. If we can deal with the backgrounds, then huge statistical gains are possible by lowering the energy threshold. Table 4.1 shows the improvement in CC and NC statistics by lowering the D_2O and salt thresholds down to 3.5 MeV[1]. With this huge increase in statistics, the measure of the integral ^8B flux can be improved significantly.

But even beyond a more precise integral flux, a lower energy threshold opens up a window on the ^8B spectrum at lower energies than have been

---

[1]Justification of this energy threshold is given in Section 4.2.7.
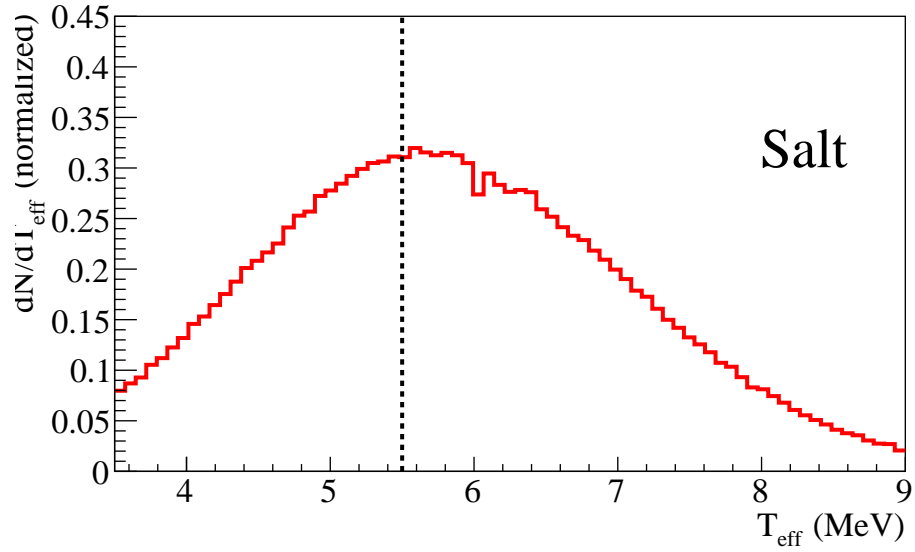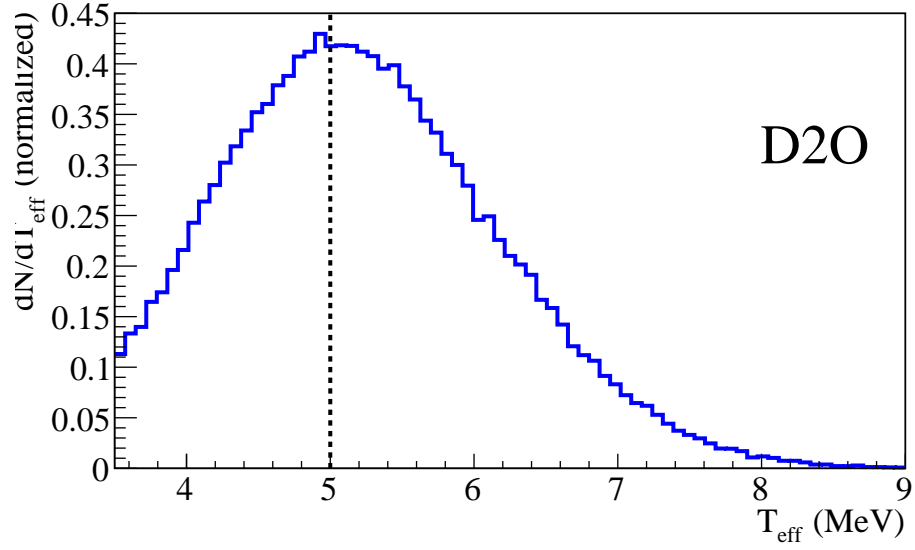
Figure 4.1: Reconstructed energy (T_{eff}) energy spectrum of Monte Carlo neutral current events for the D_2O phase (top) and the salt phase (bottom). The energy threshold of previous analyses is marked by the dashed line.
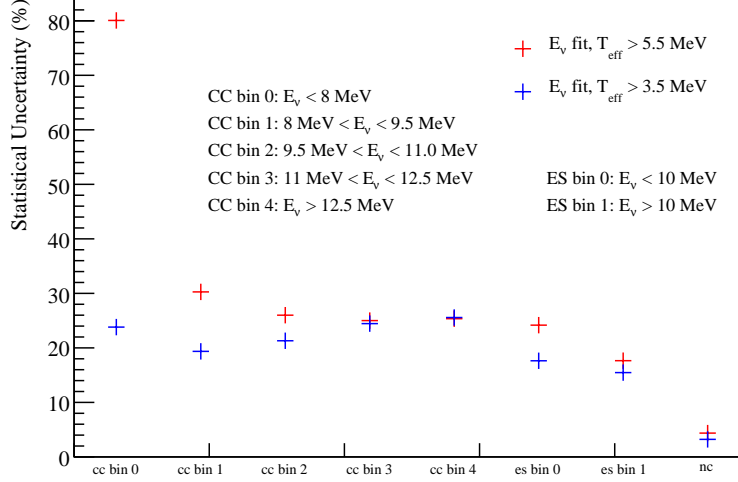
Figure 4.2: Statistical uncertainty in CC spectrum as a function of ν_e energy (not T_{eff}) for 3.5 MeV and 5.5 MeV thresholds. Data from the D_2O phase and salt phase are included for both thresholds.

observed before. The CC energy spectrum provides the most information about the underlying neutrino energy spectrum, so the analysis needs a method to obtain the model-independent energy spectrum of CC events. This CC spectrum can be then used to test the low energy distortion predicted by the neutrino oscillation model in the transition from matter to vacuum-dominated oscillation. Non-standard interactions or other new physics can alter the shape of this transition, so we would like to be as sensitive to the unexpected as possible. Figure 4.2 shows the reduction in ν_e energy spectrum uncertainties possible by lower the energy threshold from 5.5 MeV to 3.5 MeV.

Finally, we want to make a very precise estimation of the solar neutrino mixing parameters, assuming the standard oscillation model. Contours in the Δm_{21}^2, $\tan^2 \theta_{12}$ space are important for combining with other solar experiments

and also testing CPT-invariance with the KamLAND antineutrino results. Figure 4.3 shows the change in the statistics-only contours by moving from a 5.5 to 3.5 MeV threshold.

4.2 Method

Achieving the goals laid out in the previous section is no easy task. A broad spectrum of interlocking technical challenges must be overcome to produce a high precision result.

4.2.1 Signal Extraction with Backgrounds

All of the difficulty in lowering the energy threshold for the SNO analysis comes from the presence of irreducible backgrounds. With an energy threshold of 5.5 MeV, there are only 5 significant sources of events:

1. CC electrons

2. ES electrons

3. NC neutrons

4. Photodisintegration neutrons

5. "External" neutrons

The first three event types come from neutrinos, but the last two types come from intrinsic radioactivity in the detector.

Just as a neutrino with at least 2.2 MeV can split a deuteron into a proton and a free neutron, so can a gamma ray, through a process called

71
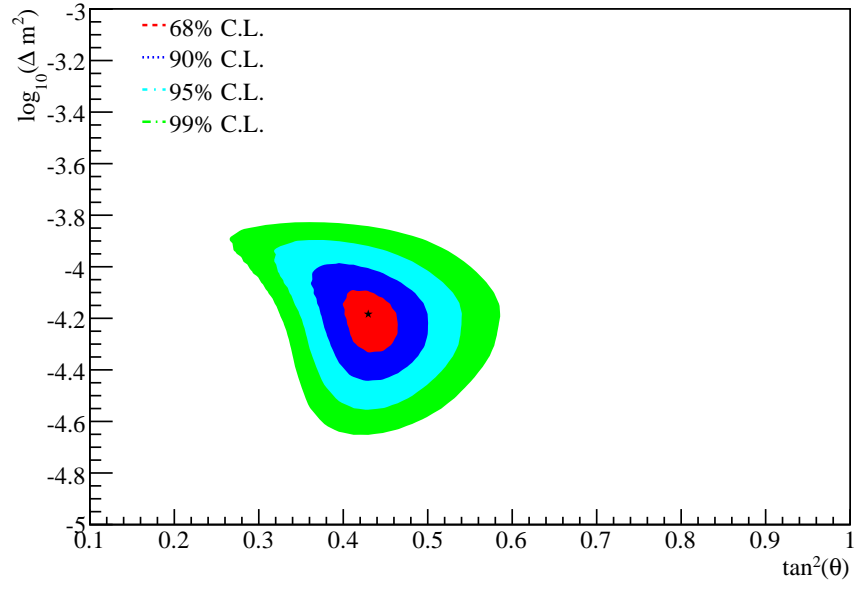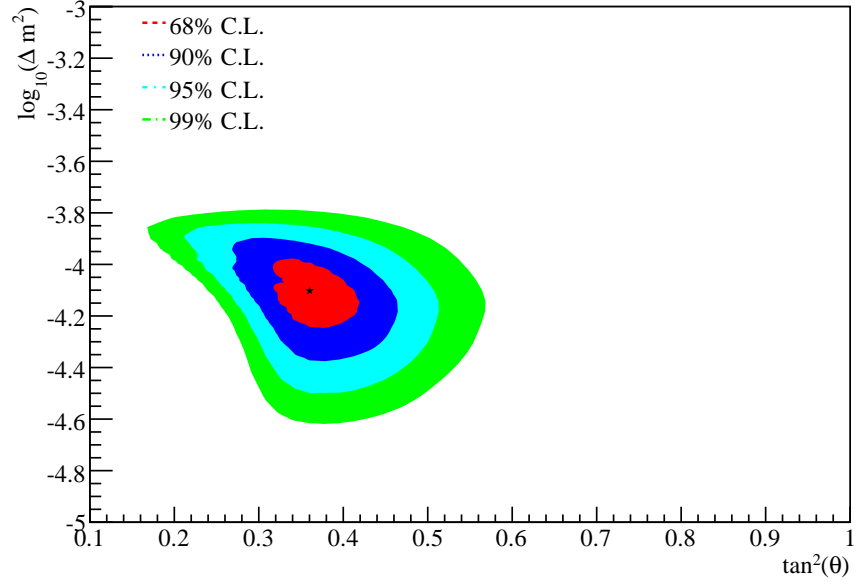
Figure 4.3: Statistics-only contours in the LMA region obtained from a two-phase fit to simulated data sets at 5.5 MeV (top) and 3.5 MeV (bottom).

"photodisintegration." Fortunately, only two beta-gamma decays in the ^{238}U and ^{232}Th chains can produce sufficiently energetic gammas to contribute to photodisintegration: ^{214}Bi (from the uranium chain) and ^{208}Tl (from the thorium chain). The beta-decay of bismuth has a Q-value of 3.3 MeV, with the two most common branches having beta endpoint energies of 3.3 MeV and 1.5 MeV. Very rarely, the decay takes a branch to a highly excited state of of ^{214}Po, which can then emit a nuclear de-excitation gamma with more than 2.2 MeV of energy. The Q-value for the decay is low enough to ensure that no gamma ray has sufficient energy to disintegrate two neutrons.

The second decay of interest, ^{208}Tl, has a much larger Q-value of 5.0 MeV. Angular momentum conservation forces ^{208}Tl to always decay to an excited state of ^{208}Pb. The gamma cascade which follows always terminates with a 2.6 MeV gamma. This gamma is the only one in the cascade which can photodisintegrate a deuteron. Similar to bismuth, a thallium decay can at most generate one photodisintegration neutron.

"External" neutrons can also come from photodisintegration from gamma rays outside the D_2O volume, but also from (α, n) reactions at the surface of the acrylic vessel (AV). Although great care was taken to clean the AV during construction, ^{222}Rn (uranium chain) in the mine air certainly came in contact with the acrylic. Alpha decays during this time can implant daughter nuclei into the acrylic surface, where they will remain after the detector is full. In particular, ^{210}Pb has a 22.3 year half-life, providing a stable source of ^{210}Po. When polonium-210 decays to ^{206}Pb, it produces a 5.4 MeV alpha, which can interact with a carbon nucleus, generating a free neutron.

Internal photodisintegration neutrons cannot be distinguished from neutrons produced by solar neutrinos, so the intrinsic radioactivity must be mea-

sured independently and its neutron production subtracted from the final NC result. External neutrons, coming from outside the D_2O volume, have a different radial profile allowing them to be distinguished and removed during the fitting procedure.

New "Electron-like" Backgrounds

When the analysis threshold is lowered below 5 MeV, a new source of backgrounds becomes significant. The 3.3 MeV beta from ^{214}Bi can be observed directly through Cherenkov light, and the 2.6 MeV gamma from ^{208}Tl can be seen when it Compton scatters an electron. The actual observation of Cherenkov light (after absorption, scattering, quantum efficiency, etc) in SNO is a rare-enough process that there are large random fluctuations in the number of detected photons. After the energy is reconstructed, this can lead to a small number of these electron-like background events having apparent kinetic energies as high as 4.5 MeV. A low energy analysis will need to deal with these ^{214}Bi and ^{208}Tl decays from the heavy water, the acrylic bulk, the light water, and the PMTs. Treating all of the "electron-like" backgrounds separately eliminates the photodisintegration and external neutron categories from before, since the neutron production is directly correlated with the beta-gamma production that will now be observed directly. An exception to this is are the (α, n) neutrons which continue to be a separate background category.

The salt phase has one additional background, ^{24}Na. Natural sodium is composed entirely of ^{23}Na, but ambient neutrons in the environment can produce ^{24}Na. Sodium-24 has a beta-gamma decay very similar to ^{208}Tl, but with a 2.7 MeV gamma at the end in 99.9% of decays. Sodium is activated in small quantities due to the deployment of radioactive sources in the detector,

74

and also during regular circulation of the heavy water out of the acrylic vessel and into the underground lab for purification.

4.2.2   Maximum Likelihood Fitting

In terms of data recorded, SNO has a very simple detector. All that is recorded is the time, charge, and identity (to lookup position) of each hit PMT. The interaction which produced a particular event cannot be inferred from the light pattern with any acceptable degree of certainty. However, the *distribution* of measurements of from many events does show identifiable patterns depending on the source of the event. The standard technique, used in previous SNO analyses as well, is to perform an *extended maximum likelihood fit*, where the free parameters are the number of each kind of event in the data sample. Section 8.2 describes the extended maximum likelihood method in detail.

Maximum likelihood fitting begins with a postulated list of all possible sources of events. Based on the discussion of backgrounds, we can say that the sources of signals are:

- Charged-current electrons

- Elastic-scattering electrons

- Neutral current neutrons

- ^{214}Bi in the D_2O volume

- ^{214}Bi in the acrylic volume

- ^{214}Bi in the H_2O volume

- ^{208}Tl in the D_2O volume

- ^{208}Tl in the acrylic volume

- ^{208}Tl in the H$_2$O volume

- ^{208}Tl in the PMT region[2]

- ^{24}Na in the D$_2$O volume (salt phase only)

- (α, n) reactions on the surface of the acrylic vessel

4.2.3   Choice of Observables

The most important decision in setting up the maximum likelihood fit is the choice of observables. Observables are the reconstructed attributes of an event, which summarize the pattern of observed light in some concise way. Usually these observables correlate with microphysical quantities of the relativistic electron which produced the light, such as the energy or direction of the particle. In principle, though, they can be anything we can derive from the hit pattern recorded by the detector.

We want the smallest number of observables, as PDF generation (Section 4.2.5) becomes more difficult the more observables we use. The collective experience of the SNO collaboration has identified four observables which are essential for separating event types in a fit. Table 4.2 summarizes these observables, which are also discussed in detail in the next sections.

Energy

Energy is obviously required to measure the spectrum of CC-produced electrons, but is also very important for separating electron-like backgrounds from

---

[2]See Chapter 6 for an explanation of why we do not include ^{214}Bi in the PMT region.

Observable	Symbol	Discriminating Power
Energy	T_{eff}	Low energy backgrounds vs. neutrinos
Radius	R^3	Internal vs. external backgrounds, CC vs. NC (weak)
Isotropy	β_{14}	^{214}Bi vs. ^{208}Tl, CC vs. NC
Direction	$\cos\theta_{\odot}$	CC vs. ES, CC vs. ^{214}Bi (weak)

Table 4.2: Four event observables used in signal extraction.

higher energy neutrino events. The reconstructed kinetic energy of an event is called T_{eff}, and the reconstruction process determines the most probable energy of a single electron that would have produced the number of hit PMT we observe. Events that come from beta-decay or neutron capture-induced gamma rays, which can scatter multiple electrons, will reconstruct with a lower electron-equivalent energy. Energy estimation is discussed in more detail in Chapter 5.

Figure 4.4 shows the T_{eff} distributions for neutron captures, ^{214}Bi betas, and ^{208}Tl gammas.

Radius

Although we reconstruct the (x, y, z) position of every event, the approximate spherical symmetry of the detector means that in the maximum likelihood fit, only the radius of the event in PSUP-centered coordinates is relevant. For convenience when making PDFs and plotting radius, we almost exclusively use a normalized cubic radius R^3, which is (radius/AV radius)^3. In these coordinates, an event at the inner surface of the acrylic vessel will have $R^3 = 1$, and events from a uniformly distributed source in volume will have a flat distribution in R^3.

Radius is important for separating events inside the heavy water region
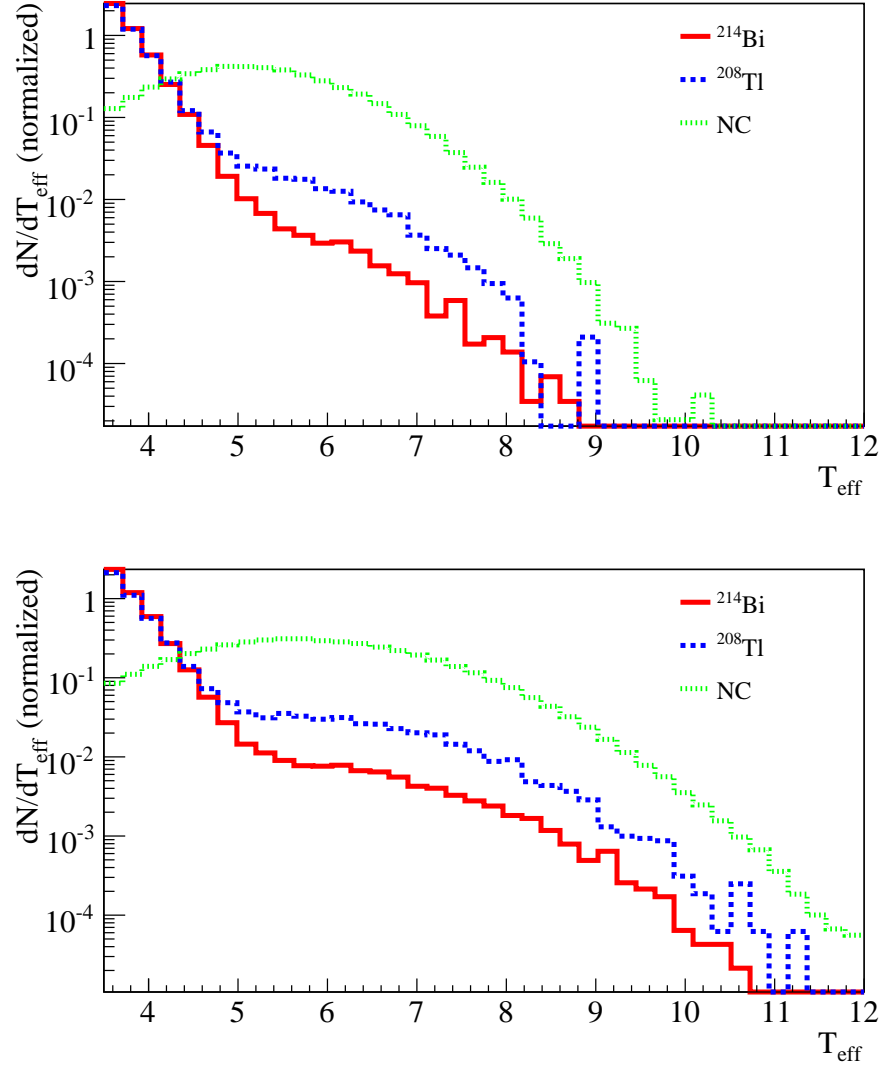
Figure 4.4: T_{eff} distributions for neutron captures, ^{214}Bi betas, and ^{208}Tl gammas in the heavy water for the D_2O phase (top), and salt phase (bottom).

78
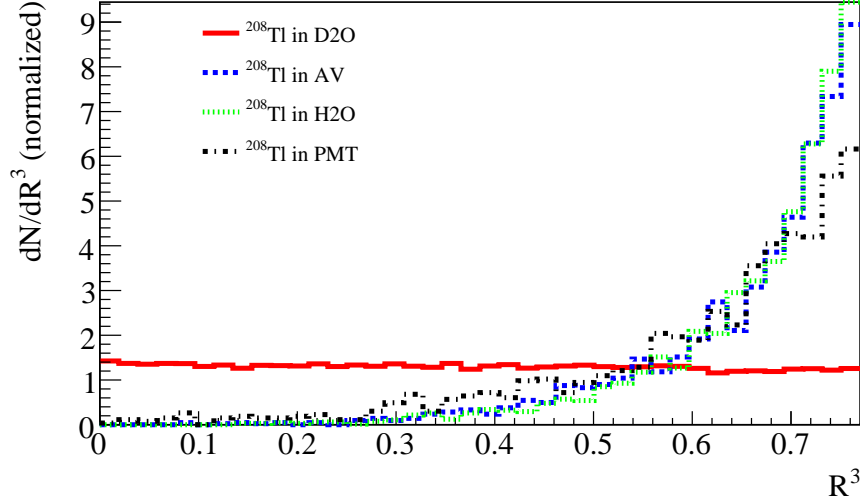
Figure 4.5: R^3 distributions for ^{208}Tl decays in various regions of the detector.

from events outside. Figure 4.5 show the radial distributions for different background regions in the salt phase. Additionally, radius provides a weak handle on NC events, as the neutron thermalizes and wanders some distance before capturing. The hydrogen in the acrylic vessel acts as a very efficient neutron sink, so the radial distribution of a uniform source of neutrons is not flat. Instead, there is a roll-off at high radius, shown in Figure 4.6. The sharpness of this roll-off depends on the overall capture efficiency of the heavy water (plus salt in second phase), with the D_2O -phase having an overall slope to distinguish it from the flat R^3 distribution of CC events.

Isotropy

The isotropy of the PMT hit pattern gives us an indication of the type of particle(s) that made initiated the event. A single electron, from CC, ES, or ^{214}Bi,
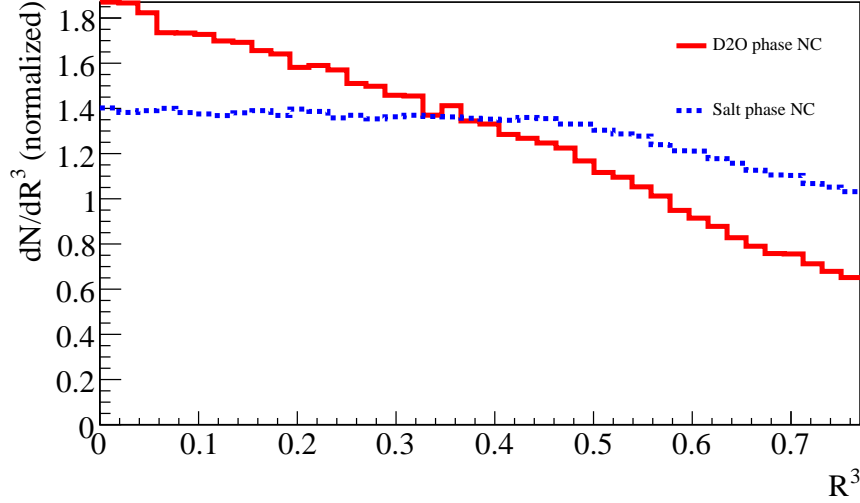
Figure 4.6: R^3 distributions for NC events in D_2O and salt phase.

will make one Cherenkov cone. Some amount of multiple scattering, especially for lower energy electrons, smears out the cone, and can even produce a few back-scattered hits. A single gamma, such as from [208]Tl or the capture of a neutron on deuterium, produces a slightly more isotropic distribution since a gamma can sometimes scatter more than one electron with energy above the Cherenkov threshold. In the salt phase, the cascade of multiple gammas when a neutron captures on [35]Cl produces an even more isotropic distribution.

Quantifying isotropy has been done in two ways historically on SNO. In D_2O phase analyses, isotropy was measured using the mean hit PMT pair angle $\langle \theta_{ij} \rangle$ as viewed from the reconstructed event vertex[44]. In the salt phase, it was found that a better measure of isotropy was a parameter called β_{14}[72].
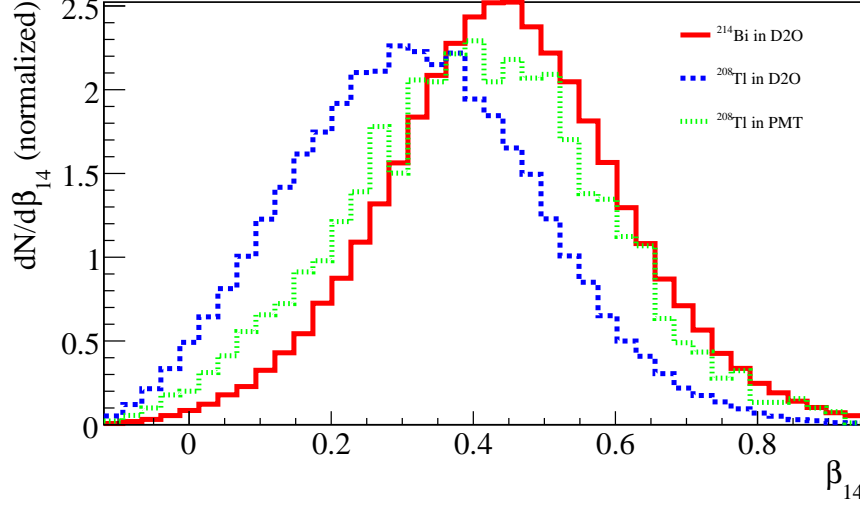
Figure 4.7: β_{14} distributions for internal ^{214}Bi in the heavy water, ^{208}Tl in the heavy water, and ^{208}Tl from the PMT region.

The β_{14} parameter is a linear combination

$$\beta_{14} = \beta_1 + 4\beta_4 \qquad (4.1)$$

with the β_l coefficients defined as

$$\beta_l = \frac{2}{N(N-1)} \left[\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} P_l(\cos\theta_{ij}) \right], \qquad (4.2)$$

where N is the number of hit PMTs in the event, P_l are the Legendre polynomials, and θ_{ij} is the angle between hit PMT i and j.

For consistency, we use β_{14} for both D_2O and salt phases in this analysis. Figure 4.7 shows the separation of backgrounds, and Figure 4.8 shows the separation of CC from NC using β_{14}.
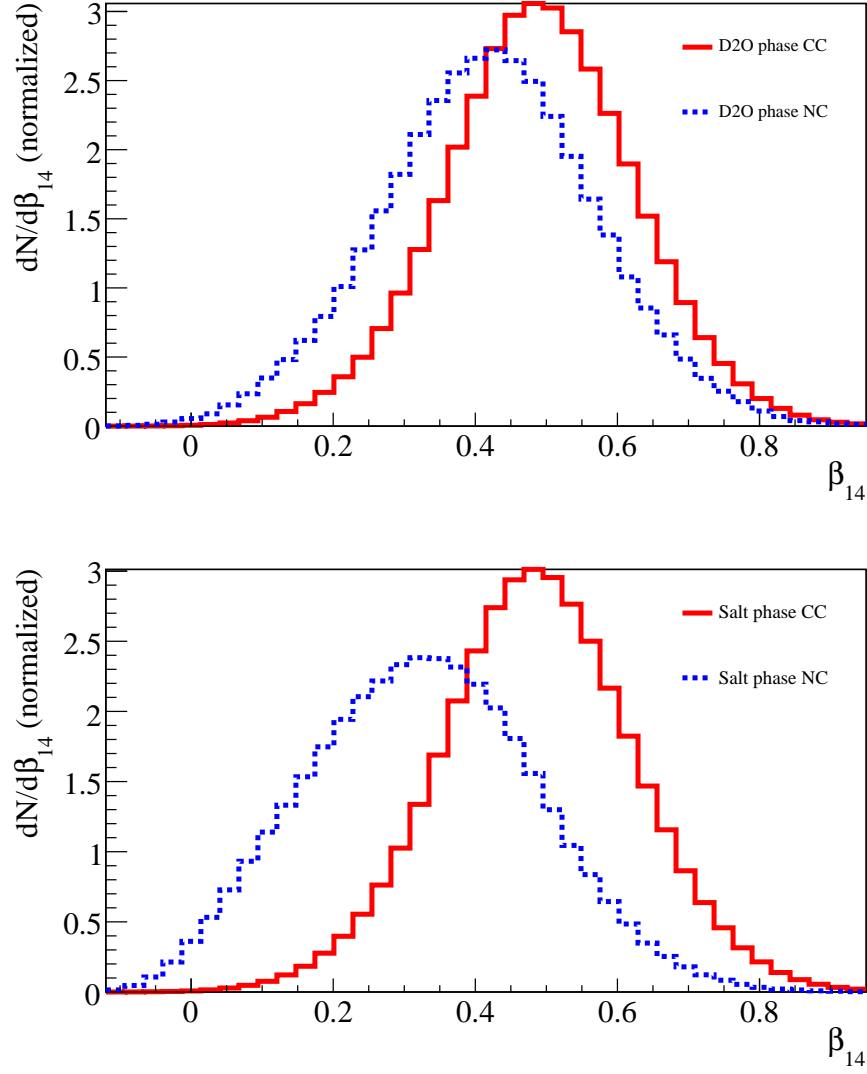
Figure 4.8: β_{14} distributions for CC and NC events in the D_2O phase (top), and salt phase (bottom).

Direction relative to Sun

The direction of the axis of the Cherenkov cone is reconstructed at the same time as the position of the event. This direction is important because it correlates with the position of the Sun for neutrino events. We quantify this with the observable $\cos\theta_\odot$ (spoken "cos theta Sun"), which is the dot product of the reconstructed electron direction vector with the unit vector pointing from the Sun to the Earth. A forward scattered electron has $\cos\theta_\odot = 1$, and a backward scattered electron has -1.

Elastic scattering of electrons has a strong forward scattering peak, which can be clearly seen in the data, indicating the source of neutrinos is in fact the Sun. The charged current electrons also have a non-flat $\cos\theta_\odot$, but is a linear function with a $-1/3$ slope, indicating a slight tendency for CC electrons to be emitted toward the Sun, rather than away from it. Figure 4.9 show the CC and ES distributions in $\cos\theta_\odot$. All other event sources, including NC, have a completely flat $\cos\theta_\odot$ distribution.

4.2.4 Joint Phase Fitting

If we can assume that the neutrino output of the Sun is stable on the several year timescale, then we can link the neutrino flux measured in the D_2O phase with the salt phase[3]. By requiring the CC, ES and NC neutrino fluxes (and CC/ES spectra) to be the same in both phases, we get an additional handle on the separation of CC from NC. In the D_2O phase, the total number of NC events is relatively small (~ 900), and the mean energy is lower than in the salt phase. One could consider the NC in the D_2O phase to be a background

---

[3]One must correct for the eccentricity of the Earth's orbit, which the two phases sample differently.
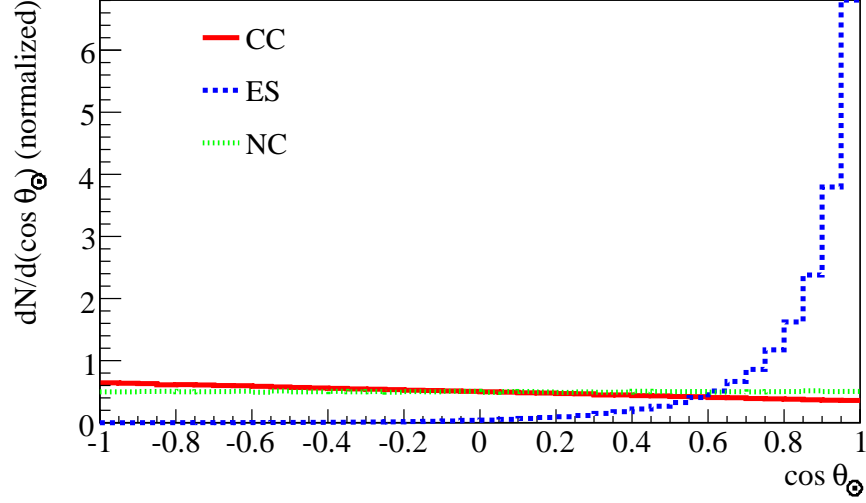
Figure 4.9: $\cos\theta_\odot$ distributions for CC and ES events in the D_2O phase. The salt phase is very similar.

to the low energy CC spectrum measurement, as β_{14} is a weak discriminator of CC and NC in that phase. However, in the salt phase, the NC statistics are 4 times larger, and the isotropic gamma cascade in chlorine means that NC is much easier to separate from CC using β_{14}. When we link the two phases together in the fit, the salt phase NC value can feed back into the D_2O phase CC spectrum by constraining the D_2O phase NC contribution. The self-consistency enforced by doing the two fits together results in a better estimate than performing each phase fit separately and combining the results. In fact, a joint phase fit is almost more powerful than β_{14} for separating CC from NC, as shown in Table 4.3

Joint D$_2$O-salt fit, use β_{14}	2.9%
Joint D$_2$O-salt fit, no β_{14}	4.9%
Salt-only fit, use β_{14}	3.1%
Salt-only fit, no β_{14}	14.1%

Table 4.3: Statistical uncertainty of NC flux for various maximum likelihood fit configurations.

4.2.5 PDF generation with Monte Carlo

The maximum likelihood method requires that we provide a PDF for each type of event we think is in the data set. Based on the list given in Section 4.2.2, we have 23 PDFs to construct, 10 for the D$_2$O phase and 11 for the salt phase. There are no calibration sources to replicate many of these signals, so instead we have adopted a three stage PDF generation strategy:

1. Test and tune SNOMAN so that simulation of calibration sources best matches the data that is observed when the source is deployed.

2. Use SNOMAN to produce a large sample of events for each PDF.

3. Observed differences between calibration source data and simulation become uncertainties for the SNOMAN-generated PDFs.

Essentially, we are using the simulation to "interpolate" between calibration sources using our knowledge of the microscopic physics. Calibration sources tell us the uncertainty of SNOMAN, and transitively, the uncertainty on our final PDFs.

85

4.2.6   Improving Energy Resolution and Systematics

The key advancement which made a low energy analysis practical was the development of a new energy estimation algorithm by M. Dunford[51]. The new algorithm included PMT hits from a much larger time window than past algorithms, increasing the photons counted by 12%. As a result, energy resolution was improved by 6%. Such a small change in resolution has an enormous impact on reducing low energy radioactive backgrounds. The steep exponential tail of backgrounds that leak to higher energies can be suppressed by up to 60%, allowing the CC neutrino events to be separated much better.

This improved energy estimator, as well other simulation upgrades, now allow us to clearly observe $\pm 3\%$ systematic difference between the data and SNOMAN as a function of the z component of the event position. Chapter 5 discusses this z-asymmetry, and shows a method for correcting it, while still providing an independent measure of systematic uncertainty.

4.2.7   Selecting an Energy Threshold

There still remains the obvious question: how low can we go? The most problematic background for low energy CC events is ^{214}Bi in the heavy water, as it is observed most often when it decays to a single energetic electron. Neither isotropy nor radius can separate the two, as both are sources of uniformly distributed single electrons. Only $\cos\theta_\odot$ provides a weak handle, as CC electrons are slightly more likely to point back toward the Sun, while ^{214}Bi are completely random.

The natural energy cutoff is then the energy at which ^{214}Bi decays dominate over the expected CC signal. For the purposes of this estimate, we
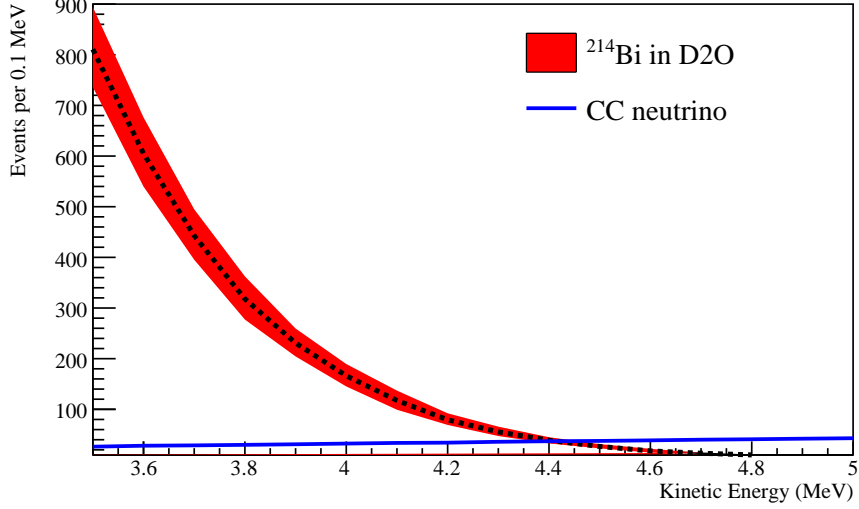
Figure 4.10: Expected T_{eff} spectra for ^{214}Bi in the heavy water and CC electrons.

can assume the CC spectrum is undistorted and suppressed to 0.35 SSM, as found in previous SNO results. External radiochemical measures of the radon levels in the water provide an independent estimate of the ^{214}Bi decay rate. Figure 4.10 shows the reconstructed energy spectrum for these two signals, normalized by their expected rates. The crossover point at $T_{\text{eff}} = 4.5$ MeV is where the signal-to-background ratio reaches 1.

However, we can push further into the exponential wall of bismuth. Ultimately, what matters is not the amount of background events, but our uncertainty in their number. If we could have perfect knowledge of the number of background events, then we could subtract them off (or allow the maximum likelihood fit to do that) the observed spectrum and be left with only CC events. For this reason, we have chosen to go an entire MeV lower, to an energy threshold of 3.5 MeV, which allows us to measure of the bismuth contribution

to better than 7%. Below 4 MeV, the CC uncertainty climbs to over 80%, but in the window 4–4.5 MeV, we can measure the CC electrons with 18% statistical uncertainty due to the bismuth constraint. Chapter 8 discusses the separation of ^{214}Bi and CC in more detail.

4.2.8 Better PMT Modeling

A 3.5 MeV threshold provides an excellent measure of bismuth, but it introduces a new background which previously had been insignificant. The PMTs, and their support hardware, are the most radioactive part of the detector. Their placement 3 meters from the fiducial volume of the detector suppresses much of this radioactivity from leaking into the analysis, but below 4 MeV, the PMT contribution becomes as large as the bismuth. The model used to build the PMT PDF becomes very important, as inaccuracies skew the measure of bismuth and bias the CC spectrum. Chapter 6 discusses the difficulty in constructing a PMT PDF in the same way as the other signals, and describes an alternate analytic form that we use instead.

4.2.9 More Sophisticated Treatment of Systematics

Reducing statistical uncertainties is pointless if the systematic uncertainties cannot also be reduced. Improvements to the SNOMAN simulation have lowered these systematic uncertainties, as have explicit corrections derived from calibration data. We can go one step further, however, and float the systematic uncertainties in the fit.

Floating uncertainties is a standard procedure for including uncertainties on PDFs in the maximum likelihood formalism. We add nuisance param-

eters to the likelihood function, which represent systematic offsets in the PDF shapes, such as an energy scaling or a convolution with a Gaussian of some width. The external estimates on these shape uncertainties we derive from calibration data become constraints on the nuisance parameters. By allowing the parameters to vary during the fit, we also use the neutrino data itself to further constrain the parameters. This can be a large improvement over the simpler "shift-and-refit" approach to propagating systematics (see Section 8.2.4), where a shape distortion is applied and the fit is repeated to see how much the neutrino flux changes. Shift-and-refit systematics ignores the change in the likelihood at the shifted point, which might in fact be better than the original fit.

Allowing systematics to float during a fit as large and complex as that described above is non-trivial. Chapter 7 discusses a PDF construction method, called kernel estimation, which makes floating systematics more computationally tractable for large, multidimensional PDFs. In Chapter 8, we show how kernel estimated PDFs with systematics are incorporated into the full maximum likelihood formalism.

4.3   Data Processing

Events are processed through an analysis pipeline shown in Figure 4.11. Due to the detailed, PMT hit-level simulation provided by SNOMAN, both Monte Carlo events and real events recorded by the detector can be processed through very similar pipelines. PDF generation is described in 7, and maximum likelihood fitting and ensemble testing in 8. The other portions of the analysis are described below.
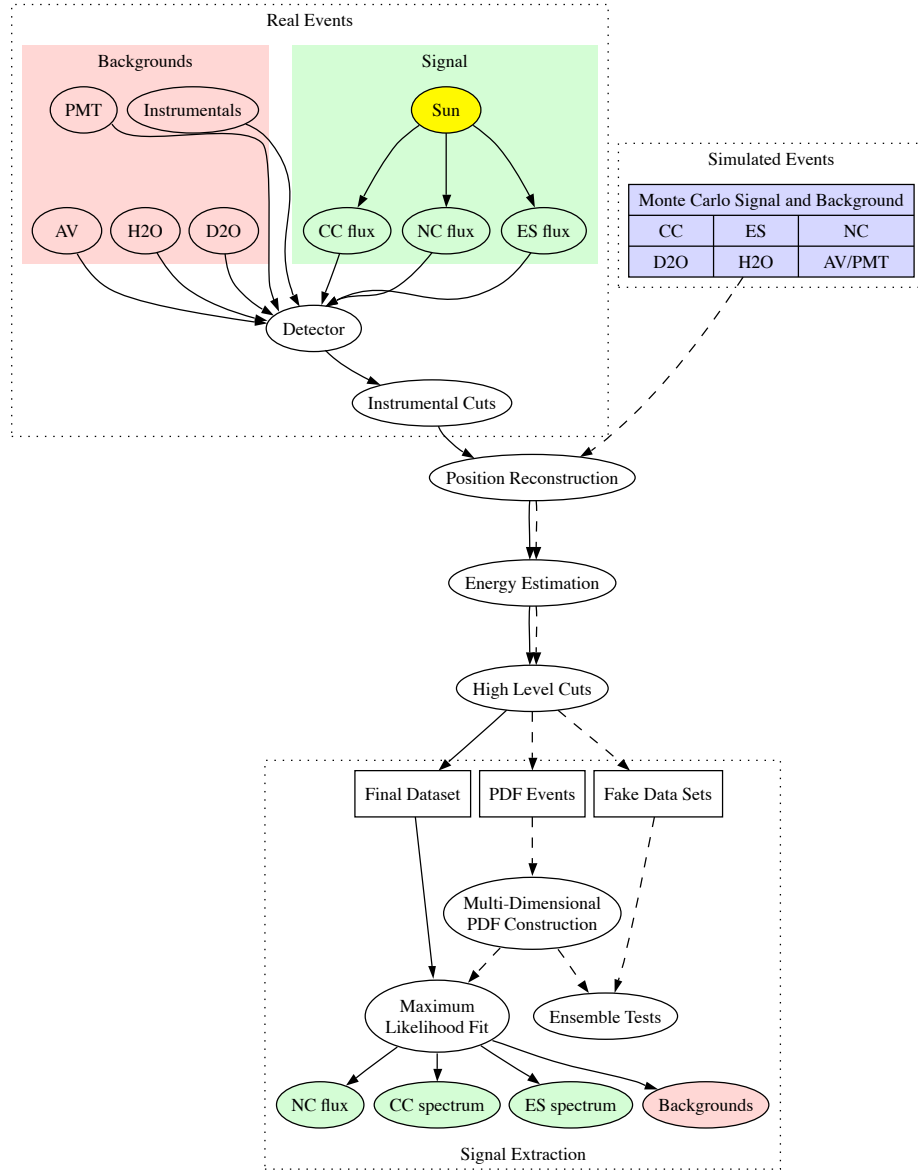
Figure 4.11: Data analysis flowchart. Dashed lines indicate Monte Carlo events.

4.3.1 Instrumental Cuts

Instrumental cuts are used to remove events which are caused by non-Cherenkov processes like electronics noise and "flashers," which are PMTs which randomly emit a short burst of light. The cuts used in this analysis are discussed in more detail in Appendix B of [51], but broadly speaking, they fall into several categories

- DAQ cuts: Eliminate events which are corrupted due to rare problems during data readout. These include PMT hits without a global trigger ("orphans") and events where a PMT is hit multiple times.

- Rate cuts: Remove bursts of events in a short time interval. This eliminates retriggers from very high energy events, like muons, as well as other sources of multiple events, like bursts of light which last for microseconds or longer.

- Charge cuts: Flashers are usually associated with a single PMT or small set of adjacent PMTs which record a very high charge. Electrical noise, on the other hand, tends to integrate to zero charge. Cutting on the pattern of charge deposited in the detector removes both extremes.

- Time cuts: Cherenkov light is very short in duration, so remove events with a very wide spread in PMT hit times.

- Hit geometry cuts: Remove events with a small group of hits on one side of the detector, and a large band or ring of light on the other side of the detector, another characteristic of flashers. Also remove events with hits in a large number of adjacent electronics channels, a sign of noise pickup.

- Light water veto: Remove events which also include hits on outward looking PMTs, or on special PMTs installed in the neck.

- Muon/Atmospheric ν cuts: Eliminate events which occur less than 20 seconds after a muon, or 250 ms after a likely atmospheric neutrino. This ensures that spallation products do not appear in the data set.

SNOMAN does not simulate instrumental events, so Monte Carlo data bypasses the instrument cut stage.

4.3.2 Position Reconstruction

The initial time, position and direction of the Cherenkov cone produced by the electron in the event is reconstructed using the PMTs hit times. The reconstruction algorithm used in the data processing of this analysis is called "FTP," and was developed by M. Neubauer[73]. FTP is a maximum likelihood fitter which optimizes a log-likelihood function with 7 free parameters, computed from the position and time of all of the hit PMTs:

$$\log \mathcal{L}(\mathbf{r}_e, \mathbf{u}_e, t_e) = \sum_{i=1}^{N_{hit}} \log P(t_i | \mathbf{r}_e, \mathbf{u}_e, t_e, \mathbf{r}_i), \qquad (4.3)$$

where $\mathbf{r}_e$ is the event position vector, $\mathbf{u}_e$ is the event direction vector, t_e is the event time, $\mathbf{r}_i$ is the position vector of hit PMT i, and t_i is the hit time of PMT i. FTP uses both the time of flight, and the angle between the direction of the event, $\mathbf{u}_e$, and the vector from the event vertex to the hit PMT, $\mathbf{r}_i - \mathbf{r}_e$, to construct the PDF $P(t_i)$. The angular portion of the PDF is strongly peaked at the Cherenkov angle, but also includes a broad tail due to multiple scattering of the electron which Cherenkov light is being emitted.

FTP achieves a position resolution of 20 cm in the x, y and z dimensions.

4.3.3   Energy Estimation

After position reconstruction is complete, then energy estimation is performed. Although events can be initially produced by a variety of particles, the energy estimation algorithm determines the most probable event energy under the assumption that it is a single electron. This is certainly a valid assumption for CC and ES electrons, but not for low energy gamma backgrounds or gammas cascades from neutron capture. Nevertheless, we do not actually need the reconstructed energy of these events to be the true energy of the interaction. We only need a consistent distribution that can be included in signal extraction.

The energy estimation algorithm used in this analysis is called "FTK," and was developed by M. Dunford[51]. FTK is described in detail in Chapter 5.

4.3.4   High-level Cuts

The term "high-level cuts" is used to describe event cuts which rely on reconstructed quantities, like event time, position, or energy, to discard events. High-level cuts are designed to remove radioactive backgrounds which have reconstructed above our energy threshold of $T_{\text{eff}} > 3.5$ MeV and inside our fiducial volume of $r < 550$ cm. Backgrounds which originate outside of the fiducial volume but reconstruct far from their true location will generally have very unusual hit patterns. The high level cuts were developed[74] to remove such events. These include:

- In-Time Ratio (ITR): Cut events where the time residual (hit time, less

93

time of flight and event time) distribution shows a large fraction of hits outside a prompt window of -2.5 to 5.0 ns.

- Charge-weighted θ_{ij}: A weighted isotropy measure which removes events with clusters of high charge.

- Cherenkov cone KS tests: Three Kolmogorov-Smirnov (KS) tests are performed on the time, azimuthal, and 2D polar-azimuthal distribution of PMT hits to see if they are consistent with the distribution expected from a Cherenkov cone. This removes misreconstructed events.

- Energy uncertainty: The energy estimator returns a most probable energy, as well as an uncertainty on that energy. Misreconstructed events tend to have very large energy uncertainties, and can be removed.

Unlike instrumental cuts, high-level cuts are applied to both Monte Carlo and real data.

4.4   Blindness

Blind analysis is a well-established technique in physics to control the unconscious tendency of an experimentalist to bias data analysis toward prior expectations[75]. Blindness schemes vary widely between experiments, but all implementations of blindness hide some aspect of the data until the data processing methods and data cleaning cuts have been finalized. Once testing on the blinded data has been completed to the satisfaction of the experimentalist, the "box can be opened." Then the unblinded data is passed through the same data analysis procedure and the result reported to the physics com-

munity. Modifications to account for unforeseen problems are still possible at this stage, but must be reported alongside the original result.

SNO has a strong commitment to blind analysis, having used it successfully in the primary ^{8}B analyses from each phase. However, circumstances make it impossible to be truly blind in this low energy threshold analysis:

- More than half of the neutrino events in this analysis have been published previously in [69] and [32].

- All of the events with $T_{\text{eff}} > 4$ MeV were analyzed in the dissertation by M. Dunford[51], whose work forms the starting point for this dissertation. These results have not yet been included in a publication by the SNO collaboration.

All of the people involved in the work described herein have read these documents extensively, and so cannot be considered blind analysts.

However, we believe in the value of the blind analysis methodology, even when applied in a weaker form. As a result, we have adopted a *pseudo-blind* approach to the low energy threshold analysis. We have implemented statistical blindness, which can be best summarized as:

> *A randomly selected 2/3 subset of the events within the analysis window will be hidden from all participants until the analysis procedure, including cuts and signal extraction code is finalized.*

No blindness is applied to calibration data, or to events which fail cuts (after cuts are finalized), or to events which fall outside the analysis window. Note that some events which have been previously analyzed are *re-blinded* by this procedure.

4.5 Summary

The SNO data set has quite a bit of information left to exploit. By lowering the analysis threshold to 3.5 MeV, we can dramatically improve the NC measurement, as well as extend the CC spectrum measurement down to lower energy. In the remainder of this dissertation, we discuss the technical challenges that must be overcome to make this measurement.

Chapter 5

Energy Estimation and Uncertainties

Energy estimation is one of the most important event reconstruction tasks. We not only need energy to determine the spectrum of CC electrons, but as a background rejection parameter. Position reconstruction can cut radioactive backgrounds from outside the fiducial volume, but there is also uranium chain (primarily from radon) and thorium chain radioactivity inside the fiducial volume. Rejecting these internal backgrounds is best done with energy. All of the background decays generate betas and gammas with energies below our energy threshold of 3.5 MeV, but the finite detector resolution produces a tail of events which leak above threshold. For this reason, we will see that minimizing the variance (i.e. improving the energy resolution) of the energy estimation algorithm is at least as important as minimizing the bias.

In this chapter we will review the energy estimation algorithm *FTK*, designed by Monica Dunford[51], and describe a revised set of energy corrections

and systematic uncertainties for the first two phases of SNO.

5.1   A General Approach to Energy Estimation

Given an accurate simulation of a detector and infinite CPU time, the best possible reconstruction algorithm would be a maximum likelihood fit which used the Monte Carlo as the likelihood function. For each possible position and energy, a high statistics simulation would produce time and charge PDFs for each PMT, which could then be used to compute the likelihood of the observed event. All optical effects would be automatically handled in such a fitter. If the simulation is designed to track time variation in the detector response (as SNOMAN is), then the reconstruction will continue to be optimal even as the detector changes.

Of course, such a fitter is, despite Moore's Law, completely impractical. Instead, we must replace full simulation with analytic models or lookup tables informed by the simulation (which is informed by actual measurements in the real detector). However, we need not abandon the Monte Carlo approach entirely. FTK is an energy estimation algorithm for SNO which uses components of the full optical simulation during the fitting process to find the most probable electron-equivalent kinetic energy, called T_{eff}, for an event, given its reconstructed position and a count of the number of hit PMTs.

Photons travel from an event vertex to the PMTs through a variety of paths. Most photons take a direct path, traveling in a straight line (not counting refraction) from the point of creation to the PMT which records
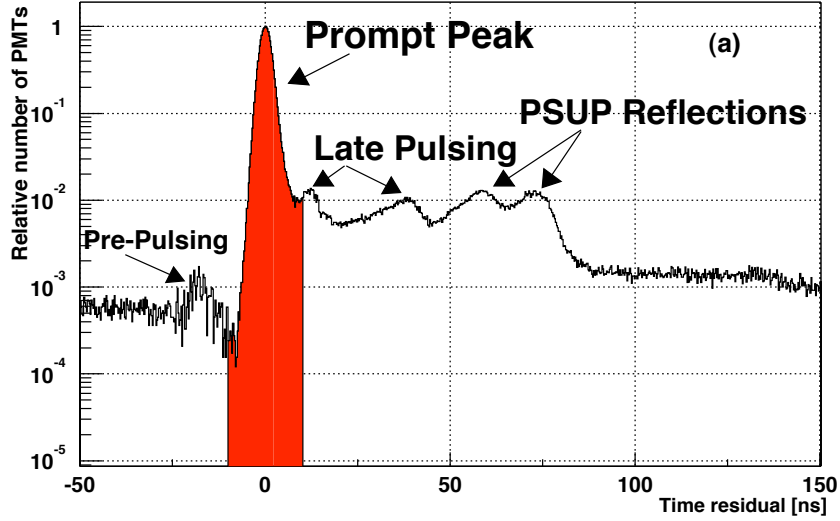
Figure 5.1: Time residual (PMT hit time - light propagation time - event time) for ^{16}N source at the center of the detector. Various sources of early and late light, including reflections from the PMT concentrators ("PSUP reflections"), are noted. The shaded region indicates the photons used by RSP, the table-lookup energy estimator. Figure from [76].

them. However, a small fraction of photons take a longer route, either being scattered or reflected before detection. Figure 5.1 shows the time distribution of hit PMTs for a source of light at the center of the detector. Previous SNO publications have used an energy estimator (called "RSP") based on a 2D lookup table which only used the number of hits in the prompt peak and the radius of the event to determine the energy. Hits more than 10 ns from the prompt peak were discarded as having too much spatial variation to be predicted by radius alone. As a result, 12% of the PMT hits were not used.

5.2 FTK: A Ray-Tracing Energy Estimator

FTK was designed to estimate the energy of the event using all of the PMT hits, not just the prompt hits. To accomplish this goal, the lookup table approach had to be abandoned in favor of a maximum likelihood method where the PDF could be generated on-the-fly for the given reconstructed event position and direction. The free parameter in the likelihood function is the energy of the event, T_{eff}. There is only one observable, N_{hit}, the number of hit PMTs. In principle, one could consider a more sophisticated likelihood function where the hit/not-hit state and hit time of every PMT in the detector was treated as a separate observable. This approach turns out to be very slow, and collapsing all of the hits down to a single number, N_{hit}, enables a number of performance improvements which are necessary to make FTK practical to run on hundreds of millions of events.

As a first step, we can define n_{exp} to be the number of *detected photons* (not PMT hits, as SNO treats two photons in one PMT as a single hit) we expect to see in the event. Breaking this number down, there are five primary sources of photons in the detector:

- n_{exp}^{dir}: Expected number of detected direct light photons which travel from the event to the PMT with only refraction at each media boundary.

- n_{exp}^{scat}: Expected number of detected photons which Rayleigh scatter once in the D_2O or H_2O before detection. Scattering in the acrylic is neglected in this approximation.

- n_{exp}^{av}: Expected number of detected photons which reflect off the inner or outer surface of the acrylic vessel due to the change in index of refraction.

- n_{exp}^{pmt}: Expected number of detected photons which reflect off the PMTs or light concentrators.

- n_{exp}^{noise}: Expected number of PMT noise hits. These are not photons, but must be included in the sum.

Using a variety of techniques, FTK computes the probability of a single photon being detected by any PMT in the detector for each of the optical cases: ρ_{dir}, ρ_{scat}, ρ_{av}, ρ_{pmt}. The direct light probability is found by tracing rays from the event vertex to each PMT, weighting each track by the attenuation probability in each medium, transmittance at each boundary, solid angle of each PMT, and detection probability given the angle of entry into the light concentrator. Scattering and reflection probabilities are found using a hybrid of ray tracing and Monte Carlo simulation of "photon bombs." Further discussion of the probability calculation can be found in Appendix C, which describes my changes to FTK to enable it to be used in the third phase of SNO, after ^{3}He proportional counters were installed into the center of the detector volume.

If we define N_γ, the number of Cherenkov photons produced in the event, then the total number of detected photons given these probabilities is

$$n_{exp}(N_\gamma) = N_\gamma \times (\rho_{dir} + \rho_{scat} + \rho_{av} + \rho_{pmt}) + n_{exp}^{noise}. \qquad (5.1)$$

To be able to compare n_{exp} to N_{hit}, we need to convert from detected photons to hits. Given the rarity of multiple photons in the same PMT, FTK corrects only dominant source term, $n_{exp}^{dir} = N_\gamma \rho_{dir}$, to account for the probability of two photons hitting the same PMT. The correction is rather involved (see Section 5.3.2 of [51]), so we will just call $N_{MPE}(N_\gamma \rho_{dir})$ the multi-photon corrected

number of PMT hits. The total number of hits is therefore

$$N_{exp}(N_\gamma) = N_{MPE}(N_\gamma \rho_{dir}) + N_\gamma \times (\rho_{scat} + \rho_{av} + \rho_{pmt}) + n_{exp}^{noise}. \qquad (5.2)$$

The likelihood of observing N_{hit} hits when expecting N_{exp} hits on average is given by a Poisson distribution,

$$\mathcal{L}(N_\gamma) = \frac{(N_{exp})_{hit}^N e^{-N_{exp}}}{N_{hit}!}. \qquad (5.3)$$

To express the likelihood in terms of energy rather than N_γ, we must integrate over the distribution of N_γ given an energy T_{eff}:

$$\mathcal{L}(T_{\text{eff}}) = \int \frac{(N_{exp}(N_\gamma))_{hit}^N e^{-N_{exp}(N_\gamma)}}{N_{hit}!} \times P(N_\gamma \,|\, T_{\text{eff}}) \, dN_\gamma, \qquad (5.4)$$

where $P(N_\gamma \,|\, T_{\text{eff}})$ is the probability density of emitting N_γ Cherenkov photons in an event with energy T_{eff}. The negative log likelihood can then be minimized using Brent's Method[77], a very efficient 1D minimization algorithm.

5.3   Energy Corrections

FTK is designed to invert the Monte Carlo simulation: the simulation takes the true electron energy produced by an event generator and maps it to a pattern of hit PMTs, whereas FTK goes the other way. Random processes and integration lose information in both directions, so we expect the inversion not to be exact. However, we do want the mapping function to be unbiased, giving the true event energy on average. Due to limitations in FTK there is some residual bias in the energy scale that must be corrected. These corrections

come in three categories, designed to achieve different goals:

- Global corrections: Correct for approximations in FTK which cause the energy scale in both data and Monte Carlo to be wrong.

- Data-only corrections: Correct for detector mismodeling in the simulation by making the energy scale in data look more like the energy scale in Monte Carlo.

- Monte Carlo-only corrections: Differences between FTK and the Monte Carlo that do not affect the data. This category of correction is extremely unusual since the Monte Carlo is tuned to reproduce the data, and FTK is tuned to reproduce the Monte Carlo. Section 5.3.3 describes a particular problem that affected the application of FTK to the D_2O phase Monte Carlo.

Data-only and Monte Carlo-only corrections must be applied first, to make the data energy scale the same as the Monte Carlo, and then the global corrections are applied to make the common energy scale more accurate.

5.3.1   Global corrections

Global corrections common to both data and Monte Carlo are required when FTK cannot reproduce the energy scale of the Monte Carlo simulation, despite using the Monte Carlo functions to compute photon detection probability. This "lack of commutativity" comes about due to the approximations and simplifications made by FTK in order to increase its speed.

The only global correction for FTK comes from its calculation of the probability of multiple photons hitting the same PMT. During fitting, FTK

only estimates the two-photon probability, ignoring the three-photon probability as a higher order effect. Above energies of 12 MeV, near the edge of the ^{8}B spectrum, this first order multi-photon correction becomes noticeably inadequate. A second order polynomial derived from comparing FTK energy to true Monte Carlo energy is used to correct this.

D$_2$O Phase

$$T_{\text{eff,corr}}^{MC,data} = -0.10872 + 1.0277\,T_{\text{eff}} - 0.0012247\,T_{\text{eff}}^2 \tag{5.5}$$

Salt Phase

$$T_{\text{eff,corr}}^{MC,data} = -0.11492 + 1.0276\,T_{\text{eff}} - 0.0012282\,T_{\text{eff}}^2 \tag{5.6}$$

5.3.2 Data Corrections

Data-only corrections to FTK correct the difference between the reconstructed energy for an event in the Monte Carlo, and the same kind of event in the data. Such differences indicate an inaccuracy in the optical model of the detector in SNOMAN.

D$_2$O Phase

Over time, the photon collection efficiency of the detector changes due to aging effects. These changes are tracked using the ^{16}N source, and a drift function is applied to the global collection efficiency in SNOMAN. Both the Monte Carlo and FTK use this drift function, which ensures simulate events incorporate the time-dependence, and that FTK properly inverts the response. However, in the D$_2$O phase, a residual drift was observed in FTK energy that was not seen

in the RSP energy, as shown in Figure 5.2. It is thought that this was caused by a gradual degradation in the concentrator reflectivity which changed the relative quantities of direct light and PMT reflections. As the drift correction was computed using only the RSP window of ± 10 ns around the prompt light peak, FTK still experienced some drift.

With only one energy fitter affected, it was decided to deal with the problem by applying data-only correction factor to FTK energy[51]:

$$
C(JD) = \begin{cases} 1.00371 & \text{if } JD < 9363 \\ 1.2308 - 2.4254 \cdot 10^{-5} \times JD & \text{otherwise} \end{cases}, \qquad (5.7)
$$

where JD is the SNO Julian Date[1] of the event.

The D$_2$O phase also shows a difference between data and Monte Carlo energies as a function of position in the detector. Section 5.4 describes this correction in detail.

Salt Phase

The SNOMAN drift correction in the salt phase accurately tracked the change in detector response for both RSP and FTK. No additional correction was required. A spatial correction for the salt phase is described in Section 5.4.

5.3.3 Monte Carlo Corrections

In order to determine the PMT detection efficiency as a function of incoming photon angle, FTK uses the phenomenological model (called "greydisk")

---

[1]SNO Julian Date is a SNO-internal time coordinate for events which is loosely based on the Julian Date used by astronomers. January 1, 2000 is SNO Julian Day 9132.
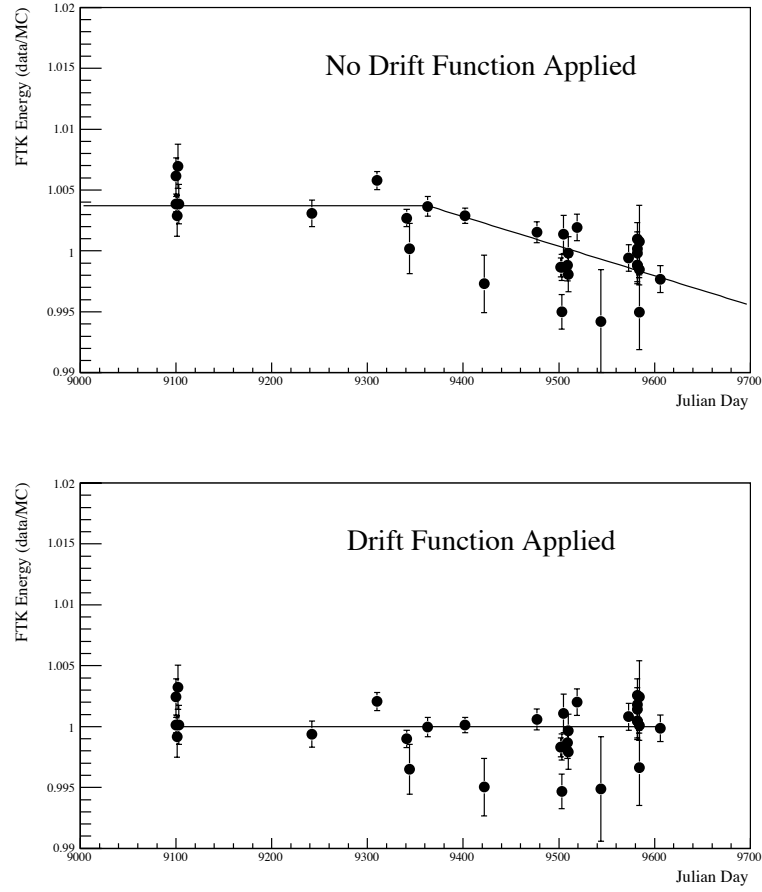
Figure 5.2: Ratio of mean FTK energy for ^{16}N source deployments at the center of the detector as a function of time. Top plot shows data/MC before the drift correction, and bottom plot shows after the correction. Figure from [51].

of PMT response derived from laserball calibrations. The full Monte Carlo, however, uses the more accurate "3D PMT model" which tracks photons individually through the entire concentrator and PMT optical unit. These two models should be the same on average, however after the processing of the D_2O phase, a difference between the 3D PMT model and the greydisk model was discovered.

To fix this problem and flatten out the energy scale as a function of radius, a radial correction factor was proposed for the D_2O phase[51]:

$$C(R^3) = 1.01159 - 0.0389943(R^3) + 0.0250065(R^3)^2, \qquad (5.8)$$

where R^3 is the normalized cubic radius $(r/600\text{cm})^3$, and $C(R^3)$ gives the ratio of FTK estimated T_{eff} over the true T_{eff} in the Monte Carlo. Originally it was thought that this correction should globally apply to both data and Monte Carlo. This assumed that the 3D PMT model used by the simulation best matched the real angular response, and the greydisk model was not correct. Studies of the energy response as a function of position in the detector (described in Section 5.4) later indicated this was not the case. As a result, this particular correction *applies only to Monte Carlo* in the D_2O phase.

5.4   Spatial Energy Correction

Prior to reprocessing the D_2O and salt data sets for the low energy analysis, a number of enhancements were made to the Monte Carlo to better reflect spatial variations in the detector. These included channel-dependent efficiencies for the PMTs, a retuned 3D PMT model, and an improved PMT timing

calibration calibration which helped to eliminate fluctuations in the origin of the coordinate system of reconstructed events over time.

While generally these changes improved agreement between data and Monte Carlo, they also highlighted discrepancies in energy response as a function of z. This is is most clearly seen in a radial plot color-coded by hemisphere[2] shown in Figure 5.3. The figure shows the difference (which we will also call a "bias") between data and Monte Carlo mean energy of each ^{16}N run as a function of source radius. When separated into top, middle and bottom runs, there is a clear divergence in the radial dependence of this bias between the top and the bottom of the detector.

Faced with an energy bias that depends upon z, we have two questions to answer:

1. Can we, and should we, do anything to correct the bias?

2. How do we incorporate this bias into our systematic uncertainties?

Previous analyses, such as SNO's salt phase results[32] and Dunford's low energy analysis[51], do not have a z-bias correction, and instead incorporate all of the discrepancy into energy scale systematic uncertainties. In particular, the z-bias had an impact on both the "radial distribution" (since R depends on z) and "detector asymmetry" uncertainties in those works. For the first low-energy analysis of the data, M. Dunford instead divided the detector into regions and computed a radial uncertainty for each region, but had no explicit detector asymmetry uncertainty.

For this analysis, I have revisited this approach, and instead use some of

---

[2]Keeping the fine tradition of naming problems after the graduate student who discovers them, this plot is called the "Dunford Plot."
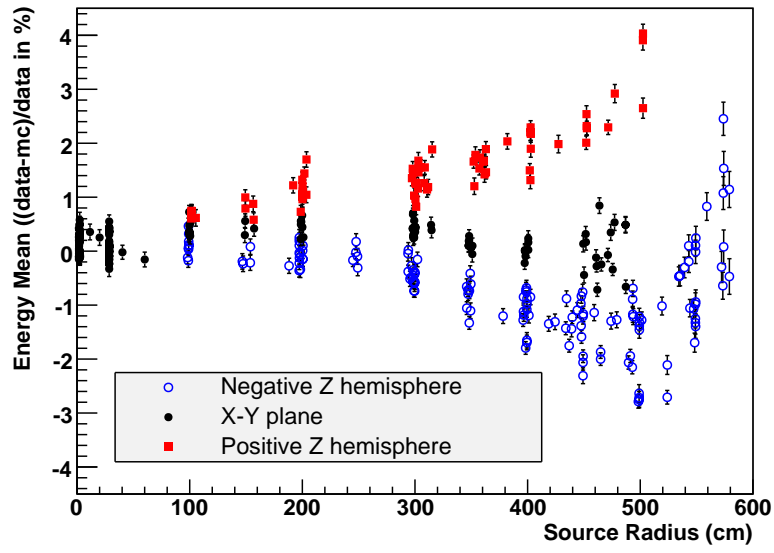
Figure 5.3: Difference between data and Monte Carlo N16 mean energy as a function of radius. ^{16}N runs in the top hemisphere, equator, and bottom hemisphere have been separated, showing a clear discrepancy in z between data and Monte Carlo.

the calibration data to generate an energy correction as a function of position. I have generated an energy correction to the data to mitigate most of the z-bias, and then computed a single "position dependence" uncertainty, which will subsume both the radial distribution and detector asymmetry uncertainties. The single position uncertainty will describe the uncertainty in the energy of events due to spatial non-uniformity.

In order to safely use calibration data for both a correction and an assessment of uncertainty, we need to split the ^{16}N into two groups within each phase. Although the detector clearly has a z-asymmetry due to design and gravity, it has high azimuthal symmetry. Calibration sources can only be deployed on the z-axis, the y-z plane or the x-z plane, so we will split the data along those groupings into set A and B. Set A consists of multi-axis runs on the $y - z$ plane and roughly half of the single axis run along the z-axis. Set B contains the $x - z$ runs, and the remaining single axis runs. Single-axis runs were split between the two sets using a random number generator. Specifically, the division criteria were:

$$|x_{\text{source}}| < 30 \text{ cm and } |y_{\text{source}}| > 30 \text{ cm} \quad \rightarrow \quad \text{set A}$$
$$|x_{\text{source}}| > 30 \text{ cm and } |y_{\text{source}}| < 30 \text{ cm} \quad \rightarrow \quad \text{set B}$$
$$|x_{\text{source}}| < 30 \text{ cm and } |y_{\text{source}}| < 30 \text{ cm} \quad \rightarrow \quad \text{set A if rand}(0,1) < 0.5, \text{ else set B}$$

The energy correction is constructed with set A, and applied to set B to determine the residual uncertainty. In this way, the correction is kept independent, and we also naturally include azimuthal variation in the energy scale in the final uncertainty. Discussion of the assessment of spatial uncertainty is found in Section 5.5.

5.4.1 Selecting Variables

Our best guess is that the z-bias is due to poor modeling of optical effects which have a z-dependence. In particular, the optics of the neck acrylic at the top, and obstructions at the bottom, such as debris and NCD anchors (see Appendix D), are thought to be the biggest contributors. Ideally, we would would use the available calibration data to tune the appropriate optical parameters in the Monte Carlo, like attenuation lengths through the neck and the bottom plates of the acrylic vessel. The difficulty in extracting these parameters from the available data, as well as the huge overhead in regenerating all of the Monte Carlo with new optics, makes this impractical. Instead, we correct this variation after the fact using a data-only correction factor which depends on some spatial observables.

Dunford introduced a coordinate to parameterize the spatial variation called $z_{intersect}$, the z coordinate of the intersection point of the acrylic vessel with a ray extending from the event position, $\mathbf{r}$, along the event direction, $\mathbf{u}$. A diagram of this is shown in Figure 5.4.

Expressed mathematically, the distance to the intersection point for an event inside the acrylic vessel is given by

$$d_{intersect} = \sqrt{(\mathbf{r} \cdot \mathbf{u}) - |\mathbf{r}^2| + r_{av}^2} - \vec{r} \cdot \vec{u}, \tag{5.9}$$

where $\mathbf{r} = (x, y, z)$ is the location of the event, $\mathbf{u} = (u_x, u_y, u_z)$ is the normalized direction of the event's Cherenkov cone, and $r_{av} = 600$cm is the radius of the acrylic vessel. The z coordinate of the intersection point is

$$z_{intersect} = z + u_z d_{intersect}. \tag{5.10}$$

111
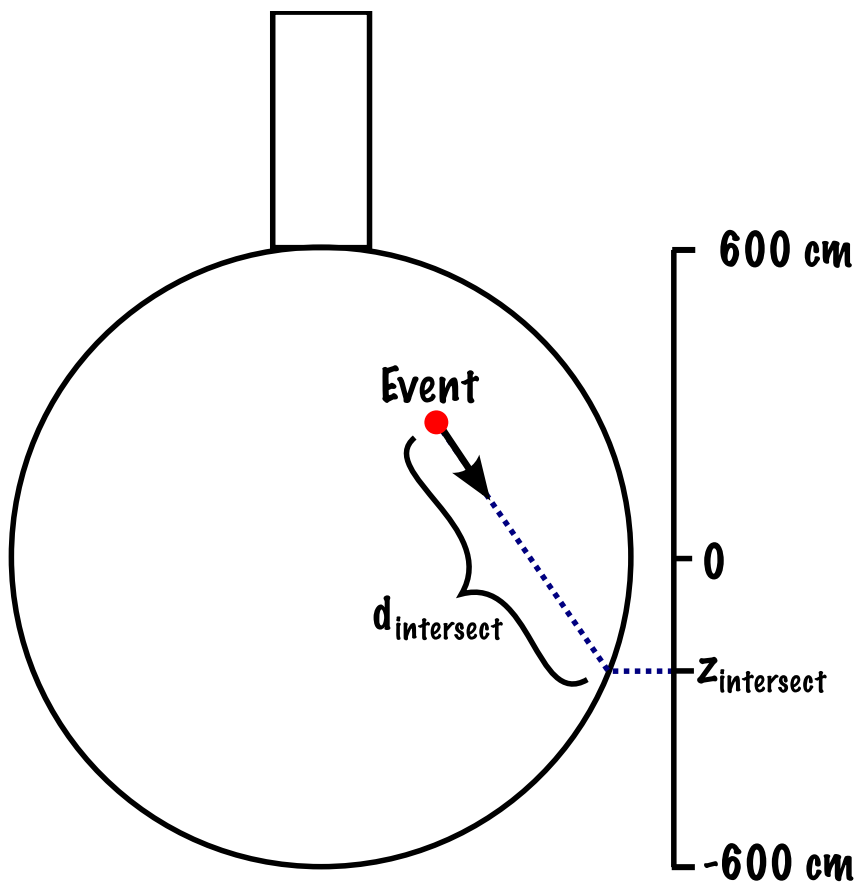
Figure 5.4: Diagram showing $d_{intersect}$ and $z_{intersect}$ for an event.

If we think the z-bias is due to the optical effects at the surface of the acrylic vessel, or approximately so (like the neck), then $(d_{intersect}, z_{intersect})$ is a very physically motivated coordinate system in which to work. We expect $z_{intersect}$ to trace optical obstructions either on the acrylic vessel, or at least projected onto the vessel, and $d_{intersect}$ will relate to how much solid angle these obstructions occupy in the event's field of view.

Breaking the vectors into components, we see that Equations 5.9 and 5.10 ultimately depend on only four independent scalars: z, u_z, ρ (the cylindrical radius coordinate) and $\mathbf{u} \cdot \hat{\mathbf{r}}$ (cosine of the angle between position and direction vectors). Testing showed that z and u_z (the z-component of the event direction) were equally-effective proxies for $z_{intersect}$ and $d_{intersect}$, so we will consider those two variables in our correction function.

A Note About z

PMT timing calibration runs using a triggered laserball were taken periodically throughout the running of SNO. Among other things, the relative timing of the PMTs is measured based upon a central run. These timing constants effectively set the origin of the coordinate system used in position reconstruction. Variations in the laserball position from run to run caused the coordinate system to move around slightly as a function of time. To fix this problem, the low energy reprocessing used a modified set of timing constants designed to make the coordinate system of reconstructed events more uniform. While eliminating nearly all of the time-related variation, reconstructed events ended up with a constant, systematic offset of 5 cm *only in real data.* Monte Carlo events have no offset. As a result, whenever z from detector-produced event

is plotted in this section, it is in fact

$$z' = z_{recon} - 5.0 \, \text{cm}. \tag{5.11}$$

5.4.2 Selecting Calibration Events

The ^{16}N source was deployed in a wide range of locations throughout data-taking, making it an excellent reference to map out the spatial dependence of the energy scale. Although runs were taken only at fixed source positions in the detector, gammas emitted by the source travel varying distances before Compton scattering an electron. We can therefore examine the detector response over a continuous range. ^{16}N events are selected from the data using several criteria:

1. Source tag, indicating a decay was observed inside the source.

2. Successful position reconstruction with $R < 600$ cm

3. Successful FTK energy fit with $T > 3.5$ MeV

4. Keep all events which reconstruct more than 70 cm from the source.

5. Keep events which reconstruct less than 70 cm from the source only if the event direction is within 45° of pointing away from the source, as shown in Figure 5.5. This prevents the direct Cherenkov cone from intersecting the calibration source manipulator and stainless steel ^{16}N decay can.

6. All standard high level cuts are applied.

There is a bit of circularity in applying the high level cuts, some of which depend on energy, to the data set from which we will extract an energy correc-
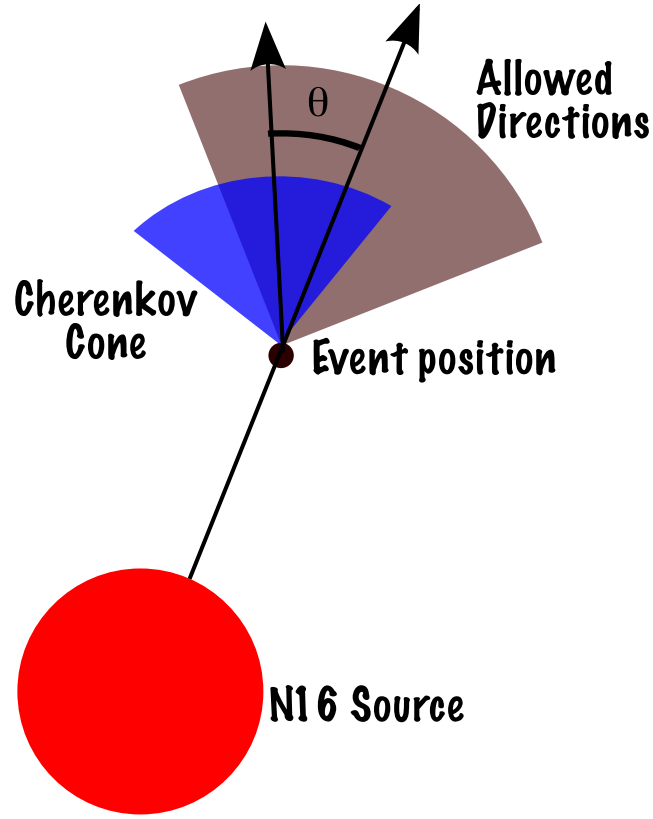
Figure 5.5: Diagram of event direction and Cherenkov cone relative to the source position. Events with $\theta < 45°$ are kept.

tion. However, this effect is minor since we are only adjusting the energy by a few percent at most. The high level cuts are necessary to remove poor energy fits at high radius which would otherwise skew the energy scale comparison with events that will ultimately be removed from the data set anyway.

Figures 5.6 and 5.7 show the distributions of the remaining calibration events in the D_2O and salt phases, respectively. The source cut is especially evident in the (u_z, z) plots, where it causes a slight tilt to the central band. Events which reconstruct below the source are more likely to be cut in the

upward direction, and vice versa

Creating the Map

FTK energies cannot be compared between data and Monte Carlo on an event-by-event basis, so instead we compare samples of events by binning them based on their observables. To be more robust against variations in the distribution tails, we perform an iterative Gaussian fit to the energy distribution in each bin. The fit is restricted to 1σ around the peak, as the full N16 distribution is known to be non-Gaussian and would generally require a more detailed treatment (see Section 5.5). Multiple iterations are used, with the $T_{eff,peak}$ and σ from the previous fit used to obtain a new fit range for the next iteration. Figures 5.8 and 5.9 show a sample of these Gaussian fits for ^{16}N data binned in (z, ρ) and (u_z, z), respectively. The peak value of the Gaussian in each bin, and its fit uncertainty, can then be placed into a standard 2D histogram to map out the observed energy of the source throughout the detector. In order to remove bad fits, we exclude cells with fewer than 500 events or have a χ^2 fit probability less than 0.01.[3]

After creating a histogram for both data and Monte Carlo, we can divide the two histograms, obtaining the ratio of the energy scales in binned form. Figure 5.10 shows a sample map created using four pairs of variables: (ρ, z), (u_z, z), $(\vec{u} \cdot \hat{r}, z)$, and $(d_{intersect}, z_{intersect})$.

In principle, we could use one of these binned map directly as a correction, but we would prefer an analytic, continuous function instead. Lacking a model for the true dependence of the bias on the observables, a reasonable

---

[3]The non-Gaussian shape of the N16 requires a narrow 1σ fit range around the peak and a loose probability cut to eliminate only the worst fits rather than all of them.
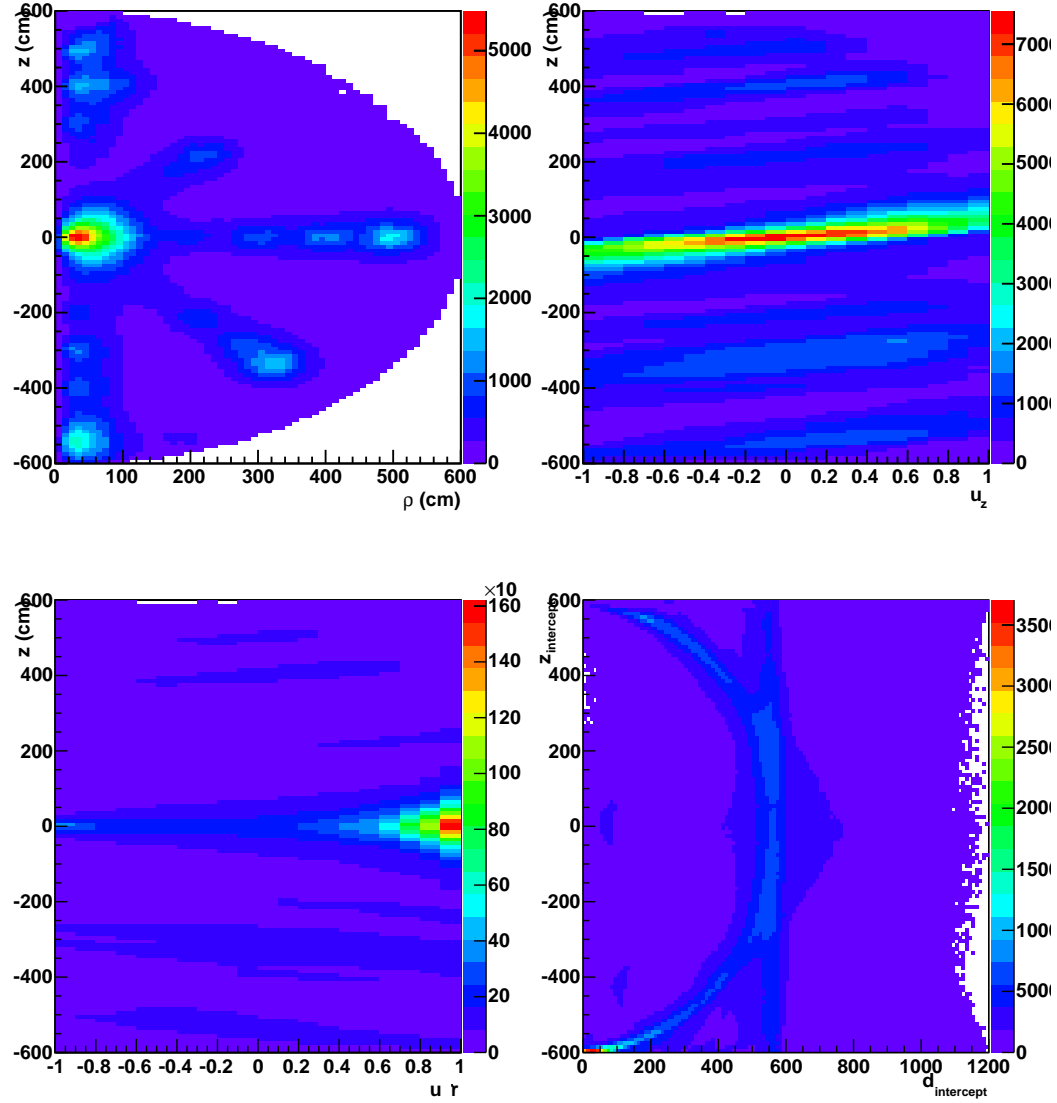
Figure 5.6: 2D projections of reconstructed D_2O phase N16 events after cutting events shadowed by the source. Upper left shows z vs. ρ (cylindrical radius coordinate), upper right shows z vs. u_z, lower left shows z vs. $\mathbf{u} \cdot \hat{\mathbf{r}}$, and lower right shows $z_{intersect}$ vs. $d_{intersect}$.
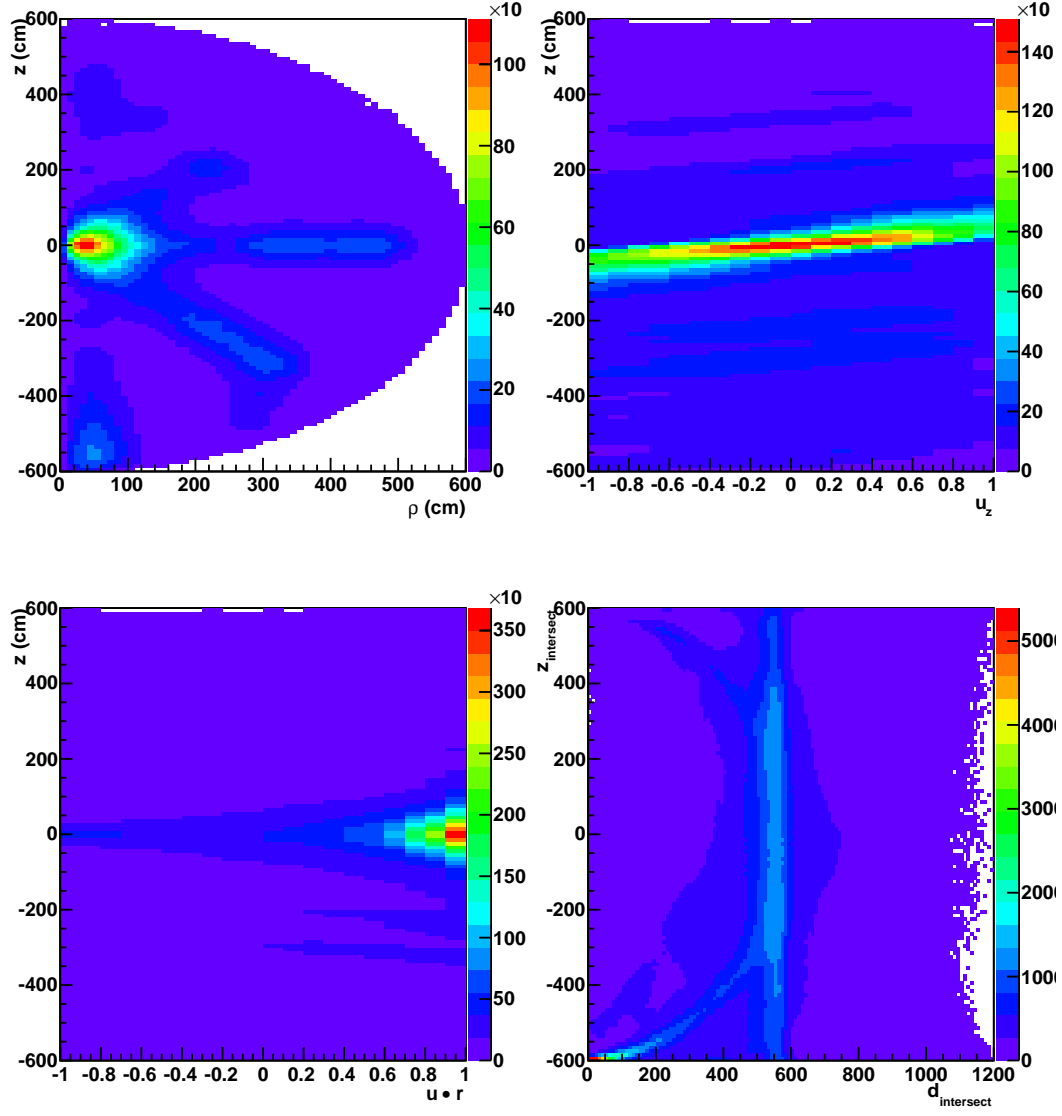
Figure 5.7: Reconstructed positions of salt phase N16 events after cutting events shadowed by the source. Upper left shows z vs. ρ (cylindrical radius coordinate), upper right shows z vs. u_z, lower left shows z vs. $\mathbf{u} \cdot \hat{\mathbf{r}}$, and lower right shows $z_{intersect}$ vs. $d_{intersect}$.

118
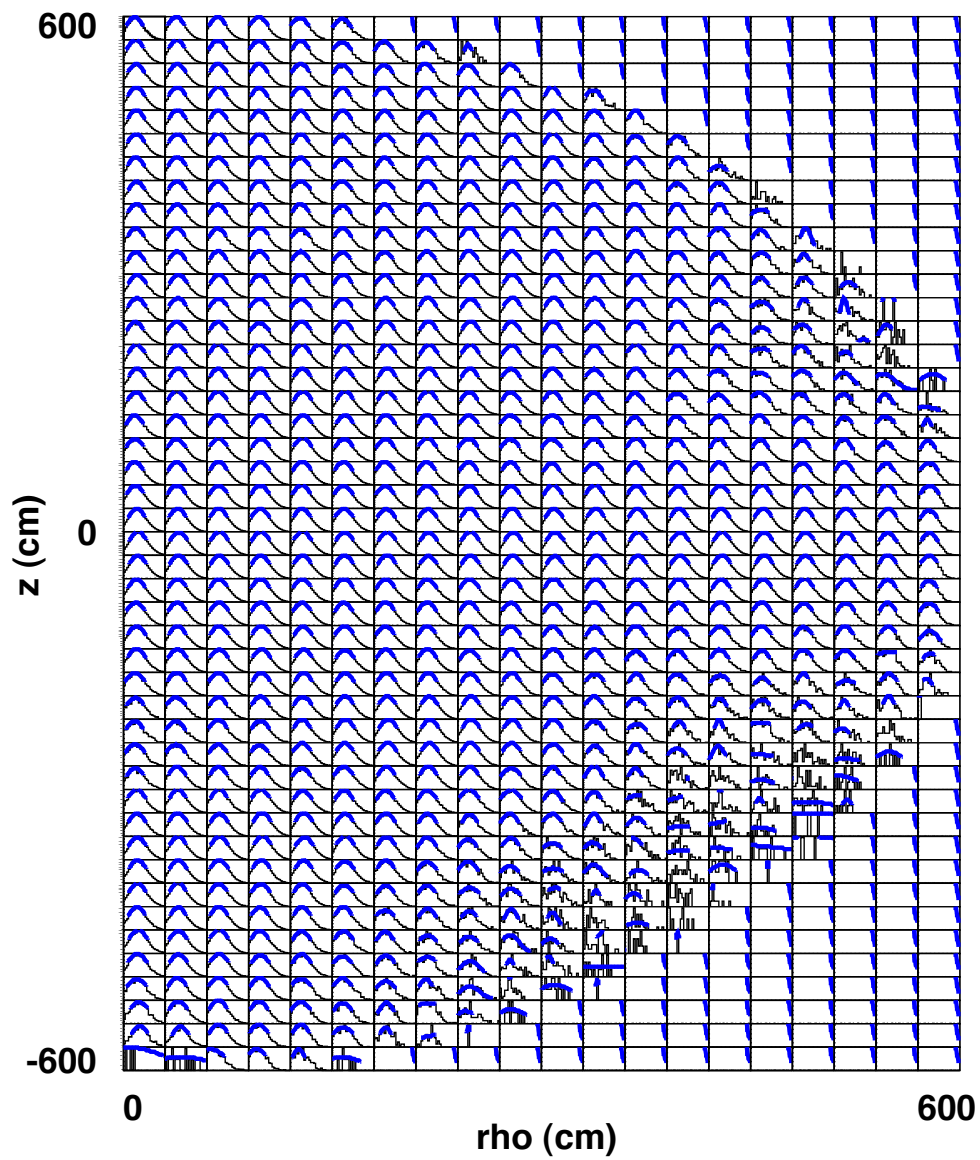
Figure 5.8: Gaussian fits to N16 data for each bin in z (vertical axis) and ρ (horizontal axis).

119
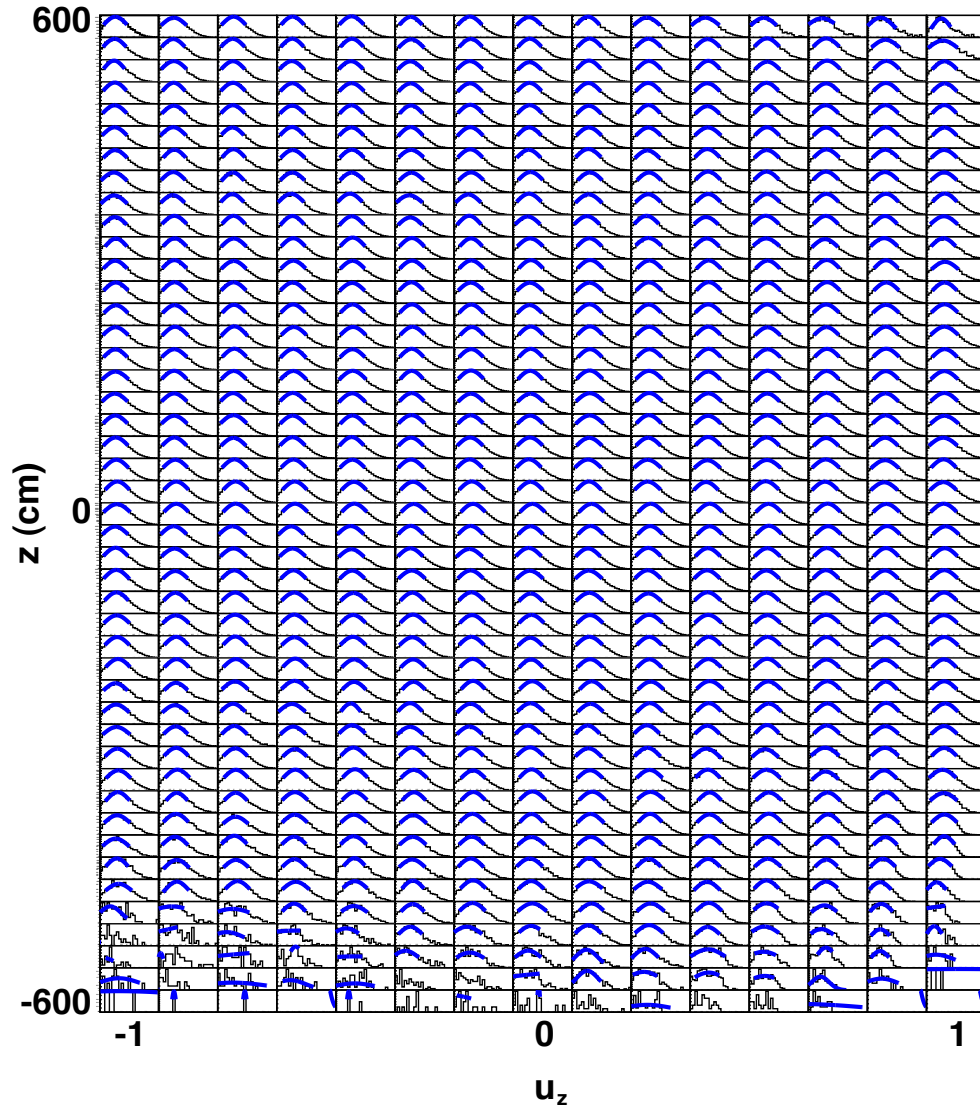
Figure 5.9: Gaussian fits to N16 data for each bin in u_z (horizontal axis) and z (vertical axis).

120
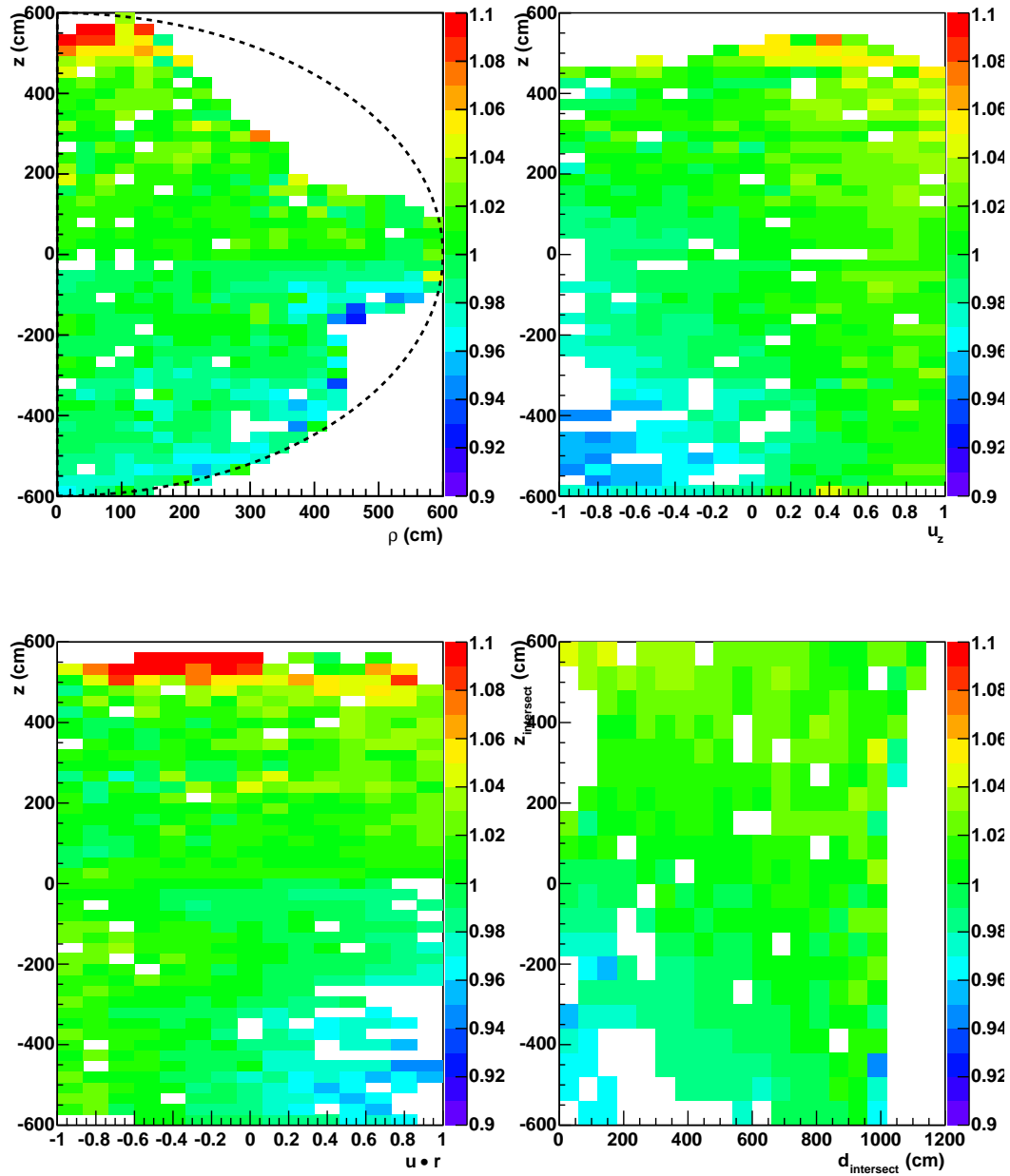
Figure 5.10: Ratio of data/MC peak N16 energy as a function of various combinations of z, ρ, $\vec{u} \cdot \hat{r}$, u_z, $d_{intersect}$, and $z_{intersect}$. White boxes indicate cells in which there was not a good Gaussian fit due to low statistics or unusual distribution shape. The black line in the upper left panel shows the inner edge of the acrylic vessel. Large gaps show areas where the source manipulator cannot reach, and so there are not enough events.

approximation would be to consider the first few terms in a Taylor expansion. To account for correlations between the two variables, at minimum, we must use a 2nd order function. For the case of (u_z, z), this looks like:

$$C(u_z, z) = A + Bu_z + Cz + Du_z z + Eu_z^2 + Fz^2. \qquad (5.12)$$

We fit this function to the 2D binned ratio map. The reconstructed energy for data can then be divided by $C(u_z, z)$ to give an energy which will be closer to what the Monte Carlo would have predicted. The uncertainties in these parameters are not needed, because we will assess the uncertainty in the correction by referencing an independent set of calibration data in Section 5.5.

As a final step in the correction, we globally scale energies for data to match the data and Monte Carlo energies of the central runs. This often makes the data-Monte Carlo agreement slightly worse in the volume-weighted comparison, but the center of the detector is our most optically well-understood region. If higher radius source runs want to pull the center out of agreement, then we would rather take that as an additional contribution to our spatial uncertainty.

5.4.3   D$_2$O Correction

Applying the method described above to the D$_2$O phase is complicated by drift in the prompt/late light ratio observed as a change in the FTK energy scale with time. The drift in Figure 5.2 clearly shows two parts, a flat region before SNO Julian Date 9363, and a decreasing response after this date. The degradation of the reflectors changes the spatial variation in the energy scale,

122

so we extract two different corrections, one for each time interval:

$$
C(u_z, z) = \begin{cases}
\begin{aligned}
& 1.00986 + 1.59412 \cdot 10^{-2}\, u_z + 2.25355 \cdot 10^{-5}\, z \\
& \quad -1.622782 \cdot 10^{-5}\, u_z z \\
& \quad +1.99929 \cdot 10^{-3}\, u_z^2 - 3.03906 \cdot 10^{-8}\, z^2 \qquad \text{if JD} < 9363 \\
\\
& 1.01028 + 2.1852 \cdot 10^{-2}\, u_z + 2.49459 \cdot 10^{-5}\, z \\
& \quad -2.46175 \cdot 10^{-5}\, u_z z \\
& \quad +1.24998 \cdot 10^{-3}\, u_z^2 - 6.24735 \cdot 10^{-8}\, z^2 \qquad \text{otherwise}
\end{aligned}
\end{cases}
\tag{5.13}
$$

The global energy scaling factor to fix the center after this correction is 0.9968. Section 8.5.1 shows how to apply all of these corrections together.

Figure 5.11 shows a comparison of the Dunford plot before and after the energy correction is applied. There is still some bias at the highest radii, but the top, middle and bottom regions of the detector are now in good agreement. It is important to note that the corrected energy plot includes the radial correction to the Monte Carlo energy scale that we do not apply to data. Figure 5.12 shows the same plot, but with the radius correction applied to both data and Monte Carlo, as was done previously, and we see that the radial bias is much worse.

5.4.4 Salt Correction

Constructing the correction for the salt phase is much more straightforward. The more accurate optical modeling of drifting detector response permits a
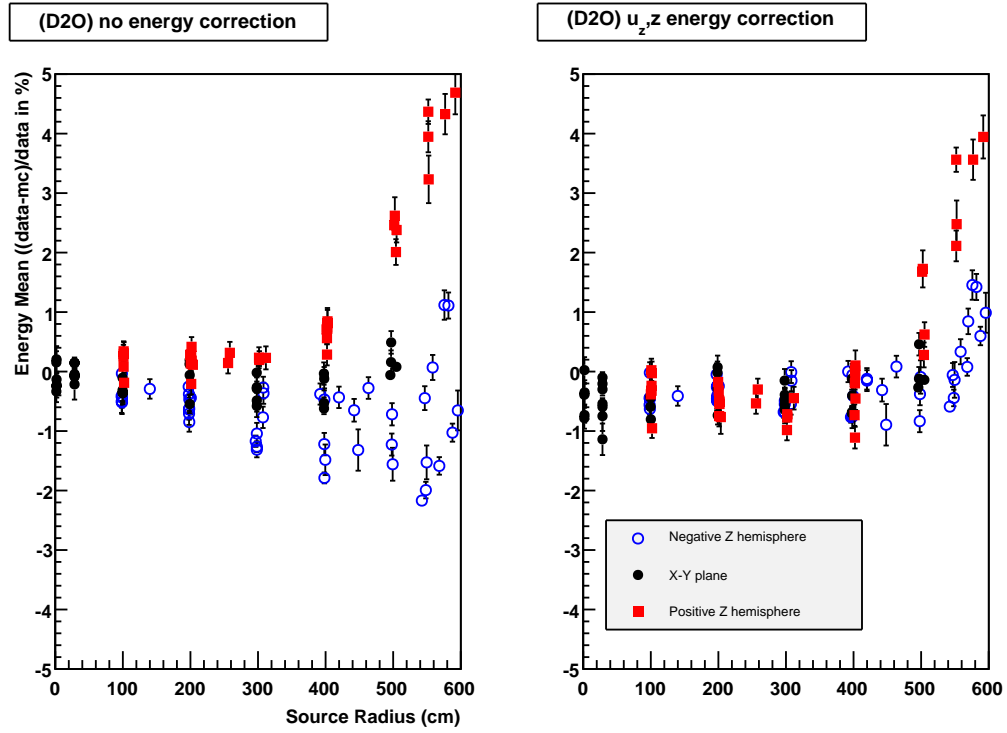
Figure 5.11: Difference between data and Monte Carlo ^{16}N source energy in the D_2O phase as a function of radius before (left) and after (right) the u_z, z energy correction.
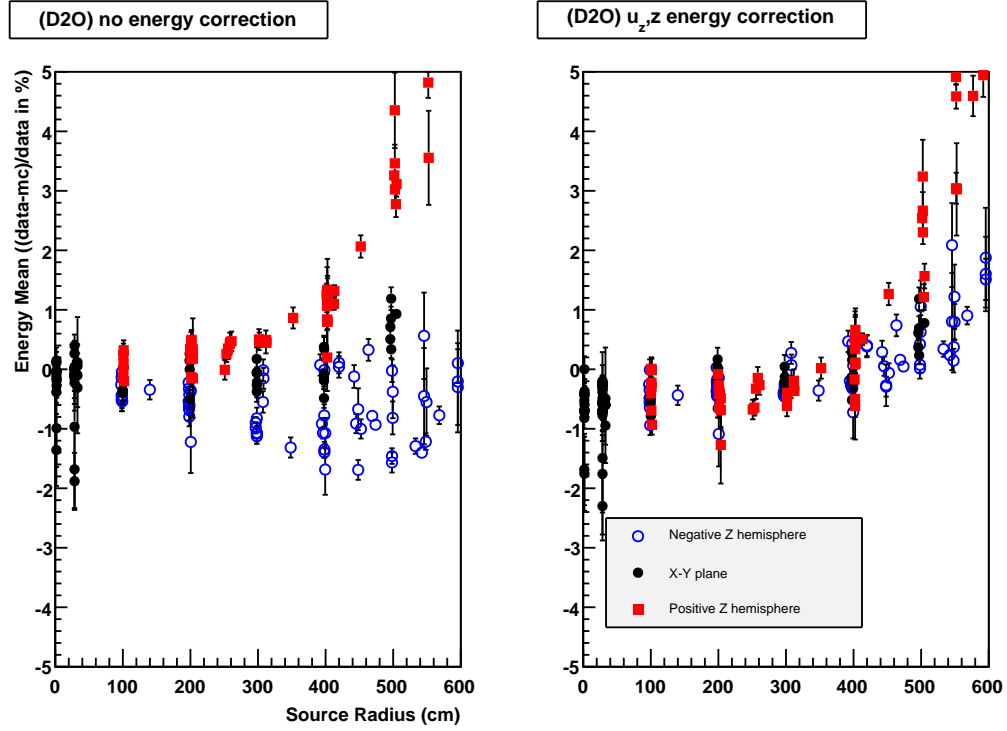
Figure 5.12: Difference between data and Monte Carlo ^{16}N source energy in the D$_2$O phase as a function of radius before (left) and after (right) the z, u_z energy correction. The radial correction in Equation 5.8 has been applied to both data and Monte Carlo in the right plot, leading to an overall radial bias.
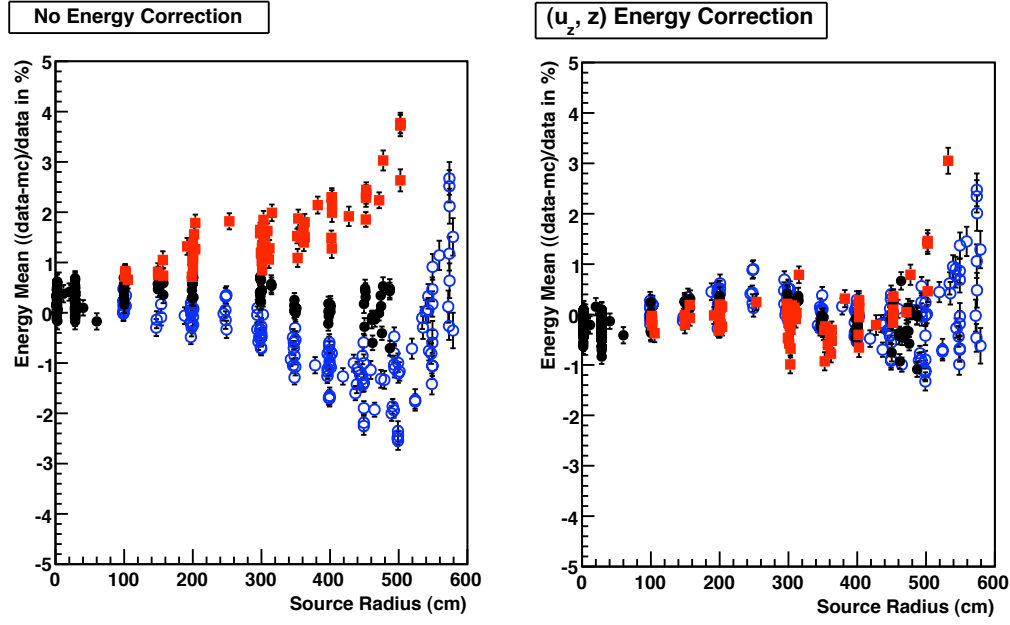
Figure 5.13: Difference between data and Monte Carlo ^{16}N source energy in the salt phase as a function of radius before (left) and after (right) the u_z, z energy correction.

single spatial correction function to apply to the entire phase:

$$
\begin{aligned}
C(u_z, z') \quad = \quad & 1.00320 + 2.19773 \cdot 10^{-2} \, u_z + 4.01820 \cdot 10^{-5} \, z' \\
& - 2.35837 \cdot 10^{-5} \, u_z z' + 1.06100 \cdot 10^{-3} \, u_z^2,
\end{aligned} \tag{5.14}
$$

The global correction factor to fix the center is 0.9979. Section 8.5.1 shows how to apply all of the salt energy corrections together. Figure 5.13 shows the Dunford plot before and after the correction.

With the more extensive source sampling in the salt phase, we can also visualize the effect of the correction in 2-dimensions. Figure 5.14 shows the uncorrected and corrected data for both set A, used to make the correction

function, and set B, which was not used to create the correction. The agreement is much improved across the entire detector, except at the neck where there is still some divergence.

During the salt phase, a set of calibration runs were taken where radon-spiked D_2O was injected into the detector and circulated to create a distributed source of ^{214}Bi decays. A spike source samples the entire detector (though not uniformly due to the circulation pattern) without any source manipulator to absorb light and potentially skew the energy measurement.

The reconstructed energy of this data can be fit using Monte Carlo of the ^{214}Bi decays, allowing a multiplicative energy scale to float. Figure 5.15 shows the results of this fit[4] with and without the salt energy correction. The energy scale systematic without the correction is $(1.8 \pm 0.6)\%$, while with the correction, the scale is $(0.0 \pm 0.6)\%$.

5.5   Spatial Energy Systematics

Having split the ^{16}N runs into two independent groups, we can use the second group to evaluate the systematic uncertainty remaining after the correction. Instead of using the iterative Gaussian approach to find the energy scale difference between data and Monte Carlo, we instead use the *most probable energy* method developed by Dunford[51] to more accurately deal with the non-Gaussian shape of the ^{16}N energy distribution.

---

[4]This fit is performed using the kernel density estimators and floating systematics techniques described in Chapter 7.
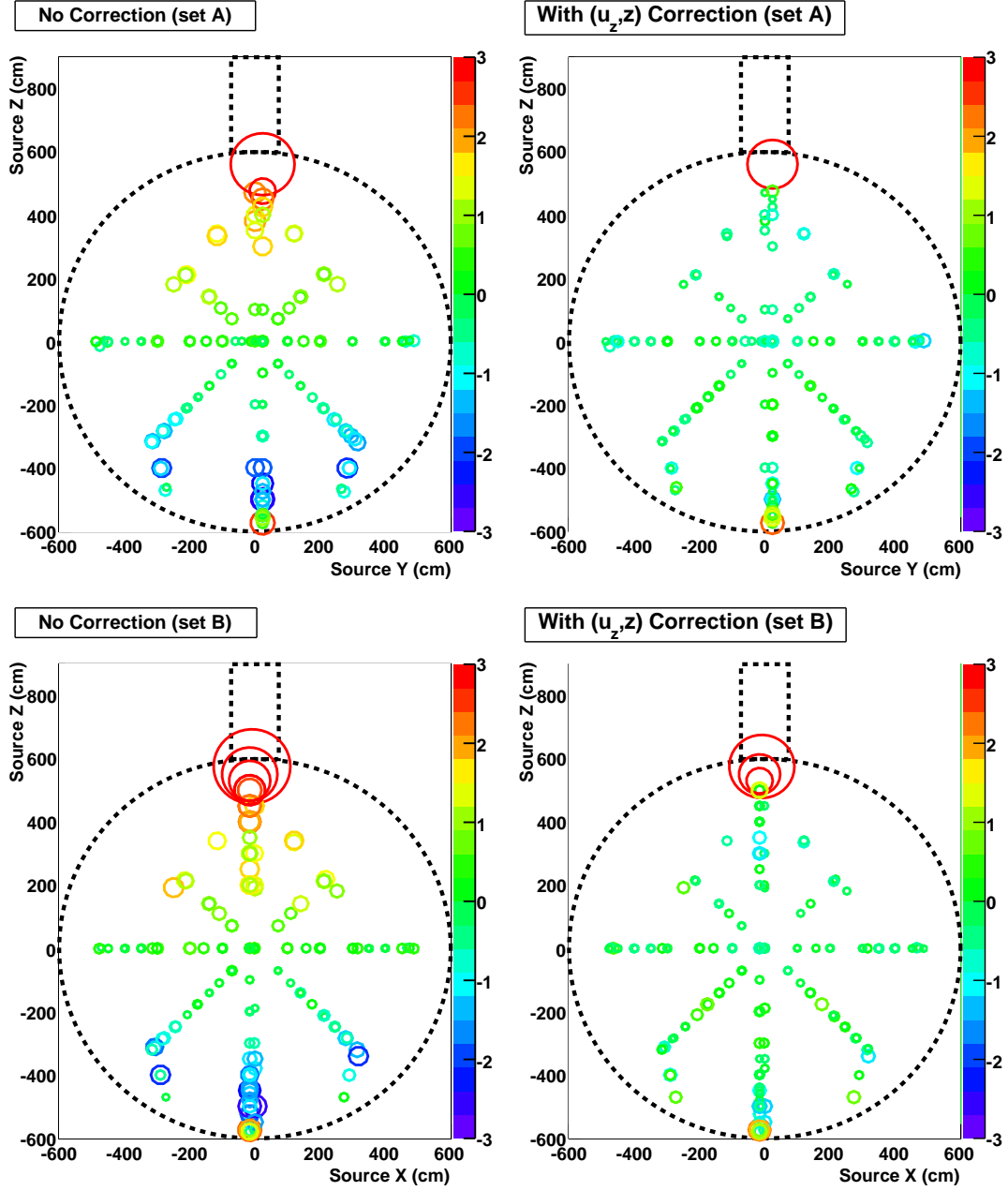
Figure 5.14: Difference in N16 peak energy between data and Monte Carlo in percent, mapped in 2D. Each circle is an N16 run, where the size of the circle indicates the absolute magnitude of the data−Monte Carlo bias. The color of the circle shows whether the bias is positive or negative, and also the magnitude (in percent) according to the color scale. Good agreement between data and Monte Carlo is therefore indicated by small green circles. The top plots show set A, which is on the y-z plane, and the bottom plots show set B, which is on the x-z plane.
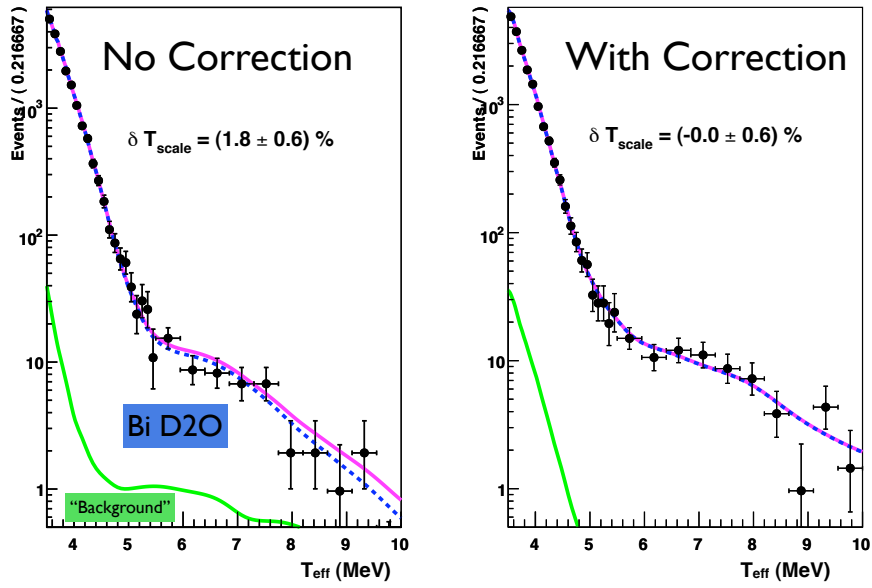
Figure 5.15: Fit of Monte Carlo energy distribution to 9000 events from a radon spike performed in the salt phase. The energy scale, number of ^{214}Bi decays, and number of "background" events were allowed to float in the fit without constraint. The background event PDF comes from unused normal runs prior to the injection of the radon.

5.5.1 Most Probable Electron Energy Method

Although ^{16}N produces a mono-energetic gamma ray, to an energy reconstruction algorithm, it is a variable energy source. A gamma ray can Compton scatter multiple electrons, and each electron must be above the Cherenkov energy threshold to produce light. The visible light from a single gamma ray therefore takes a distribution of possible values, with gammas that scatter multiple electrons appearing to have lower energy than those which deposit most of their energy into a single electron. SNOMAN can generate a prediction of this effective electron energy distribution by outputting the number of Cherenkov photons produced in each simulated ^{16}N event. The Cherenkov photon number distribution is then mapped to the *most-probable* FTK energy distribution using the same function that FTK uses in its likelihood calculation. This gives the distribution shown in Figure 5.16, which is the energy distribution of the ^{16}N source *before* the detector optical response adds further smearing.

The observed energy distribution is the convolution of the most probable energy distribution $P_{prob}(E)$ with a detector resolution function, which we can describe as Gaussian with an energy dependent width

$$\sigma = p_1 + p_2\sqrt{E}, \tag{5.15}$$

and an energy offset p_3. The convolution integral is:

$$P(T_{\text{eff}}) = N \int P_{prob}(T_{\text{eff}}) e^{\left(\frac{T_{eff} - E - p_e}{2\sigma^2}\right)} dE. \tag{5.16}$$

By fitting $P(T_{\text{eff}})$ to the observed energy distribution of ^{16}N in data and Monte
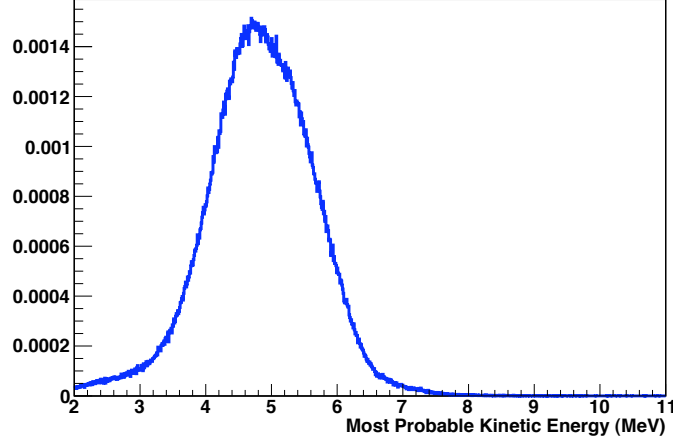
Figure 5.16: Most-probable FTK energy distribution for ^{16}N before detector resolution effects. Data provided by M. Dunford.

Carlo for each run, we can obtain p_1, p_2, and p_3, and compare them run by run. This gives the spatial variation of both the energy scale and the resolution. The fractional systematic uncertainty due to spatial variation is then the volume-weighted average of

$$\delta = \frac{p_{3,data} - p_{3,MC}}{p_{3,MC}} \tag{5.17}$$

over all runs in set B.

Based on the asterisk-pattern in which calibration runs are taken, we binned the the detector in 2D radial-polar bins shown in Figure 5.17 and Figure 5.19. Reduced ^{16}N sampling in the D_2O phase necessitates fewer bins. In bin i, the average δ_i and RMS spread σ_i is computed. We use the RMS as a conservative estimate of the uncertainty since the number of runs in a bin can be as small as 4, where the standard deviation of the mean would be a suspect measure of variation.

To account for differences in the volume of each bin along with the varying uncertainties for the energy scale in each bin, the average of all bins is found with a weighted χ^2 minimization:

$$\chi^2 = \sum_{i=0}^{n} w_i \frac{(\delta_i - \bar{\delta})^2}{\sigma_i^2}. \tag{5.18}$$

Comparing this to the standard unweighted χ^2 fit, where $w_i = 1$, we see that the constraint on w_i is

$$\sum_{i=0}^{n} w_i = n. \tag{5.19}$$

To weight the fit by volume, we use

$$w_i = \frac{n v_i}{V}, \tag{5.20}$$

where n is the number of bins, v_i is the volume of bin i and V is the total volume in all bins. This is equivalent to rescaling the bin errors so that

$$\sigma_i' = \frac{\sigma_i}{\sqrt{n v_i / V}} \tag{5.21}$$

where σ_i' is the new uncertainty on the ith bin, and σ_i is the original uncertainty.

5.5.2   D$_2$O Phase

Figure 5.18 shows δ_i and σ_i for the D$_2$O phase unrolled into a 1D profile histogram, where the width of each bin is proportional to the volume of that bin, and the bin color is the same as the region colors used in Figure 5.17.

No runs in
this bin
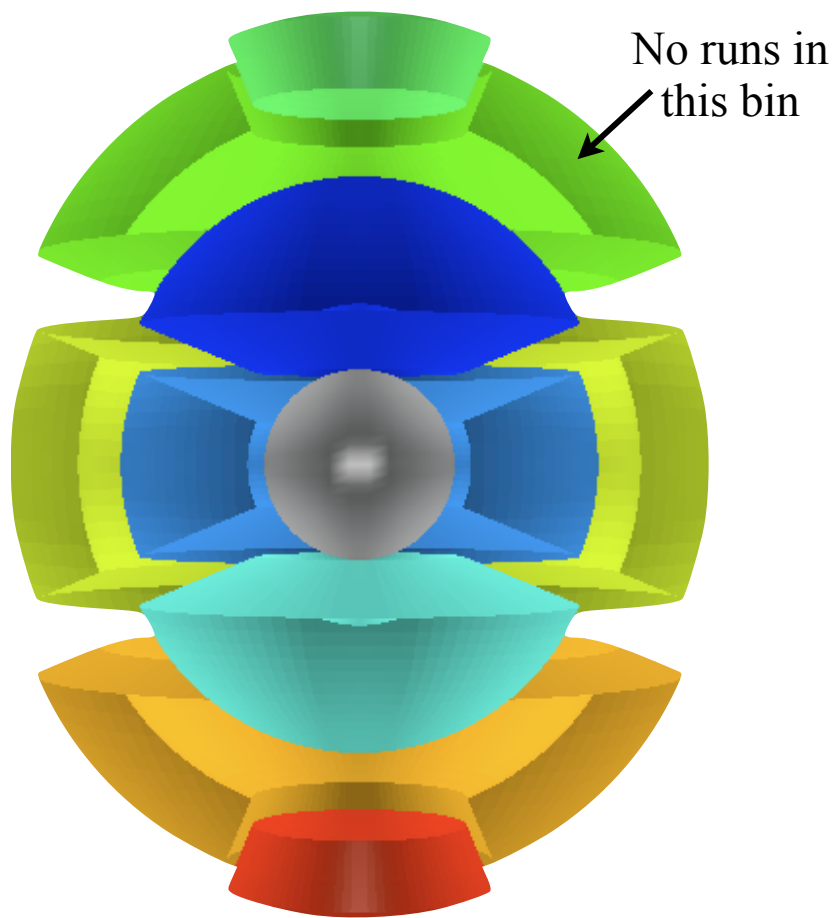
Figure 5.17: Exploded view of binning used for volume weighting in the D_2O phase. Due to limitations of the calibration manipulator system, no source run can be take in the outermost diagonal bin.
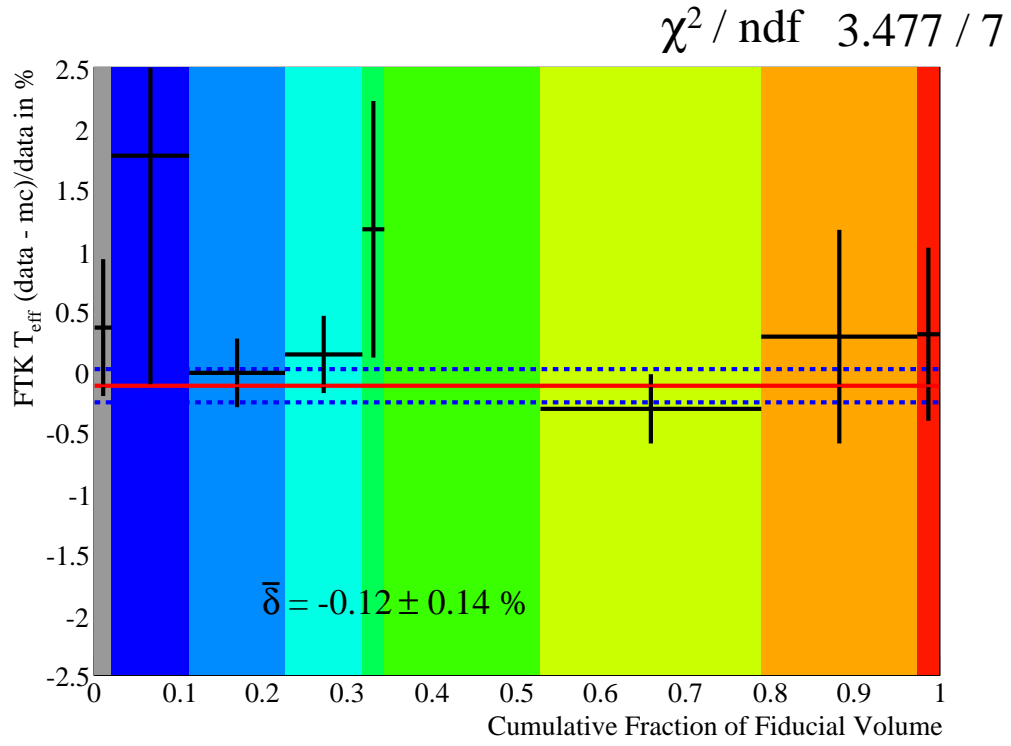
Figure 5.18: D$_2$O phase mean and RMS energy data-MC scale difference in each of the 3D bins shown in Figure 5.17. The volume-weighted chi-square best-fit line is shown. Note that error bars on points do not reflect the volume-weighting factor.

5.5.3 Salt Phase

Figure 5.20 shows δ_i and σ_i for the D_2O phase unrolled into a 1D profile histogram, where the width of each bin is proportional to the volume of that bin, and the bin color is the same as the region colors used in Figure 5.19.

5.6 Summary of Energy Scale Systematics

Spatial variation is of course not the only source of energy scale uncertainties in the detector. There is also uncertainty from:

- PMT status: The energy estimators must know how many PMTs are actively functioning in order to scale the detector collection efficiency appropriately.

- Electronics threshold/gain: The channel thresholds and PMT gains determine the probability that a photoelectron is actually detected at the front end DAQ. This impacts the expected N_{hit} for a given energy.

- Rate effects: ^{16}N source data is usually taken at a high event rate to minimize calibration time relative to normal neutrino data taking. The response of the detector is slightly different depending on the event rate.

- Time stability: We assume (after drift corrections) that the energy scale in the detector is constant over time. This systematic measures our confidence in that.

- ^{16}N source modeling: The precise energy response of the source depends upon how well we model the construction of the decay chamber and attached hardware, as well as the physics of the ^{16}N decay.
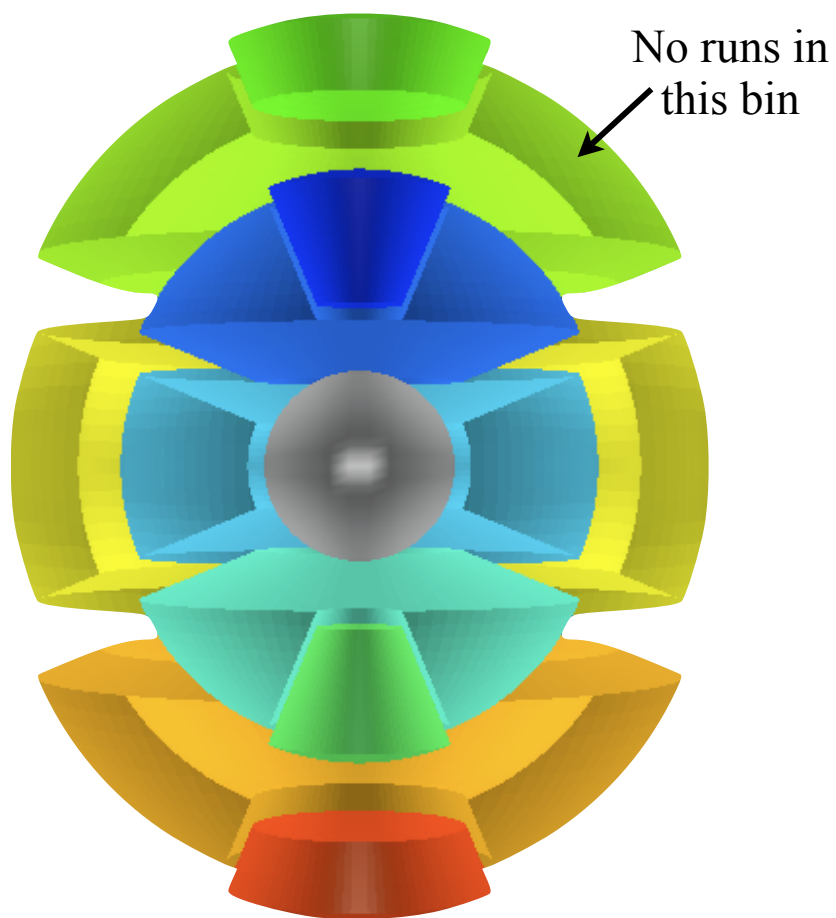
Figure 5.19: Exploded view of binning used for volume weighting in the salt phase. Due to limitations of the calibration manipulator system, no source run can be take in the outermost diagonal bin.
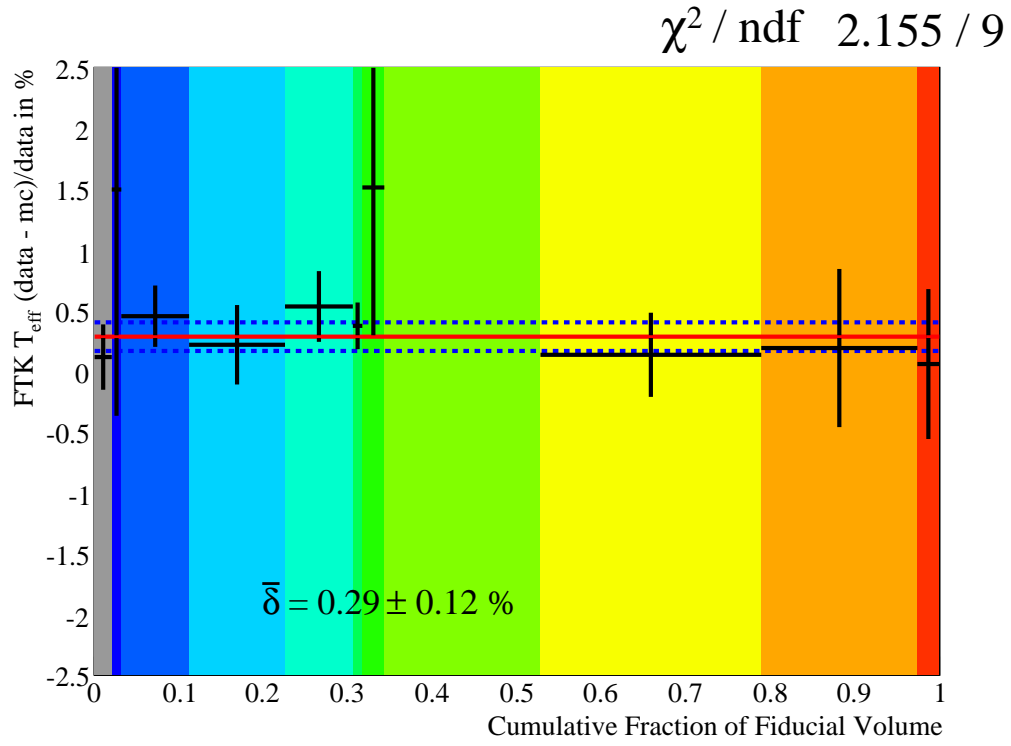
Figure 5.20: Salt phase mean and RMS energy data-MC scale difference in each of the 3D bins shown in Figure 5.19. The volume-weighted chi-square best-fit line is shown. Note that error bars on points do not reflect the volume-weighting factor.

Uncertainty	D$_2$O Phase (%)	Salt Phase (%)	Correlation	Reference
Detector PMT status	0.01	0.01	Uncorrelated	[51]
Electronics Threshold	0.20	0.20	Uncorrelated	[51]
Electronics Gain	0.40	0.40	Uncorrelated	[51]
Rate Effects	-0.69 ± 0.17	-0.29 ± 0.09	Correlated	[51]
^{16}N Source Modeling	0.4	0.4	Correlated	[78]
Spatial Variation	0.2	0.3	Uncorrelated	

Table 5.1: Summary of energy scale uncertainties. The correlation column indicates whether the source of the uncertainty is correlated or uncorrelated between phases.

Table 5.1 shows the values of these other uncertainties, along with the spatial uncertainty described in this chapter. We have taken the total spatial uncertainty to be sum in quadrature of the offset in scale, a measure of azimuthal asymmetry from x-z and y-z plane differences, and the the uncertainty in the offset, a measure of our statistical uncertainty in the energy scale due to finite sampling.

5.7 Energy Resolution

Using p_1 and p_2 from Equation 5.15, we can also compute the energy resolution as a function of energy. The resolution quantity we are interested in, though, is not the detector resolution, but the monoenergetic electron resolution. The electron resolution also includes variation in the number of Cherenkov photons produced by an electron of a particular energy, which is the analog of Figure 5.16 for a monoenergetic electron source. Figure 5.21 shows the most probable energy distribution for two electron energies.

For a given energy, we can find the electron resolution by convolving the
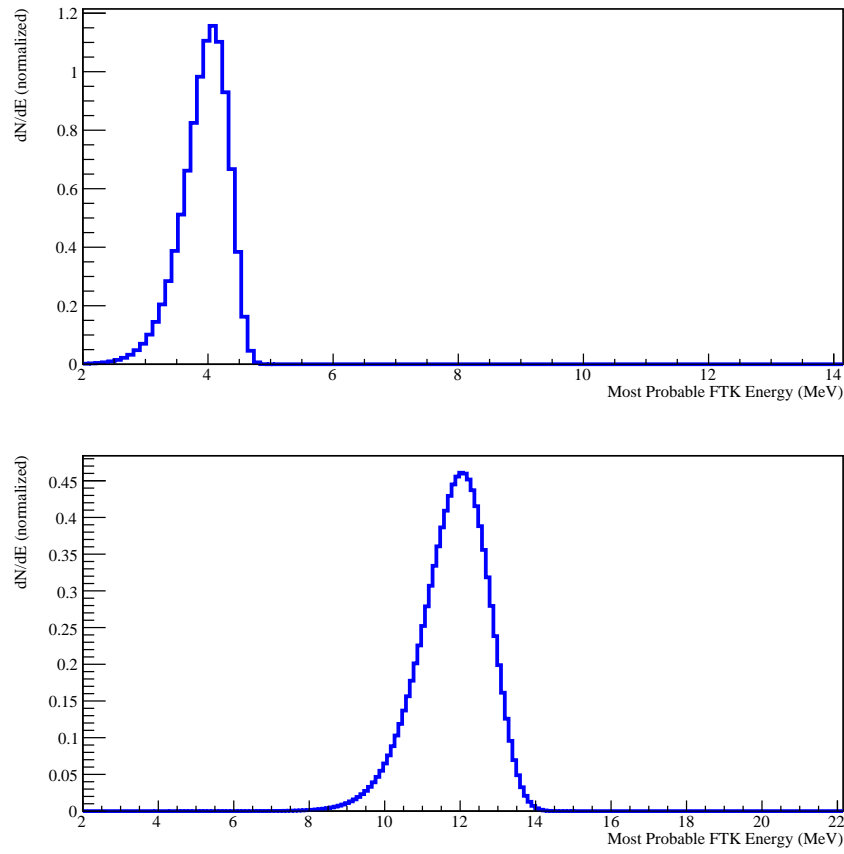
Figure 5.21: Most probable FTK energy distribution for monoenergetic electrons with kinetic energy 4 MeV (top) and 12 MeV (bottom).

139

most probable FTK energy distribution with the energy-dependent detector resolution Gaussian extracted from a particular ^{16}N calibration source run. Then we can combine and volume-weight these electron resolutions from all of the source runs in precisely the same way that we did for energy scale and get a detector-wide volume-weighted energy resolution. Figure 5.22 shows these distributions for the D$_2$O and salt phases.

Ultimately, what we care about is our uncertainty as to how closely the Monte Carlo matches the resolution of the data. Let

$$\sigma^2_{extra} = \sigma^2_{data} - \sigma^2_{MC} \tag{5.22}$$

be our measure of the resolution difference between data and Monte Carlo. Then we can construct a fully volume-weighted average of σ^2_{extra} over all of the runs. Figure 5.23 shows the volume-weighted σ^2_{extra} as a function of energy for the D$_2$O and salt phases. Section 6.4 compares the detector resolution systematic computed this way with the other low-energy calibration sources. There we find that multiple sources show a systematic positive σ^2_{extra}, indicating a slightly worse detector response in data than in Monte Carlo.

5.8 Summary

The FTK energy estimator is a critical component of the low-energy analysis. Previously, significant spatial variation in the FTK energy scale throughout the detector made the energy scale uncertainty both large and difficult work with. The new correction method described in this chapter solves the problem while still providing a way to independently assess energy scale uncertainty
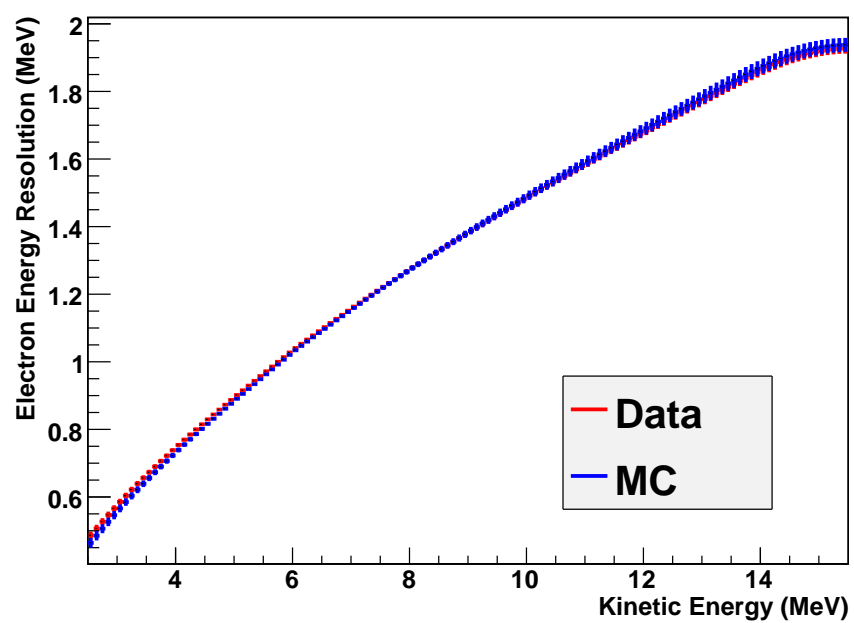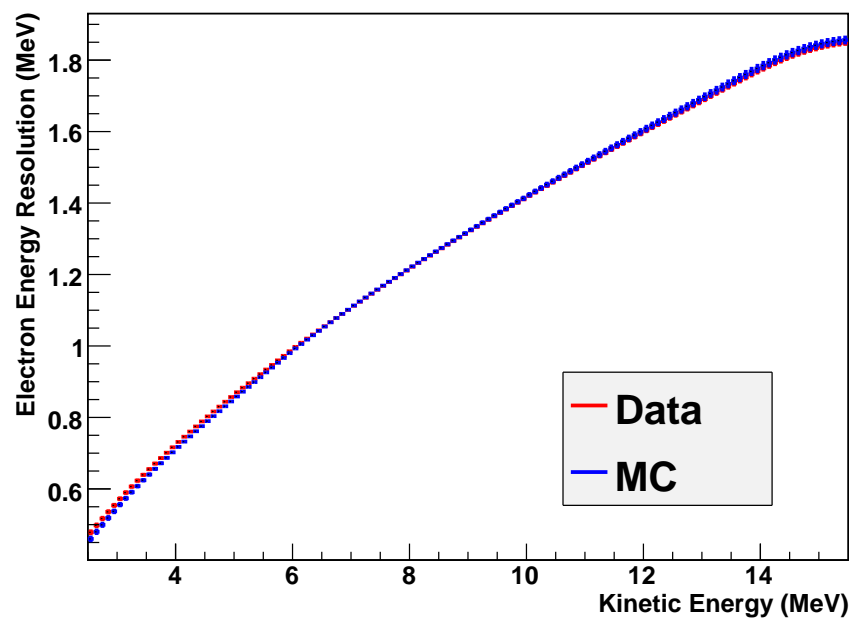
Figure 5.22: Monoenergetic electron resolution as a function of electron energy for D$_2$O phase (top) and salt phase (bottom). Note that values outside the 3.5–6.5 MeV range are extrapolations of the resolution function obtained from ^{16}N.
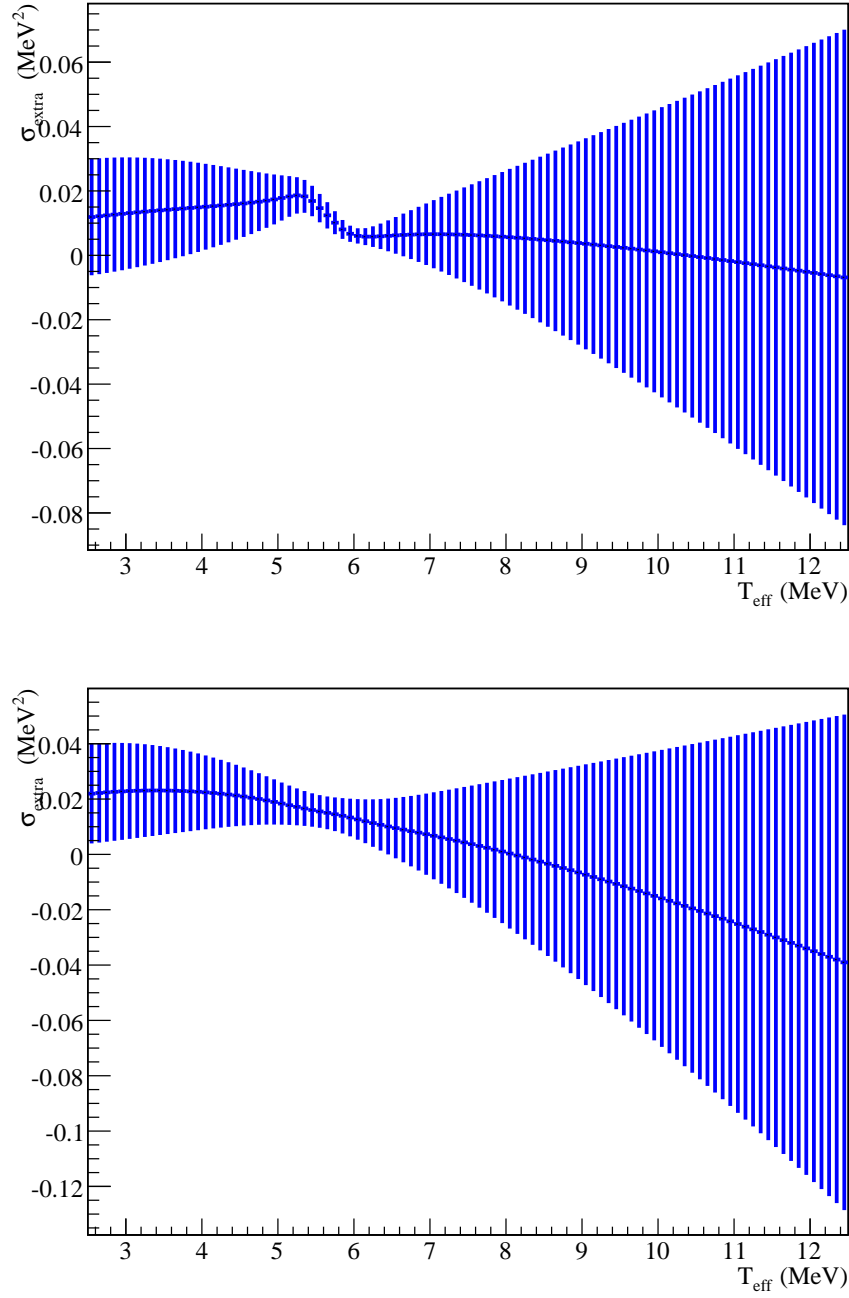
Figure 5.23: σ^2_{extra} as a function of energy for D_2O phase (top) and salt phase (bottom). Note that values outside the 3.5–6.5 MeV range are extrapolations of the resolution function obtained from ^{16}N.

using ^{16}N calibration data. The 2D radial-polar binning of runs to measure the uncertainty allows the uncertainties to be volume-weighted in a more accurate manner than previously.

Chapter 6

Background Verification

Given the variety of backgrounds to consider in the low energy signal extraction, we are forced to rely on the SNOMAN Monte Carlo to a greater extent than previously. Testing the simulation at lower energy is important to verify that our systematic uncertainties cover both the signals and the backgrounds. In this chapter, we look at two radon spikes performed during the salt phase as distributed sources of low energy events. We also discuss various options to model the PMT events, and the relative benefits of each.

6.1   Internal ^{214}Bi

During the salt phase, 81 ± 4 Bq of ^{222}Rn[32] was mixed with D_2O and deliberately injected into the D_2O volume as a low energy calibration source. Found in the ^{238}U chain, ^{222}Rn has a 3.8 day half-life. After mixing, the ^{222}Rn becomes a distributed source of ^{214}Bi decays in the detector. Figure 6.1 shows the distribution of events above 3.5 MeV and inside the fiducial radius of 550 cm.
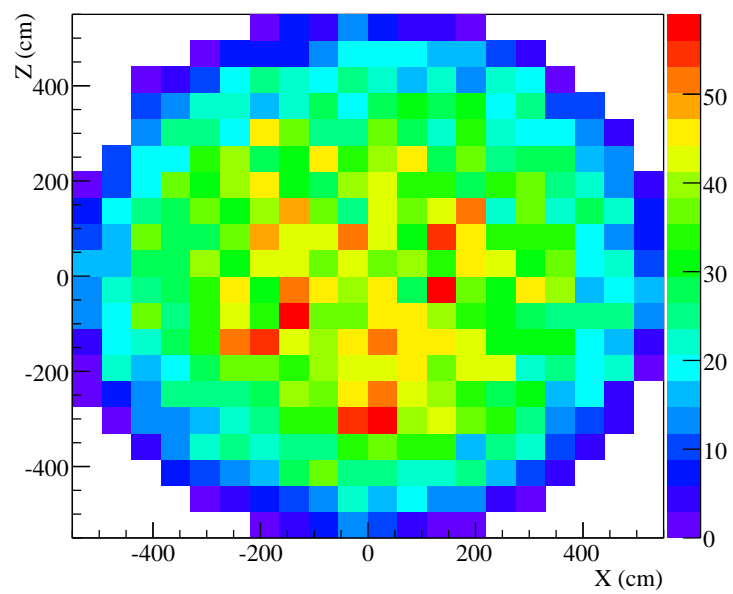
Figure 6.1: Distribution of events in the internal radon spike.

Energy resolution (σ_{extra})	0.08 ± 0.03 MeV
β_{14} Electron Scale	$-0.6 \pm 0.4\%$
β_{14} Neutron Scale	$-0.2 \pm 4.2\%$

Table 6.1: Systematic parameters from internal radon spike fit.

To test the Monte Carlo, we can perform a maximum likelihood fit using two signals: ^{214}Bi and "quiet data." Quiet data is a sample of normal neutrino and background events from standard run conditions which are also present in the spike data. The ^{214}Bi PDF comes from Monte Carlo of ^{214}Bi decays distributed uniformly in the detector. The spike data, after the energy correction described in the previous chapter, is then fit with this two PDF model, using the observables T_{eff} and β_{14}. In addition, the energy resolution of the ^{214}Bi PDF is floated in the form of a Gaussian convolution, where σ_{extra} is a free parameter in the fit. The mean β_{14} for the electron and photodisintegration neutron pieces of the PDF are also allowed to vary in the fit in the form of a multiplicative scaling. Figure 6.2 shows the 1D projections of the best fit. The best fit energy resolution and β_{14} scaling parameters are listed in Table 6.1.

6.2   External ^{214}Bi

A second spike was performed during the salt phase by injecting ^{222}Rn into the light water region between the acrylic vessel and the PMT support structure. Figure 6.3 shows the event distribution above 3.5 MeV and also inside the fiducial radius of 550 cm. It is important to note that nearly all of these events are by definition misreconstructed, since the radon is outside 605 cm.
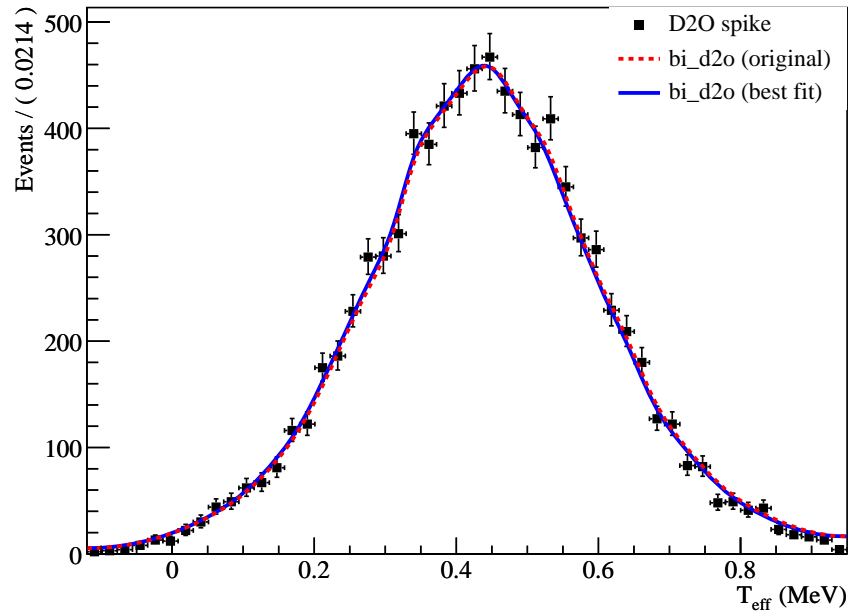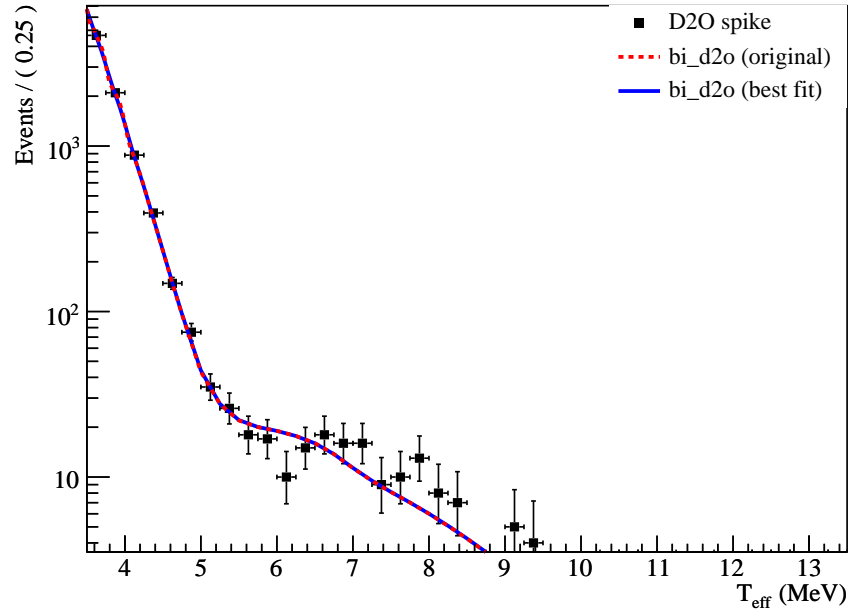
Figure 6.2: One-dimensional projections of Monte Carlo PDFs fit to internal radon spike data. Values of systematic parameters shown in Table 6.1.
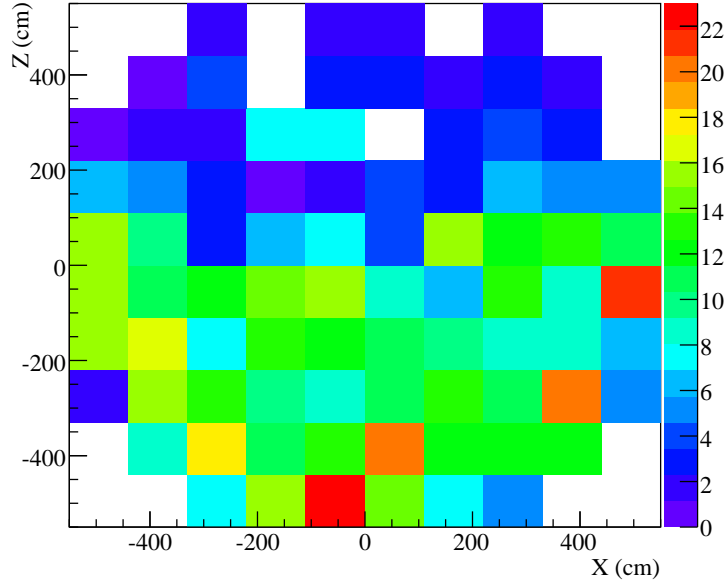
Figure 6.3: Distribution of events in the external radon spike.

Unlike the internal radon spike, the external spike was not as well mixed, as evidenced by the strong up/down asymmetry in the event distribution. However, the spatial energy correction helps to ensure the energy scale throughout the detector is uniform enough to compare these events to the Monte Carlo.

The fit for the external ^{214}Bi is identical in structure to the internal ^{214}Bi fit. The PDF is made from uniformly distributed ^{214}Bi decays in the H2O region, and the data is fit in two dimensions, T_{eff} and β_{14}. A floating energy resolution, β_{14} scale for electrons and neutrons are also included. Figure 6.4 show the results of this fit, and Table 6.2 shows the resulting parameters.

148
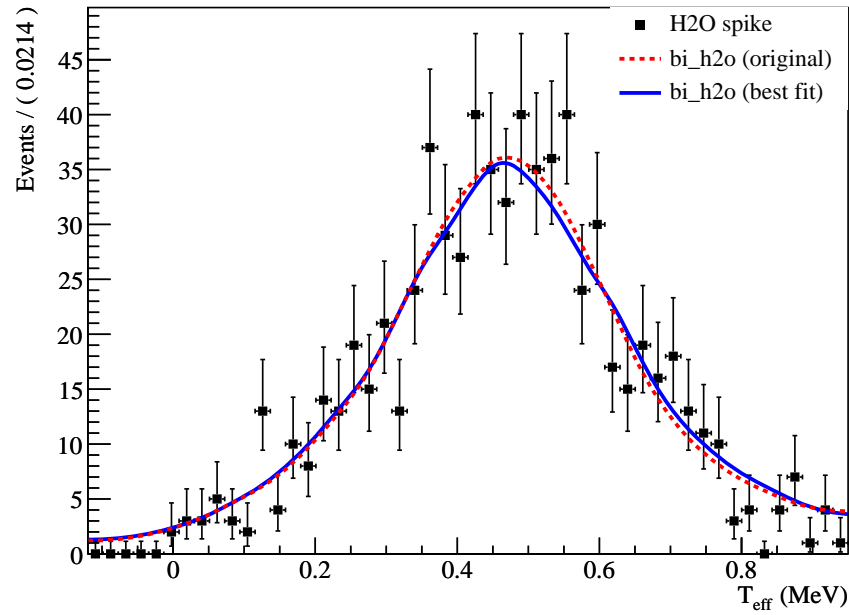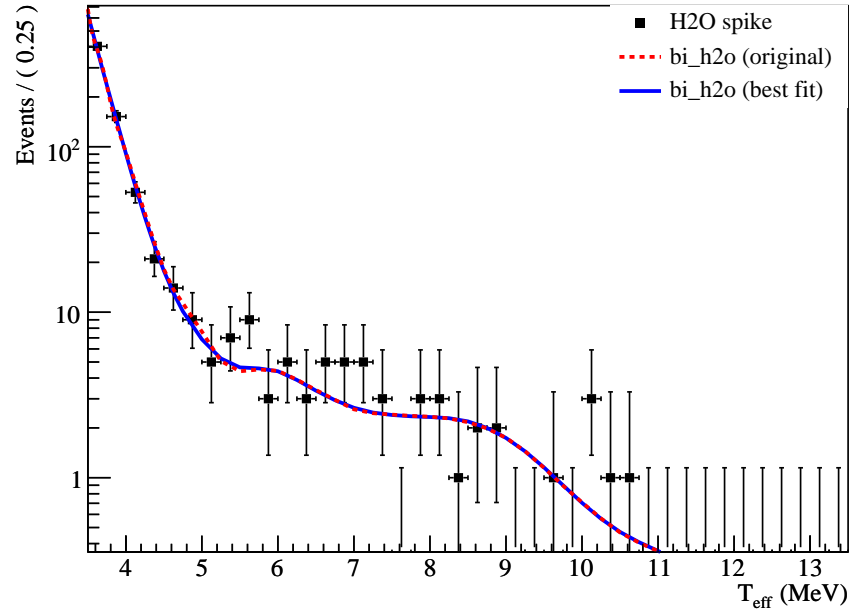
Figure 6.4: One-dimensional projections of Monte Carlo PDFs fit to external radon spike data. Values of systematic parameters shown in Table 6.2.

Energy resolution (σ_{extra})	0.26 ± 0.04 MeV
β_{14} Electron Scale	$1.8 \pm 2.0\%$
β_{14} Neutron Scale	$9.5 \pm 10.6\%$

Table 6.2: Systematic parameters from external radon spike fit.

6.3   PMTs

The PMTs are a challenging source of background events to model and measure. Radioactive decays come from the uranium and thorium in the PMT glass itself, as well as the resistor chain on the back of the tube and the steel support structure holding the PMTs. SNO was constructed in order to minimize the impact of PMT events on the analysis as much as possible. As mentioned in Section 3.3, the forward-viewing PMT concentrators provide a natural cut in the acceptance of high radius events. In addition, the distance between the PMTs and the edge of the fiducial volume is 3 meters, which is many times the Compton scattering length for multi-MeV gamma rays and many times the resolution of event position reconstruction. Thus, an energy and radius cut are the two most effective ways to eliminate PMT events.

Nevertheless, the decay rate of uranium and thorium in the PMT glass is high enough that a tail of events still can leak into an energy window from 3.5 to 4.0 MeV (reconstructed electron equivalent). In that energy window, inside a radius of 550 cm, there are more PMT events than any single other event type. Few of these events leak above 4 MeV, the energy at which we can start to make a useful measurement of CC events. However, the huge number of PMT events below 4 MeV complicates the estimation of internal ^{214}Bi decays, as discussed in Chapter 4. An accurate, unbiased measure of internal ^{214}Bi is important for the 4 MeV CC measurement, which leads to our interest in the

properties of PMT events. This chapter discusses our understanding of PMT backgrounds, and compares various models of PMT events, ultimately settling on an analytic model as being the best compromise.

6.3.1 Origin of PMT Events

In order to model PMT events we need to know where they come from. The most likely candidates are ^{214}Bi (uranium chain) and ^{208}Tl (thorium chain) decays in the PMT glass and support structure. The decay of ^{214}Bi deposits much of its energy through a beta emission with a 3.3 MeV endpoint, whereas the dominant component of a ^{208}Tl decay is the 2.6 MeV gamma ray.

A sensitive discriminator between these two cases is the isotropy of the light pattern observed in the detector. While generally we use β_{14} as our standard isotropy measure, previous SNO analyses have also used a parameter called θ_{ij}, which is the average pairwise angle between hit PMTs, as viewed from the reconstructed event vertex. If we look outside the fiducial volume of the neutrino analysis, the number of PMT events grows rapidly, and we can get a very pure sample of them. B. Heintzelman[79] investigated the mean θ_{ij} of various background models compared to data, as a function of other observables. Figures 6.5, 6.6, 6.7 shows the θ_{ij} as a function of NHITS, R^3, and $\hat{U} \cdot \hat{R}$. The last variable is the cosine of the angle between the position and direction vectors of the event, and is an excellent discriminator of PMT events. PMT events peak very strongly at negative $\hat{U} \cdot \hat{R}$, which corresponds to inward traveling events. In all three figures, we see that the Monte Carlo simulation of ^{208}Tl decays in the PMTs best matches the data. As we construct our PDF of PMT events, we will then assume the PMT radioactivity is purely the result
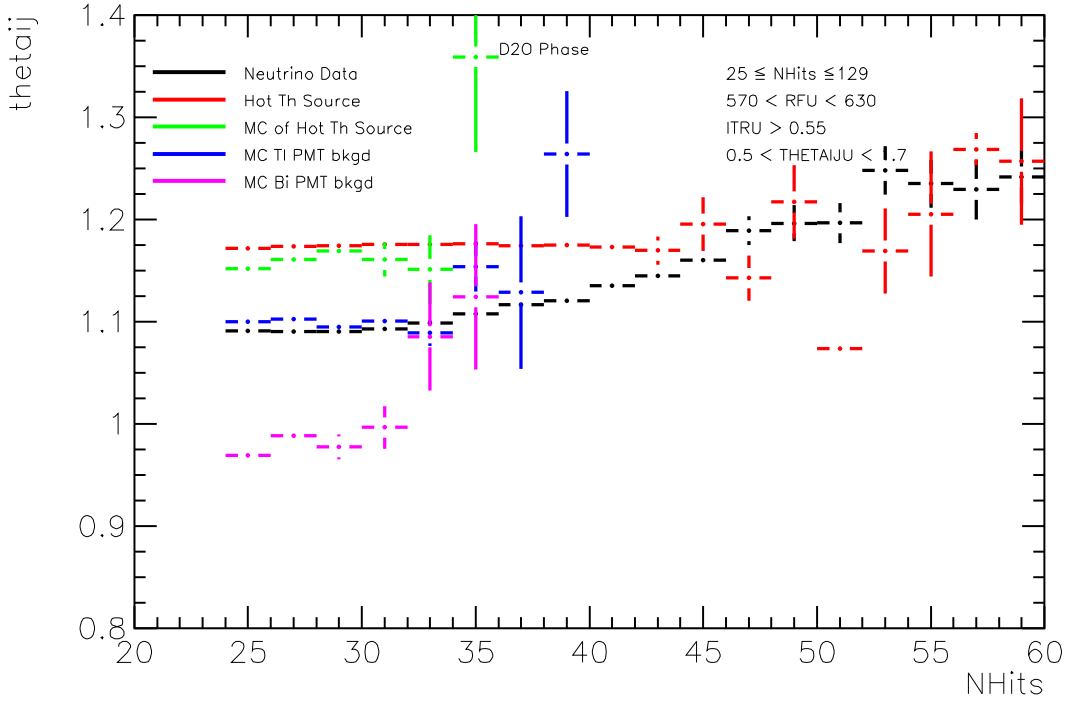
Figure 6.5: Mean θ_{ij} for data and various models of PMT radioactivity as a function of NHITS. The "Hot Th" source is described in Section 6.3.4. Figure courtesy of B. Heintzelman[79].

of thorium-chain decays in the PMT glass.

6.3.2   Monte Carlo Model

Based on the results from the last section, our most accurate model of PMT events comes from the SNOMAN Monte Carlo. The simulation seeds events randomly in the glass regions of the detector, which correspond to the PMT envelopes. Each decay contains the emission of a 2.6 MeV gamma ray as well as several smaller betas which generally fall below the Cherenkov threshold. Figures 6.8 shows the D_2O and salt phase distributions of T_{eff}, β_{14}, and R^3 for
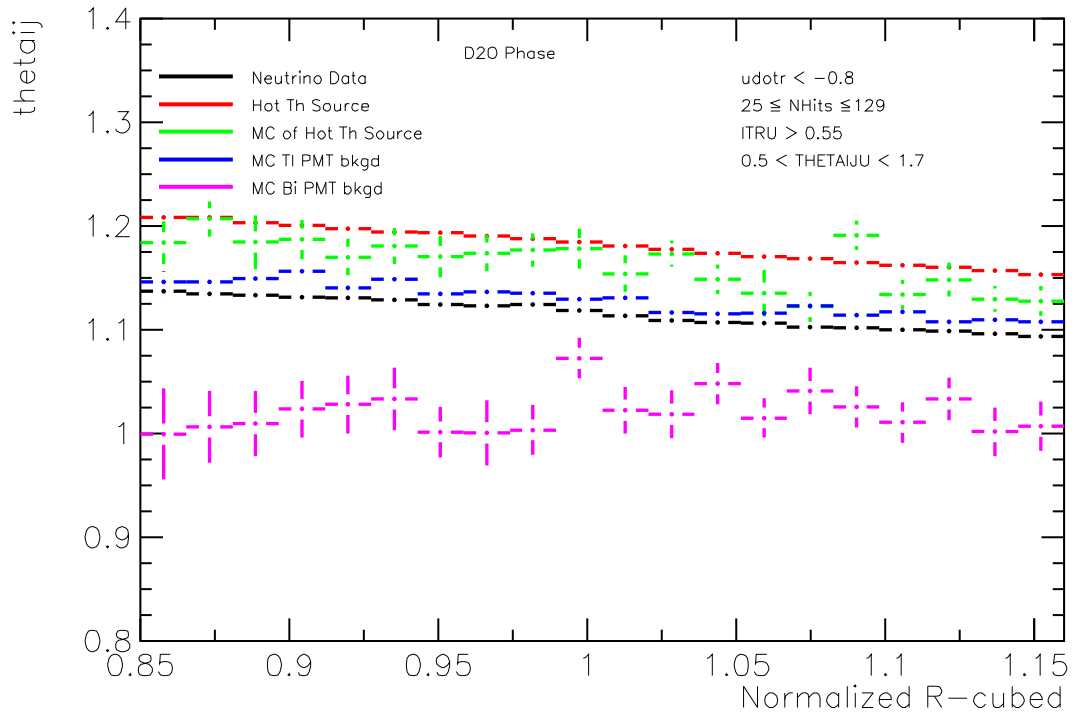
Figure 6.6: Mean θ_{ij} for data and various models of PMT radioactivity as a function of R^3. The "Hot Th" source is described in Section 6.3.4. Figure courtesy of B. Heintzelman[79].
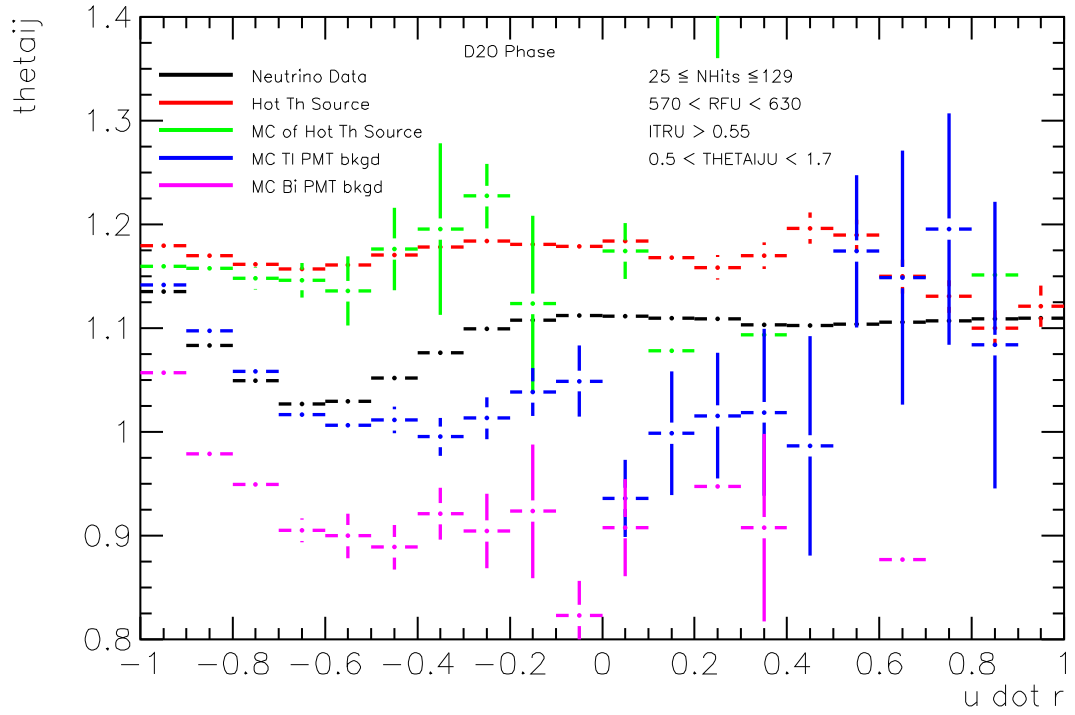
Figure 6.7: Mean θ_{ij} for data and various models of PMT radioactivity as a function of $\hat{U} \cdot \hat{R}$. The "Hot Th" source is described in Section 6.3.4. Figure courtesy of B. Heintzelman[79].

these events.

The main obstacle to using this model to construct a kernel estimated PDF (Chapter 7) is the incredibly low acceptance of PMT events. Very, very few decays actually produces an event which reconstructs inside the fiducial volume and above 3.5 MeV. Using all of the computing power available to the SNO collaboration, we were able to generate the equivalent of 700 Bq of ^{208}Tl decays over the entire livetime of SNO in the first two phases. Independent measures of the PMT radioactivity indicate that the activity of the PMTs in the real detector is in the range of 2 to 7 kBq[80], vastly larger than our Monte Carlo sample. In terms of events, 700 Bq amounts to 292 events in the D$_2$O phase and 1374 events in the salt phase. This is barely adequate for a salt phase PDF, and really not enough for a D$_2$O phase PDF.

6.3.3 Data/Monte Carlo Comparisons

In order to check the quality of the Monte Carlo simulation of PMT events, we need a sample of real data from the detector which is an almost pure sample of PMT events. Going outside the fiducial volume gives a large sample of PMT events, but the accuracy of those events do not necessarily reflect the accuracy inside the fiducial volume.

Fortunately, we have a source of PMT events *inside* the fiducial volume as well. A cut called the "QPDT cut"[74] is designed specifically to remove events which originate in the PMT glass by looking for a high charge hit early in the event. The expectation is that this early, high charge hit is in the PMT where the decay occurred, or perhaps the PMT where the decay gamma first Compton scattered an electron.
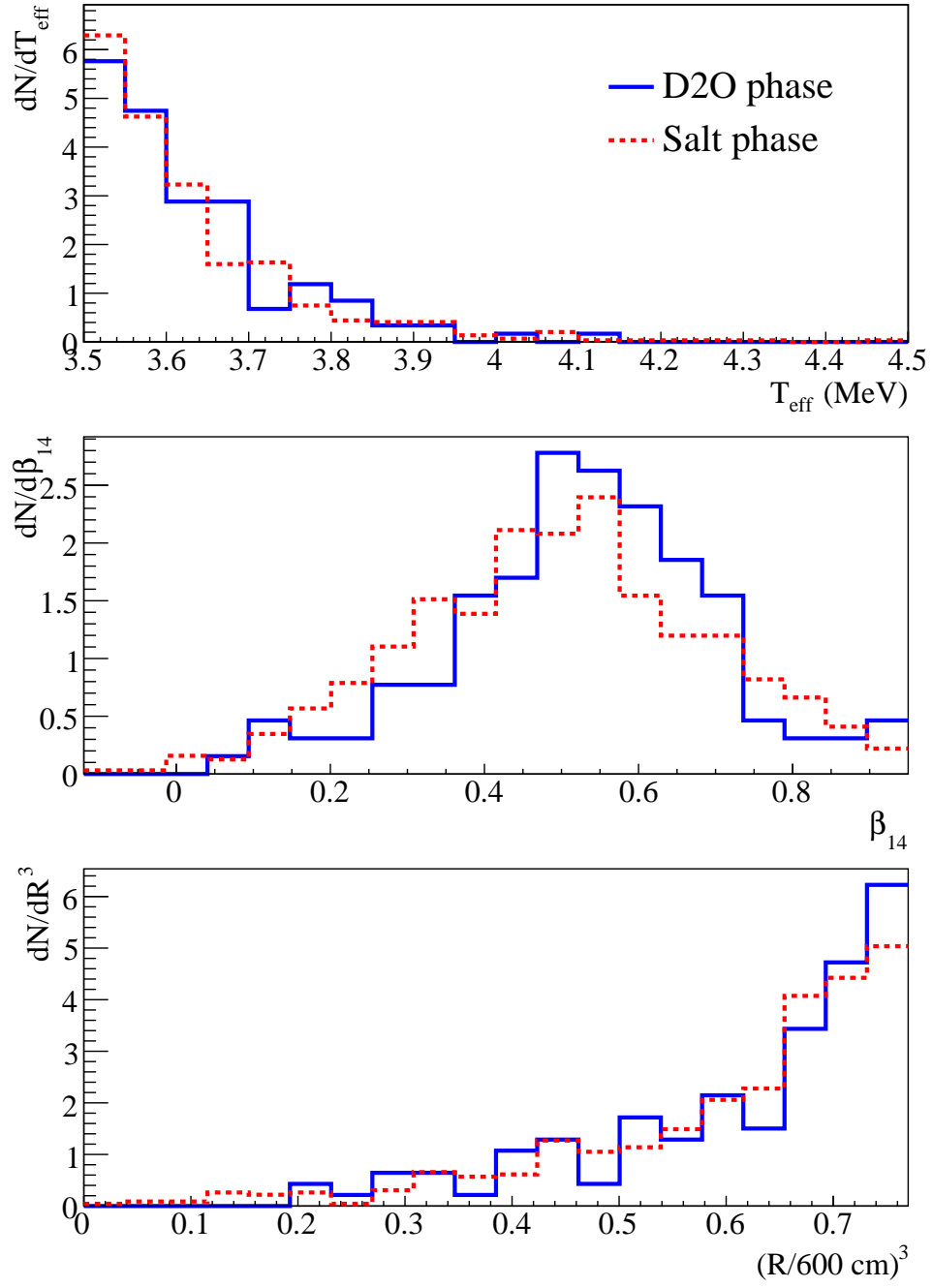
Figure 6.8: Reconstructed observables for Monte Carlo PMT events in the D$_2$O and salt phases.

By looking at both Monte Carlo and data which fail this cut, but pass all other cuts, PMT events can be preferentially selected from the event list. These events do not appear in the normal neutrino analysis, so this is an outside-the-box study. Orebi Gann[81] compared Monte Carlo and data which failed this QPDT cut, shown in Figure 6.9, finding relatively good agreement. The difference between data and Monte Carlo were characterized by the systematics:

- Energy convolution of 0.216 MeV

- β_{14} scale uncertainty of 5%

- R^3 scale uncertainty of 4.5%

These systematics are included in the PDF uncertainty during the final signal extraction.

6.3.4 Calibration Source Model

Scarcity of CPU time originally drove us to consider using data from a calibration source for the PDF, as was done previously[51].

In the D_2O phase, a 20 μCi thorium[1] source was deployed through a guide tube at the radius of the PMT support structure, and in the salt phase, a 25 μCi source was used[82]. Both sources were encased in a stainless steel can, which absorbs all the betas in the decay chain. The branching fraction to the ^{208}Tl decay in the chain is 36%, so these sources provided a 266 kHz

---

[1]This "thorium" source is in fact a ^{232}U source. However, the decay chain of ^{232}U eventually merges with the ^{232}Th chain before ^{208}Tl, the background we are interested in, is produced. To avoid confusion with the ^{238}U chain we call this ^{232}U source a "thorium" source.
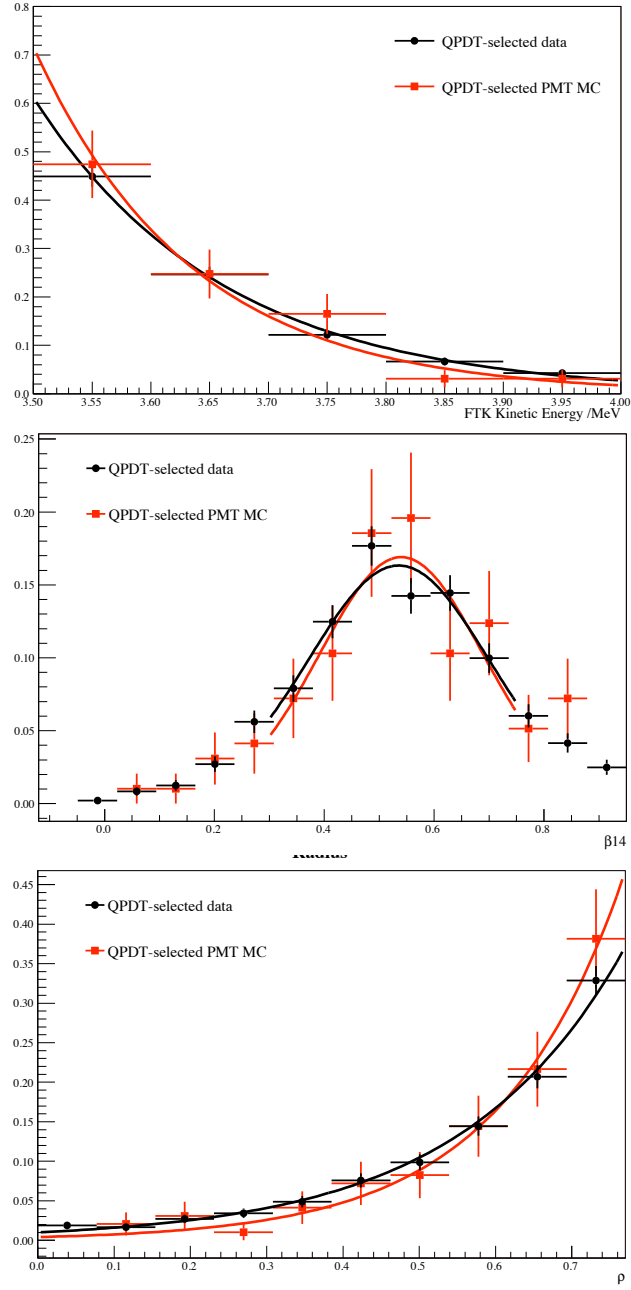
Figure 6.9: Comparison of QPDT-failing events from the salt phase data set, and from Monte Carlo of ^{208}Tl decays in the PMT glass. Top plot shows reconstructed event energy (T_{eff}), middle plot shows isotropy (β_{14}), and bottom plot shows normalized R^3. Figures courtesy of G. Orebi Gann[81].

158

^{208}Tl decay rate	333 kHz
Rate of two gammas within 400 ns	44.4 kHz
Rate of single gamma not followed by second in 400 ns	244 kHz
Acceptance of single gamma	$5.37 \cdot 10^{-8}$
Acceptance of double gamma	$6.43 \cdot 10^{-7}$
Rate of single gamma events passing all cuts	0.0131 Hz
Rate of double gamma events passing all cuts	0.0285 Hz
Relative fraction of single events	32%

Table 6.3: Pileup rates and acceptance for the 25 μCi thorium source deployment in the salt phase. Acceptance is defined to be number of events passing all cuts per decay.

and 333 kHz source of predominately 2.6 MeV gammas. The high decay rate in these sources was intended to compensate for the very low acceptance of low energy PMT-like events inside the fiducial volume. The deployments of these sources yielded 4,000 events in the D$_2$O phase and 10,000 events in the salt phase within our energy and fiducial volume window, after the standard high-level cuts (Section 4.3.4) except the NHIT burst cut were applied.

While the statistics from the source deployments are quite respectable, there is a fundamental flaw in these data sets. The rates are so high that pileup becomes the dominant contribution to the final PDF. To quantify this effect, we performed a Monte Carlo study of single 2.6 MeV gammas and coincident pairs of 2.6 MeV gammas within a 400 ns window. Table 6.3 summarizes the results of the study, and shows that 32% of the events in the LETA analysis window are expected to be the product of single gammas, and the rest will be two gammas within 400 ns of each other. With such high decay rates, even triple gammas will be a non-trivial contribution, which has not been included here.
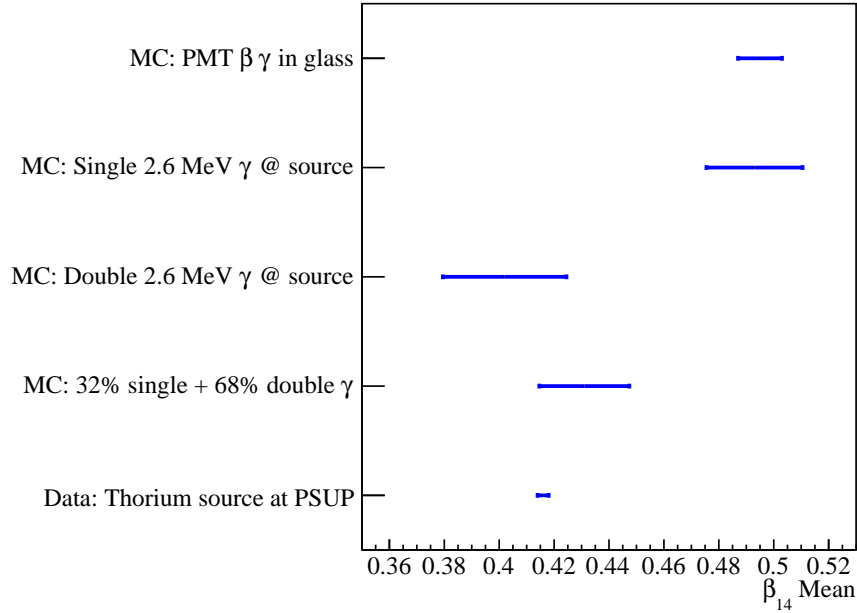
Figure 6.10: Comparison of mean β_{14} for various Monte Carlo samples to the observed data from thorium source at the PMTs. The error bars show statistical uncertainty only.

Two gamma events will be more isotropic than single gamma events, and therefore will have a much lower β_{14} mean. Figure 6.10 shows a comparison of the β_{14} mean for calibration source data compared to various Monte Carlo scenarios. Monte Carlo of single gamma events and PMT beta-gamma events have the same β_{14} mean, whereas double gamma events and the source data have a lower mean. If we take the linear combination of singles and doubles predicted by the Monte Carlo pileup study, then we expect a mean salt phase β_{14} of 0.431 ± 0.016, and the observed mean of the source data is 0.416 ± 0.002. The pileup hypothesis is consistent with the data, though the comparison is statistically limited by the amount of Monte Carlo we can run.

Based on this comparison, we conclude that the thorium source is unsuitable for use in constructing a PMT PDF, and the pileup rate makes it suspect even for estimating the systematic uncertainties of the simulation by comparing source data to source Monte Carlo. We were forced to abandon the source and find an alternative method for building and verifying a PMT PDF.

6.3.5   Analytic Model

If we cannot generate enough Monte Carlo to make a PDF, and pileup poisons the calibration source data, then we are left with trying to construct an analytic model of the PMT distribution in the four observables of interest: T_{eff}, β_{14}, R^3, and $\cos\theta_\odot$. Background decays have no correlation with the position of the Sun, so we can set $P(\cos\theta_\odot)$ to a constant. For the other three dimensions, we propose the following phenomenological model:

$$P(T_{\text{eff}}, \beta_{14}, R^3, \cos\theta_\odot) = e^{A\,T_{\text{eff}}} \times \mathcal{N}(\beta_{14}\,|\,\bar{\beta}_{14} = B + CR^3, \sigma = D) \times (F + e^{E\,R^3}),$$

(6.1)

where $\mathcal{N}(x|\bar{x}, \sigma)$ is a Gaussian distribution with the given mean and width. Exponentials are a reasonable model for a reconstruction tail in position or energy, and the β_{14} distribution is well approximated by a Gaussian. The radial dependence of the β_{14} mean has also been included as it is the most significant correlation between the observables.

By fitting Equation 6.1 to the generated PMT Monte Carlo for each phase separately, we can obtain model parameters and uncertainties on the model. Tables 6.4 and 6.5 show the fit parameters and their uncertainties. The model is overlaid on the Monte Carlo distributions in Figures 6.11 and 6.12. Parameters B and C are almost 100% anti-correlated, so their uncertainties

A	$-7.65 \pm 0.46 \ \mathrm{MeV}^{-1}$
B	0.445 ± 0.056
C	0.157 ± 0.086
D	0.182 ± 0.008
E	7.17 ± 0.45
F	3.31 ± 1.43

Table 6.4: D_2O phase PMT model parameters. See Equation 6.1 for an explanation of the meaning of the parameters.

A	$-7.25 \pm 0.20 \ \mathrm{MeV}^{-1}$
B	0.366 ± 0.023
C	0.249 ± 0.036
D	0.195 ± 0.005
E	6.10 ± 0.18
F	1.07 ± 0.66

Table 6.5: Salt phase PMT model parameters. See Equation 6.1 for an explanation of the meaning of the parameters.

should be applied in opposite directions. An additional uncertainty is included from the comparison of QPDT-selected PMT events from data and Monte Carlo in the generation of the final uncertainties shown in Table 8.7. It should be noted that the flat component of the PMT model, F, fits to zero on the Monte Carlo PMT events, but fits to a non-zero value on the QPDT-selected data. We have included the fit value of F from the data[81], with its large uncertainty, in the PMT model.

6.4 Energy Resolution Systematics

With the ^{16}N fits from the previous chapter, and the low energy background fits described in this chapter, we have 3 separate measurements of the energy
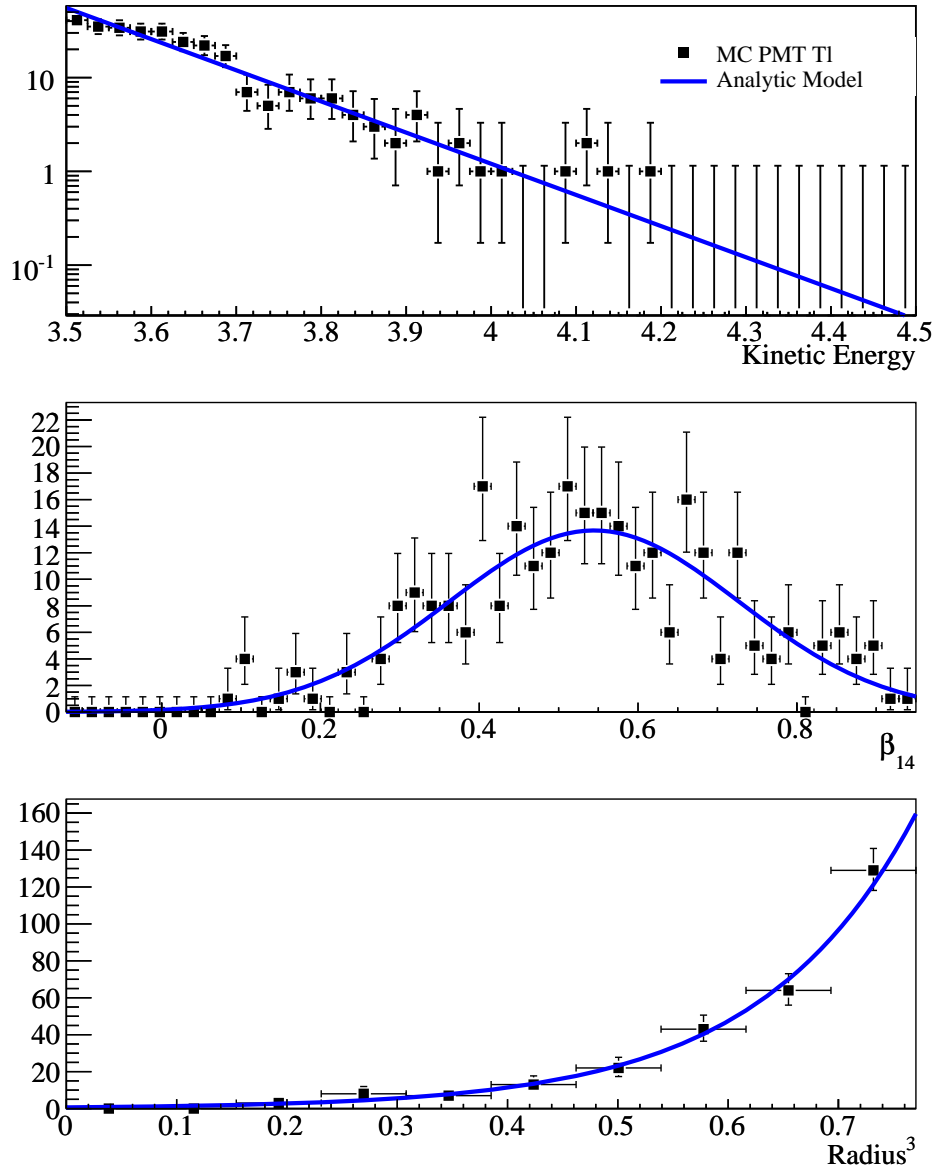
Figure 6.11: Analytic model of D_2O phase PMT distribution compared with Monte Carlo events. Table 6.4 lists the model parameters.
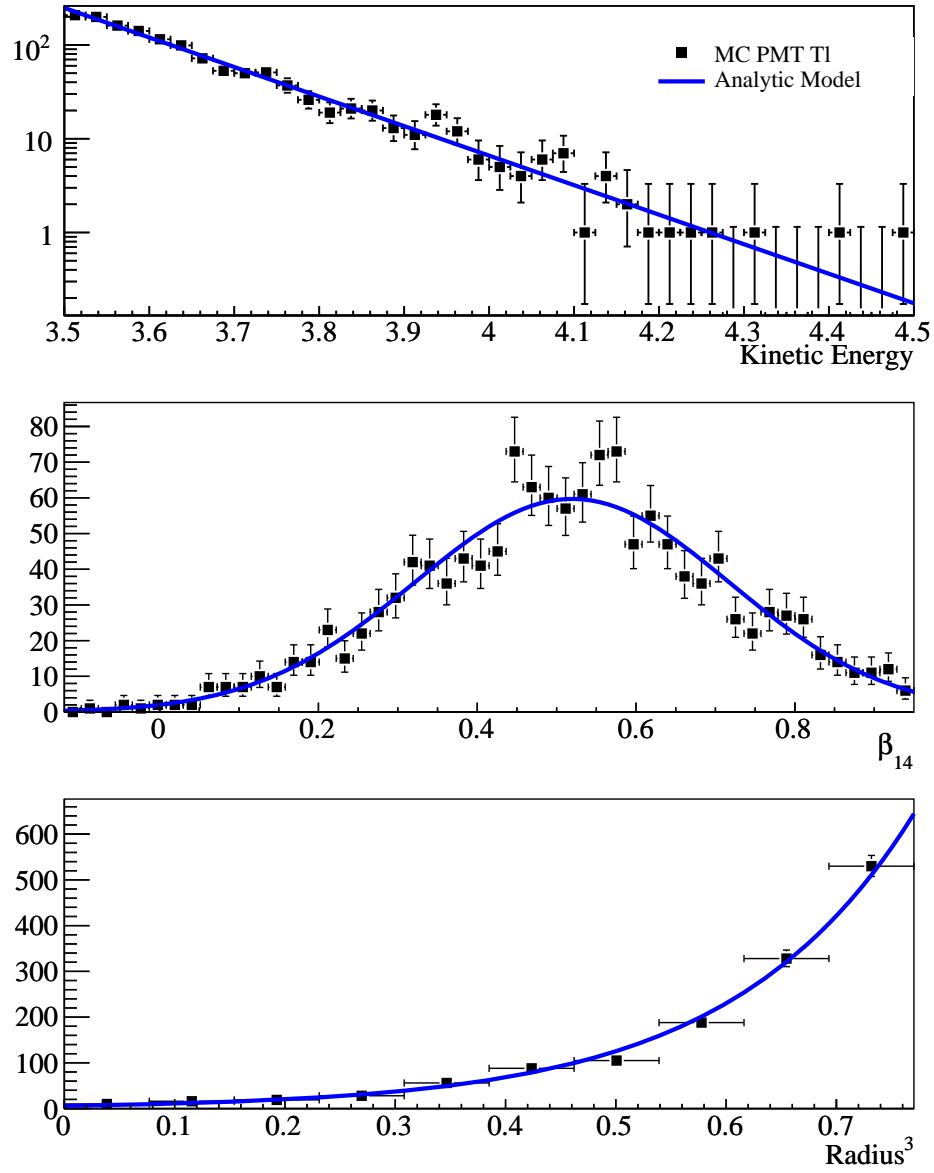
Figure 6.12: Analytic model of salt phase PMT distribution compared with Monte Carlo events. Table 6.5 lists the model parameters.
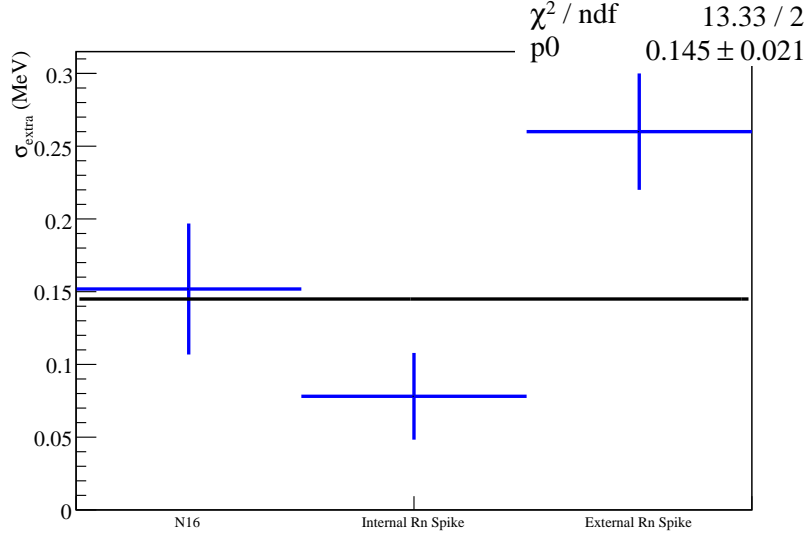
Figure 6.13: Salt phase energy resolution systematic σ_{extra} at $T_{\text{eff}} = 3.5$ MeV as measured by ^{16}N, and the internal and external radon spikes.

resolution systematic uncertainty in the salt phase. Figure 6.13 compares the results from the different sources and shows weighted average. The large χ^2 indicates there are clearly some systematic issues remaining. Following the PDG prescription for combining inconsistent measurements (pg. XX of [10]), we can penalize the error bars of all points until the χ^2 is 1. This makes the constraint on the energy systematic worse, which is appropriate due to the non-uniform sampling of the radon spikes. Based on this, the salt phase Monte Carlo needs a convolution with a Gaussian of $\sigma_{extra} = 0.145 \pm 0.053 MeV$. The D$_2$O phase has no other spike measurements, but based on the ^{16}N alone, the convolution required is $\sigma_{extra} = 0.155 \pm 0.031$.

6.5   Summary

The internal and external radon spikes, as well as the QPDT-selected PMT events, show good agreement with the Monte Carlo. These spikes also help determine the energy resolution systematic uncertainty, along with the ^{16}N fits from the previous chapter. The PMT background PDF is difficult to construct from the limited Monte Carlo statistics available, but we can use an analytic model fit to the Monte Carlo to produce a smooth, continuous distribution.

Chapter 7

Kernel Density Estimation

As described in Section 4.2.2, the maximum likelihood method requires that we supply a PDF, $P_j(\mathbf{x})$, for each signal j that occurs in our data set. We also have a Monte Carlo simulation which can generate a random sample of events for any desired signal. In effect, we want to infer a continuous probability distribution from a finite collection of discrete samples drawn from that distribution.

This is a common problem in statistical analysis which has two general classes of solutions: *parametric* and *non-parametric* density estimators. There is no rigorous distinction between the two, but generally parametric methods involve analytic functions with a small number of free parameters. Often, these analytic functions are motivated by the underlying physics which produces the event and the detector response. For example, the PDF for a resonance in an invariant mass distribution can be described with a Breit-Wigner distribution convolved with a Gaussian detector resolution. The free parameters in this PDF can be extracted from a data sample using a parameter estimation technique such as the maximum likelihood method, or χ^2 minimization.

The main problem with parametric density estimation is that it requires a mathematical model for the distribution *a priori*. With multi-dimensional PDFs, it can be difficult to produce an analytic function which includes all of the correlations between observables, and still accurately describes the PDF over a wide range of $\mathbf{x}$. Instead, we would prefer a non-parametric density estimator, which does not assume a particular functional form for the distribution. The simplest and most commonly used non-parametric density estimator is a binned histogram. Simply counting the number of events in a bin and dividing by the bin volume provides a convenient approximation to the PDF.

Histograms can be problematic for some applications as they are fundamentally piecewise continuous functions. Even in truly constant regions of the PDF, adjacent histogram bins will be discontinuous since each bin will have random statistical fluctuations due to finite sample size. The continuity problem can be mitigated by increasing the sample size and using non-uniform binning, as well as performing some kind of interpolation between bin centers. As the number of dimensions in the PDF increases, statistical fluctuations become a greater concern, but interpolation and variable binning also become more complicated.

An alternative to histograms is *kernel density estimation* (also called kernel estimation). Kernel estimation is also a non-parametric method, but it is unbinned. A function, called the *kernel*, is centered over each sample from the PDF, and the sum of these kernels is used as an approximation to the true PDF. If the kernel function is continuous and differentiable, then the PDF, being a sum of kernels, will also be continuous and differentiable. Kernel estimation is also very straightforward to extend to multiple dimensions through the use of multidimensional kernel functions.

7.1 Theory of Kernel Estimation

7.1.1 One-Dimensional Kernel Estimators

A one-dimensional kernel estimator derived from a set of PDF samples t_i ($i \in [1, n]$) is

$$\hat{P}(x) = \frac{1}{n} \sum_{i=1}^{n} K_i(x - t_i), \tag{7.1}$$

where n is the number of samples, and $K_i(x)$ is the kernel for sample i. The $K_i(x)$ are normalized so that

$$\int K_i(x) \, dx = 1, \tag{7.2}$$

which ensures that $\hat{P}(x)$ is a properly normalized PDF. To make analysis of the kernel estimator more tractable, we can restrict the $K_i(x)$ to all have the same functional form, $K(x)$, and only depend on a rescaled coordinate, $(x - t_i)/h_i$:

$$\hat{P}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{x - t_i}{h_i}\right). \tag{7.3}$$

The scale parameters h_i are called the kernel *bandwidths*. There are many possible options for the kernel function $K(x)$, but the most common choice is the Gaussian kernel,

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{7.4}$$

A Gaussian has many nice properties, such as being continuous and continuously differentiable for all x. Additionally, it is the most studied kernel[1] in the

---

[1]The Epanechnikov kernel, $\frac{3}{4}(1 - x^2)$ for $|x| < 1$, can be shown[83] to be optimal in general, but the discontinuous derivative at $x = \pm 1$ leads to kinks in the full PDF, which are undesirable for other reasons. The Gaussian is close enough to optimal to be an acceptable
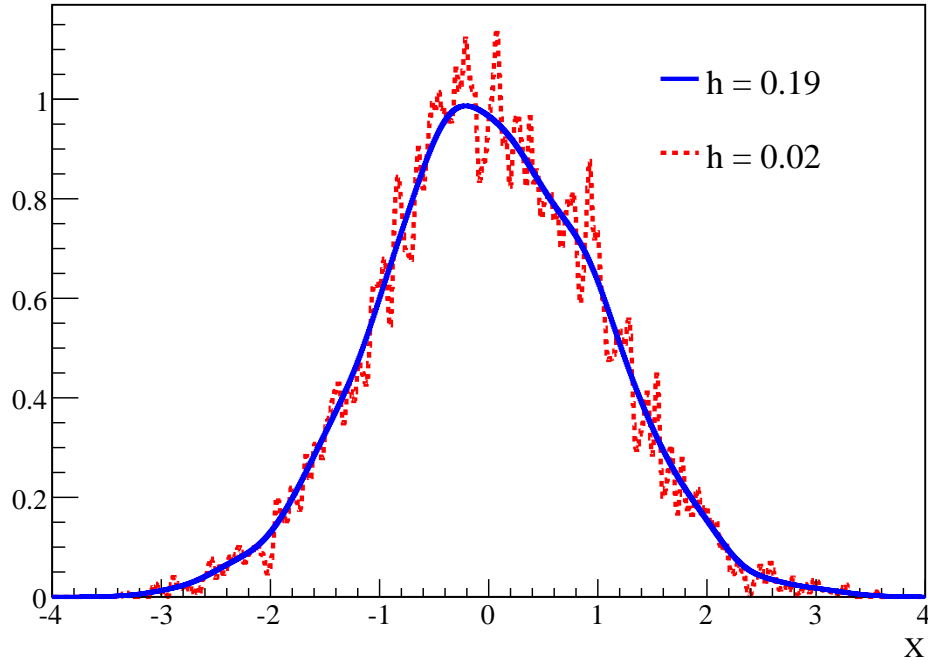
Figure 7.1: A one-dimensional kernel estimator of a Gaussian distribution derived from 5000 samples. The bandwidth $h = 0.19$ is the optimal choice, as computed from Equation 7.12.

literature.

All of the "magic" in the kernel estimator is in the choice of the bandwidths, h_i. The bandwidths act as smoothing parameters, determining the local variability of the PDF. Kernel estimators with single bandwidth, $h_i = h$ for all i, are called *fixed bandwidth* estimators, whereas estimators with h_i that vary for different i are called *adaptive bandwidth* estimators. Figure 7.1 shows a fixed bandwidth kernel estimator for two different choices of h.

substitute.

7.1.2 Error Analysis

In order to select appropriate bandwidths, we need a measure of how close the estimated PDF, $\hat{P}(x)$ is to the true PDF, $P(x)$. The most common choice is the integral square error (ISE),

$$\text{ISE}(\hat{P}) = \int (\hat{P}(x) - P(x))^2 \, dx. \tag{7.5}$$

However, the ISE measures the error for the kernel estimator derived from a particular event sample drawn from the true distribution, $P(x)$. We actually want the expectation value of the ISE, called the mean integral square error (MISE),

$$\text{MISE} \equiv \left\langle \text{ISE}(\hat{P}) \right\rangle = \left\langle \int (\hat{P}(x) - P(x))^2 \, dx \right\rangle. \tag{7.6}$$

Other measures of error are possible, though the MISE is the most commonly used because it is easiest to analyze.

Wand and Jones[83] show that the MISE for a fixed bandwidth, 1D kernel estimator is

$$\text{MISE} = (nh)^{-1} \int K^2(x) \, dx + (1 - n^{-1}) \int (K_h * P)^2(x) \, dx$$
$$- 2 \int (K_h * P)(x)P(x) \, dx + \int P^2(x) \, dx, \tag{7.7}$$

where $*$ is the convolution operator, and $K_h(x) \equiv (1/h)K(x/h)$. This equation reduces to a much simpler form if we consider the large n limit[2], called the

---

[2]Formally, we are taking $h \to 0$ and $n \to \infty$, such that $nh \to \infty$.

asymptotic mean integral square error (AMISE),

$$\text{AMISE} = (nh)^{-1} \left(\int K^2(x) \, dx \right) + \frac{1}{4} h^4 \left(\int x^2 K(x) \, dx \right)^2 \left(\int (P''(x))^2 \, dx \right),$$
$$(7.8)$$

where $P''(x)$ is the second derivative of $P(x)$.

Assuming that $K(x)$ is the Gaussian kernel from Equation 7.4, we can compute the first two integrals in the AMISE to get the expression

$$\text{AMISE}_{\text{Gauss}} = \frac{1}{2\sqrt{\pi} \, nh} + \frac{1}{4} h^4 \int (P''(x))^2 \, dx. \qquad (7.9)$$

As $n \to \infty$, $h \to 0$ and $nh \to \infty$, the AMISE will asymptotically approach zero. The kernel estimator is therefore unbiased as the sample size goes to infinity. This functional form also highlights the bias-variance tradeoff which is inherent to kernel estimation. A small h makes the kernel estimator fluctuate rapidly, but on average, more closely approximate the true PDF, minimizing bias. A large h smoothes the estimated PDF, which minimizes variance. However, making h too small or too big will increase the overall mean integral square error, so a balance of bias and variance is required.

We can solve for the h which minimizes the AMISE,

$$h_{\text{AMISE}} = \left(\frac{1}{2\sqrt{\pi} n \int (P''(x))^2 \, dx} \right)^{1/5}. \qquad (7.10)$$

With this equation, we can find the optimal bandwidth h_{AMISE} for any PDF, assuming the true PDF shape is known. For example, if the true PDF, $P(x)$,

is a Gaussian of width σ, then

$$\int (P''(x))^2\, dx = \frac{3}{8\sqrt{\pi}\sigma^5}, \tag{7.11}$$

and the optimal bandwidth is

$$h_{\text{AMISE}} = \left(\frac{4}{3n}\right)^{1/5} \sigma. \tag{7.12}$$

Using the optimal h_{AMISE} with Equation 7.10, we find that to leading order, the AIMSE scales like $n^{-4/5}$. For comparison, a maximum likelihood estimate of a Gaussian distribution using a two parameter (μ, σ) Gaussian model[84] has

$$\text{AMISE} = \frac{7}{16\sqrt{\pi}} n^{-1}. \tag{7.13}$$

Thus we find that the convergence of the non-parametric kernel estimator is slightly worse than a parametric estimator.

A similar error analysis can be applied to histograms[85], using the bin width b as an analog to the kernel bandwidth. The optimal bin width is

$$b_{\text{AMISE}} = \left(\frac{6}{n \int (P'(x))^2\, dx}\right)^{1/3}. \tag{7.14}$$

For a Gaussian distribution, the optimal choice of bin size yields an AMISE which is proportional to $n^{-2/3}$. Asymptotically, the histogram has slower convergence to the true distribution than a kernel estimator, and both are worse than a parametric estimate with a functional model of the true distribution.

Of course, this analytic approach to computing the MISE/AMISE is only useful in understanding the general properties of various estimators. Find-

173

ing the optimal h_{AMISE} given a sample from an arbitrary distribution in this way requires that the functional form of the true PDF already be known. But if that were always possible, we would not need kernel estimation in the first place! Section 7.2.1 discusses some practical techniques for bandwidth selection when the true distribution is not known.

7.1.3 Multidimensional Kernel Estimators

Kernel estimation can be extended to PDFs defined on a multidimensional space $\mathbb{R}^d$. All that is required is to select an appropriate multidimensional kernel, $K(\mathbf{x})$. A d-dimensional analog to the Gaussian kernel defined in Equation 7.4 is the multivariate Gaussian with unit width:

$$K(\mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}} \right)^d e^{-\mathbf{x}^T \mathbf{x}/2}. \tag{7.15}$$

Now we introduce a real, symmetric, invertible $d \times d$ bandwidth matrix, $\mathbf{H}_i$, and write the generic multivariate kernel estimator

$$\hat{P}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\det \mathbf{H}_i} K\left(\mathbf{H}_i^{-1}(\mathbf{x}\mathbf{t}_i) \right), \tag{7.16}$$

where $\mathbf{t}_i$ is the vector of observables describing each event i used to build the PDF.

A spectral decomposition of the bandwidth matrix,

$$\mathbf{H}_i = \mathbf{V}_i^{-1} \mathbf{D}_i \mathbf{V}_i, \tag{7.17}$$

helps to illuminate the structure. The orthonormal matrix $\mathbf{V}_i$ describes a

local coordinate rotation, and the matrix $\mathbf{D}_i$ is diagonal with elements we will label h_{ij}, where $j = 1 \ldots d$. In the general case, one would want to select the bandwidth matrix such that V_i performs a local rotation into a coordinate system where the observables are uncorrelated, to match the structure of the kernel in Equation 7.15. With sufficient PDF statistics, each kernel is relevant to a very small region of space, and we can neglect local correlations. Setting $\mathbf{V}_i$ to the identity matrix, the kernel estimator then simplifies to

$$\hat{P}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\det \mathbf{D}_i} K \left(\mathbf{D}_i^{-1}(\mathbf{x} - \mathbf{t}_i) \right). \qquad (7.18)$$

Since our multidimensional kernel is just a product of 1D Gaussians, we can rewrite the estimator in component form,

$$\hat{P}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \frac{1}{h_{ij}} K \left(\frac{x_j - t_{ij}}{h_{ij}} \right). \qquad (7.19)$$

7.1.4 Adaptive Kernel Estimators

Section 7.1.2 showed the error analysis of a fixed bandwidth kernel estimator. One might expect that allowing a different h_i for each Gaussian kernel would lead to an improved estimate of the true PDF.

To motivate this idea, consider a toy example. Suppose we have only one event, with observable value $x = t_1$, with which to build our kernel estimator. This event gives us very little information. The underlying PDF could be a delta function or a flat distribution with equal likelihood. It is reasonable then, to assign a large value for h_i since it is still quite possible for the PDF to be non-zero in a wide range around t_1, as shown in Figure 7.2.
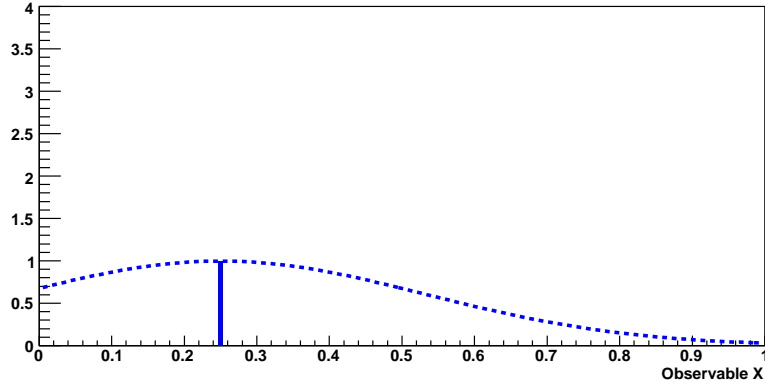
Figure 7.2: A kernel estimator derived from one event.

Now suppose we obtain a second event for building our PDF, with value $x = t_2$. If t_2 is very close to t_1, then it is more likely that we have a PDF distribution peaked near t_1 and t_2 than a flat distribution. We should then reduce h_i to narrow the kernel around these events, obtaining the kernel estimator shown in Figure 7.3.

However, if t_2 is far away from t_1, then we may still have a flat distribution, or perhaps a bimodal one. This ambiguity means that each event carries less information, and so we should use a wide kernel, as shown in Figure 7.4.

Generalizing to a large number of events, we can say that the width of the kernel, and therefore h_i, should be smaller when the local density of events is large. Low density regions will have relatively larger statistical fluctuations due to finite sampling, and so the increased kernel bandwidth smears out such variation. In the limit of infinite event statistics for our PDF, any reasonable kernel function will approach a delta function and the overall kernel estimator will approach the true probability density.

Trying to optimize n different h_i turns out to be a very difficult in gen-
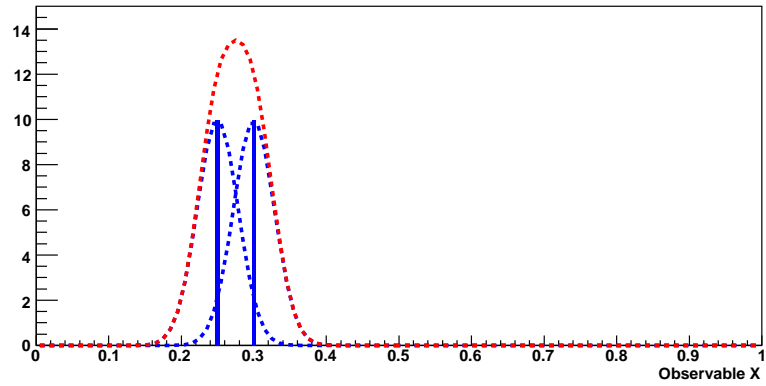
Figure 7.3: A kernel estimator derived from two events close together. The individual event contributions are shown in blue, and the sum in red.
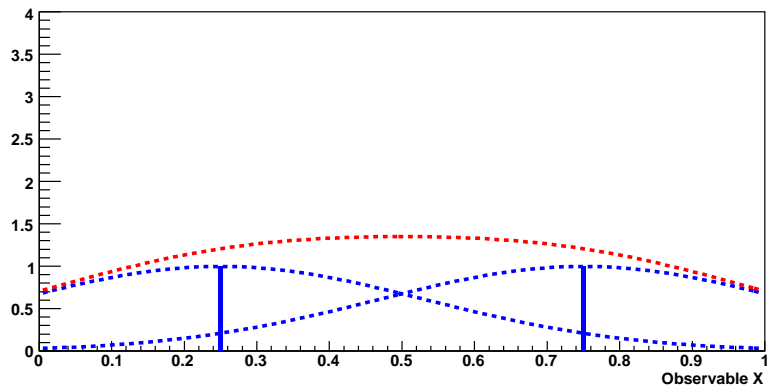
Figure 7.4: A kernel estimator derived from two events far apart.

eral. As such, there is not a universally preferred technique for bandwidth selection, with many of the methods relying on heuristic arguments. The earliest work on adaptive kernel estimation was done by Abramson[86]. In particular, Abramson proposed an inverse square-root dependence to the bandwidth for 1D kernel estimators:

$$h_i = \frac{h}{\sqrt{\hat{P}_0(t_i)}}, \tag{7.20}$$

where h is some fixed bandwidth and t_i is a point from the set of PDF samples, as was used in Equation 7.3. The subscript on $\hat{P}_0$ indicates that it does not have to be the true PDF $P(x)$, but rather just an estimate of it. It is common to use a fixed bandwidth kernel estimator as the pilot distribution $\hat{P}_0$ for the purposes of computing the h_i.

In a particle physics context, Cranmer[87] suggests an adaptive bandwidth heuristic based on the scaling the optimal fixed bandwidth for a Gaussian distribution. In d-dimensions, the bandwidth in dimension j is

$$h_{ij} = \rho \left(\frac{4}{d+2} \right)^{1/(d+4)} \sigma_j n^{-1/(d+4)} \left(\frac{1}{\sigma \hat{P}_0(\vec{t_i})^{1/d}} \right), \tag{7.21}$$

where ρ is a scale factor usually set to 1, σ_j is the standard deviation of the PDF sample in dimension j, and σ is the geometric mean $\left(\prod_{j=1}^{d} \sigma_j \right)^{1/d}$. At first glance, the origin of this formula is rather mysterious, so it is instructive to break it down factor-by-factor:

- $n^{-1/(d+4)}$ — With more events, the bandwidth should get smaller, but preserving the limit $nh \to \infty$. The dependence on d shows the "curse of dimensionality." As the number of dimensions in the PDF grows, our total information about the PDF drops. In the case $d = 1$, this reduces

178

to the $-1/5$ exponent in Equation 7.10.

- σ_j — The standard deviation of the events in dimension j provides a length scale for the kernel bandwidth. The bandwidth will automatically scale with linear coordinate transformations of the PDF observables, and will carry the correct units.

- $\left(\frac{4}{d+2}\right)^{1/(d+4)}$ — This coefficient comes from minimizing the AMISE for a d-dimensional Gaussian PDF. Compare to Equation 7.12 for $d = 1$. As an approximation, assuming a Gaussian shape is a reasonable starting point for unimodal distributions.

- $\sigma \hat{P}_0(\mathbf{t}_i)^{1/d}$ — The bandwidth is inversely proportional to this dimensionless quantity. The geometric mean σ is included in the product to cancel the units of $P_0^{1/d}$, so that this factor is invariant to linear coordinate transformations. Roughly, this factor is proportional to the average distance between points in the PDF sample in the neighborhood of $\mathbf{t}_i$. Note that for $d = 1$, this prescription differs from Abramson's inverse square-root formula.

- ρ — This scaling constant is included simply to highlight that this particular bandwidth formula is not optimal for all PDF shapes. Some PDFs may benefit from adjust ρ to be different than 1. For example, bimodal distributions should probably have $\rho < 1$, as σ_j will tend to overestimate the length scale of PDF structure.

7.2 Implementation Techniques

The biggest drawback to kernel estimation is speed: both in bandwidth selection and in evaluation. A straightforward implementation of Equation 7.3 requires computing n Gaussians to obtain the value of the PDF at a single point. When the PDF is composed of 1 million events, and the PDF needs to be evaluated at 18,000 points, kernel estimation quickly becomes a computationally intractable problem. The situation is even worse if the PDF shape is being varied (see Section 7.2.5) during optimization of the likelihood function. Every time the optimization algorithm (i.e. MINUIT) changes a shape parameter, all 18,000 points will need to be re-evaluated.

No existing kernel estimator implementation was fast enough[3] to perform this task, so we were forced to incorporate a number of shortcuts to speed the process up. These will be described in the following sections.

7.2.1 Bandwidth Selection

We perform the bandwidth selection for each PDF according to Equation 7.21. Traditionally, the pilot estimate of the PDF, $\hat{P}_0(\mathbf{x})$, is taken to be a fixed bandwidth kernel estimator. However, the evaluation of this pilot estimate requires $O(n)$ calls to the exponential function for each $\mathbf{t}_i$. The computational complexity of bandwidth selection is therefore $O(n^2)$, which is unacceptable for our usage. The Monte Carlo event sample size for our PDFs is typically $O(10^5)$, with some PDFs reaching $O(10^6)$. Bandwidths only need to be computed once for a particular PDF event sample, so in principle one could cache

---

[3]A Gaussian version of the fast multipole method does exist for efficient evaluation of kernel estimators, and is described in [88]. However, the method does not allow systematics to be propagated to the PDF on the fly, as described in 7.2.5, so we do not use it here.
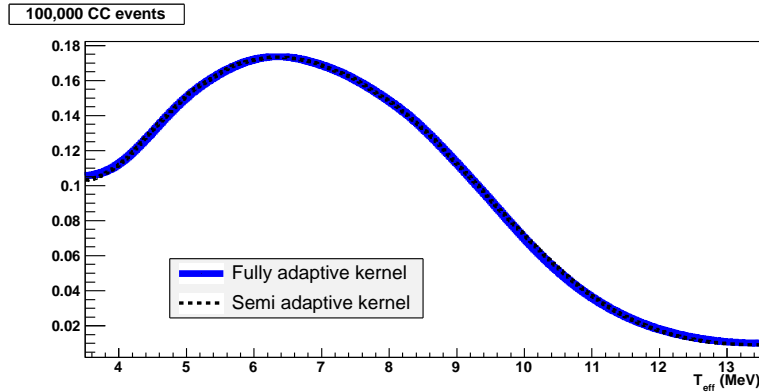
Figure 7.5: A comparison of the fully-adaptive and semi-adaptive kernel estimators derived from 100,000 CC events. The generation time was 18 minutes for the fully adaptive kernel and 26 seconds for the semi-adaptive kernel.

these bandwidths on disk, and reuse them on every run. This makes it very cumbersome to change the PDF event sample (for example, to look at a different energy range), so instead we adopted a technique to make bandwidth calculation much faster.

Instead of using a fixed-bandwidth kernel estimator for the pilot estimate, we use a binned histogram to bootstrap the process. As a result, for all i where $\mathbf{t}_i$ are contained in the same $\hat{P}_0$ bin, the h_{ij} are identical. We call the kernel estimator constructed this way a *semi-adaptive* kernel estimator. The bandwidths do vary from event to event, but only in discrete steps. In general, this optimization has no significant effect on the PDF estimate, but a huge impact on the initialization time, as shown in Figure 7.5.

After working so hard to eliminate bins from the PDFs, reintroducing them might seem counterintuitive. The binning only causes discontinuities in the distribution of h_{ij}, but the actual kernel estimator $\hat{P}(\mathbf{x})$ remains continuous. For most distributions, we are relatively insensitive to the bin size. One

181
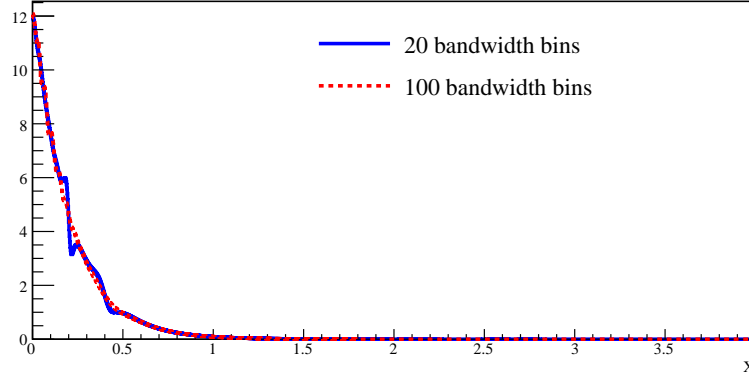
Figure 7.6: A comparison of two semi-adaptive kernel estimators with different binning used in the pilot estimate. Scooping in the estimator is apparent for exponential distributions if the bin size is too large compared to the PDF decay length.

exception to this rule is a PDF which is just an exponential tail in one of its dimensions. Such a fast-changing distribution can experience "scooping" when the pilot histogram has too few bins. Figure 7.6 shows an example of this. The problem goes away if more bins are added to the pilot distribution in the bandwidth calculation.

7.2.2 Boundary Correction

In practice, the PDFs we use are always defined (if only for computational convenience) over a finite range in the space of observables. This presents a problem for events near the boundary, since a kernel centered on an event close to the edge has significant overflow into the disallowed region. Some of the probability associated with that kernel will be lost, effectively de-weighting events near boundaries. To avoid underestimating the PDF at the edges, there

are several options:

- Construct the PDF from events which cover a wider range than the range used in the fit, and "crop" the PDF. If the boundary is a physical boundary, then this is not possible. For example, you cannot have an event with negative radius.

- Use a kernel which becomes asymmetric near the boundaries in order to conserve probability inside the allowed region. There are difficulties in doing this in an analytically simple way.

- Mirror the $\mathbf{t}_i$ points across the boundary. Reflected points are given the same bandwidth h_i as the original point, and can be added to the list of points which are used to compute the kernel estimator. The reflected points produce an event sample which extends beyond the PDF boundaries by assuming the PDF is periodic in space. If the derivative of the PDF at the boundary is large in magnitude, then the assumption of periodicity tends to cause the kernel estimator to slightly over or undershoot the true PDF.

We find the last option to be the most practical approach. In one dimension, the reflection process is straightforward, but in d-dimensions, there are $3^d - 1$ reflection points, as shown in Figure 7.7. Memory usage can be limited by only keeping reflection points which are close to the boundary. "Close" is quantified in Section 7.2.4, where the σ-cut is described.

Figure 7.8 shows a one-dimensional kernel estimator with and without the boundary correction.
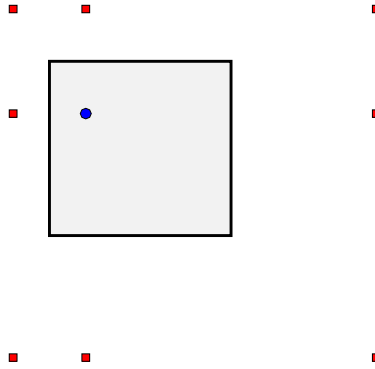
Figure 7.7: Boundary correction points generated by reflection. The circle marks the original point, and the $3^2 - 1 = 8$ small squares show the points generated by reflecting over combinations of boundaries in each dimension.
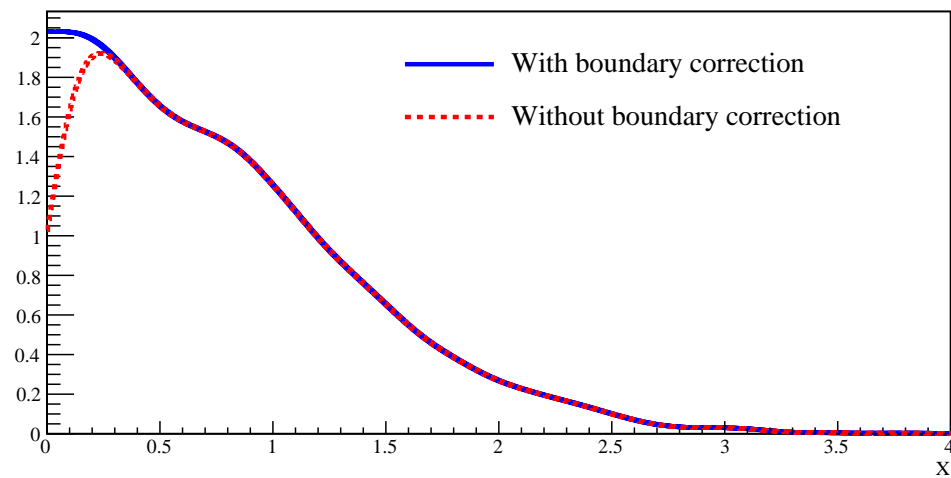
Figure 7.8: A sample kernel estimator ($n = 5000$) with the boundary correction turned on and off. The true PDF is a half-Gaussian with mean of 0 and width of 1.

7.2.3 Normalization

The kernel estimator as defined in Equation 7.3 normalized, by construction, for the interval $(-\infty, \infty)$. However, as mentioned in the previous section, we use the PDFs over a finite interval. Therefore we need to divide the PDF by its integral to ensure it is normalized in the actual fit range. This can be done analytically for the Gaussian kernel using the error function

$$\mathrm{erf}(b) \equiv 2 \int_0^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = 2 \int_0^b K(x) \, dx. \qquad (7.22)$$

The integral of the adaptive 1D kernel estimator in the interval $[a, b]$ is

$$\int_a^b \hat{P}(x) \, dx = \frac{1}{2n} \sum_{i=1}^n \left[\mathrm{erf} \left(\frac{b - t_i}{h_i} \right) - \mathrm{erf} \left(\frac{a - t_i}{h_i} \right) \right], \qquad (7.23)$$

and the integral of the d-dimensional estimator in the (hyper)-cubic volume V with corner points $\mathbf{a}$ and $\mathbf{b}$ is

$$\int_V \hat{P}(x) \, d\mathbf{x} = \frac{1}{2^d n} \sum_{i=1}^n \prod_{j=1}^d \left[\mathrm{erf} \left(\frac{b_j - t_{ij}}{h_{ij}} \right) - \mathrm{erf} \left(\frac{a_j - t_{ij}}{h_{ij}} \right) \right]. \qquad (7.24)$$

It is important to note that n also now includes the reflected events added to the event sample after the boundary correction procedure.

7.2.4 Evaluation

Evaluation of the kernel estimator is by far the critical performance path. For each likelihood calculation with $O(10)$ PDFs and a data set of $O(10^4)$ events, Equation 7.19 must be computed. In the most complicated fits we

have tested, optimization of the likelihood function requires $O(10^4)$ iterations. All together, a single fit will require $O(10^9)$ PDF evaluations. In this section we will describe some strategies for speeding up kernel estimation.

σ-Cut

The most expensive operation in Equation 7.19 is the exponential function used in the definition of the Gaussian kernel. Even highly optimized implementations of the exponential require 75 clock cycles per function call[4].

As written, the equation requires nd exponential evaluations, but the d-dimensional Gaussian kernel can be trivially transformed to only required n exponentials:

$$\hat{P}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \left(\prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}h_{ij}} \right) \exp \left(-\frac{1}{2} \sum_{j=1}^{d} \left(\frac{x_j - t_{ij}}{h_{ij}} \right)^2 \right). \qquad (7.25)$$

We can further reduce the number of exponentials by accepting a small amount of error in the estimator. The sum which yields the value of the PDF at a particular point $\mathbf{x}$ is dominated by a handful of PDF points $\mathbf{t}_i$ in the immediate vicinity of $\mathbf{x}$. If we call the distance between $\mathbf{x}$ and event $\mathbf{t}_i$,

$$\sigma_i(\mathbf{x}) = \sqrt{\sum_{j=1}^{d} \left(\frac{x_j - t_{ij}}{h_{ij}} \right)^2}, \qquad (7.26)$$

then we can look at the distribution of σ for a particular $\mathbf{x}$ and PDF shape. Figure 7.9 shows the σ distribution for a Gaussian PDF $\mathcal{N}(0,1)$ at two points: one at the peak and one in the tail at $x = 2$. As we move away from the peak

---

[4] `fastexp()` from the AMD Math Core Math Library[89].
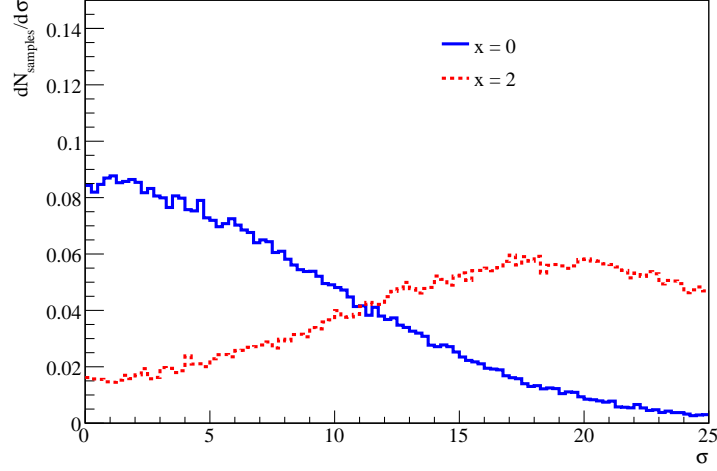
Figure 7.9: Distance of 100,000 PDF samples from two evaluation points $x = 0$ and $x = 2$. Samples are used in a kernel estimator of a unit Gaussian distribution, and distance is shown in units of σ, defined in Equation 7.26. In the $x = 2$ case, 15% of the samples are more than 25σ from the evaluation point.

of the PDF, the majority of the PDF samples are far away from $\mathbf{x}$ in units of σ.

Points with a very large σ contribute little to the final sum, so we can shortcut the exponential with an approximation:

$$
\exp_{\text{fast}}\left(-\frac{\sigma_i^2}{2}\right) = \begin{cases} 0 & \text{if } \sigma_i > \sigma_{\text{cut}} \\ \exp\left(-\frac{\sigma_i^2}{2}\right) & \text{otherwise} \end{cases} . \tag{7.27}
$$

Figure 7.10 shows the absolute error as a function of σ_{cut} and also the relationship between the error and the number of full exponentials which need to be performed. Floating point numbers have a limited precision, and relative errors smaller than 10^{-15} are lost in the round-off error of double precision

arithmetic[90]. Thus, when $\hat{P}(x)$ is $O(1)$, there should be no observable effect to employing $\sigma_{\text{cut}} = 8$. We chose $\sigma_{\text{cut}} = 7$, as 10^{-12} error is acceptable for our application, and it allows us to skip at least half, and often much more than half, of the exponential function calls.

Partitioning

Evaluation of background PDFs which peak far from the bulk of the events are accelerated enormously by the σ-cut. The new critical performance path becomes the actual computation of σ_i in order to choose a branch in Equation 7.27. The calculation of σ_i involves only a handful of simple floating point operations, though. In fact, we find that the bottleneck is not arithmetic, but rather the memory bandwidth required to move the t_{ij} and h_{ij} from system memory onto the CPU in order to compute Equation 7.26. Evaluating a 3D kernel estimator composed of 1,000,000 events requires transferring 48 MB of data from system memory to the CPU. As this exceeds the size of the fast on-CPU cache memory, the data will be read over and over again, each time the PDF is evaluated at a single $\mathbf{x}$.

In order to limit the amount of unnecessary memory transfers, we can *partition* the domain of the PDF $\hat{P}(\mathbf{x})$ into d-dimensional rectangular regions V_k. Each volume k is defined by two points $\mathbf{a}_k$ and $\mathbf{b}_k$ which are on diagonally opposite corners of the box, such that $a_{kj} < b_{kj}$ for all $j \in [1, d]$. We can construct a set of kernel estimators $\hat{P}_k(\mathbf{x})$ such that

$$\hat{P}(\mathbf{x}) = \hat{P}_k(\mathbf{x}) \text{ if } \mathbf{x} \in V_k. \tag{7.28}$$

Following the logic from the previous section, we only need to include
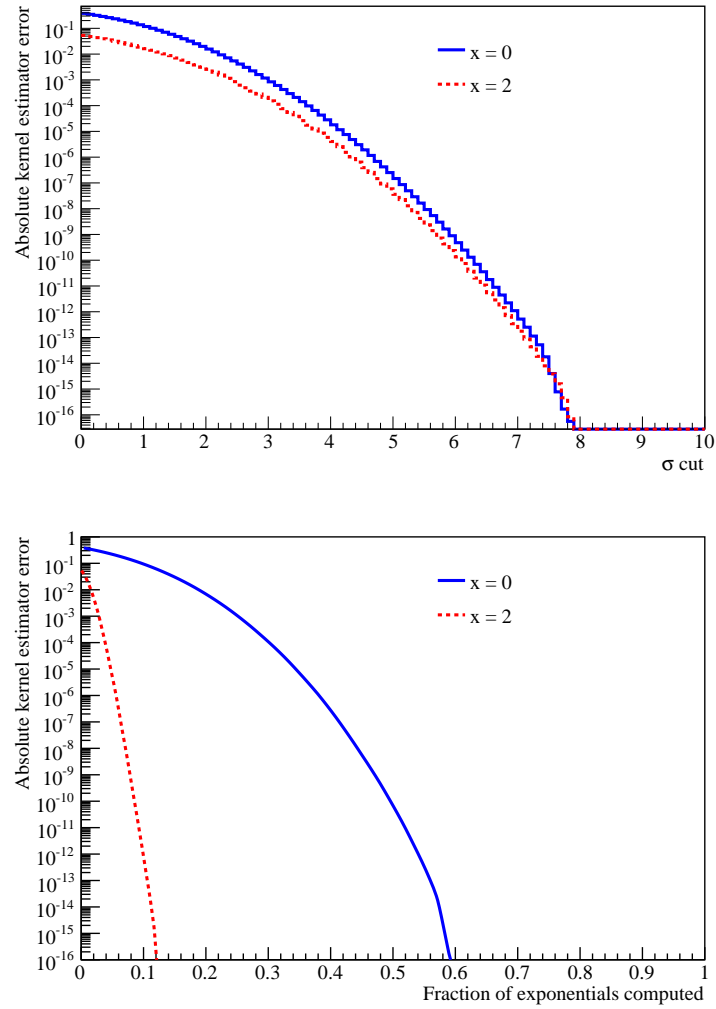
Figure 7.10: Absolute error on $\hat{P}(x)$ at $x = 0$ and $x = 2$ with a σ-cut. Top plot shows the error as a function of the cut parameter, and the bottom plot shows the error as a function of the fraction of exponential function evaluations that actually need to be performed. These curves are computed for $n = 100,000$.
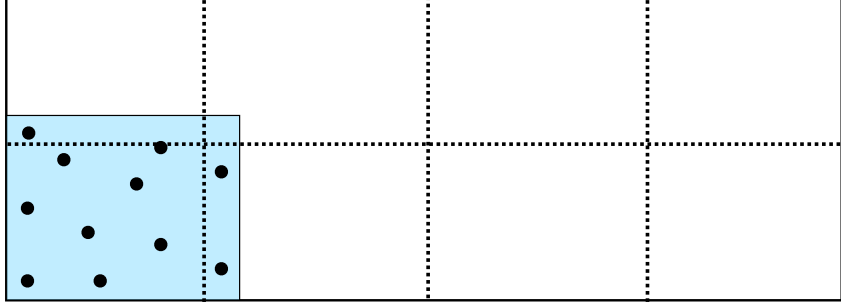
Figure 7.11: Schematic of how PDF samples are assigned to a partition. The partition in the lower left is valid only for x inside the dashed boundary, but includes all PDF samples in the shaded region.

PDF samples in $\hat{P}_k$ which contribute to the final sum after the σ-cut. The set of samples, T_k associated with partition k is

$$T_k = \{ \ \mathbf{t}_i \ | \ \forall i \text{ if } \exists \mathbf{x} \in V_k \text{ such that } \sigma_i(\mathbf{x}) < \sigma_{\text{cut}} \}. \tag{7.29}$$

All PDF samples $\mathbf{t}_i \in V_k$ will be in T_k, but also points just outside of T_k as well. These boundary points ensure the continuity of the distinct kernel estimator fragments across partition boundaries. Figure 7.11 shows a diagram of the partitioning for a 2D kernel estimator. Each of these $P_k(\mathbf{x})$ can be treated as a totally independent kernel estimator and normalized over the restricted region V_k as described in Section 7.2.3.

Unlike the σ-cut optimization, where we trade accuracy for speed, the partitioning optimization trades memory for speed. Samples near the partition boundaries appear in multiple T_k sets, increasing the overall memory usage of the kernel estimator while reducing the amount of memory accessed in any given evaluation. In order to pick a suitable partition size in each of the d dimensions, we have found this equation for the number of partition slices

along dimension j to be useful:

$$N_{\text{part},j} = \frac{b_j - a_j}{b_{kj} - a_{kj}} = \left\lceil f \times \frac{1}{n} \sum_{i=1}^{n} \frac{b_j - a_j}{\sigma_{\text{cut}} h_{ij}} \right\rceil, \qquad (7.30)$$

where b_j and a_j are the boundaries of the entire PDF domain and f is a tunable parameter that trades memory for speed. When $f = 1$, the size of each partition is roughly the size of the bandwidth times σ_{cut}. This is the point of diminishing returns. Making $f < 1$ will increase memory usage rapidly as samples start to appear in the boundary regions of several consecutive partitions, without substantially improving performance by decreasing the number of samples in each partition. Although $f = 1$ sounds optimal, we find that the memory usage is too great for current computers with 2–4 GB of memory. A reasonable tradeoff is found at $2.5 < f < 3.5$ on such systems[5].

Calculation with 3D Graphics Hardware

As a final speed improvement, we discovered that the massively parallel floating point hardware found on modern 3D graphics cards is very efficient at evaluation of kernel estimators. A detailed discussion of the architecture of graphics processor units (GPUs) and the mapping of the kernel estimator algorithm to them is given in Appendix A. With the GPU, we are able to evaluate a single PDF approximately 1500 times per second, on average. This speed is sufficient to allow moderately complex fits to finish in hours on a single computer.

---

[5]In the more restricted memory environment of graphics cards (see Appendix A) with 512–768 MB of memory, the memory factor f must be set to 4.5.

7.2.5 Floating Systematic Uncertainties

Kernel estimation requires that all PDF samples are retained in memory at all times. This makes a kernel estimator far more resource-hungry than a histogram, but at the same time, opens up the doors to a very natural treatment of systematic uncertainties.

As discussed in Section 4.2.9, systematic uncertainties on PDF shapes are included in the maximum likelihood formalism by augmenting the parameters of the model with a vector of systematic parameters, $\boldsymbol{\Delta} = (\Delta_1, \Delta_2, \ldots, \Delta_s)$. The systematic parameters are nuisance parameters which map out a continuous family of PDF shapes, which we postulate contains the true PDF that accurately describes our data. Being able to vary these systematics continuously opens up the possibility of floating these parameters during a maximum likelihood fit. The PDF can actually morph on the fly to seek the best PDF shape given the data. In addition, this allows correlations between signal parameters and nuisance systematic parameters to be treated consistently.

To achieve this goal in the kernel estimator framework, we formally define a systematic to be a triplet of functions,

$$S : \mathbb{R}^d \times \mathbb{R}^s \quad \rightarrow \quad \mathbb{R}^d \tag{7.31}$$

$$C : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^s \quad \rightarrow \quad \mathbb{R}^d \tag{7.32}$$

$$W : \mathbb{R}^d \times \mathbb{R}^s \quad \rightarrow \quad \mathbb{R}. \tag{7.33}$$

The first function $S(\mathbf{t}_i, \boldsymbol{\Delta}) \rightarrow \mathbf{t}'_i$, is called the *transformation function*. It displaces the PDF sample $\mathbf{t}_i$ by some amount that depends on some parameters in $\boldsymbol{\Delta}$, but also might depend on multiple observables in $\mathbf{t}_i$. For example, an

energy-dependent radial scaling of samples would be implemented this way.

The second function $C(\mathbf{t}_i, \mathbf{h}_i, \mathbf{\Delta}) \rightarrow \mathbf{h}_i'$ is called the *convolution function*. It modifies the kernel bandwidth for each sample. As the name suggests, it can be used to implement resolution systematics through convolution. In 1D, for example, we can do a fixed convolution using the systematic

$$C(t_i, h_i, \sigma_{\text{conv}}) = \sqrt{h_i^2 + \sigma_{\text{conv}}^2}. \tag{7.34}$$

This is another advantage to using the Gaussian kernel $K(x)$: convolution has a compact, analytic form.

The third function $W(\mathbf{t}_i, \mathbf{\Delta})$ is called the *weight function*. It allows the systematic to reweight the events in the PDF. This is useful for systematics which involve detection efficiencies, which may or may not depend on the observables themselves. We will also float neutrino survival probabilities this way in Section 8.9.

Multiple systematics can be combined through composition,

$$(S_2 \circ S_1)(\mathbf{t}_i, \mathbf{\Delta}) \quad \equiv \quad S_2(S_1(\mathbf{t}_i, \mathbf{\Delta}), \mathbf{\Delta}) \tag{7.35}$$

$$(C_2 \circ C_1)(\mathbf{t}_i, \mathbf{h}_i, \mathbf{\Delta}) \quad \equiv \quad C_2(S_1(\mathbf{t}_i, \mathbf{\Delta}), C_1(\mathbf{t}_i, \mathbf{h}_i, \mathbf{\Delta}), \mathbf{\Delta}) \tag{7.36}$$

$$(W_2 \circ W_1)(\mathbf{t}_i, \mathbf{\Delta}) \quad \equiv \quad W_2(S_1(\mathbf{t}_i, \mathbf{\Delta}), \mathbf{\Delta}) \times W_1(\mathbf{t}_i, \mathbf{\Delta}). \tag{7.37}$$

Composition is not commutative, although in the limit of small systematics, or systematics which operate on disjoint observables, we can treat systematics as nearly commutative.

The complete kernel estimator with floating systematics has the form

$$\hat{P}(\mathbf{x}, \boldsymbol{\Delta}) = \frac{1}{\sum_{i=1}^{n} W(\mathbf{t}_i, \boldsymbol{\Delta})} \sum_{i=1}^{n} W(\mathbf{t}_i, \boldsymbol{\Delta}) \times \prod_{j=1}^{d} \frac{1}{h_{ij}} K \left(\frac{x_j - (S(\mathbf{t}_i, \boldsymbol{\Delta}))_j}{(C(\mathbf{t}_i, \mathbf{h}_i, \boldsymbol{\Delta}))_j} \right).$$

(7.38)

The normalization formula is also changed to

$$\int_{V} \hat{P}(x, \boldsymbol{\Delta}) \, d\mathbf{x} = \frac{1}{2^d \sum_{i=1}^{n} W(\mathbf{t}_i, \boldsymbol{\Delta})}$$
$$\sum_{i=1}^{n} W(\mathbf{t}_i, \boldsymbol{\Delta}) \times \prod_{j=1}^{d} \left[\mathrm{erf} \left(\frac{b_j - (S(\mathbf{t}_i, \boldsymbol{\Delta}))_j}{(C(\mathbf{t}_i, \mathbf{h}_i, \boldsymbol{\Delta}))_j} \right) \right.$$
$$\left. - \mathrm{erf} \left(\frac{a_j - (S(\mathbf{t}_i, \boldsymbol{\Delta}))_j}{(C(\mathbf{t}_i, \mathbf{h}_i, \boldsymbol{\Delta}))_j} \right) \right].$$

(7.39)

If we apply the *identity systematic*,

$$S(\mathbf{t}_i, \boldsymbol{\Delta}) \equiv \mathbf{t}_i \tag{7.40}$$

$$C(\mathbf{t}_i, \mathbf{h}_i, \boldsymbol{\Delta}) \equiv \mathbf{h}_i \tag{7.41}$$

$$W(\mathbf{t}_i, \boldsymbol{\Delta}) \equiv 1, \tag{7.42}$$

then Equation 7.38 reduces to Equation 7.19.

7.3   Examples

This section contains a selection of examples designed to illustrate the difference between binned histograms and kernel estimators, using the SNO Monte Carlo event sample.

7.3.1 1D

The simplest approach to signal extraction is to factorize the PDF into 1D components,

$$P(T_{\mathit{eff}}, R^3, \beta_{14}, \cos\theta_{\mathrm{sun}}) = P(T_{\mathit{eff}}) \times P(R^3) \times P(\beta_{14}) \times P(\cos\theta_{\mathrm{sun}}). \quad (7.43)$$

It is well-known that this factorization causes significant bias in the fit, but it is helpful to visualize the difference between binned and unbinned PDFs in 1D. Figure 7.12 shows the CC PDFs derived from 1000 Monte Carlo events using both methods, where the binned PDF has the bin sizes used in previous SNO analyses[91] While neither method produces an accurate PDF, the unbinned kernel estimator captures the basic features of the distributions and would be very usable in a fit.

Of course, we have substantially more than 1000 Monte Carlo events from which to build a PDF. Figure 7.13 shows the results with half of the events from the LETA data set (the other half being set aside for fake data sets). The two methods become indistinguishable. Clearly, the kernel estimator is overkill in 1D given the amount of Monte Carlo statistics we have.

7.3.2 2D

Once we move to 2D PDFs, the factorization is

$$P(T_{\mathit{eff}}, R^3, \beta_{14}, \cos\theta_{\mathrm{sun}}) = P(T_{\mathit{eff}}, \beta_{14}) \times P(R^3) \times P(\cos\theta_{\mathrm{sun}}), \quad (7.44)$$

which accounts for correlations between isotropy and energy. Figure 7.14 shows the 2D factor in the PDF as generated using 1000 NC events, and Figure
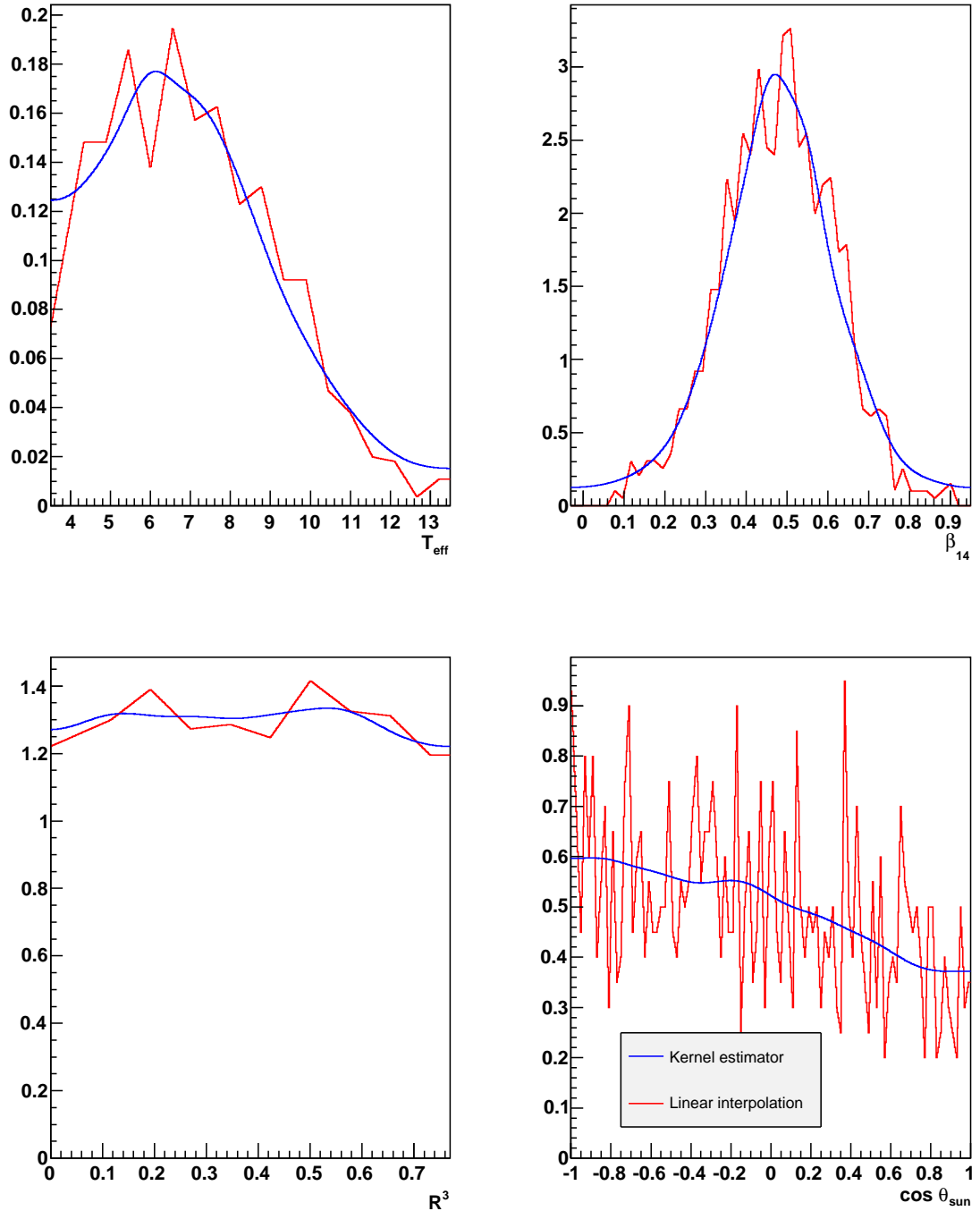
Figure 7.12: Binned and unbinned PDF estimates of the 1D components of the CC signal. Both PDFs are generated from the same 1000 CC events. The bin size used for the binned PDF is that of [91].
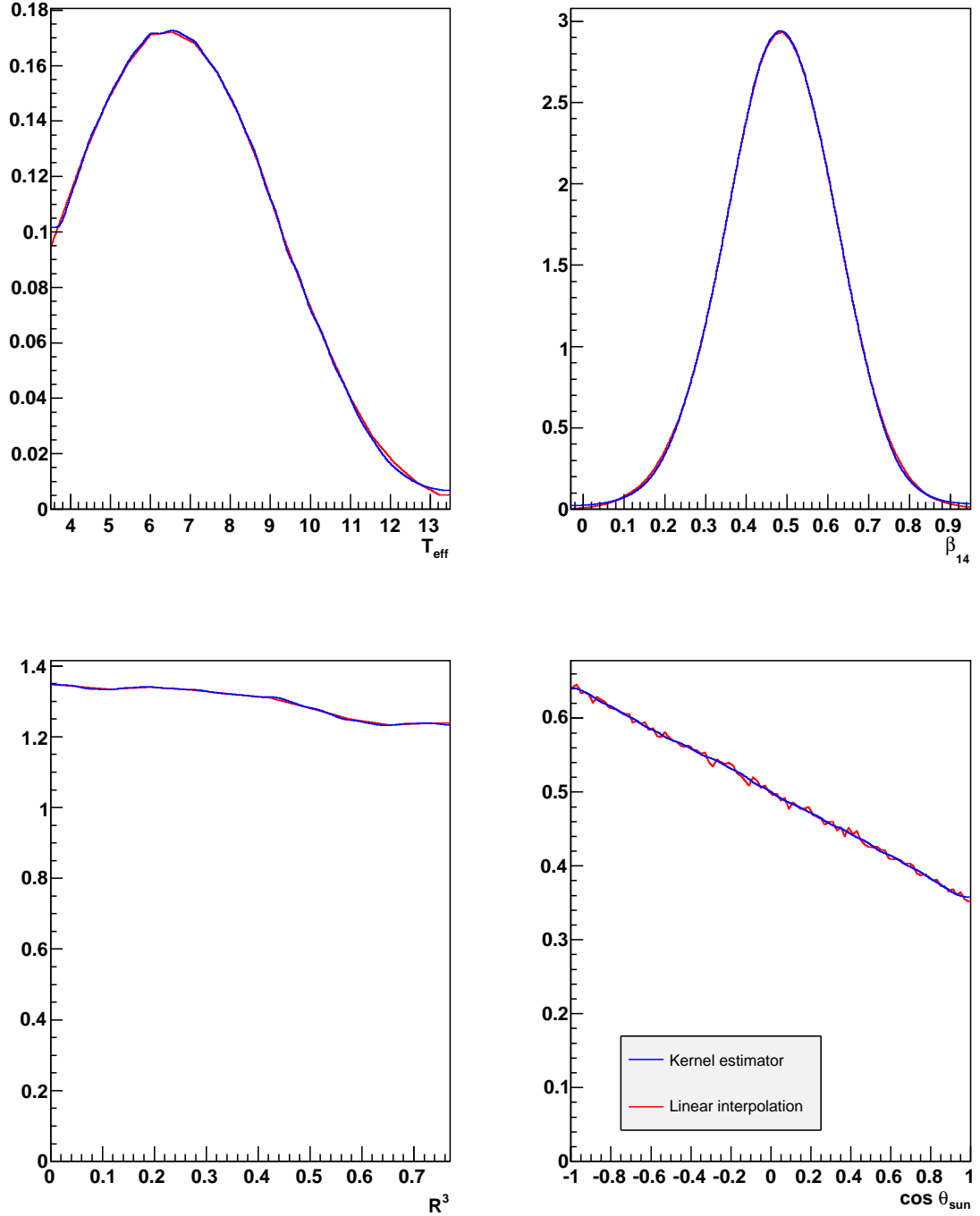
Figure 7.13: Binned and unbinned PDF estimates of the 1D components of the CC signal. Both PDFs are generated from 1.6 million events.

7.15 shows the same PDFs generated from half of the LETA Monte Carlo. Again, with half of the available Monte Carlo, there does not appear to much benefit to the kernel estimator. The contour plot looks slightly more smooth for the kernel estimator, but the longer energy axis in the plot makes the overestimation of the PDF at the low energy boundary easy to spot. At the other 3 boundaries, the PDF naturally decays to zero, so the overestimation problem is not present.

To really evaluate the local smoothness of the PDF, we should examine some 1D slices. Figure 7.16 shows a few samples where the smoothing benefits of the unbinned estimator are evident, although we can also see the shape distortion in the slices in the high tails of energy and β_{14}. The broadening of the distribution in the tails is a "feature," representing our reduced knowledge, though the size of the effect can be tuned with the ρ parameter in the bandwidth formula.

7.3.3 3D

In 3D, we can merge radius into one of the factors,

$$P(T_{\mathit{eff}}, R^3, \beta_{14}, \cos\theta_{\mathrm{sun}}) = P(T_{\mathit{eff}}, R^3, \beta_{14}) \times P(\cos\theta_{\mathrm{sun}}). \qquad (7.45)$$

In three dimensions, displaying anything but a PDF slice is nearly impossible. Figure 7.17 shows some example 1D slices through the distribution. The kernel estimator stands out here as much smoother and less susceptible to statistical fluctuations, although shape distortion is especially evident in the tails.

Figure 7.14: Binned and unbinned PDF estimates of the of the NC signal. Both PDFs are generated from the same 1000 NC events.

Kernel Estimator
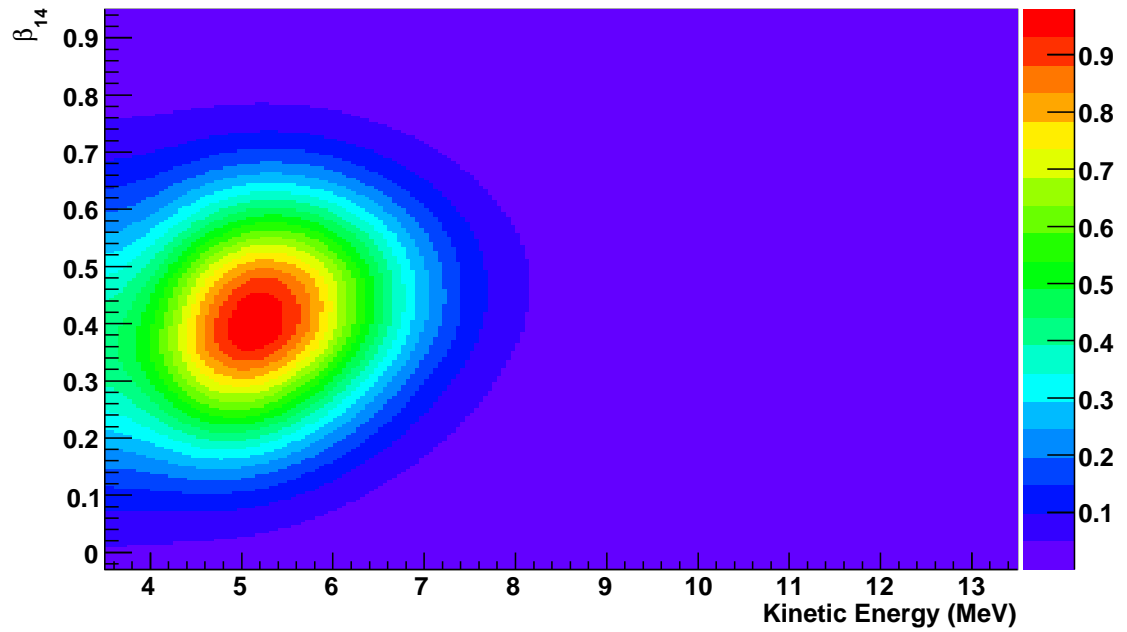
Linear Interpolation
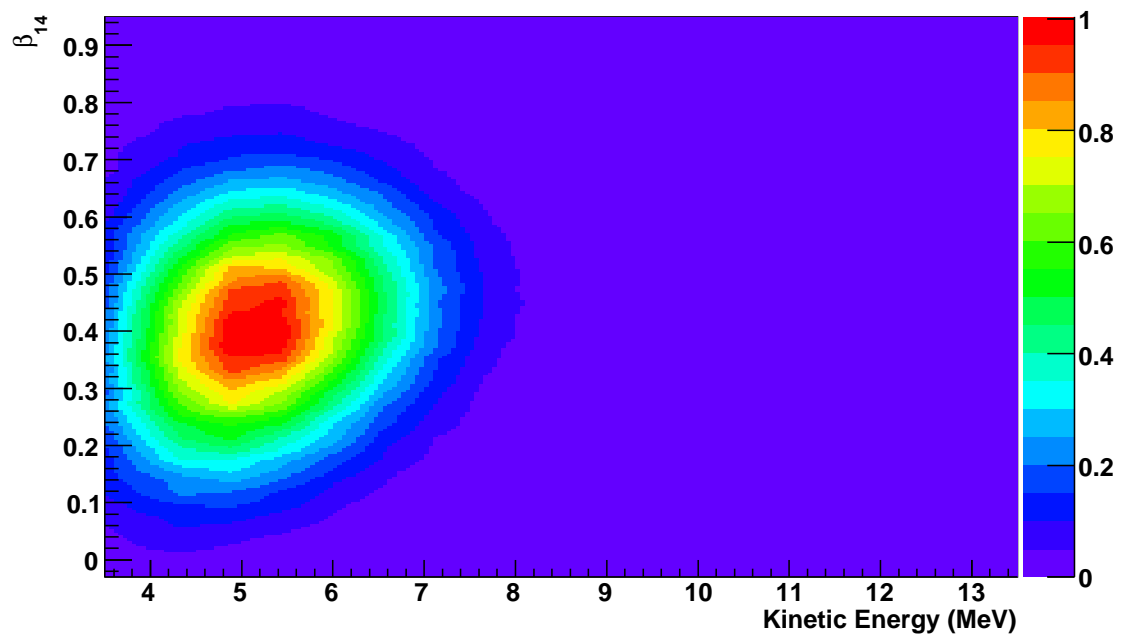
Figure 7.15: Binned and unbinned PDF estimates of the of the NC signal. Both PDFs are generated from 241,000 NC events.
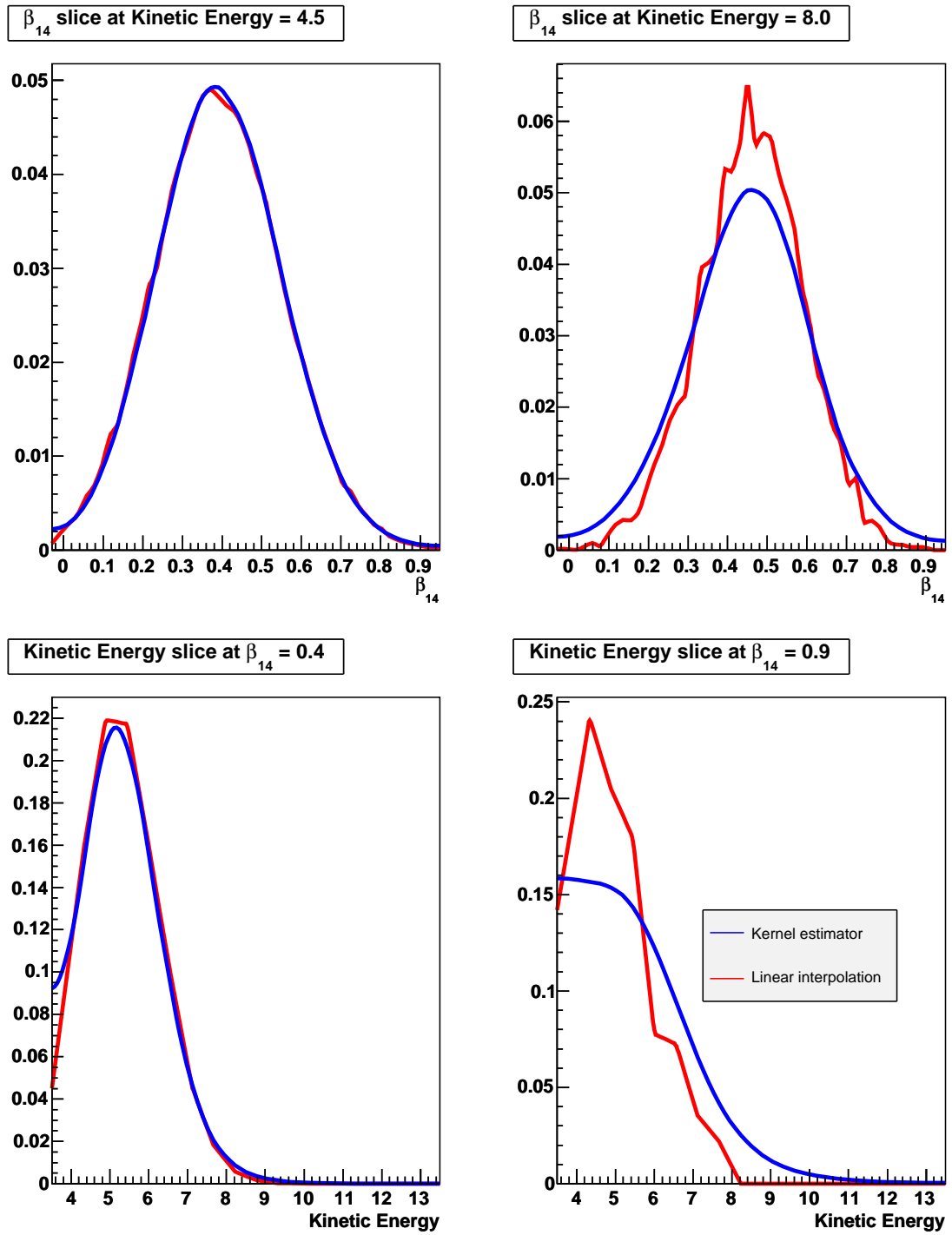
Figure 7.16: Slices of the binned and unbinned 2D PDF estimates shown in Figure 7.15.

201
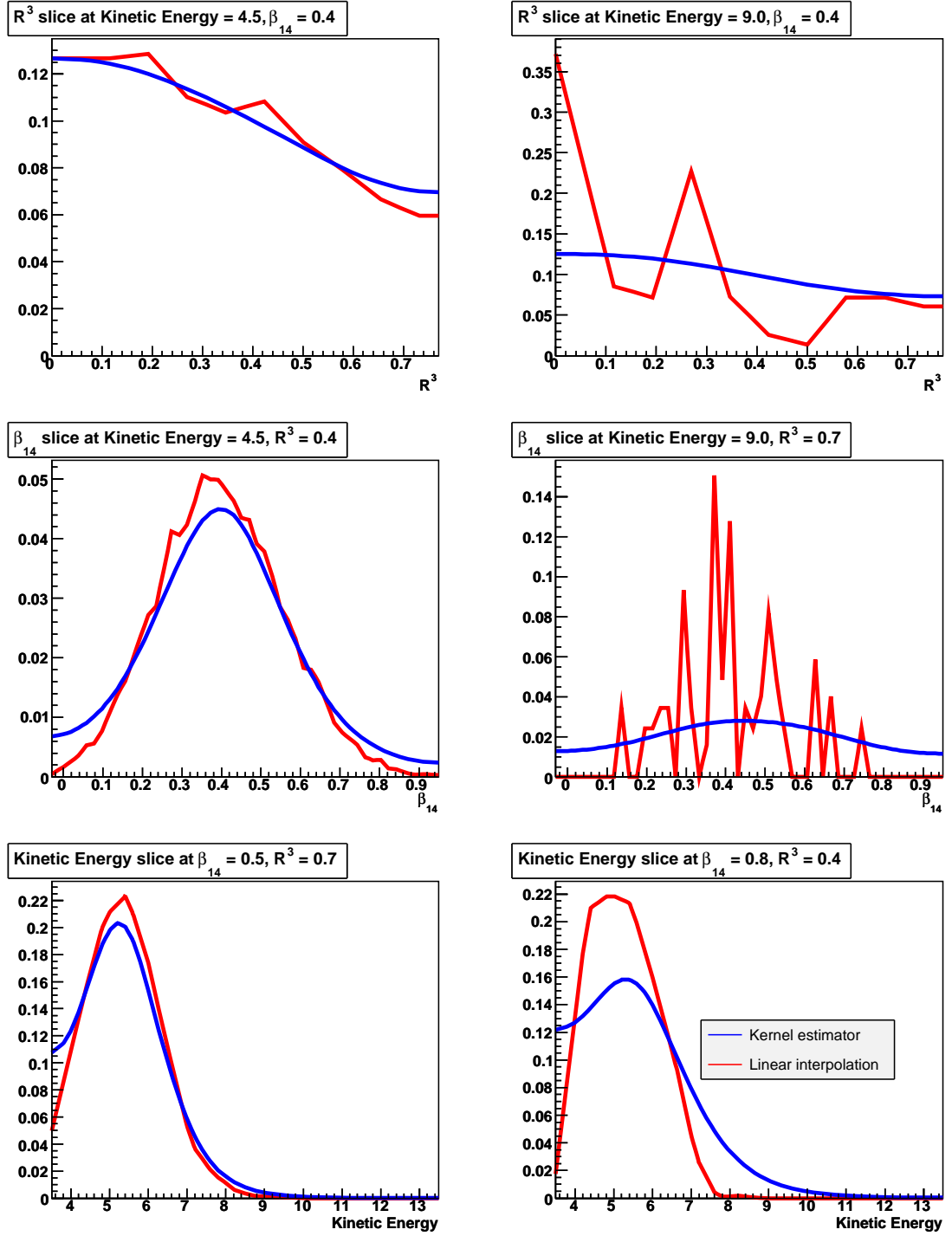
Figure 7.17: Slices of the binned and unbinned 3D PDF estimates generated from 240,000 NC events.

7.4 Summary

Kernel estimation of PDFs is a very effective technique to handle the problems inherent to binned PDFs generated from finite data samples. When the option is available, we always do better with more statistics, regardless of method. But by eliminating binning discontinuities, kernel estimators ensure our PDF is smooth and continuous for any amount of statistics.

There is an important caveat that cannot be emphasized enough: We have not gotten something for free. Instead we have traded the practical simplicity of histograms for the continuity and flexibility of the kernel estimator formalism.

However, once we have accepted and overcome the implementation hurdles, kernel estimation opens up some new possibilities with little additional work. Keeping all of the PDF samples in memory enables a direct approach to floating PDF systematics through the transformation, convolution, and weight functions. We will use this feature in the signal extraction chapter both to more consistently treat detector systematics, but also to perform more novel fits, such as directly fitting for the neutrino survival probability.

Chapter 8

Signal Extraction

Having cut, reconstructed, and kernel estimated the Monte Carlo events into PDFs, we are now prepared for the final step in the analysis chain: signal extraction. Formally, this is just a maximum likelihood fit (or set of fits), but there are a number of kinds of fits we can perform. In this chapter, we discuss the maximum likelihood formalism, different signal extraction options, the software used, and the verification procedure.

8.1   Goals

Ultimately, we want two different pieces of information from the SNO data set:

- A contour plot showing the confidence region in Δm_{21}^2 and $\tan^2 \theta_{12}$, taking into account matter effects in the Sun and the Earth.

- A ν_e spectrum and total ^{8}B neutrino flux which is independent of the oscillation model.

The model-dependent contours provide the strongest constraint on the neutrino mixing parameters, and are also needed by the physics community in global fits to all solar neutrino experiments and KamLAND. SNO provides the best measure of the mixing angle, θ_{12}, in this case. The ^{8}B spectrum is important for testing of new models without having to re-run signal extraction entirely from scratch.

8.2   Maximum Likelihood Method

Like previous SNO analyses, we use the generalized (or "extended") maximum likelihood method to estimate the contribution of several pure signals to a mixed sample of events. In this formalism, we have a data sample of N events, each event i described by a d-dimensional vector, $\mathbf{x}_i$ of observable quantities, such as energy or position. We postulate that this set of events is a combination of M different signals and no others. A "signal" in this context is a particular source of events, such as charged-current interactions. The term also applies to nuisance backgrounds, such as beta decays of ^{214}Bi. Both kinds of events are considered signals in our discussion of the maximum likelihood method. Each signal j has a probability density function (PDF), $P_j(\mathbf{x})$, which gives the joint distribution of observables for that particular kind of event. As a probability density, the P_j are normalized such that

$$\int P_j(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1. \tag{8.1}$$

We wish to estimate the expectation value[1] of the number, $\tilde{N}_j$ for signal

---

[1]Distinguishing between the expectation value of the number of events, $\tilde{N}_j$, and the actual number of events in the data set, N_j, is important to obtain the correct form of the

j, of each kind of event in the data set. The sum of these expectation values gives the expected total number of events

$$\tilde{N} \equiv \sum_{j=1}^{M} \tilde{N}_j. \tag{8.2}$$

The observed number of events in the data set, N, is random variable drawn from a Poisson distribution with mean $\tilde{N}$, with probability

$$\mathrm{Poisson}(N \,|\, \tilde{N}) = \frac{\tilde{N}^N e^{-\tilde{N}}}{N!}. \tag{8.3}$$

The probability density function which describes the distribution of observables for events in the experiment, assuming $(\tilde{N}_1, \tilde{N}_2, \ldots \tilde{N}_M)$ is

$$P(\mathbf{x} \,|\, \tilde{N}_1, \tilde{N}_2, \ldots \tilde{N}_M) = \sum_{j=1}^{M} \frac{\tilde{N}_j}{\tilde{N}} \times P_j(\mathbf{x}). \tag{8.4}$$

The $\tilde{N}_j/\tilde{N}$ factor gives the probability that a randomly chosen event is of a particular signal j, and P_j provides the probability density for the observables $\mathbf{x}$, assuming the event is of type j.

With Equations 8.3 and 8.4, we can construct the likelihood function for our data with free parameters $\tilde{N}_1, \tilde{N}_2, \ldots \tilde{N}_M$:

$$\mathcal{L}(\tilde{N}_1, \tilde{N}_2, \ldots \tilde{N}_M) = \frac{\tilde{N}^N e^{-\tilde{N}}}{N!} \prod_{i=1}^{N} P(\mathbf{x_i} | \tilde{N}_1, \tilde{N}_2, \ldots \tilde{N}_M). \tag{8.5}$$

If we expand P into a sum over P_j and pull the $1/\tilde{N}$ factor out of the sum

---

extended likelihood function.

and product, we get

$$\mathcal{L}(\tilde{N}_1, \tilde{N}_2, \ldots \tilde{N}_M) = \frac{\tilde{N}^N e^{-\tilde{N}}}{N!} \tilde{N}^{-N} \prod_{i=1}^{N} \left(\sum_{j=1}^{M} \tilde{N}_j \times P_j(\mathbf{x_i}) \right) \qquad (8.6)$$

$$= \frac{e^{-\tilde{N}}}{N!} \prod_{i=1}^{N} \left(\sum_{j=1}^{M} \tilde{N}_j \times P_j(\mathbf{x_i}) \right) \qquad (8.7)$$

The functional form of the $P_j(\mathbf{x})$ which go into the model can be very complicated, so generally we must maximize $\mathcal{L}$ with the aid of a computer using some kind of directed search algorithm. For reasons of floating point precision, the problem is usually formulated as a minimization of the negative log likelihood

$$-\log \mathcal{L}(\tilde{N}_1, \tilde{N}_2, \ldots \tilde{N}_M) = \tilde{N} - \log N! - \sum_{i=1}^{N} \log \left(\sum_{j=1}^{M} \tilde{N}_j \times P_j(\mathbf{x_i}) \right), \qquad (8.8)$$

which we will often call the *NLL* for short.

A constant offset will not change the location of the minimum of $-\log \mathcal{L}$, so we can drop the term that depends on N. If we also expand $\tilde{N}$ to show the sum over parameters, then we get the equation in the form which is actually used during minimization:

$$-\log \mathcal{L}(\tilde{N}_1, \tilde{N}_2, \ldots \tilde{N}_M) = \sum_{j=1}^{M} \tilde{N}_j - \sum_{i=1}^{N} \log \left(\sum_{j=1}^{M} \tilde{N}_j \times P_j(\mathbf{x_i}) \right). \qquad (8.9)$$

8.2.1   Multiple Data Sets

We need to combine two different data sets in this analysis: the D_2O and salt phases. The approach we use (also used by RooFit) is to augment the vector of

observables $\mathbf{x}_i$ for each event i with an additional discrete variable, p_i, which indicates the phase from which the event came[2]. At the same time, we let the j index in P_j range over PDFs from all phases, but modify the PDFs to accept this phase variable. If signal j does not correspond to the phase p_i, then $P_j(\mathbf{x}, p_i) = 0$, otherwise $P_j(\mathbf{x}, p_i)$ takes the normal value.

8.2.2 Correlated Event Rates

Equation 8.9 does not yet include our knowledge of correlations between the rates of some kinds of events. In particular, we assume that the time-averaged solar neutrino rate (accounting for the eccentricity of Earth's orbit) and energy spectrum are constant across both phases of SNO. Therefore not all of the $\tilde{N}_j$ parameters are independent. The true free parameters in our fit are event source rates, which can directly determine the number of events in multiple phases. With this approach, the $\tilde{N}_j$ are now functions of a potentially smaller list of parameters $r_{j'}$:

$$\tilde{N}_j(\mathbf{r}) = \sum_{j'=1}^{M'} \epsilon_{jj'} \times r_{j'}, \tag{8.10}$$

where $\epsilon_{jj'}$ is a matrix of detection efficiencies which converts from the units of the rate parameter to a number of events we would expect to see in the data set. Although this formula generically allows multiple physical sources (indexed by j') to map onto a single kind of event (indexed by j), our particular usage is much more restricted. We only allow one $\epsilon_{jj'}$ to be non-zero for any given j, restricting a particular kind of event to only have one source. This could be called the "one Sun hypothesis." Table 8.1 shows a concrete

---

[2][$D_2O = 1$, Salt $= 2$] or [D_2O day $= 1$, D_2O night $= 2$, Salt day $= 3$, Salt night $= 4$], for example.

	CC rate	ES rate	NC rate
D$_2$O phase CC events	2150 events/SSM	0	0
D$_2$O phase ES events	0	366 events/SSM	0
D$_2$O phase NC events	0	0	909 events/SSM
Salt phase CC events	3006 events/SSM	0	0
Salt phase ES events	0	513 events/SSM	0
Salt phase NC events	0	0	3369 events/SSM

Table 8.1: An example $\epsilon_{jj'}$ matrix for a joint, two-phase fit with only neutrino events, where CC, ES and NC are treated as independent sources. The row is indexed by j and the column is indexed by j'.

example of the efficiency matrix for a two-phase, neutrino-only fit. We derive all of these detection efficiencies using the Monte Carlo, which we have used to generate events with a known source rate, uniformly distributed over the detector livetime.

The choice of units for the rate parameters is arbitrary, so long as the units of $\epsilon_{jj'}$ cancel them out. For the neutrino signals, CC, ES and NC, we work in units where 1.0 equals the neutrino production rate of the undistorted BP2000 Standard Solar Model, as this is the model which was implemented in the Monte Carlo event generator. The source parameters are then mapped to a number of events in the two phases which are directly correlated with each other.

For radioactive backgrounds, we assume a distinct source term for each phase, and work in units of Becquerel. Certainly, liquid media like the D$_2$O and H$_2$O will not have the same radioactivity in each phase as the water is circulated out of the detector for purification, and can be exposed to a varying amount of radon contamination in the process. Solid media, like the acrylic vessel and the PMTs, in principle could have correlated source terms.

However, as we are not certain whether the dominant source of radiation is surface or bulk contamination, we have chosen to allow these sources terms to also be independent between phases. This allows for the possibility of leeching of contaminants from the surface into the water over time, where they would then be continuously removed by the water purification process.

The NLL in terms of these new rate parameters, also including the phase variable discussed previously, is

$$
- \log \mathcal{L}(\mathbf{r} = r_1, r_2, \ldots r_{M'}) = \sum_{j=1}^{M} \tilde{N}_j(\mathbf{r}) - \sum_{i=1}^{N} \log \left(\sum_{j=1}^{M} \tilde{N}_j(\mathbf{r}) \times P_j(\mathbf{x_i}, p_i) \right).
$$

(8.11)

8.2.3 Constraints

Some of the rate parameters, in particular the radioactive backgrounds, are constrained by external measurements. Both the D_2O and H_2O radioactivity were continuously monitored during the data collection period, providing a completely independent measure of the decay rates that we would like to include in the likelihood function. These constraint measurements are PDFs themselves, with the rate parameter representing a measure of the radioactivity. Taking our estimate to be Gaussian distribution, we can add a constraint term to the NLL,

$$
- \log \mathcal{L}_{\text{constraint}}(\mathbf{r}) = \frac{1}{2} \sum_{j'=1}^{M'} \frac{(r_{j'} - \bar{r}_{j'})^2}{\sigma_{\bar{r}_{j'}}^2},
$$

(8.12)

where $\bar{r}_{j'}$ is the value of the external measurement, and $\sigma_{r_{j'}}$ is the uncertainty on that measurement[3]. Parameters which are unconstrained have $\sigma_{r_{j'}} \to \infty$, which zeros out the contribution from that term in the constraint.

8.2.4 Systematic Uncertainties

The maximum likelihood method as described so far has no place to input uncertainties in the PDFs. Our Monte Carlo is not perfect, and each PDF has some kind of systematic uncertainty associated with in. We have quantified these uncertainties by specifying a set of *parameterized transformations* for our PDFs. The space of transformations maps each PDF to a family of related PDFs which contains the true PDF, in our belief, based upon the tests we have performed. The free parameters in the transformation functions have confidence intervals derived from calibration data which delineate our 68% confidence interval in the space of possible PDFs.

For a concrete example, consider the energy scale systematic uncertainty. Since energy scale depends on a product of many efficiencies, we parameterize our uncertainty on the energy scale with a multiplicative correction,

$$T'_{\text{eff}} = (1 + \alpha)T_{\text{eff}}, \tag{8.13}$$

where α is a free parameter which we have constrained to be $0^{+0.0044}_{-0.0095}$. In 1D, this means the true PDF with 68% confidence is $P((1+\alpha)T_{\text{eff}})$, for some value of α in that interval. Two ways to propagate these parameterized uncertainties to the rate parameters $\mathbf{r}$ are discussed in the next section.

---

[3]The constant term in the normalization of the Gaussian has been dropped in the NLL as it only produces a constant offset with no impact on the location or shape of the minimum.

Shift-Refit Systematics

The simplest approach, and that used by SNO in previous papers, is to prop-
agate PDF systematics using the "shift-and-refit" technique. The fit is per-
formed with all the systematic parameters at their nominal values (usually
zero). Then the fit is repeated with each systematic varied to its $+1\sigma$ extreme,
and then again at the -1σ extreme. Each systematic is done separately, ex-
cept in cases where the systematics are believed to be correlated, in which case
they are treated together as a single systematic. The change in the central
value of the rate parameters, $\mathbf{r}$, returned by the fit is taken as a systematic
uncertainty on the rate parameters. These differences (separate for the posi-
tive and negative direction) can be added together in quadrature to obtain a
composite systematic uncertainty.

While easy to compute, shift-and-refit systematics propagation is prob-
lematic for three reasons:

- It assumes that the central fit is at the minimum of the augmented
  likelihood space that includes both rate and systematics dimensions. It
  is possible that the 1σ transformed PDF is in fact a better fit to the data
  than the central value.

- It ignores any further constraint on the systematics that could be offered
  by the data itself. For example, the neutrino data has a huge sample of
  ^{214}Bi events, which give us an excellent measure of the energy resolution
  of the detector.

- It neglects correlations between systematics induced by the data in the
  fit. As another example, consider a systematic offset in the mean β_{14} of

neutrons. This should not be correlated between the D_2O and salt phases due to the completely different neutron capture signature. However, both systematics affect the NC rate, which we constrain by construction to be the same in both phases. That assumption indirectly links the β_{14} uncertainty in both phases.

These issues tend to inflate the final systematic uncertainties on the rates, and can even lead to counter-intuitive situations where the systematic uncertainty on a two-phase fit is the average of the separate single-phase fits[51]. One would normally expect the joint fit to be at least as good as either fit, if not better than both considered separately.

Nevertheless, we do use the shift-and-refit method for some systematics as a practical convenience, but we would like to treat a handful of the dominant systematics in a more consistent way.

Floating Systematics

A better approach to systematics includes the systematic parameters as additional dimensions in the likelihood space. The minimization algorithm is free to vary the systematic distortions right along with the event rate parameters. While we are not interested in the values of these nuisance parameters, allowing them to vary during the minimization of the NLL solves the problems outlined before:

- Uncertainty in the PDF shape is propagated directly to the neutrino flux parameters.

- The best minimum is found, even if it requires a non-zero systematic.

- The data itself may better constrain the systematics than our calibrations, leading to a smaller overall uncertainty.

- Correlations between rate and systematic parameters come directly from the likelihood space.

Floating systematics has been used before on SNO by J. Wilson[91], although she implemented it using interpolated histograms and inverse transformations applied to the data points, rather than direct transformations to the PDF. This scheme was difficult to apply as it required a computationally expensive numerical integration to renormalize each PDF after any systematic was changed. As a result, the method was limited to PDF factors which were 1D or 2D. PDFs with 3 dimensions could not be integrated quickly enough with sufficient accuracy for minimization.

An alternative signal extraction technique was applied in the third phase ("NCD phase") of SNO[92]. Rather than gradient descent, a Monte Carlo Markov Chain method was used to statistically sample the full likelihood space in both flux and systematic dimensions. The high dimensionality of the space requires the use of a large computing cluster to ensure a sufficient statistics for an analysis.

We solve the systematics problem by taking advantage of the flexibility of kernel estimators described in Section 7.2.5. The analytic integral allows the PDF to be renormalized very quickly with high accuracy as systematics

change. So, we can introduce systematic parameters $\boldsymbol{\Delta}$ into the likelihood,

$$
-\log \mathcal{L}(\mathbf{r}, \boldsymbol{\Delta}) = \sum_{j=1}^{M} \tilde{N}_j(\mathbf{r}, \boldsymbol{\Delta}) - \sum_{i=1}^{N} \log \left(\sum_{j=1}^{M} \tilde{N}_j(\mathbf{r}, \boldsymbol{\Delta}) \times P_j(\mathbf{x_i}, p_i, \boldsymbol{\Delta}) \right)
$$
$$
+ \frac{1}{2} \sum_{j'=1}^{M'} \frac{(r_{j'} - \bar{r}_{j'})^2}{\sigma_{\bar{r}_{j'}}^2} + \frac{1}{2} \sum_{k=1}^{s} \frac{(\Delta_k - \bar{\Delta}_k)^2}{\sigma_{\Delta_k}^2}. \quad (8.14)
$$

The likelihood now depends on rate and systematic parameters, and both kinds of parameters can have constraints. For example, an β_{14} scale systematic has a constraint based on observations of the reconstructed energy of calibration sources.

PDFs take a systematic argument, as first shown in Equation 7.38, but now the $\tilde{N}_j$ functions which convert rates to numbers of events also depend on the systematics. This is because the systematic uncertainties can affect the number of events we would expect to observe for a fixed source intensity, by moving events into or out of the analysis window. This can be parameterized by adding an extra renormalization term to Equation 8.10,

$$
\tilde{N}_j(\mathbf{r}, \boldsymbol{\Delta}) = \left(\sum_{j'=1}^{M'} \epsilon_{jj'} \times r_{j'} \right) \times \frac{|\{\mathbf{t}_i | \forall i \text{ s.t. } S(\mathbf{t}_i, \boldsymbol{\Delta}) \in V\}|}{|\{\mathbf{t}_i | \forall i \text{ s.t. } \mathbf{t}_i \in V\}|} \times \frac{\sum_{i=1}^{n} W(\mathbf{t}_i, \Delta)}{n},
$$
$$
(8.15)
$$

where V is the volume in the space of observables being analyzed, S and W are the systematic transformation and weighting functions from Section 7.2.5, and the $|\ |$ operation counts the number of elements in the set. The middle fraction counts what fraction of events are still in the analysis window after systematics, and the last fraction counts how much the detection efficiency has

been affected by reweighting[4].

8.2.5   Ensemble Testing

As a general principle, we do not want to develop and test our signal extraction method on real data. This can lead to bias, but it also makes troubleshooting problems more difficult. Instead, we would like to have an ensemble of "fake" data sets for testing which contain known event compositions and systematic distortions. Then we can extract the event rates and systematics for many fake data sets and determine if the fit procedure is correct.

We obtain these fake data sets from the Monte Carlo simulation. When doing this, it is very important to ensure that the event samples used to build the PDFs are *independent* from the events placed into the fake data samples. It is a common practice to generate fake data by sampling from the PDFs themselves, however we go one step further. Instead, we take the starting pool of Monte Carlo events and divide them into two, non-overlapping sets. One set is used to make PDFs, and the other set is used to construct fake data. The advantage to this approach is that it allows us to also test if shape distortions introduced into the PDF by the kernel estimation process have affected the overall fit. Events drawn from the kernel-estimated PDF would be automatically consistent with any distortion, and the test fits would look too good.

Two metrics are useful for testing the correctness of the fit: pull and bias. We start with N_{set} fake data sets, each containing an average $\bar{r}_{j'}$ signal for

---

[4]Reweighting systematics are used to account for uncertainties in efficiencies or cross sections, such as deuteron photodisintegration uncertainties or neutron capture efficiencies. Variations in these quantities actually change the number of expected events.

all signal types j, but with the actual number of events randomly drawn from a Poisson distribution around that mean. Each maximum likelihood fit returns an optimal value for the parameters $r_1, r_2, \ldots r_{M'}$, as well as an uncertainty for each parameter $\sigma_{r_1}, \sigma_{r_2}, \ldots \sigma_{r_{M'}}$. The *pull* for parameter j' is

$$\text{Pull}(r_{j'}) = \frac{r_{j'} - \bar{r}_{j'}}{\sigma_{r_{j'}}} \qquad (8.16)$$

and the bias is

$$\text{Bias}(r_{j'}) = \frac{r_{j'} - \bar{r}_{j'}}{\bar{r}_{j'}}. \qquad (8.17)$$

The distribution of the pulls for many fits should tend toward a normal distribution with a mean of 0 and a width 1. A correct pull distribution demonstrates that the fit is unbiased, and that the fit uncertainty predicts the variation that would be observed if the experiment were repeated many times. A pull distribution with width significantly different than 1 may indicate that the error-estimation procedure is not accurate. For example, if the negative log-likelihood space is not sufficiently parabolic, the errors computed by the HESSE procedure in MINUIT will be inaccurate. The pull distribution can indicate this, suggesting that a more rigorous error estimation procedure, such as MINOS, is required. Note that the inclusion of constraints in the likelihood when there are strongly correlated parameters can actually cause the width of the pull distribution to be less than 1. Appendix B discusses the origin of this problem, as well as a solution.

The distribution of the bias does not have any particular shape, but the mean bias should also be zero for unbiased fits (as the name would suggest). While pull is more statistically meaningful, the bias is more intuitive since it shows how far the average fit is from the true value as a fraction or percentage.

If the average bias is much less than the typical relative uncertainty $\sigma_{r_{j'}}/r_{j'}$ returned by the fitter, then we can safely ignore the bias.

8.3   sigex.py

Sigex.py is a signal extraction program that was written by the author to use kernel estimation and floating systematics in the maximum likelihood formalism described above. The program is implemented in the interpreted programming language, Python[93], with performance-critical elements in C++. PyROOT[94] provides the bridge between the two languages. Much of the C++ portion of the program leverages the RooFit[95] framework. RooFit was developed by the BaBar collaboration for a variety of maximum likelihood fitting tasks, and has been included in the main ROOT[96] distribution since version 5.12.

Figure 8.1 shows the main stages of the signal extraction process. All stages are controlled by a configuration file which describes the type of fit to be performed, including the observables, signals, constraints, systematics, and transformations. After reading the configuration file, sigex.py goes through the following process:

- Load Monte Carlo PDF events from disk.

- Apply PDF transformations: Shift-and-refit systematics are implemented at this stage. Monte Carlo corrections, like those described in Section 8.5 are also applied here.

- Cut to PDF window: The events are chopped to the relevant observable window. To avoid boundary problems, the PDF is constructed from

218

events over a wider observable range than the data.

- Compute Efficiency Matrix: The number of events in each PDF is counted, and combined with the source rate used in the Monte Carlo to generate the $\epsilon_{jj'}$ matrix that maps signal rates to numbers of events. A list of events is kept in memory so that $\epsilon_{jj'}$ can be recomputed as systematics vary, and events move into and out of the analysis window.

- Construct Kernel Estimator: The PDF event lists are fed into a kernel estimator function which computes the bandwidth for each event and mirrors events across PDF boundaries. This data is then passed to the desired kernel estimator implementation code, either CPU or graphics card. If the graphics card is used, the events are loaded into graphics memory at this point.

- Load data set from disk.

- Apply data transformations: This is only used for systematic tests where a fake data set is given a deliberate systematic offset to see if the fit can reproduce it.

- Cut data to analysis window: Data events are discarded whose observables fall outside the specified range. This range is smaller than the PDF range.

- Build likelihood function: The PDFs and data set are brought together, and a special fast negative-log likelihood data structure is built.

- Perform minimization: The minimization strategy is specified by the user, and can include gradient descent or simulated annealing.
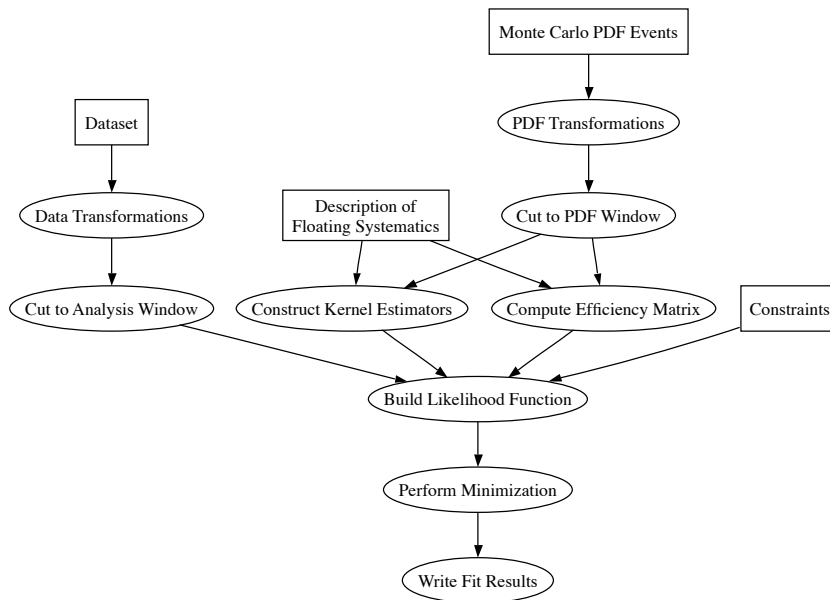
Figure 8.1: Data flow for sigex.py application.

- Write fit results: The best fit parameters, uncertainties, covariance matrix, and fit configuration are written to an output file. The file can be reloaded later to resume the fit, for example to estimate errors with a more accurate algorithm.

8.3.1 Minimization Algorithm

The actual minimization of the NLL function is performed by TMinuit, a direct C++ transliteration of the original FORTRAN version of MINUIT[5]. MINUIT uses a variable-metric method with inexact line search[97] to descend toward the minimum, using the local function gradient as a guide. Sigex.py also sup-

_____

[5]TMinuit should not be confused with Minuit2, a complete re-implementation of MINUIT in C++ by the original MINUIT author.

ports simulated annealing[6] and a hybrid minimization which combines gradient descent with simulated annealing. If the likelihood has multiple local minima in some dimensions, those can be searched using simulated annealing where each iteration performs a complete gradient descent in the other dimensions. This method can be much faster than simulated annealing in all dimensions.

8.3.2 Kernel Estimation

Although RooFit includes a kernel estimation implementation, it was found to be far too slow and limited to use for these fits. An entirely new kernel estimation implementation was written with the features discussed in Chapter 7, including the ability to float systematics. Both CPU and graphics-card accelerated versions of the algorithms are included, allowing the software to run on a variety of systems. The graphics-card version of the code was used for all fits in this analysis.

One additional feature of this kernel estimator implementation is the inclusion of *event attributes*. Event attributes are like observables, but they do not appear in the dimensions of any PDF. They are extra bits of information that can be used when floating systematics to determine how to perturb the event. For example, each Monte Carlo CC and ES event in the PDF can be tagged with the energy of the neutrino that produced it (as computed by SNOMAN). In Section 8.9, we use the E_ν attribute to reweight the CC and ES PDFs by the electron neutrino survival probability on the fly during the fit. This allows θ_{12} and Δm_{12}^2 to be treated as systematic (though hardly nuisance) parameters in the fit via an event weighting systematic.

---

[6]From the MathMore library in ROOT.

8.3.3   Fast NLL Computation

As part of the minimization process, MINUIT will stop after each iteration and estimate the local partial derivatives by taking small steps around the current best fit point. Most of the dimensions in the likelihood space are event rates rather than systematics. Changing a rate does not change the shape of the PDF, and that provides an opportunity for optimization. If we cache the PDF evaluations from the previous iteration, we can reevaluate the likelihood very quickly. In particular, we precompute a matrix

$$F_{ij} = P_j(\mathbf{x}_i) \tag{8.18}$$

which contains the value of every PDF at every point in the data set in the model. This table can easily be less than 2 MB in size for our fits, enabling most or all of it to fit into the fast on-chip cache of current CPUs. Looping over the matrix to compute

$$-\log \mathcal{L}_{partial} = -\sum_{i=1}^{N} \log \left(\sum_{j=1}^{M} \tilde{N}_j(\mathbf{r}, \boldsymbol{\Delta}) F_{ij} \right) \tag{8.19}$$

can be done very fast. It avoids unnecessary calls to the slow kernel estimation code, and even turns out to be 100 times faster than the built-in RooFit caching[7].

Whenever a systematic parameter changes, F_{ij} must be recomputed. We keep track of the dependencies between the PDFs and the systematics, so that if a systematic parameter changes, we only update the columns of F_{ij}

---

[7]RooFit's caching system is designed for arbitrary PDFs, whereas the fast matrix optimization only works for our particular likelihoods.

Observable	Min	Max
T_{eff}	3.5 MeV	12 MeV
β_{14}	-0.12	0.95
R^3	0	$(550 \text{ cm}/600 \text{ cm})^3$
$\cos\theta_\odot$	-1	1

Table 8.2: Analysis window for observables used in all fits.

which have changed. In a fit with no free systematic parameters, F_{ij} only needs to be computed once, and the minimization can be completed in about 60 seconds. This is critical for the generation of LMA contours described in Section 8.9.

8.4 Fit Observables

The fits described in Sections 8.8, 8.9 8.10 share the same Monte Carlo events for PDF construction, and use the same four primary observables: T_{eff}, β_{14}, R^3, and $\cos\theta_\odot$. Table 8.2 shows the allowed range for these observables in the data set. These observables all have corrections and systematic uncertainties associated with them which we review in the next section.

8.5 Corrections to Observables

Based on studies of the calibration sources, we have developed corrections for both reconstructed energy (T_{eff}) isotropy (β_{14}) and position. These justification for the energy corrections are discussed in Sections 5.3. Some corrections are applied to just data, or just Monte Carlo, or in a few cases, both, with the overall goal of bringing data and Monte Carlo into better agreement.

8.5.1 D$_2$O Data Corrections

The z position is first corrected so that z', the fixed position, is

$$z' = z - 5.0 \text{ cm.} \tag{8.20}$$

Then we compute the time drift correction for energy,

$$C(JD) = \begin{cases} 1.00371 & \text{if } JD < 9363 \\ 1.2308 - 2.4254 \cdot 10^{-5} \times JD & \text{otherwise} \end{cases}, \tag{8.21}$$

where JD is the SNO Julian date of the event, and the 2D spatial correction

$$C(u_z, z') = \begin{cases} 1.00986 + 1.59412 \cdot 10^{-2}\, u_z + 2.25355 \cdot 10^{-5}\, z' \\ \quad -1.622782 \cdot 10^{-5}\, u_z z' \\ \quad +1.99929 \cdot 10^{-3}\, u_z^2 - 3.03906 \cdot 10^{-8}\, z'^2 \qquad \text{if JD} < 9363 \\ \\ 1.01028 + 2.1852 \cdot 10^{-2}\, u_z + 2.49459 \cdot 10^{-5}\, z' \\ \quad -2.46175 \cdot 10^{-5}\, u_z z' \\ \quad +1.24998 \cdot 10^{-3}\, u_z^2 - 6.24735 \cdot 10^{-8}\, z'^2 \qquad \text{otherwise} \end{cases} \tag{8.22}$$

where u_z is the direction cosine along the z-axis. Using these two factors, the reconstructed energy for D$_2$O data is then

$$T'_{data} = -0.10872 + 1.0277 \left(\frac{T_{data}}{0.9968C(u_z, z')C(JD)} \right) - 0.0012247 \left(\frac{T_{data}}{0.9968C(u_z, z')C(JD)} \right)^2. \tag{8.23}$$

The quadratic polynomial compensates for inaccuracies in the reconstruction at high energy as multiple photons become more likely to strike the same

PMT.

D$_2$O Monte Carlo Corrections

The energy of Monte Carlo events in the D$_2$O phase has a radial correction factor,

$$C(R^3) = 1.01159 - 0.0389943(R^3) + 0.0250065(R^3)^2, \qquad (8.24)$$

where R^3 is the unitless quantity (radius/600 cm)^3. This gives the energy for D$_2$O Monte Carlo,

$$T'_{MC} = -0.10872 + 1.0277 \left(\frac{T_{MC}}{C(\rho)} \right) - 0.0012247 \left(\frac{T_{MC}}{C(\rho)} \right)^2 \qquad (8.25)$$

Note that the multi-photon polynomial is the same as for data.

The D$_2$O Monte Carlo also has an isotropy correction,

$$\beta'_{14} = 0.9919 \times \beta_{14}, \qquad (8.26)$$

which is applied to all Monte Carlo events, regardless of signal type, except for the analytic PMT model.

Salt Data Corrections

Similar to D$_2$O , salt data has a position correction $z' = z - 5.0$ cm. However, the salt phase has no time dependence correction, as the energy reconstruction tracks the detector variation directly. There is still a 2D spatial correction

225

factor

$$
\begin{aligned}
C(u_z, z') &= 1.00320 + 2.19773 \cdot 10^{-2}\, u_z + 4.01820 \cdot 10^{-5}\, z' \\
&\quad -2.35837 \cdot 10^{-5}\, u_z z' + 1.06100 \cdot 10^{-3}\, u_z^2,
\end{aligned}
\tag{8.27}
$$

giving a corrected energy of

$$
T'_{data} = -0.11492 + 1.0276 \left(\frac{T_{data}}{0.9979 C(u_z,\, z')} \right) - 0.0012282 \left(\frac{T_{data}}{0.9979 C(u_z,\, z')} \right)^2.
\tag{8.28}
$$

Salt Monte Carlo Corrections

Salt phase Monte Carlo only has one energy correction for the multi-photon effect (same as data):

$$
T'_{MC} = -0.11492 + 1.0276\, T_{MC} - 0.0012282\, T_{MC}^2
\tag{8.29}
$$

Salt Monte Carlo also has an isotropy correction which only applies to neutron events,

$$
\beta'_{14} = 0.9856 \times \beta_{14} + \mathrm{Gauss}(0, 0.015)
\tag{8.30}
$$

where $\mathrm{Gauss}(m, \sigma)$ is a random value from a Gaussian distribution of mean m and width σ which is independently drawn for every event. This correction fixes both the mean and the width of the isotropy distribution for neutrons to match calibration data. It is important to note that this correction not only applies to pure neutron signals, such as NC and AV neutrons, but also to photodisintegration neutrons in PDFs for radioactive backgrounds, like ^{214}Bi

226

and ^{208}Tl.

8.6   Normalization Corrections

In addition to corrections to PDF observables, there are also overall normalization corrections required to match up the event acceptance between data and Monte Carlo. These include:

- hep neutrinos. Less than 1% of neutrinos in the ^{8}B energy range are hep neutrinos. We subtract these out by fixing hep neutrino PDFs at 1 SSM in the maximum likelihood fit.

- Conversion of BP2000 SSM to flux units. BP2000, used in generating the Monte Carlo, predicts $5.15 \times 10^{-6}\,\mathrm{cm^{-2}s^{-1}}$. The results from signal extraction should therefore be divided by 1.105 to convert to units of BS2005. This scale factor is not needed when reporting neutrino flux in the physical units, $\mathrm{cm^{-2}s^{-1}}$.

- Aborted events in the simulation. A small number of events are aborted by the Monte Carlo due to tracking errors while propagating photons through the detector. Compensating for these failures requires reweighting events by $W = 1 + 0.0006238 T_{\mathrm{eff}}$. Due to a performance problem with sigex.py, this scaling was not included in the fits, but instead had to be propagated as a systematic. It makes a negligible contribution to the overall uncertainties.

- Target density. The number of deuterons and electrons in the detector is estimated to be slightly different than the simulation, changing the

overall interaction rate. CC and NC results must be divided by 1.0122, and ES by 1.0131[81]. These are applied after the fitting process.

- Additional target nuclei. In the salt phase, neutrinos can also interact with sodium and chlorine nuclei. This only occurs 0.02% of the time[81], so a correction is unnecessary.

- Livetime. The burst cuts used to remove spallation events from muons also generate some deadtime relative to the Monte Carlo. These effects are included directly in the generation of the efficiency matrix $\epsilon_{jj'}$ by scaling D_2O entries by 0.986 and salt entries by 0.989.

- Sacrifice. The instrumental and high level cuts also induce some sacrifice of neutrino events. Similar to livetime, $\epsilon_{jj'}$ are scaled by 0.9924 for D_2O phase events, 0.9930 for salt phase "electron-like" events, and 0.9954 for salt phase neutrons[81]. Sacrifice also has its own systematic uncertainty associated with it.

8.7   Systematics

Tables 8.3, 8.4, 8.5, 8.6 list each of these systematics, whether they are correlated between phases, and the mathematical form of their application to the PDF. The analytic PMT model has its own systematic uncertainties, described in Section 6.3.5, which are summarized in Table 8.7. Finally, there are a handful of reweighting systematics which relate to efficiencies in Table 8.8.

All of the systematics have been parameterized with one or two free parameters. The tables show the 1σ confidence interval for each systematic parameter in each phase. If a systematic is correlated between phases (the

"Corr?" column), then the $+1\sigma$ and -1σ displacements should be applied to both phases in the same direction simultaneously. If systematic is uncorrelated between the phases, then the systematics can be applied separately and independently. The "Application" column shows the mathematical form of the transformation. As a shorthand, systematics can use the $\star$ operator to indicate convolution. These can be applied either as a convolution operation on the kernel estimator, or by randomly jittering PDF samples by a Gaussian displacement during the load stage, before constructing the PDF. To avoid the need for deconvolution to test the negative side of a resolution systematic, instead we test only the positive direction, but take the displacement of the fit parameters as a two-sided uncertainty.

8.8 Unconstrained T_{eff} Fit

The unconstrained T_{eff} fit has been traditionally how SNO has obtained spectral information from the neutrino data. The CC and ES PDFs, which carry energy information about the neutrinos, are chopped up into 0.5 MeV segments. The CC and ES rate for each energy bin are allowed to freely vary without constraint from neighboring bins. Other PDFs, like NC and backgrounds, are not segmented, and retain their energy shape.

There are a number of drawbacks to the unconstrained T_{eff}. Firstly, it is difficult to interpret. The CC spectrum obtained is not the neutrino energy spectrum, nor is it even the energy spectrum of CC electrons. Adjacent CC bins are correlated through the smearing of neutrino energy, first by the CC differential cross-section, and again by the detector energy resolution. A mono-energetic electron neutrino which interacts via the charged-current reaction has

| | Δ parameter | | | | Corr? | Application |
Type	D$_2$O+	D$_2$O–	Salt+	Salt–		
Scale	0.444	-0.953	0.444	-0.953	Y	$T' = T \times (1 + \Delta/100)$
Scale	0.37	-0.45	0.2	-0.2	N	$T' = T \times (1 + \Delta/100)$
Resolution	0.181	0	0.199	0	N	$T' = T \star \mathcal{N}(0, \Delta)$
Non-linearity		-0.0137		-0.0137	Y	$T' = T \times (1 + \Delta(T - 5.05)/(19.0 - 5.05))$

Table 8.3: Reconstructed energy (T_{eff}) systematic uncertainties. See Section 8.7 for an explanation of the columns. Note that energy resolution is taken as ranging uniformly from 0 to the limit given above. No Gaussian distribution of the parameter is assumed.

Type	Δ parameter				Corr?	Application
	D_2O+	D_2O-	Salt+	Salt−		
Scale (e^-, D_2O n)	0.42	-0.42	0.24	-0.24	Y	$\beta'_{14} = \beta_{14} \times (1 + \Delta/100)$
Scale (salt n)			0.38	-0.22	N	$\beta'_{14} = \beta_{14} \times (1 + \Delta/100)$
Width (e^-, D_2O n)	0.42	-0.42	0.54	-0.54	N	$\beta'_{14} = \beta_{14} + (\beta_{14} - \beta_{14}) \times (\Delta/100)$
Width (salt n)			0.0045		N	$\beta'_{14} = \beta_{14} \star \mathcal{N}(0, \Delta)$
Energy dependence	0.00069	-0.00069	0.00069	-0.00069	Y	$\beta'_{14} = \beta_{14} \times (1 + (0.002756 + \Delta) \times (T_{\text{eff}} - 5.589))$

Table 8.4: Isotropy (β_{14}) systematic uncertainties.

Type	Δ parameter				Corr?	Application
	D_2O+	D_2O-	Salt+	Salt–		
Scale: x	0.08	-0.57	0.04	-0.34	N	$x' = x \times (1 + \Delta/100)$
Scale: y	0.1	-0.52	0.04	-0.26	N	$y' = y \times (1 + \Delta/100)$
Scale: z	0.5	-0.5	0.07	-0.59	N	$z' = z \times (1 + \Delta/100)$
Offset: x	1.15	-0.13	0.62	-0.07	N	$x' = x + \Delta$
Offset: y	2.87	-0.17	2.29	-0.09	N	$y' = y + \Delta$
Offset: z	2.58	-0.15	3.11	-0.16	N	$z' = z + \Delta$
Resolution: x	3.3	N/A	3.1	N/A	N	$x' = x \star \mathcal{N}(0, \Delta)$
Resolution: y	2.2	N/A	3.4	N/A	N	$y' = y \star \mathcal{N}(0, \Delta)$
Resolution: z	1.5	N/A	5.3	N/A	N	$z' = z \star \mathcal{N}(0, \Delta)$
Edep Fid Vol	0.85	-0.49	0.41	-0.48	N	$W = 1 + (\Delta/100) * (T - 5.05)$

Table 8.5: Position ($R^3 = (x^2 + y^2 + z^2)^3/(600\text{cm})^3$) systematic uncertainties. The W variable represents the weight of the event.

	Δ parameter				
Type	D_2O+	D_2O-	Salt+	Salt–	Corr?
Resolution	0.11	-0.11	0.11	-0.11	N
	Application: $(\cos\theta_\odot)' = 1 + (\cos\theta_\odot - 1)(1 + \Delta/100)$				

Table 8.6: $\cos\theta_\odot$ systematic uncertainties. Note that if the transformation moves the observable outside the range of $[-1, 1]$, the observable is given a random value from that interval.

	Δ parameter				
Type	D_2O+	D_2O-	Salt+	Salt–	Corr?
T exp	0.0024	-0.0024	0.008	-0.008	N
R^3 exp	0.04	-0.04	0.04	-0.04	Y
R^3 flat	1.43	-1.43	0.66	-0.66	Y
β_{14} intercept	0.04	-0.04	0.04	-0.04	Y
β_{14} slope	0.04	-0.04	0.04	-0.04	Y
β_{14} width	0.04	-0.04	0.04	-0.04	Y

Table 8.7: Systematic uncertainties to PMT analytic model parameters. See Section 6.3.5 for more discussion.

	Δ parameter					
Type	D_2O+	D_2O-	Salt+	Salt–	Corr?	Application
Neutron Capture	0.0024	-0.0024	0.008	-0.008	N	$W = 1 + \Delta$
Photodisintegration	0.04	-0.04	0.04	-0.04	Y	$W = 1 + \Delta$

Table 8.8: Other systematic uncertainties. The W variable indicates an reweighting of PDF samples. Neutron capture uncertainties apply to all neutron events (including NC), but photodisintegration uncertainties only apply to neutron events in ^{214}Bi and ^{208}Tl PDFs.

a width of 1.4 MeV in T_{eff}, which spans three energy bins. The unconstrained fit makes no use of this information, and these correlations do not appear in the covariance matrix of the CC bins.

The biggest challenge in the unconstrained T_{eff} fit is clearly separating CC events from ^{214}Bi in the D_2O volume. Both kinds of events are characterized by a single electron, so isotropy cannot distinguish them. CC events and ^{214}Bi are distributed uniformly throughout the D_2O volume, so radius cannot separate them either. The only handle left to prevent total covariance between ^{214}Bi and CC is $\cos\theta_\odot$.

Since the energy shape of CC is free to vary, small PDF biases can cause adjacent CC bins to move in a correlated way to mimic a background. Early on in the analysis, we used a CC PDF factorization within each energy bin of $P(\beta_{14}, R^3) \times P(\cos\theta_\odot)$ and found bias in the lowest two CC energy bins (3.5–4.0 and 4.0–4.5MeV) when performing ensemble tests on fake data sets.

The solution was to merge the 2D and 1D factors into a full 3D PDF in order to capture the 3-way correlation between β_{14}, R^3 and $\cos\theta_\odot$. Figure 8.2 shows the negatively sloping $\cos\theta_\odot$ for CC events in three different energy/radius regions for high β_{14} (less isotropic) and low β_{14} (more isotropic events). At low energies, events near the center of the detector with high β_{14} have a steeper $\cos\theta_\odot$ slope, better separating the CC from the internal backgrounds. If we look at higher radius or at higher energy, the effect goes away. Only a fully 3D PDF will capture this effect, so we have adopted it in our PDF parameterization, shown in Table 8.9, for the unconstrained T_{eff} fit.
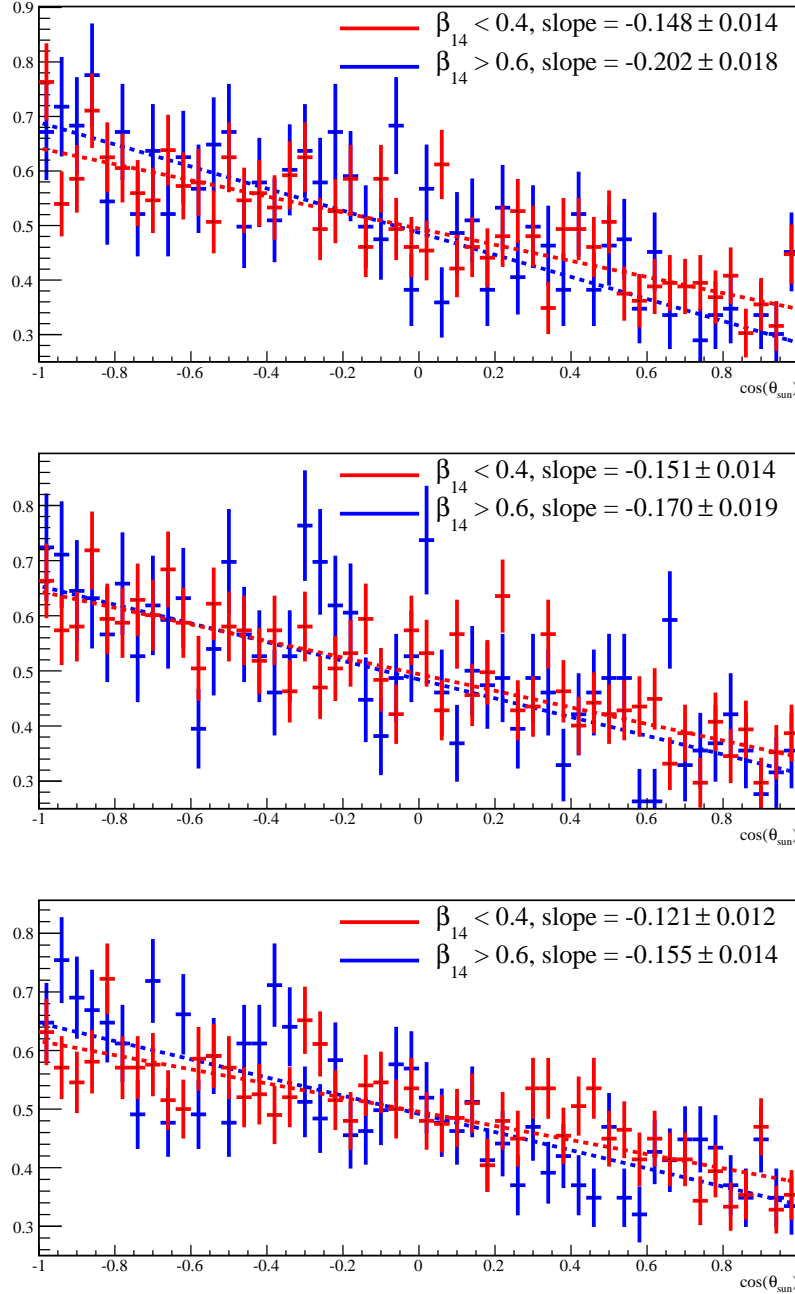
Figure 8.2: $\cos\theta_\odot$ distribution for Monte Carlo CC events in different energy and radius regimes. The top plot shows events with $R < 250$ cm and $3.5 < T_{\text{eff}} < 4.0$ MeV, which have the best separation from a flat ^{214}Bi background at high β_{14}. The middle plot shows the outer range $532 < R < 550$ cm and $3.5 < T_{\text{eff}} < 4.0$. The bottom plot shows higher energies in the central region, with $R < 250$ cm and $5.5 < T_{\text{eff}} < 6.0$. Both higher radius and higher energy events have worse background separation.

Signal	PDF factorization
CC, ES bins	$\text{Window}_k(T_{\text{eff}}) \times P(\beta_{14}, R^3, \cos\theta_\odot)$
NC	$P(T_{\text{eff}}, \beta_{14}, R^3) \times P(\cos\theta_\odot)$
PMT	Analytic Model (see Chapter 6)
Other backgrounds	$P(T_{\text{eff}}, \beta_{14}, R^3) \times \text{Flat}(\cos\theta_\odot)$

Table 8.9: PDF factorization used in unconstrained T_{eff} fit. $\text{Window}_i(T_{\text{eff}})$ is a boxcar function of width 0.5 MeV spanning energy bin k and $\text{Flat}(\cos\theta_\odot)$ is a constant spanning the entire -1 to 1 observable range.

8.8.1 Ensemble Tests

The first stage of verification is an ensemble test on the 15 fake data sets we generated from a separate pool of Monte Carlo events. Figures 8.3, 8.4, and 8.5 show the pull distributions for the CC, ES, and NC+backgrounds, respectively. There are no major problems with the neutrino signals, although some of the constrained backgrounds show a narrow pull distribution, while the unconstrained backgrounds show a pull inconsistent with zero. The bias plots in Figures 8.6, 8.7, and 8.8 show the bias as a percentage of the true signal rate we put into the fit. Here we can see the first CC bin shows a 40% bias. This is not a significant problem as the neutrino signal in that bin has 65% statistical uncertainty, and similarly large systematic uncertainties. The first CC bin is not particularly interesting as a neutrino flux measurement. We only include it because we want to measure the background PDFs down to 3.5 MeV, where we get much larger statistics and therefore a better constraint at higher energy.

236
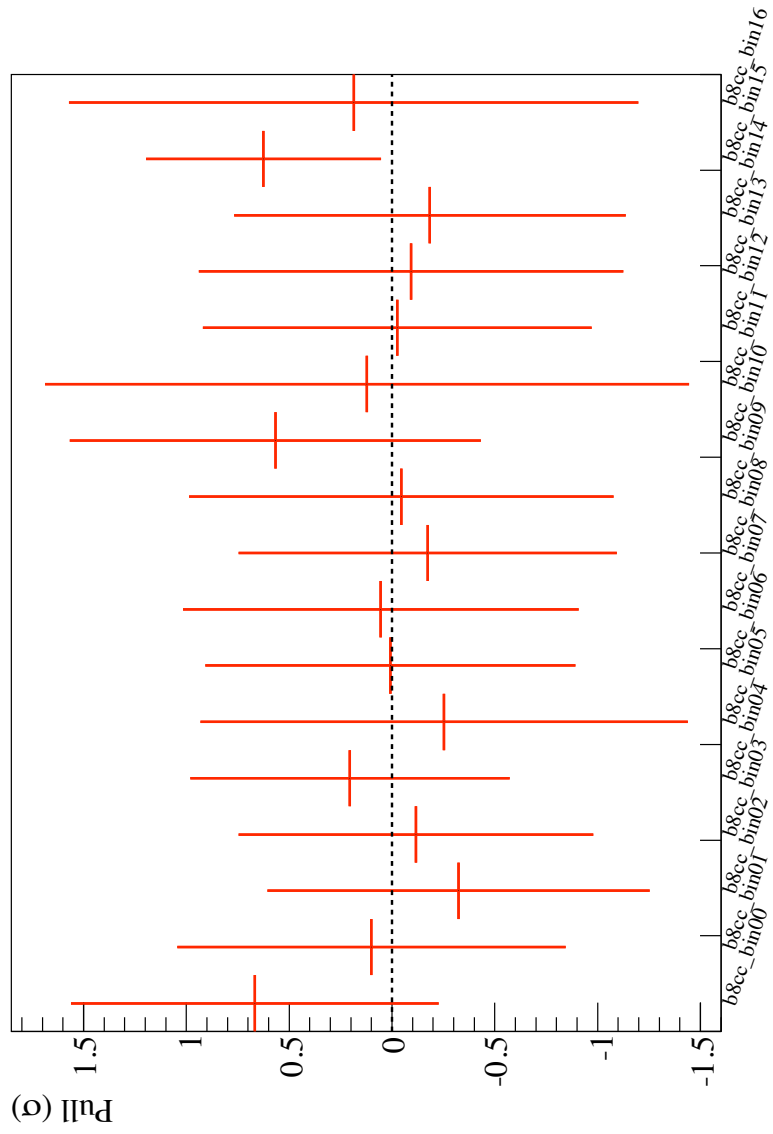
Figure 8.3: Pull distribution for CC events in ensemble test. Each CC bin is 0.5 MeV, with lower edge starting at 3.5 MeV.
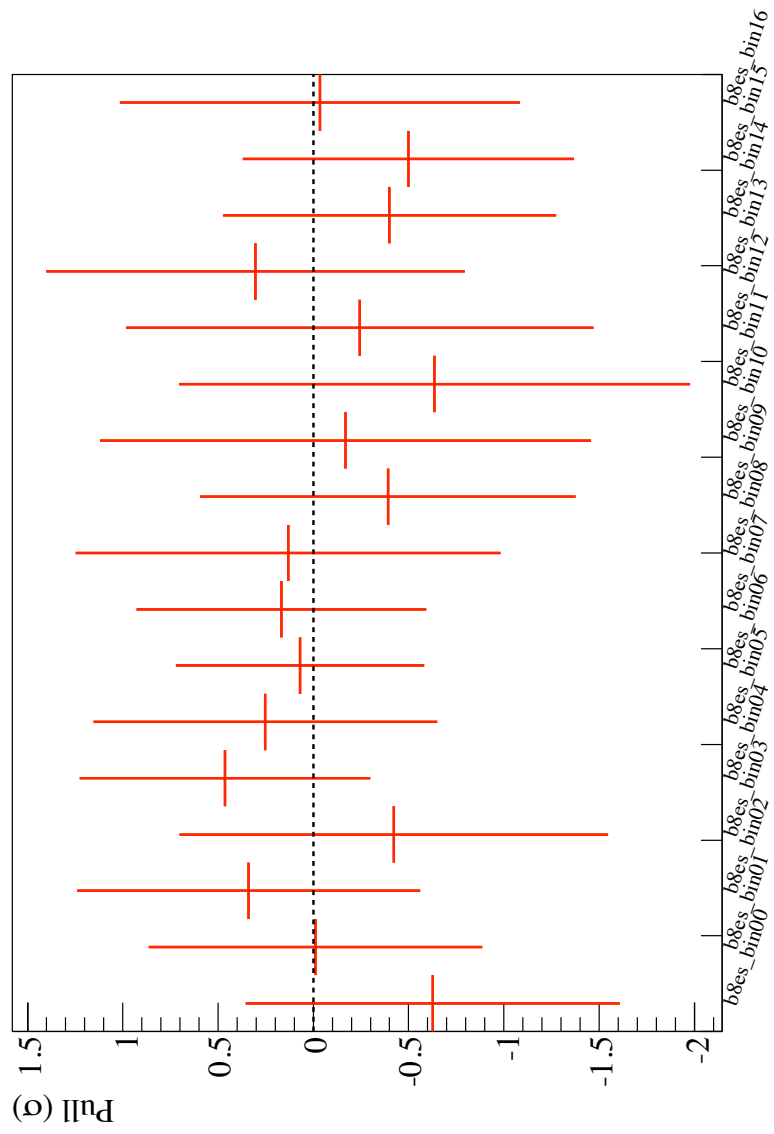
Figure 8.4: Pull distribution for ES events in ensemble test. Each ES bin is 0.5 MeV, with lower edge starting at 3.5 MeV.
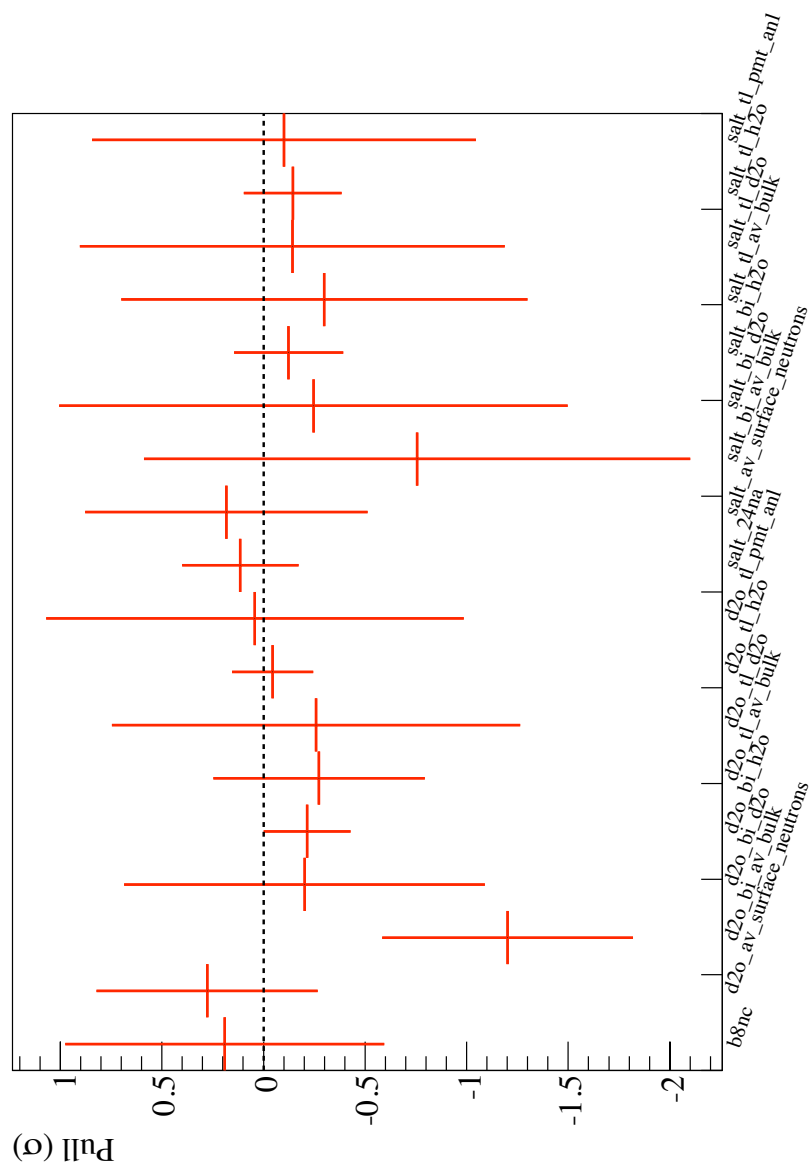
Figure 8.5: Pull distribution for NC and background events in ensemble tests.
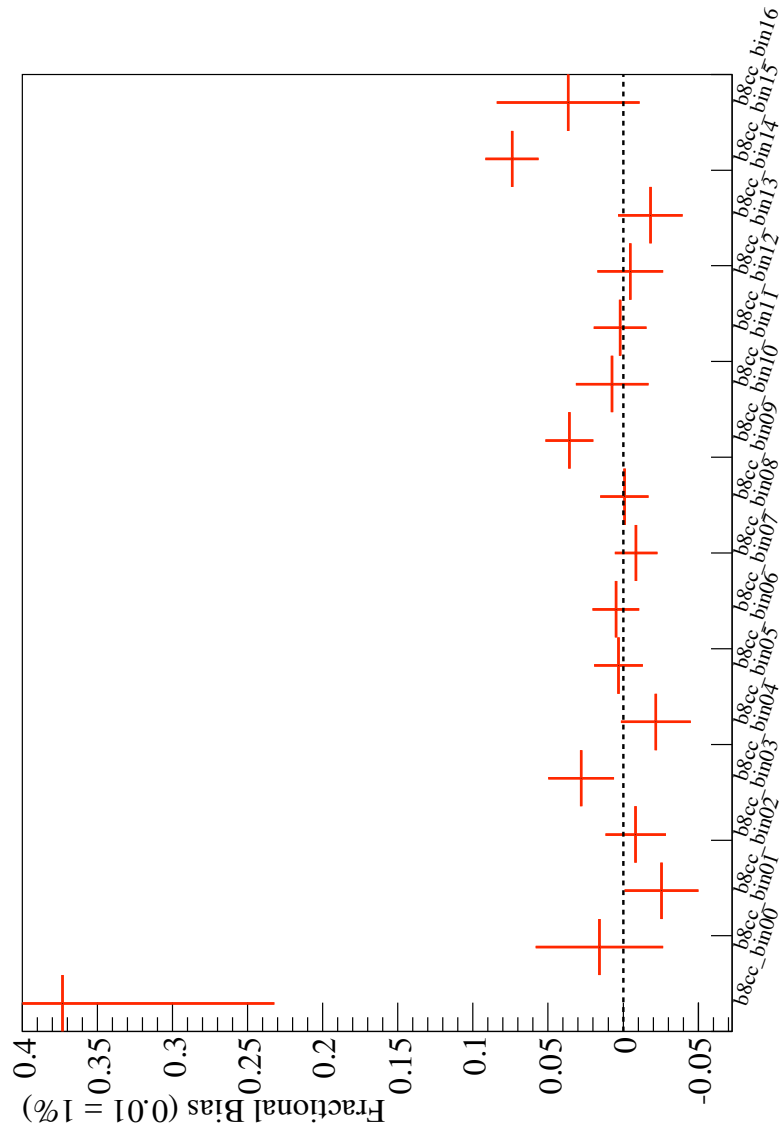
Figure 8.6: Bias distribution for CC events in ensemble test. Each CC bin is 0.5 MeV, with lower edge starting at 3.5 MeV.
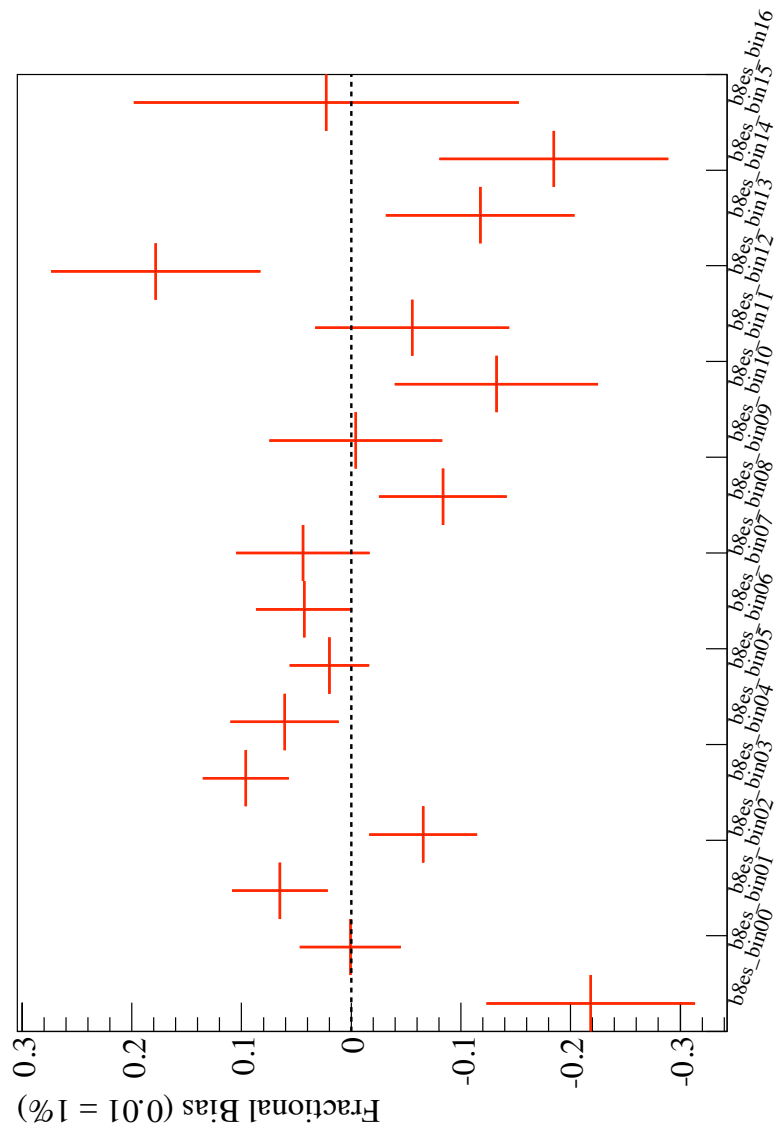
Figure 8.7: Bias distribution for ES events in ensemble test. Each ES bin is 0.5 MeV, with lower edge starting at 3.5 MeV.

Figure 8.8: Bias distribution for NC and background events in ensemble tests.
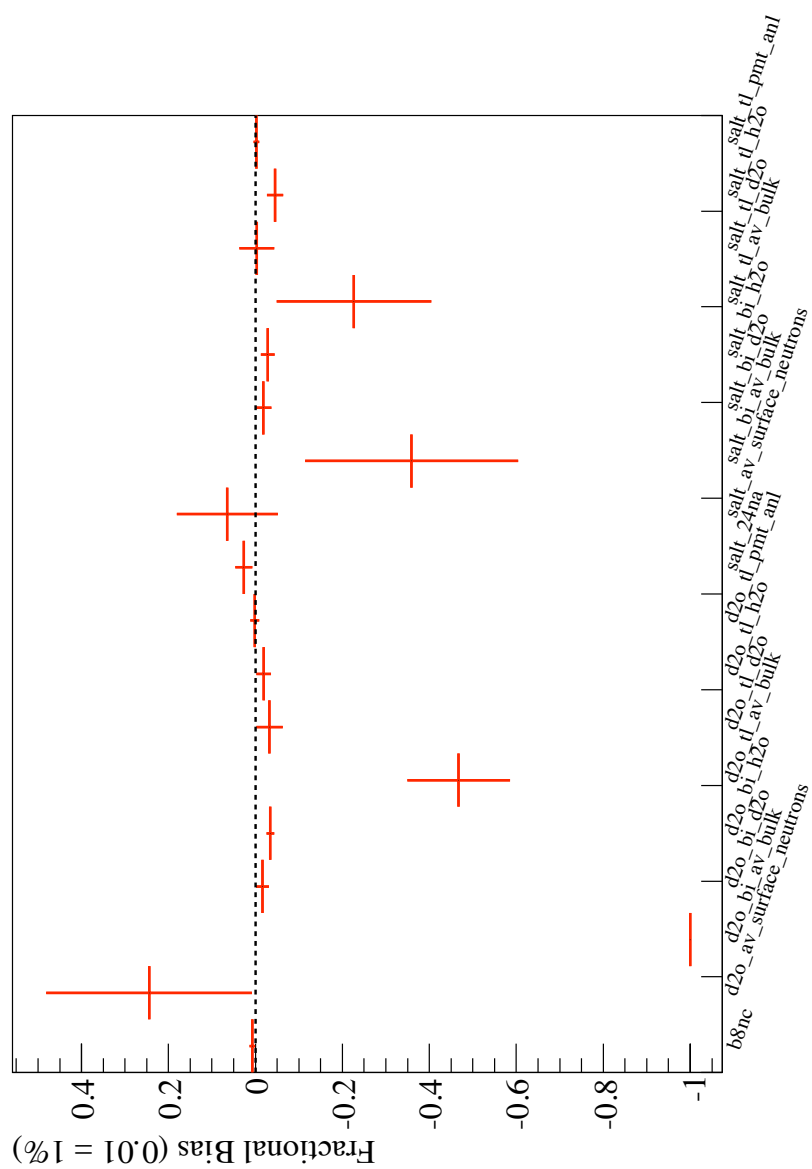
8.9 Fitting for Mixing Parameters

If we assume the three neutrino oscillation model and input θ_{23}, $|\Delta m_{32}^2|$, and approximate density profiles for the Sun and Earth, then we can numerically integrate forward the ν_e states through the Sun and Earth to get the survival probability $P_{\nu_e \to \nu_e}$ as a function of E_ν, θ_{12} and Δm_{21}^2. To get sensitivity to day-night asymmetries, we can split this table into a day component and a night component. Figure 8.9 shows these tables. The daytime survival probability shows the "MSW triangle" where total conversion from ν_e to $\nu_\mu + \nu_\tau$ occurs, and the nighttime table shows the ν_e regeneration that happens as neutrinos pass through the Earth. The reweighting is applied to CC and ES separately with the functions:

$$W_{CC} = P_{\nu_e \to \nu_e}(E_\nu) \tag{8.31}$$

$$W_{ES} = P_{\nu_e \to \nu_e}(E_\nu) + 0.1558846(1 - P_{\nu_e \to \nu_e}(E_\nu)). \tag{8.32}$$

$$\tag{8.33}$$

The second term in the ES weighting accounts for the partial sensitivity to ν_μ and ν_τ in that reaction. This sort of direct extraction of contours in signal extraction has not been done by SNO before. Instead, contours were obtained from the unconstrained T_{eff} spectrum provided the signal extraction method described in Section 8.8.

The PDF factorization in this fit is listed in Table 8.10. We divide the data into four phases: D$_2$O-day, D$_2$O-night, Salt-day and Salt-night. The background rates are allowed to freely vary between day and night, and between D$_2$O and salt phase. The signal parameters include:
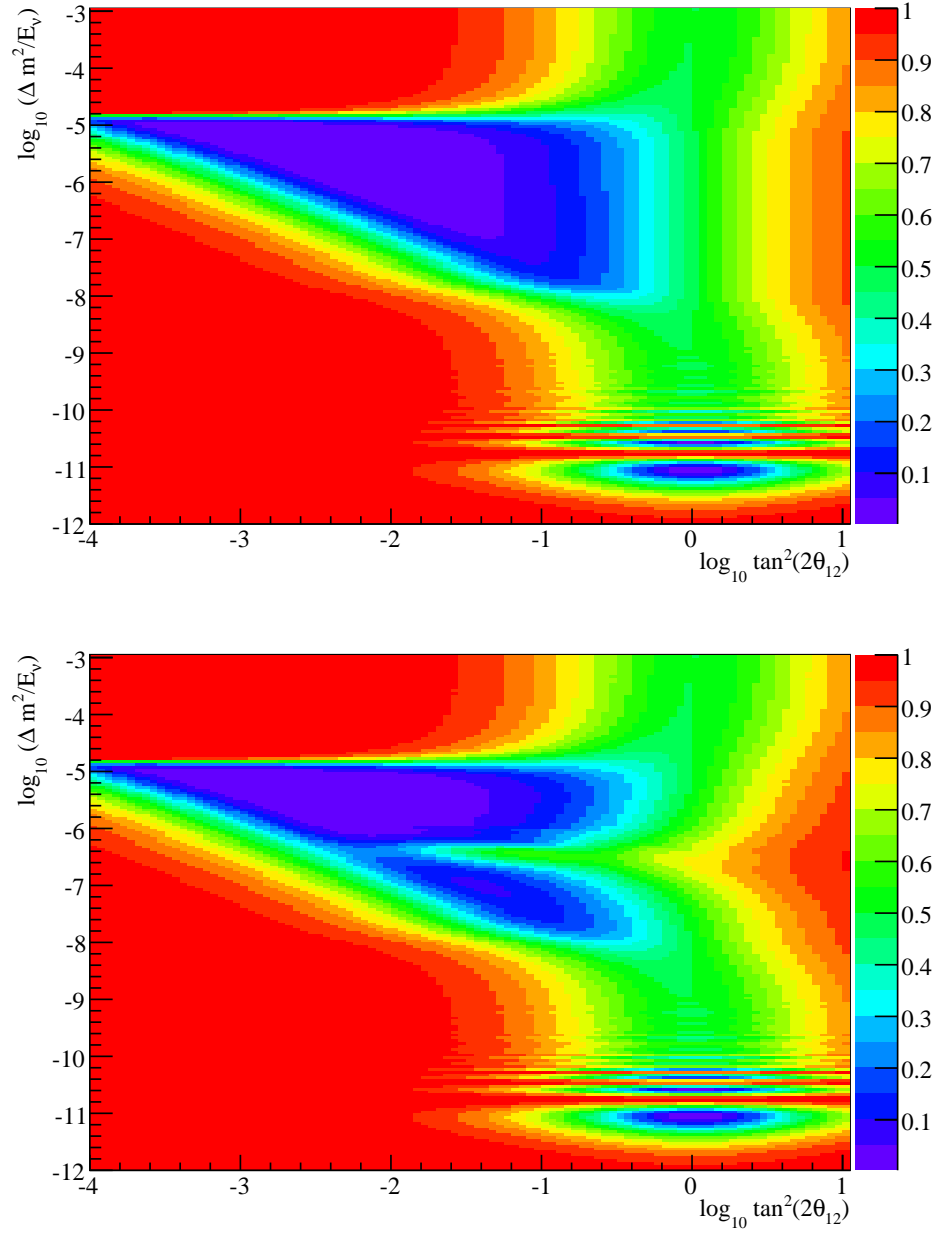
Figure 8.9: D$_2$O phase ν_e survival probability (color axis) during the day (top) and night (bottom) as a function of $\log_{10}\tan^2(2\theta_{12})$ and $\log_{10}(\Delta m^2/E_\nu)$. Tables computed by O. Simard[20].

Signal	PDF factorization
CC, ES, NC	$P(T_{\text{eff}}, \beta_{14}, R^3) \times P(\cos\theta_\odot)$
PMT	Analytic Model (see Chapter 6)
Other backgrounds	$P(T_{\text{eff}}, \beta_{14}, R^3) \times \text{Flat}(\cos\theta_\odot)$

Table 8.10: PDF factorization used in the mixing parameter fit and the polynomial survival probability fit. $\text{Flat}(\cos\theta_\odot)$ is a constant spanning the entire -1 to 1 observable range.

- f_{8B}: Total neutrino flux from the Sun. This sets the overall scale of CC, ES and NC events in all four phases.

- $\log_{10}\tan^2(2\theta_{12})$

- $\log_{10}(\Delta m^2)$

The last two parameters are treated as reweighting systematics and applied to the CC and ES events using the 2D table shown in Figure 8.9 (and the equivalent table for the salt phase). While we can certainly optimize the mixing parameters through gradient descent in the likelihood space, it is probably more interesting to be able to map out the 1, 2, 3 and 4σ contours. The speed of sigex.py allows us to rapidly grid scan over a specific region of parameter space, fixing the mixing parameters, and performing a full signal extraction allowing f_{8B} and the backgrounds to float. Each grid point takes about 2 minutes to evaluate, including rebuilding the NLL table and running MINUIT.

Figure 8.10 shows the contours for a fake data set that has no distortion or day-night asymmetry. The figure also compares the contours to those generated without a separation of day and night events. The day-night effect gives SNO a much better handle on Δm^2.
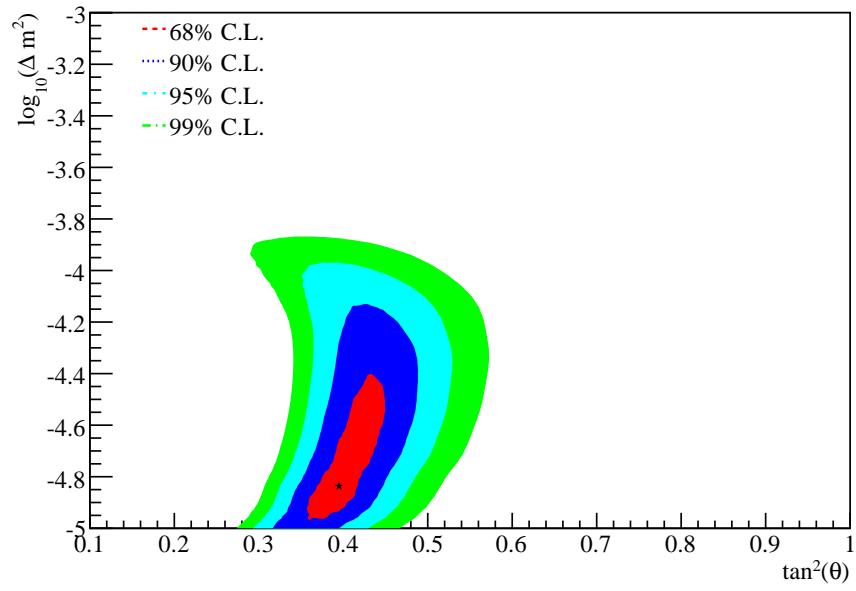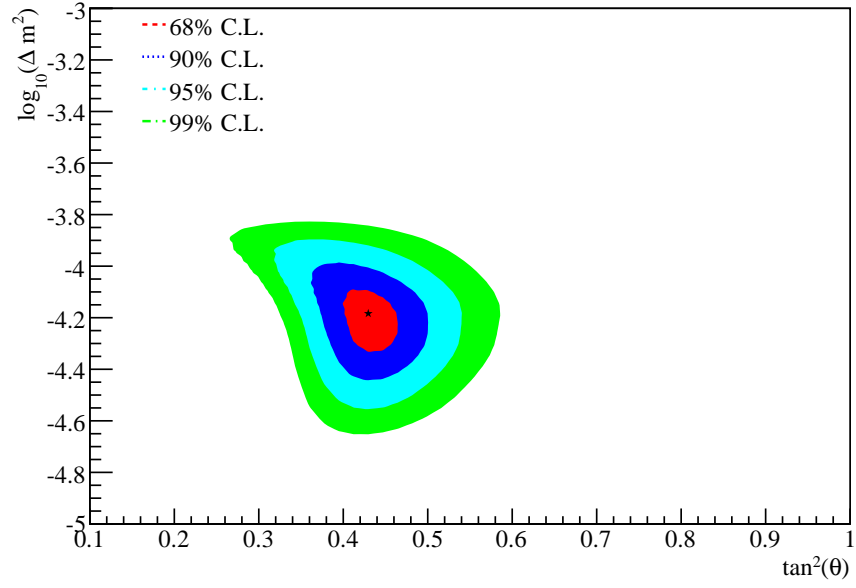
Figure 8.10: Mixing parameter contours in the LMA region extracted from a fake data set including day-night variation (top) and without separating day-night (bottom). Statistical uncertainties only.

8.10 Fitting for Survival Probability

Fitting for mixing parameters is very powerful, but it gives no information about new physics that might change the transition from vacuum to matter dominated-oscillations. To see unexpected distortion in the spectrum we need to remove the constraints of the model. However, as we saw in the unconstrained T_{eff} and E_ν fits, too many free parameters can be a bad thing, especially when correlations are involved. However, if we postulate a simple phenomenological form:

$$P_{\nu_e \to \nu_e}(E_\nu) = p_0 + p_1(E_\nu - 10.0\text{MeV}) + p_2(E_\nu - 10.0\text{MeV})^2. \qquad (8.34)$$

The choice of 10 MeV as the origin of the quadratic expansion is because that energy is right in the region of greatest sensitivity for SNO. Expanding the function about that point helps to decorrelate p_0 from the other parameters.

In the survival probability fit, our free parameters are:

- f_{8B}: Total neutrino flux from the Sun. This sets the overall scale of CC, ES and NC events in all phase.

- $p_{0,day}, p_{1,day}, p_{2,day}$: Survival probability for events during the day.

- $p_{0,night}, p_{1,night}, p_{2,night}$: Survival probability for events during the night.

These seven parameters, along with their uncertainties and covariance matrix, should make it relatively easy to test or constrain any model of new physics which would affect the neutrino survival probability.

The PDF factorization in the survival probability fit same as before, as is the division of the D_2O and salt phases into separate day and night
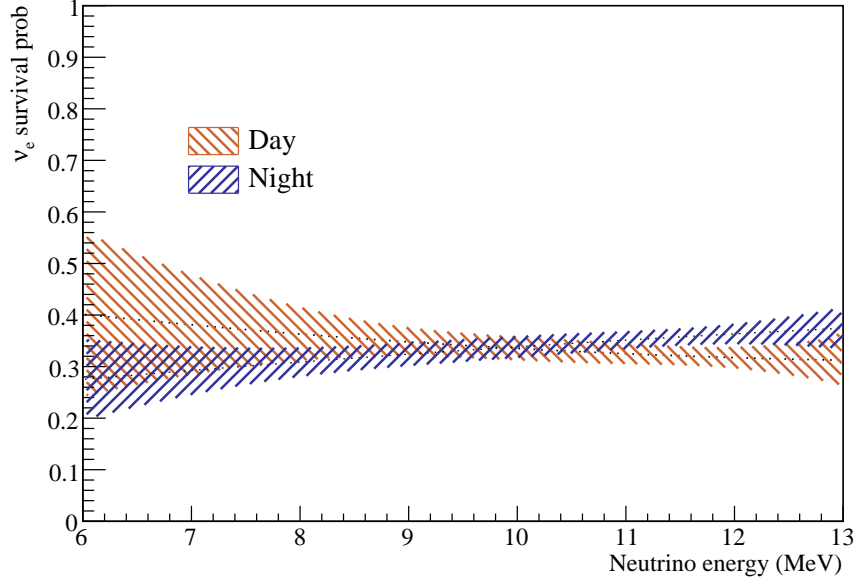
Figure 8.11: Extracted survival probability for day and night from a fake data set with no spectrum distortion and no day-night asymmetry. Error bands are statistical only.

components. In order to display the survival probability as a function of energy with errors, we can draw randomly from the a Gaussian distribution of the p_i parameters, including correlations, and map out a 1σ error band, like that shown in Figure 8.11 for a fake data set.

8.11 Summary

Thanks to the flexibility of kernel estimated PDFs, we have the freedom to perform a wide variety of signal extraction methods, each with their own strengths and weaknesses. For comparison with previous SNO results, the unconstrained T_{eff} is important, but the most difficult to perform at low energy.

Fitting for mixing parameters or survival probability polynomial coefficients is much more powerful and useful for analysts outside SNO.

Chapter 9

Results and Conclusion

This chapter shows the application of the signal extraction techniques described in the previous chapter to 1/3 of the D_2O and salt phase data sets.

9.1 Unconstrained T_{eff} Energy Fit

For this fit, we chose to float the following systematics:

- Energy resolution: D_2O and salt phase uncorrelated

- PMT energy exponential parameter: D_2O and salt uncorrelated

- PMT R^3 flat parameter: D_2O and salt uncorrelated

- β_{14} scale, electrons: D_2O and salt uncorrelated

- β_{14} scale, neutrons: D_2O and salt uncorrelated

- β_{14} energy dependence: D_2O and salt correlated

All other systematics discussed in Section 8.7 are propagated with the shift and smear technique.

Figures 9.1, 9.2, 9.3, 9.4 show 1D projections from the fit to the 1/3 data set. Table 9.1 lists the neutrino flux parameters in units of SSM (BP2005), with normalization corrections described in Section 8.6. Extracted backgrounds, in units of events are listed in Table 9.2, and values of floated systematic parameters are shown in Table 9.3. The CC and ES T_{eff} spectra are plotted in Figure 9.5.

9.2 Other Fits

While the unconstrained T_{eff} fit is the primary result, we can also get a preview of how the other signal extraction techniques will perform when their implementation is finalized. These other methods will play a significant role in the result to be published in the coming months.

9.2.1 Mixing Parameter Scan

Figure 9.6 shows the contours obtained by scanning the mixing parameter space with the 1/3 data set, separating events into day and night zenith bins. Backgrounds day and night have been allowed to independently vary. Ultimately, we will need to reassess day-night asymmetry in the detector response to propagate systematic uncertainties to these contours.
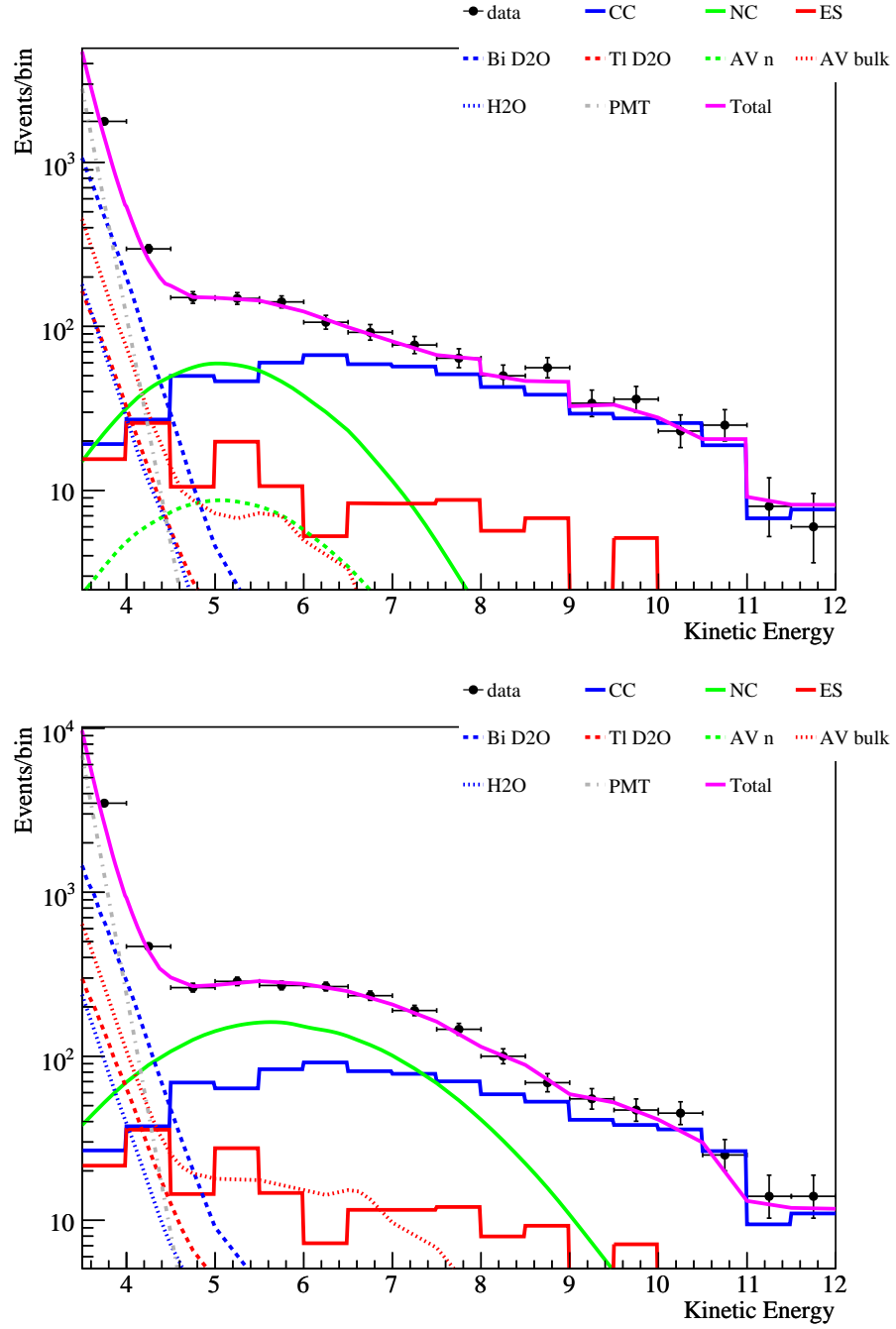
Figure 9.1: 1D T_{eff} projection of unconstrained T_{eff} fit with 1/3 data. D_2O phase is shown on top, salt phase on bottom.
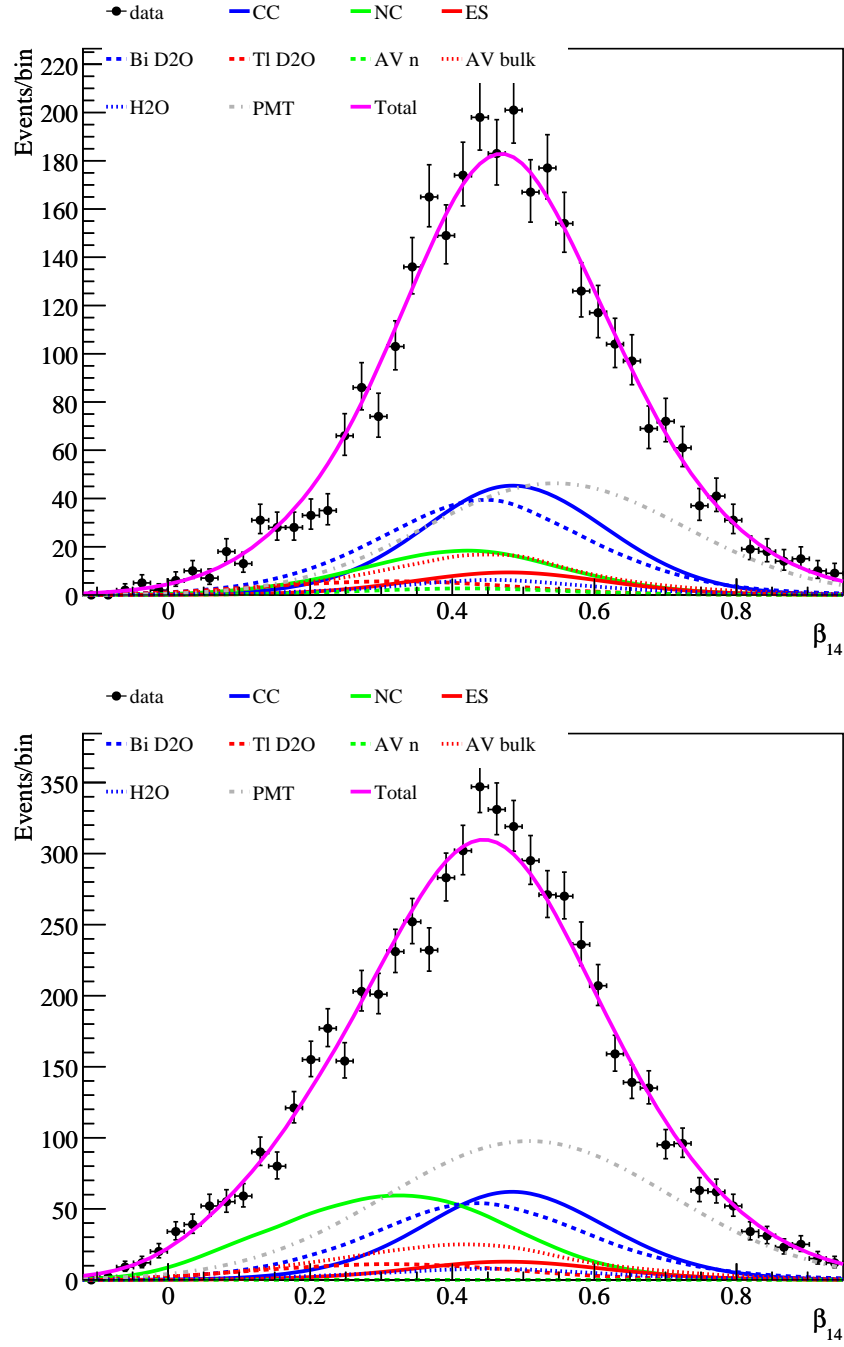
Figure 9.2: 1D β_{14} projection of unconstrained T_{eff} fit with 1/3 data. D_2O phase is shown on top, salt phase on bottom.
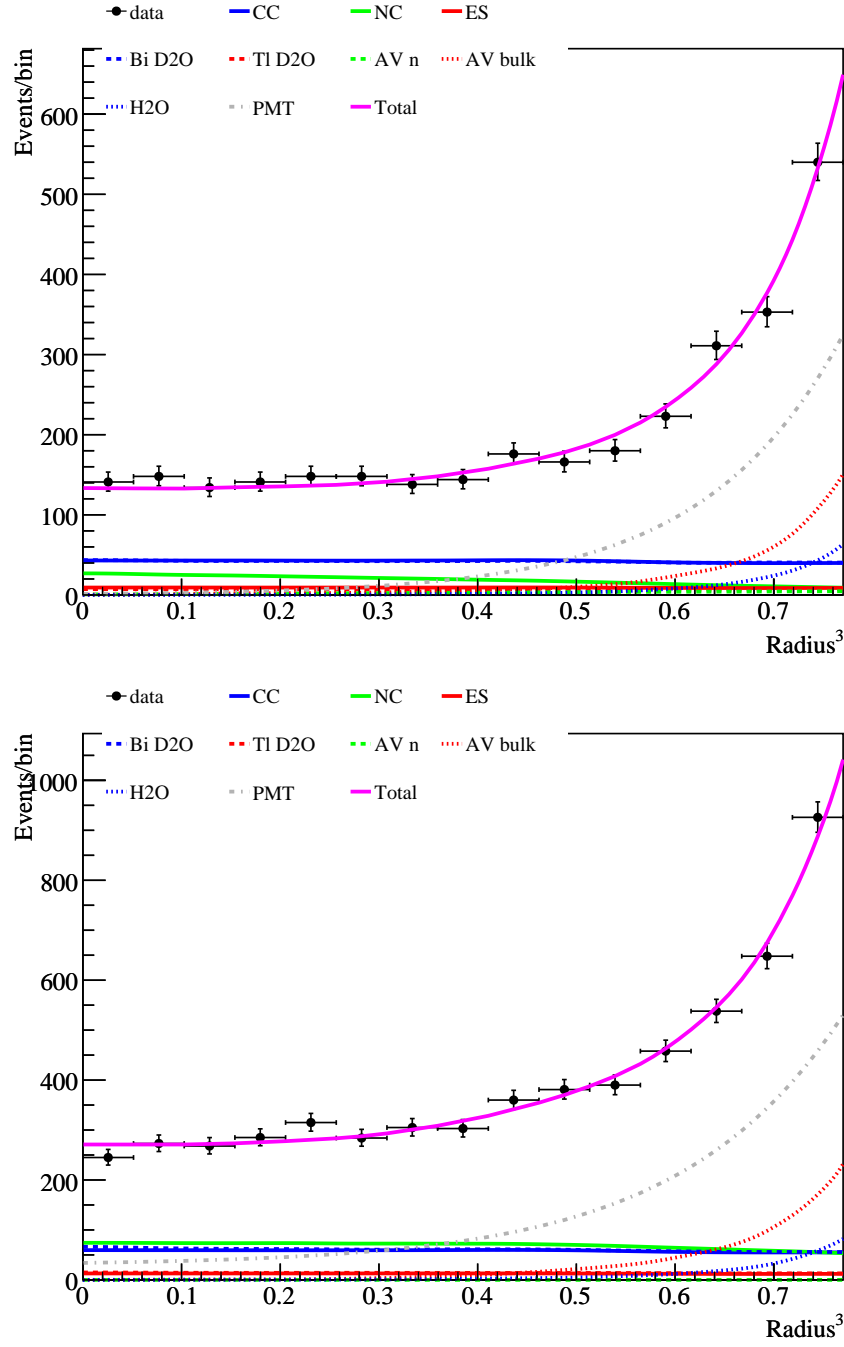
Figure 9.3: 1D R^3 projection of unconstrained T_{eff} fit with 1/3 data. D_2O phase is shown on top, salt phase on bottom.
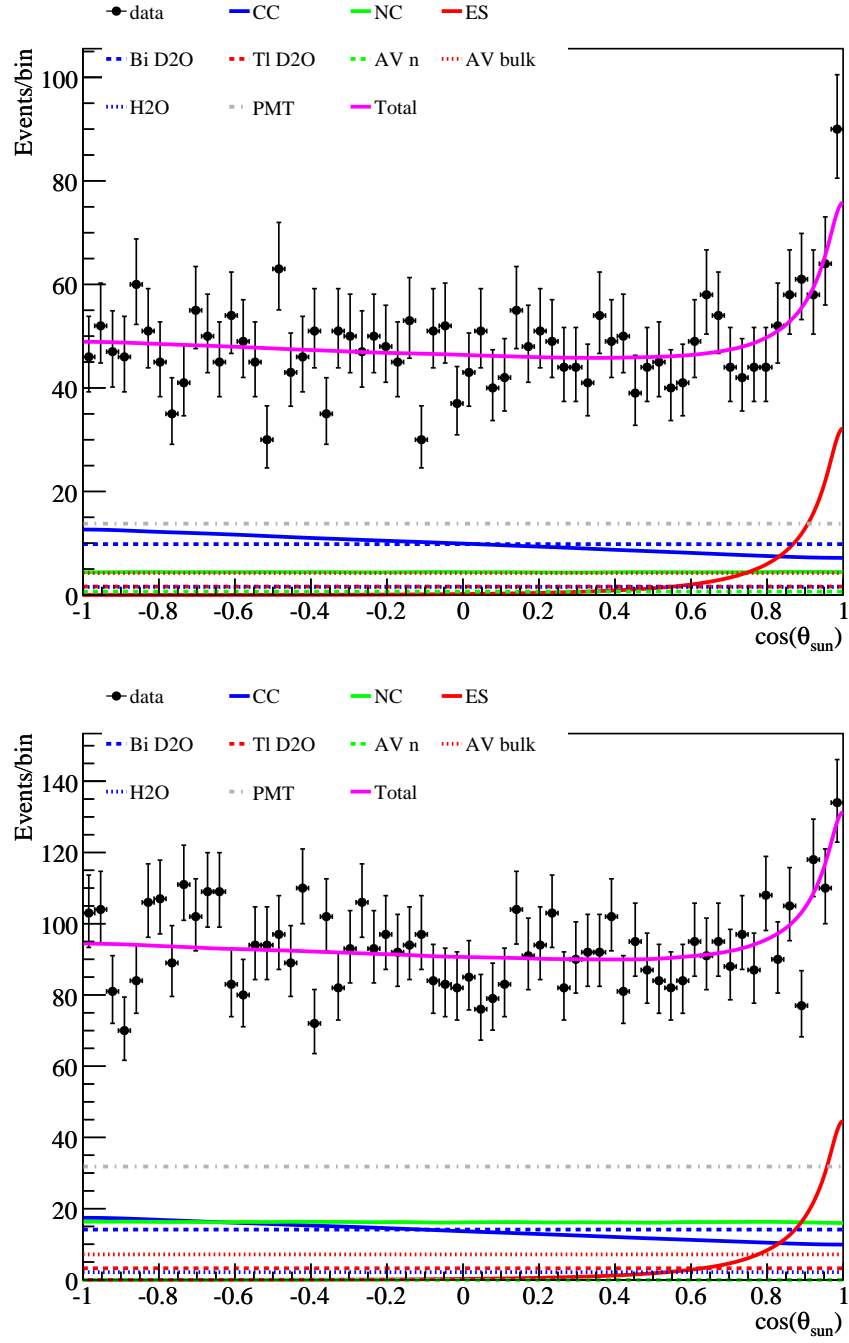
Figure 9.4: 1D $\cos\theta_\odot$ projection of unconstrained T_{eff} fit with 1/3 data. D_2O phase is shown on top, salt phase on bottom.

Signal	Value	Fit Uncert	Syst Uncert
cc3.5	0.163	±0.237	$^{+0.136}_{-0.147}$
cc4.0	0.191	±0.073	±0.045
cc4.5	0.307	±0.048	±0.010
cc5.0	0.256	±0.041	±0.008
cc5.5	0.312	±0.039	±0.005
cc6.0	0.344	±0.037	±0.003
cc6.5	0.300	±0.032	±0.006
cc7.0	0.298	±0.031	±0.005
cc7.5	0.286	±0.029	±0.005
cc8.0	0.261	±0.028	±0.004
cc8.5	0.269	±0.029	±0.005
cc9.0	0.247	±0.030	±0.004
cc9.5	0.285	±0.035	±0.002
cc10.0	0.346	±0.043	±0.002
cc10.5	0.342	±0.050	±0.004
cc11.0	0.177	±0.044	±0.001
cc11.5	0.303	±0.067	$^{+0.031}_{-0.002}$
es3.5	0.419	±0.263	±0.046
es4.0	0.744	±0.151	±0.030
es4.5	0.326	±0.111	±0.017
es5.0	0.679	±0.142	±0.025
es5.5	0.404	±0.130	±0.012
es6.0	0.232	±0.123	±0.015
es6.5	0.414	±0.131	±0.014
es7.0	0.481	±0.157	±0.021
es7.5	0.593	±0.166	±0.007
es8.0	0.461	±0.167	±0.019
es8.5	0.659	±0.209	$^{+0.021}_{-0.031}$
es9.0	0.296	±0.175	±0.023
es9.5	0.786	±0.261	±0.016
es10.0	0.280	±0.251	±0.021
es10.5	0.291	±0.294	$^{+0.017}_{-0.050}$
es11.0	0.652	±0.351	$^{+0.028}_{-0.014}$
es11.5	0.000	±0.140	±0.000
nc	0.842	±0.039	±0.022

Table 9.1: Neutrino flux extracted from 1/3 data set fit in units of BS2005 SSM ($5.69 \times 10^6 \, \mathrm{cm^{-2} s^{-1}}$). "Fit Uncert" includes both statistical uncertainty and the uncertainty from floated systematics.

Bkg	Events	Fit Uncert	Syst Uncert
D_2O , AV n	42	± 29	$-9/+7$
D_2O , ^{214}Bi AV	0	± 58	$-0/+0$
D_2O , ^{214}Bi D_2O	623	± 50	$-64/+86$
D_2O , ^{214}Bi H_2O	65	± 10	$-1/+1$
D_2O , ^{208}Tl AV	269	± 46	$-60/+80$
D_2O , ^{208}Tl D_2O	104	± 33	$-36/+32$
D_2O , ^{208}Tl H_2O	33	± 10	± 1
D_2O , ^{208}Tl PMT	873	± 47	$-161/+225$
Salt, ^{24}Na	117	± 32	$-3/+3$
Salt, AV n	0	± 19	± 0
Salt, ^{214}Bi AV	6	± 2243	$-28/+62$
Salt, ^{214}Bi D_2O	908	± 84	$-176/+213$
Salt, ^{214}Bi H_2O	91	± 24	$-2/+2$
Salt, ^{208}Tl AV	454	± 276	$-64/+44$
Salt, ^{208}Tl D_2O	209	± 79	$-130/+125$
Salt, ^{208}Tl H_2O	49	± 14	$-1/+1$
Salt, ^{208}Tl PMT	2047	± 387	$-249/+563$

Table 9.2: Number of background events extracted from 1/3 data set. "Fit Uncert" includes statistical and floating systematic uncertainties. The large uncertainty for salt ^{214}Bi AV is due to MINUIT's estimate of the error being skewed by the boundary.
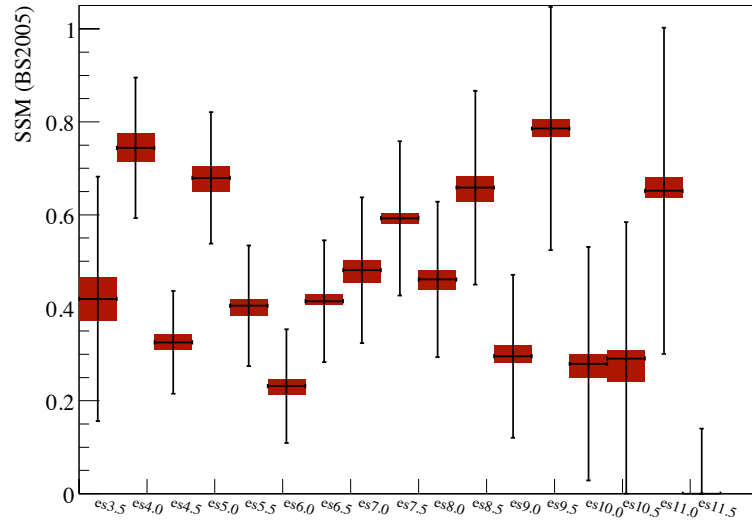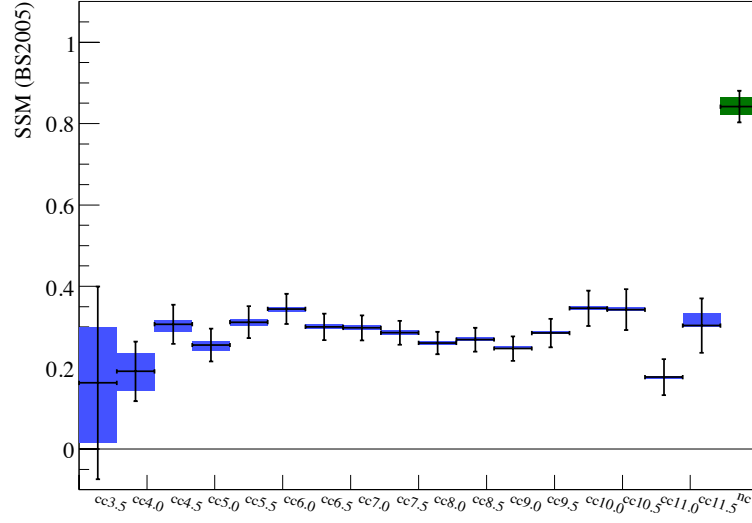
Figure 9.5: T_{eff} spectra for 1/3 data fit. Thin error bars show the statistical uncertainty, and the wide colored boxes show the systematic uncertainties.
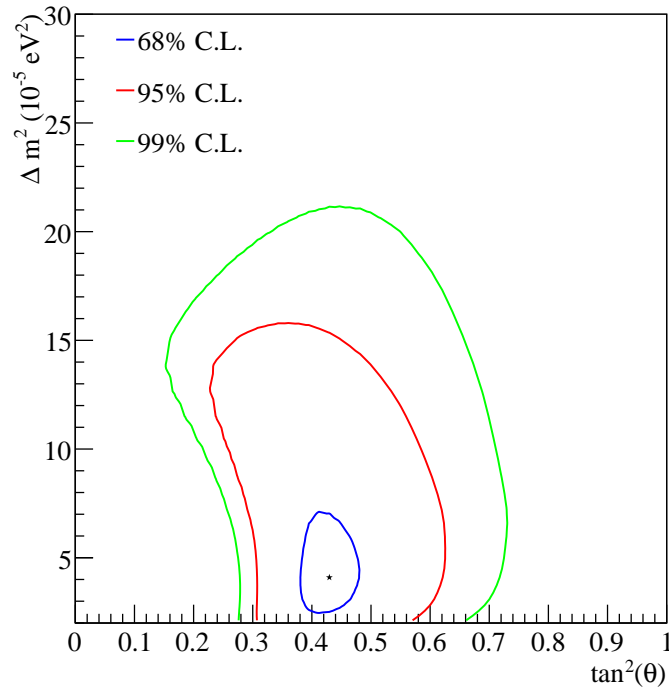
Figure 9.6: Statistics-only contours from 1/3 data set, using day-night variation as an additional constraint. Note that both axes are linear for easier comparison to Figure 2.11. The 99% C.L. is shown here rather than the 99.73% C.L., as that line is mostly off scale.

Parameter	Fit Value	Original Constraint
D$_2$O Phase, T_{eff} resolution	0.026 ± 0.014 MeV	(None)
D$_2$O Phase, β_{14} e^- scale	0.0001 ± 0.0036	0 ± 0.0042
D$_2$O Phase, β_{14} n scale	-0.0002 ± 0.0022	0 ± 0.0024
D$_2$O Phase, PMT T_{eff} exp	-6.38 ± 0.42	-7.65 ± 0.77
D$_2$O Phase, PMT R^3 flat	0 ± 0.42	3.31 ± 1.43
Salt Phase, T_{eff} resolution	0.015 ± 0.019 MeV	(None)
Salt Phase, β_{14} e^- scale	0.0 ± 0.0022	0 ± 0.0024
Salt Phase, β_{14} n scale	0.0003 ± 0.0035	0 ± 0.0038
Salt Phase, PMT T_{eff} exp	-6.7 ± 1.0	-7.25 ± 0.49
Salt Phase, PMT R^3 flat	6.32 ± 2.08	1.07 ± 0.66
β_{14} Energy Slope	$(-4.1 \pm 5.2) \times 10^{-4}$	$(0 \pm 6.9) \times 10^{-4}$

Table 9.3: Best fit values of floated systematic parameters.

Parameter	Day	Night
p_0	0.349 ± 0.041	0.323 ± 0.038
p_1	0.010 ± 0.014	0.005 ± 0.013
p_2	-0.0117 ± 0.0097	-0.0007 ± 0.0087
^{8}B flux (SSM, BP2000)	0.991 ± 0.056	

Table 9.4: Parameters from polynomial survival probability fit to 1/3 data. Definition of parameters given in Equation 8.34.

9.3   Survival Probability Polynomial Fit

The polynomial survival probability fit extracts a quadratic survival probability for day and night events separately. Figure 9.7 shows the extracted probability for the 1/3 data set and the 1σ statistical error bands. Table 9.4 lists the fit parameters from Equation 8.34 and Table 9.5 shows the correlation matrix.
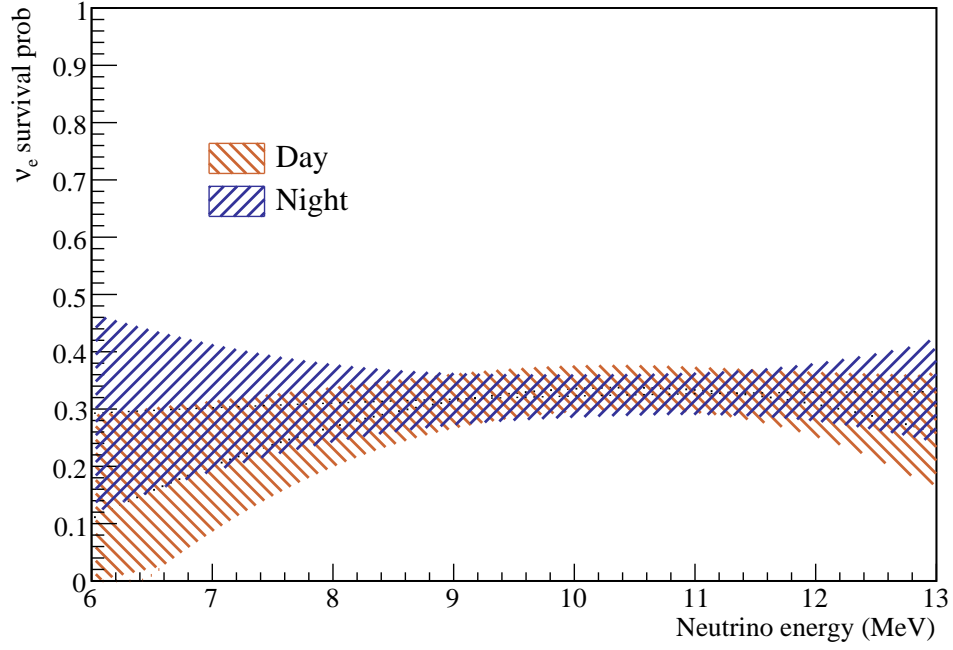
Figure 9.7: Polynomial survival probability fit to 1/3 data set and 1σ error band. Only statistical errors show.

	$p_{0,\text{day}}$	$p_{1,\text{day}}$	$p_{2,\text{day}}$	$p_{0,\text{night}}$	$p_{1,\text{night}}$	$p_{2,\text{night}}$	$^8\text{B flux}$
$p_{0,\text{day}}$	$+1.000$	-0.313	-0.693	$+0.353$	-0.200	$+0.058$	-0.558
$p_{1,\text{day}}$	-0.313	$+1.000$	$+0.092$	-0.166	$+0.094$	-0.027	$+0.262$
$p_{2,\text{day}}$	-0.693	$+0.092$	$+1.000$	$+0.026$	-0.015	$+0.004$	-0.042
$p_{0,\text{night}}$	$+0.353$	-0.166	$+0.026$	$+1.000$	-0.353	-0.598	-0.633
$p_{1,\text{night}}$	-0.200	$+0.094$	-0.015	-0.353	$+1.000$	$+0.037$	-0.359
$p_{2,\text{night}}$	$+0.058$	-0.027	$+0.004$	-0.598	$+0.037$	$+1.000$	-0.105
$^8\text{B flux}$	-0.558	$+0.262$	-0.042	-0.633	-0.359	-0.105	$+1.000$

Table 9.5: Parameter correlations from polynomial survival probability fit to 1/3 data. Definition of parameters given in Equation 8.34.

9.4 Conclusion

Neutrino physics is entering the precision era. The vast quantity and variety of neutrino data being collected is giving us the evidence needed to put the three neutrino oscillation model through even stricter verification. In the solar neutrino sector, the transition from vacuum to matter-dominated oscillation as a function of energy is a powerful indicator of new physics that couples to neutrinos. Phenomena as diverse as non-standard interactions or mass-varying neutrinos can modify the electron neutrino survival probability in this critical transition region.

The Sudbury Neutrino Observatory contributed significantly to the discovery and verification of solar neutrino oscillation in previous papers. Using the experience we have gained since then, we are now pushing the SNO data set to explore energies lower than previously though possible. Including events with reconstructed kinetic energies as low as 3.5 MeV nearly doubles the NC statistics over previous analyses, and quadruples the number of observed CC events generated by neutrinos with less than 8 MeV of energy. Together, these advances will enable SNO to make the most precise measurement of the mixing angle θ_{12} and the total flux of ^{8}B neutrinos.

In order to analyze data that low in energy, a complete top-to-bottom overhaul of the analysis had to be undertaken by the SNO collaboration. In Chapter 5, I described a spatial correction to the reconstructed energy which makes the energy scale much more uniform. The correction, along with a new method for assessing energy systematics, reduces the contribution of detector non-uniformity to the final energy scale systematic uncertainty to 0.2% in the D$_2$O phase and 0.3% in the salt phase.

The proliferation of backgrounds at low energy require us to rely more heavily on the Monte Carlo to generate PDFs for all these sources of events. The background verification described in Chapter 6 shows that the Monte Carlo does accurately reproduce the background data we see in the detector, although the PMT PDFs must be constructed from an analytic model due to a lack of Monte Carlo events.

I have applied the kernel estimation formalism to PDF construction, which has not been used on SNO before. This well-established statistical technique allows non-parametric, continuous PDFs to be constructed from an event sample without binning. PDFs in this form can be distorted continuously, which opens up a natural way to float systematic uncertainties on non-parametric PDFs. A new kernel estimator implementation strategy, including both algorithmic improvements and the use of commodity 3D graphics cards as coprocessors, was described in Chapter 7. Together, these modifications have improved the performance of kernel estimation by several orders of magnitude, enabling complex maximum likelihood fits to be performed on a standard computer workstation.

The flexibility of kernel estimators opens up new signal extraction options. Traditionally, SNO has fit for an unconstrained CC and ES T_{eff} spectra, and then analyzed that spectra to obtain neutrino mixing parameters and contours. I have also performed this kind of signal extraction for the low energy analysis, but also demonstrated two other methods. Both the mixing parameter fit and the polynomial survival probability fit treat the electron neutrino survival probability as a *systematic*. The fast kernel estimator code allows the likelihood function to be evaluated over and over as this probability function varies, either within the three neutrino oscillation model, or using a completely

263

phenomenological quadratic function.

The polynomial fit is especially useful because it allows direct testing of new models on the SNO data set. One only needs to Taylor expand a prediction of $P_{\nu_e \to \nu_e}$ around 10 MeV (E_ν), and compare coefficients to the fit values and covariance matrix. Previously, testing a model required applying the survival probability to the ^{8}B spectrum and propagating that distribution through the detector response. Then the predicted T_{eff} could be compared to the published data. With the polynomial fit, we have removed the detector response already, simplifying this procedure greatly.

Using 1/3 of the SNO data from the D_2O and salt phases, we have performed all three of these fits. Full systematics have been propagated through the T_{eff} fit, given a measured NC flux of $4.79 \pm 0.22 \, (\text{stat}) \pm 0.13 (\text{syst}) \times 10^6 \, \text{cm}^{-2} \text{s}^{-1}$. This is in excellent agreement with the previously published D_2O and salt phase results:

$$\text{D2O phase NC} = 5.09^{+0.44}_{-0.43}(\text{stat}) \, ^{+0.46}_{-0.43}(\text{syst}) \times 10^6 \, \text{cm}^{-2}\text{s}^{-1}$$
$$\text{Salt phase NC} = 4.94^{+0.21}_{-0.21}(\text{stat}) \, ^{+0.38}_{-0.34}(\text{syst}) \times 10^6 \, \text{cm}^{-2}\text{s}^{-1}$$

Additionally, no significant distortion of the CC spectrum is observed in either the T_{eff} fit or the polynomial survival probability fit.

In future work we will apply the signal extraction described in this dissertation to the full D_2O and salt phase data sets. In addition, we will propagate detector systematics through the polynomial fit to provide the most accurate day and night survival probabilities to the physics community. With that data in hand, we will either add another result in support of the neutrino oscillation model, or perhaps pry open a window to new physics.

Appendix A

General Purpose Computing with Graphics Cards

Like all consumers of computing resources, physicists have benefited greatly from the rapid technological advancement of inexpensive computers. This growth is usually expressed in terms of Moore's Law[98], which initially predicted a doubling every year of the number of transistors which could be economically placed into a single integrated circuit.[1] CPU designers find themselves with an exponentially growing number of transistors per chip which they apply to the real problem: producing an exponential growth in CPU performance.

Converting more transistors to more performance is a non-trivial task. Most common programming languages express computation as a linear sequence of arithmetic and memory operations, whose flow is controlled by branching and looping statements. To run sequential programs faster, CPU

---

[1]Moore later revised his prediction to a doubling every 2 years, though it is frequently quoted as doubling every 18 months[99].

designers have had to increase clock rates, and decrease the number of clock ticks required to complete each instruction. Moore's Law says nothing about CPU clock rates, but a side-effect of packing more transistors onto a silicon wafer is that the transistors naturally get smaller. With shorter paths to travel, signaling times and power usage can be reduced, allowing clock rates to be increased.

However, system memory speeds have not kept up with CPU clock rates, so modern CPUs now include a *memory cache* to keep recently used data on-chip[100]. Caching has become so important that many CPUs now spend half of their total transistor count on various layers of memory caches. CPUs also use a large number of transistors to discover and exploit instruction-level parallelism in a program. When a sequence of instructions operate on different registers or memory locations, the instructions can be reordered, or processed partially in parallel to keep all parts of the CPU actively doing work. Identifying register dependencies and pipelining instructions through the various CPU stages is a challenging task, and requires a lot of transistor logic to achieve.

A.1   The Graphics Processing Unit

Along side the evolution of consumer CPUs, the graphics processing unit (or GPU) has grown in capability and complexity. The GPU controls the video monitor on a workstation or laptop, acting as an intermediary between the CPU and the analog or digital signaling (VGA or DVI) to the monitor. Originally the GPU was not much of a processor, just a chip which copied 2D array of pixels to the screen at some fixed rate. However, the GPU slowly

acquired more and more functionality, until today, where it now acts as a full 3D rendering coprocessor.

GPU designers have access to the same silicon fabrication technology as CPU designers, but a different computing task to solve. Rendering 3D environments to raster display devices is inherently a parallel activity, and therefore demands a different type of processor. Modern GPUs emphasize large numbers of identical, programmable floating point units with limited looping and branching ability, fed directly by a wide memory bus with little or no caching[2]. Transistor count per execution unit is conserved by having a reduced instruction set, and sharing instruction decoders between arithmetic units, thereby requiring that some or all units be doing exactly the same operations at the same time.

In the past few years, two major advancements have made GPUs potentially usable for general purpose computation. First, texture[3] processing was expanded to include nearly IEEE-compliant 32-bit floating point operations. Second, *shader languages* were standardized which allowed application developers, rather than than graphics driver authors, to directly program the floating point units. The purpose of these languages was to give developers the freedom to create new lighting and surface effects without having to wait for such functions to be implemented directly in hardware. These languages include all standard math operations and transcendental functions, as well as 3D vector operations. In principle, they could be applied to any calculation task, and not just pixel shading. Thus was born the field of general purpose

---

[2] The memory access pattern in 3D rendering would be unlikely to benefit from cache memory unless the cache were nearly the size of the main graphics memory.

[3] A *texture* is a 2D raster image which is mapped onto the surface of a 3D object to give it color and texture, like that of a brick wall for example.

	Athlon 64 X2 (CPU)	GeForce 8800 GTX (GPU)
Processor clock rate	2.4–3.0 GHz	1.35 GHz
Execution cores	2	96–128
Transistor count	164–243 million	681 million
On-chip Memory	1–2 MB	256 kB
Off-chip Memory Bus	128 bit	384 bit
Off-chip Memory Capacity	4 GB (or more)	768 MB
Approximate GFLOPS	5–10	346

Table A.1: Comparison of example CPU and GPU characteristics. (Note that GFLOPS can be a very inaccurate measure of performance and are shown here just as an illustration of the dramatically different floating point capabilities of the devices.) Data from [101] and [102].

GPU (GPGPU) computing.

Modern GPUs have phenomenal computing power compared to CPUs, as shown in Table A.1. Ultimately, this comes from the different needs of the GPU. The enormous cache memory of the CPU is not necessary on the GPU, allowing those transistors to be "spent" on additional floating point arithmetic units. Figure A.1 shows a comparison of the die for an AMD Quad Core Opteron CPU, and the NVIDIA G80 GPU.

A.2 NVIDIA's Compute Unified Device Architecture

NVIDIA, one of the major manufacturers of high-end GPUs, released a revised GPU architecture in November 2006 which supports both standard 3D rendering tasks, as well as general purpose computation. Using their "Compute Unified Device Architecture" (CUDA) development tools, the GPU execution
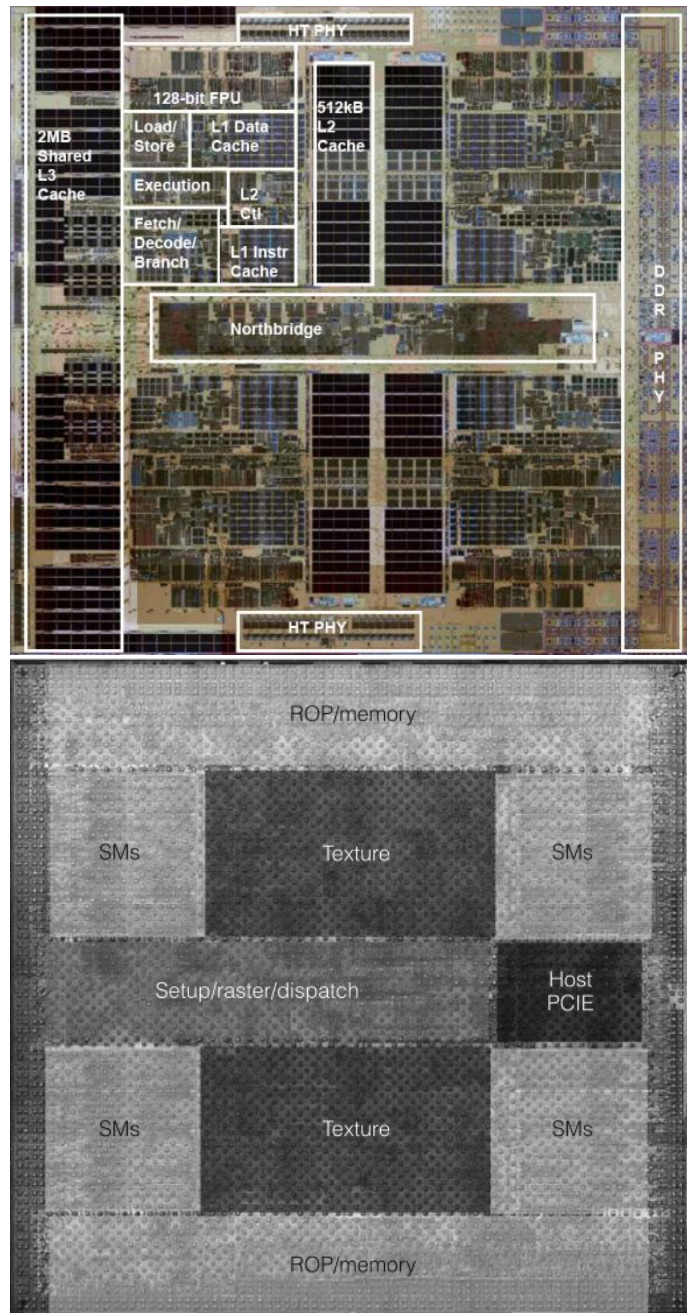
Figure A.1: AMD Quad Core Opteron CPU die image (top) compared to the NVIDIA G80 GPU (bottom). The area devoted to arithmetic is denoted "FPU" ("Floating Point Unit") in the CPU image and "SM" ("Streaming Multiprocessor") in the GPU image. A much greater area on the GPU is used for floating point arithmetic. Images from [103] and [104].

units and device memory can now be directly accessed by the user-written programs[101].

CUDA is a programming model for GPUs manufactured by NVIDIA geared toward *data-parallel* processing. There are many parallel programing models, and data parallelism [105] is an approach that treats a computation as a task that to be applied to many data elements independently and simultaneously. This is very similar to the vector operations that were supported by older supercomputers, most notably the Cray-1[106].

Beyond just a programming interface, CUDA also describes general hardware architecture to be implemented by the GeForce 8 series and future GPUs. Previous GPU designs divided the chip resources into specific vertex shaders, rasterizers, interpolators, and pixel shaders. In a break with this tradition, CUDA is a more homogenous architecture, which allows the same computing resources to be used for different and potentially non-graphical tasks.

A CUDA-capable GPU consists of 1 or more *multiprocessors* connected to off-chip memory called *device* or *global memory*. Device memory is on the same circuit board as the GPU, and is distinct from the main system memory used by the CPU. The multiprocessors all access device memory through a single memory bus, which can supply up to 104 GB/sec in some devices. While much faster than normal system memory, device memory can have hundreds of clock cycles of latency. The device memory is most effectively used when many consecutive read operations are issued at once, and can be coalesced by the hardware into several large reads.

Current NVIDIA products place anywhere from 1 to 16 multiprocessors on a single chip. A multiprocessor contains several logical units (Figure A.2):

- Instruction decoder—Low level instructions are converted into control signals to activate different operations. There is one instruction decoder per multiprocessor.

- Stream processors—Each stream processor has its own set of local registers and contains an integer and floating point arithmetic logic unit (ALU). Current devices have 8 stream processors per multiprocessor.

- Shared memory—Memory space that is visible to all stream processors within a multiprocessor. Different multiprocessors cannot access each other's shared memory. Current devices have 16 kB of shared memory per multiprocessor. Shared memory is useful for algorithms that need to exchange data elements between parallel calculations.

- Constant cache—An 8 kB cache which provides quick access to global constants.

- Texture cache—An 8 kB cache which optimizes reads to 1D, 2D or 3D arrays which are read in a spatially localized manner. To support graphical applications, the texture cache also provides hardware linear interpolation between array elements, which is useful in applications that want to interpolate a lookup table.

To control the hardware, the user writes a function called a *kernel*[4]. The user then specifies how many copies of the kernel to execute simultaneously by defining a *thread grid*. The thread grid organizes threads of execution into groups called *blocks*. A block of threads all run on the same multiprocessor,

---

[4]Not to be confused with the Gaussian kernel in kernel estimation.
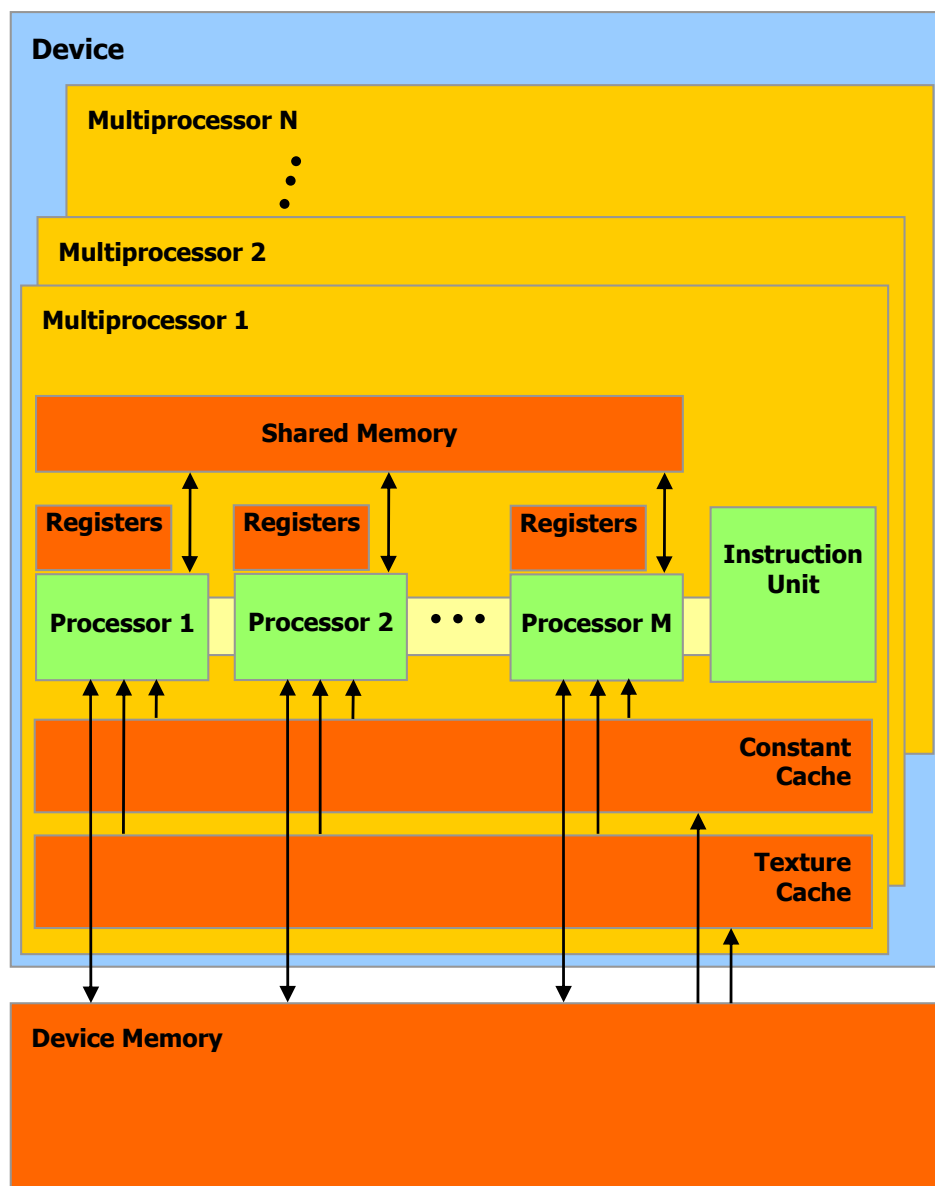
Figure A.2: Hardware design of a CUDA device. Figure from [101], pg 14.

and so can exchange data with each other during execution through the shared memory on the multiprocessor.

It is important to note that threads on a CUDA GPU are not like threads on a CPU. Threads are handled in groups of 32, called *warps*. All threads in a warp must be executing the exact same instruction. At an if-statement, if half the threads in a warp take one branch, and the other half take the other branch, the thread scheduler will split the group into two half-full warps. The blank slots in each warp will cause the ALUs to be underutilized in this case. This is a significant limitation to the kinds of algorithms which can be implemented efficiently in CUDA.

Also in contrast to threads on the CPU, CUDA threads are extremely lightweight and have essentially zero scheduling overhead. A normal, compute-heavy CPU program with 100 threads would have horrible performance due to the operating system overhead of switching between threads. CUDA programs typically have 10,000 threads in-flight at any given time. A significant change in perspective is required to effectively program for GPU, rather than CPU, parallelism.

A.3 Kernel Estimation in CUDA

Kernel estimation is a trivially data-parallel task. Equation 7.38 can be implemented by computing each of the exponentials in parallel, then summing the terms together using a *reduction* stage. Reduction can be efficiently performed in parallel by doing pair-wise additions of adjacent elements, reducing the list of values to sum by half in each iteration.

In the context of a maximum likelihood fit, there is an even higher

layer of parallelism in the calculation. Not only does evaluation of the kernel estimator at given $\mathbf{x}$ require summing over n Gaussians, but we also need evaluate it at N different $\mathbf{x}$, corresponding to the actual events observed in the data set. In fact, it is much more efficient to compute the value of the kernel estimator at 64 different $\mathbf{x}$ at a time, rather than repeating the kernel call for each $\mathbf{x}$ individually.

A.3.1 Setup

Communication between the GPU and the CPU take place over a (relatively speaking) slow interconnect, the PCI-Express bus. To minimize the usage of this bus, we load all of the PDF events and the events from the data set into the device memory of the card once at the beginning of the program. At this stage, we can also apply the partitioning optimization described in Section 7.2.4. The GPUs used in this analysis had between 512 MB and 768 MB of on-card memory, necessitating the use of fewer partitions to conserve space.

Floating point numbers have three parts: a sign bit s, a *mantissa m*, and an *exponent p*. Single precision numbers use 1 bit for the sign, 23 bits for the mantissa[5], and 8 bits for the exponent, using 32 bits in total. The real number which corresponds to a single precision floating point number is defined to be

$$f = -1^s \times m \times 2^{p-127}, \tag{A.1}$$

where 127 is called the *exponent bias*. The scientific notation format of floating point gives it a large dynamic range, suitable for representing physical

---

[5]By requiring non-zero floating point numbers to be normalized in a particular way, the IEEE 754 standard can define an implicit leading mantissa bit of 1. This gives the single precision mantissa effectively 24 bits of precision while only using 23 in the storage format.

quantities of vastly different magnitudes.

To further save space in device memory, a special "fixed point" representation of PDF events was used instead of floating point. Fixed point representations eliminate p, leaving just a fixed exponent bias. To represent the observable t_{ij} on the GPU, I used the fixed point format:

$$f = m \times 2^{-11} \tag{A.2}$$

where m is a signed, 16-bit integer between -32768 and 32767. The fixed point format can represent real numbers between -16 and 15.999 with roughly 5 significant (decimal) figures of precision. To map the observable range of t_{ij} to this interval, we use the mean $\bar{t}_j$ and the RMS σ_j of the entire PDF as a basis. Then each observable takes the form:

$$t_{ij} = \bar{t}_j + \sigma_j \times (m_{ij} \times 2^{-11}).s \tag{A.3}$$

On the GPU, we only need to store the m_{ij} for each event, and the $(\bar{t}_j, \sigma_j)$ for each PDF, cutting the total memory requirement in half compared to normal floating point. In exchange, we have introduced some quantization error into the PDF events. Five significant figures is sufficient to represent any observable in SNO with realistic accuracy, so this is a negligible effect.

Since we are using the semi-adaptive bandwidth method, there are only a small number of unique bandwidths h_{ij} in a given PDF. We only need to store the partition number of the event to be able to retrieve the bandwidths from a lookup table. A 16-bit integer is sufficient for this as well. Thus, the GPU storage required for each PDF event is $2(d+1)$ bytes, rather than $8d$

bytes with 32-bit floating point values. This is not only a huge savings in device memory, but also a huge savings in device memory bandwidth, as each PDF event needs to be copied from device memory to the individual stream processors during the calculation.

Compared to the tens of millions of PDF events across all of the different signals, there are relatively few data events (roughly 27,000). The observables for data events are stored in the standard single precision floating point format.

A.3.2 Evaluation

With the PDF and data events preloaded onto the card, the evaluation of the kernel estimator for all of the data events proceeds as shown in Figure A.3. The mapping of work onto the CUDA thread-block structure is as follows:

- A block contains 64 threads, with each thread responsible for accumulating the sum of Gaussians at the location of a single data event.

- A total of 96 blocks are launched, with 1/96 of the total number of PDF events assigned to each block. Different blocks read disjoint sets of PDF events.

Prior to the beginning of the GPU code, the CPU code loads all of the current systematics parameters into the constant memory on the GPU. Constant memory is chosen because it is very efficient at broadcasting small amounts of data to every thread.

Load Bandwidth Table

In the first stage of GPU execution, the threads in a block work together to copy the small bandwidth table into the fast shared memory of each multi-
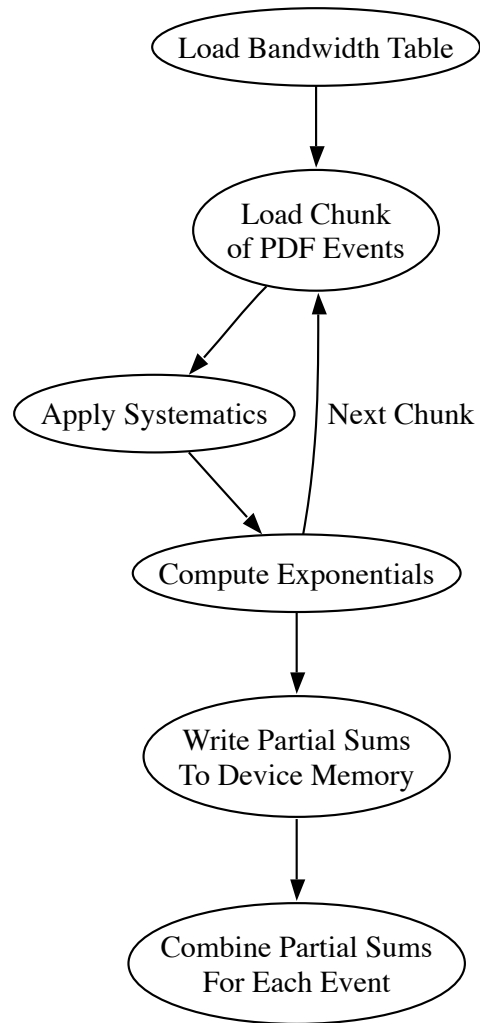
Figure A.3: Flowchart for evaluation of the kernel estimator on the GPU.

processor. These values are frequently accessed so copying them closer to the stream processors gives a major speed boost.

Load PDF Chunk

The main body of the evaluation code is a sequence of load, transform, compute stages. In the load stage, the 64 threads in each block work together to copy a "chunk" of 64 PDF events into the shared memory. During this process, the event observables are converted from the fixed point representation to the usual floating point representation.

Apply Systematics

The current parameters for systematics are read from constant memory, and the transformations are applied to the PDF events in shared memory. To allow systematics to alter the bandwidths, the corresponding bandwidth vector for each PDF event is also looked up by index and copied to a temporary location in shared memory before systematics are applied.

As a bonus, the GPU hardware provides free linear interpolation in 1D, 2D, and 3D. Systematics that use lookup tables, like the reweighting events by the $P_{\nu_e \to \nu_e}$ survival probability table, make use of this hardware to sample between table entries with no additional arithmetic.

Compute Gaussians

At this stage, the role of each thread in the block switches. Before, each thread was responsible for loading and transforming a PDF event into shared memory. Once that is complete, the thread then becomes responsible for accumulating

the sum of exponentials for a single data event. Every thread in the block loops over the 64 PDF events, computes the exponential as shown in Equation 7.38, and updates the sum in a thread-local register.

A kernel estimator is a large sum of potentially millions of terms, which raises the problem of roundoff error. Floating point arithmetic, being based on finite precision scientific notation, is very good at limiting the growth of fractional error in multiplication. When performing addition, errors can accumulate if operands of very different magnitudes are added together. For example, suppose we are working in a base-10 floating point system with 4 significant digits. The sum,

$$1.000 \times 10^1 + 3.142 \times 10^0 = 1.314 \times 10^1, \tag{A.4}$$

has an absolute error of 2×10^{-3}, a reasonable error relative to the precision of our floating point numbers. However, suppose we perform the sum,

$$1.000 \times 10^1 + 3.142 \times 10^{-4} = 1.000 \times 10^1. \tag{A.5}$$

The error now is 3.142×10^{-4}, also a small value, but the answer is the same as one of the operands. If this addition is being performed 1 million times, the total sum will be:

$$((1.000 \times 10^1 + 3.142 \times 10^{-4}) + 3.142 \times 10^{-4}) \ldots + 3.142 \times 10^{-4} = 1.000 \times 10^1. \tag{A.6}$$

Due to the intermediate rounding at each stage, the answer is the same as before, and the error in this sum is 3.142×10^2! The easiest solution to this problem is to use floating point numbers with more significant digits. For this

reason, double precision arithmetic, which expands the mantissa from 23 to 52 bits, is most commonly used in scientific computing. The GPUs used in this analysis have no double precision hardware, and software emulation of double precision on single precision hardware is very slow.

A simpler solution was described by Kahan[107] in 1965, now called *Kahan summation*. In this algorithm, two accumulator registers are used: a sum register, and a *compensation* register. Let `sum` be the summation register, `c` be the compensation register, and `terms[]` be an array (indexed from 0) of size `N` of values to sum. The Kahan summation algorithm is:

```
sum = terms[0]
c = 0.0
for i = 1..N-1
begin
    y = terms[i] - c
    t = sum + y
    c = (t - sum) - y
    sum = t
end
```

By continually accumulating the error in `c` and reintroducing it into the sum, the loss of precision in large sums is reduced significantly. This algorithm requires four times as many additions as normal summation, which is much better than the 11-fold increase in additions required to emulate double precision using the DSFUN90 library[108].

Write Sums

Once all of the PDF chunks assigned to this thread block have finished, each thread in the block writes to device memory the partial sum it has accumulated.

Reduction

Finally, to obtain the PDF values for the data events being evaluated, a second GPU function is called which adds together the 96 partial sums (one for each thread block) for each data event in parallel. Once this process is complete, the value of the PDF for every event in the data resides in device memory, and can be copied back to the CPU for use in calculating the likelihood function.

A.4 Summary

The production of inexpensive, commodity CPUs has fueled the rapid expansion of computational power available to physicists. We are seeing the first stages of a secondary expansion of scientific computing power, driven by massively data-parallel hardware, like graphics processing units. For highly parallel tasks, current GPU hardware offers a ten-fold increase in floating point capability that can be added to any recent workstation for $300 or less.

The development tools required to harness this power are still in their infancy, though. NVIDIA's CUDA toolkit requires rethinking and rewriting code to run calculations on a GPU. Having invested the effort, we have found that kernel estimation is particularly amenable to GPU parallelization. The techniques described in this appendix give an additional 8x increase in speed

over our most highly optimized CPU kernel estimation code. A task like gradient descent maximum likelihood fitting does not distribute well over many computers, so boosting the speed of a single computer with a GPU is the ideal solution.

Appendix B

Constraints and Correlations: Is Your Pull Normal?

The discussion of pull distributions in Section 8.2.5 notes that the presence of constraints and strongly correlated parameters can cause the pull distribution to have a width less than 1.

To illustrate this problem, we will consider a toy model in which we have two signals A and B, and we wish to measure the number of events N_A and N_B in our data set. Let's suppose we were late on the day analysis projects were handed out, and got stuck with with the fit no one wanted to do, as shown in Figure B.1. Signals A and B have nearly identical PDFs in the observable x, and are therefore very hard to separate. With a correlation coefficient between N_A and N_B of -0.99, the fit uncertainty on the two quantities is likely to be large.

But suppose we are clever (or desperate) and discover an independent way to measure N_B and include it in our fit. We go to verify our fit proce-
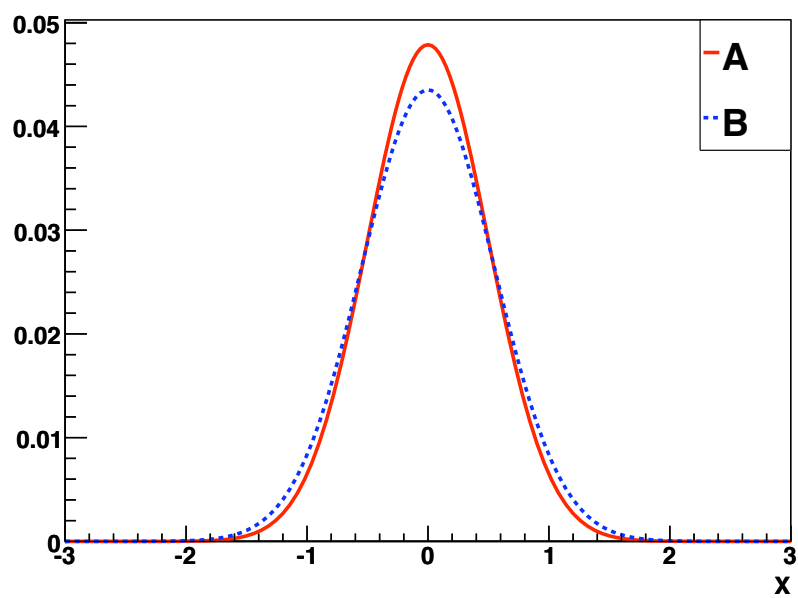
Figure B.1: PDFs for the two signal types, A and B, defined over a hypothetical observable x.

dure with 200 test data sets, and suddenly disaster strikes! Figure B.2 shows that adding the penalty term makes the pull distribution too narrow for both parameters, but N_B much narrower than N_A.

B.1 Discussion

To understand what is going on here, it is instructive to consider the extreme case. Suppose A and B have identical PDFs. Then the negative log-likelihood without a constraint has no unique minimum point. Instead there is a minimizing line,

$$N_A = N_{data} - N_B, \tag{B.1}$$

where N_{data} is the number of events in the data set being fit. The addition of a external constraint $(\tilde{N}_B, \tilde{\sigma}_B)$ picks out a unique minimum,

$$N_A = N_{data} - \tilde{N}_B \tag{B.2}$$

$$N_B = \tilde{N}_B. \tag{B.3}$$

According to our Poisson fluctuation procedure, the total number of events N_{data} varies from set to set, so N_A will as well. But N_B is uniquely defined by our constraint, and is independent of the data set. Thus, we get the pull distribution shown in Figure B.3. The pull width of N_B is zero, and the pull width of N_A is close to $1/2$.

Is our fitting procedure incorrect? Signals A and B are identical, so our data set can only estimate the sum $N_A + N_B$. The external constraint drives the estimation of N_B entirely. Not surprisingly, the uncertainty on N_B returned by the fit is 10%, precisely the uncertainty on our constraint. So,

285
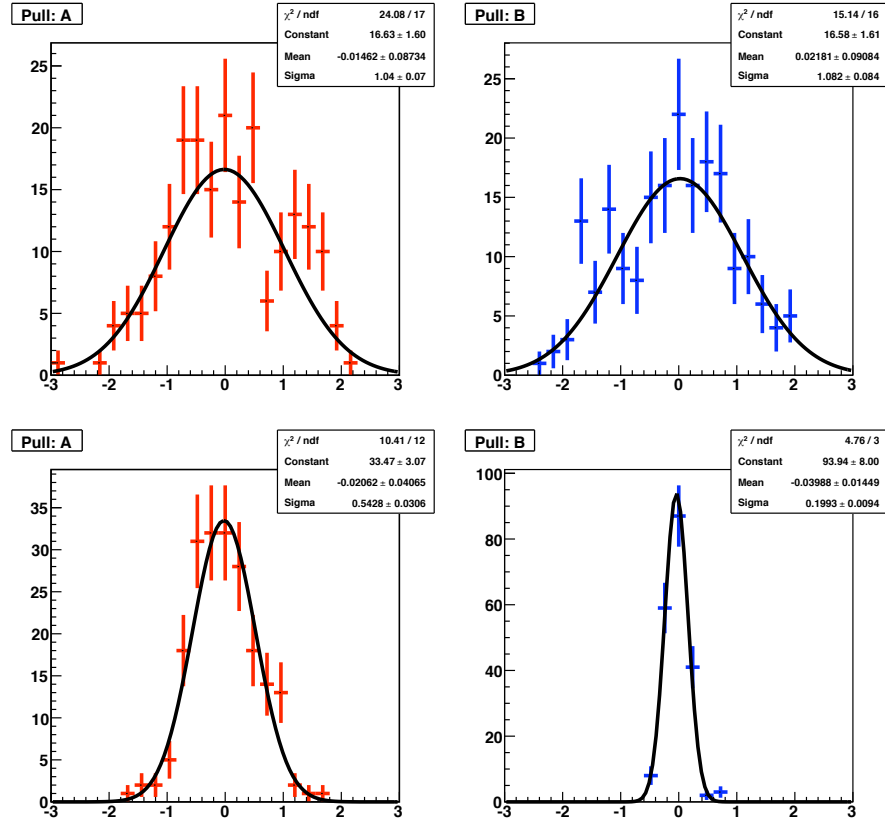
Figure B.2: The top row shows the pull distributions for N_A and N_B with no external constraint. The bottom row shows the pull distribution after a 10% constraint on N_B has been added to the likelihood. The average value for N_A and N_B is 500 events, so the Poisson fluctuations are 4.5%.
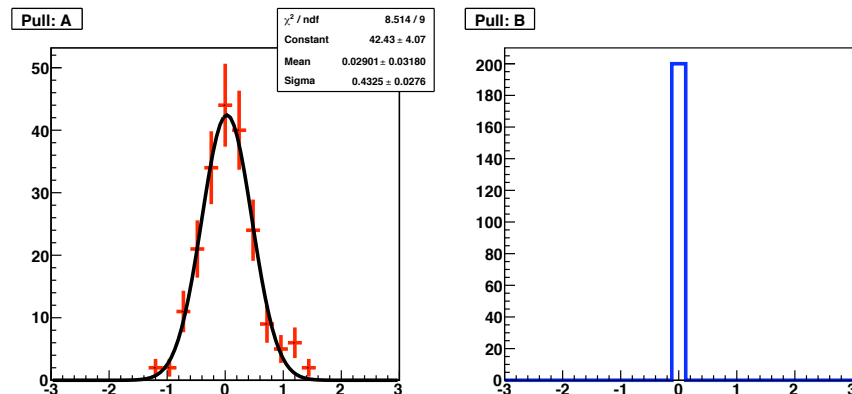
Figure B.3: The pull distribution for A and B when both signals have identical PDFs, and B has a 10% external constraint.

we have a fitting procedure which returns an unbiased estimate of N_B with a meaningful fit uncertainty, and yet the pull distribution is wrong.

B.2   Solution

The solution to this problem emerges upon closer consideration of the pull procedure. Like good frequentists, we generate many test data sets in order to verify the statistical distribution of an ensemble of hypothetical experiments. If we were to run SNO 100 times, would our signal extraction yield an NC flux normally distributed around the true value, with a width consistent with the uncertainty reported by the fit? This is why each test data set needs a randomized selection and number of events.

However, our constraints are *also* measurements. Putting the penalty terms into the likelihood makes the constraint appear to be independent of the data, but this is not true. A hypothetical repetition of the SNO experi-

ment would re-record all the neutrino runs, but it would also repeat all of the calibration data and water radio-assays. The constraints are really a part of the SNO data set, and as such, should be varied between test data sets.

The pull procedure can be augmented with a step for each fit where the central value of each constraint $\tilde{N}_i$ is independently Gaussian-fluctuated according to the uncertainty $\tilde{\sigma}_i$. Note that the constraint term is being changed between fits, but does not directly affect the data selection itself. The number of each event type is still drawn from a Poisson distribution. In the case where the fit parameter is a systematic, such as energy scale, the procedure is the same. The energy scale of the events is not changed from fit to fit, but instead the constraint term is varied.

Figures B.4 and B.5 show the pull distributions for both toy models of signals A and B, including the constraint-variation step. The width of the pull distribution in the strong anti-correlation case is now consistent with 1. Even the degenerate case where A and B are indistinguishable now has a correct pull width. The procedure solves the problem, and provides a way to continue to use the pull distribution as a diagnostic tool in fits with both external constraints and highly correlated parameters.
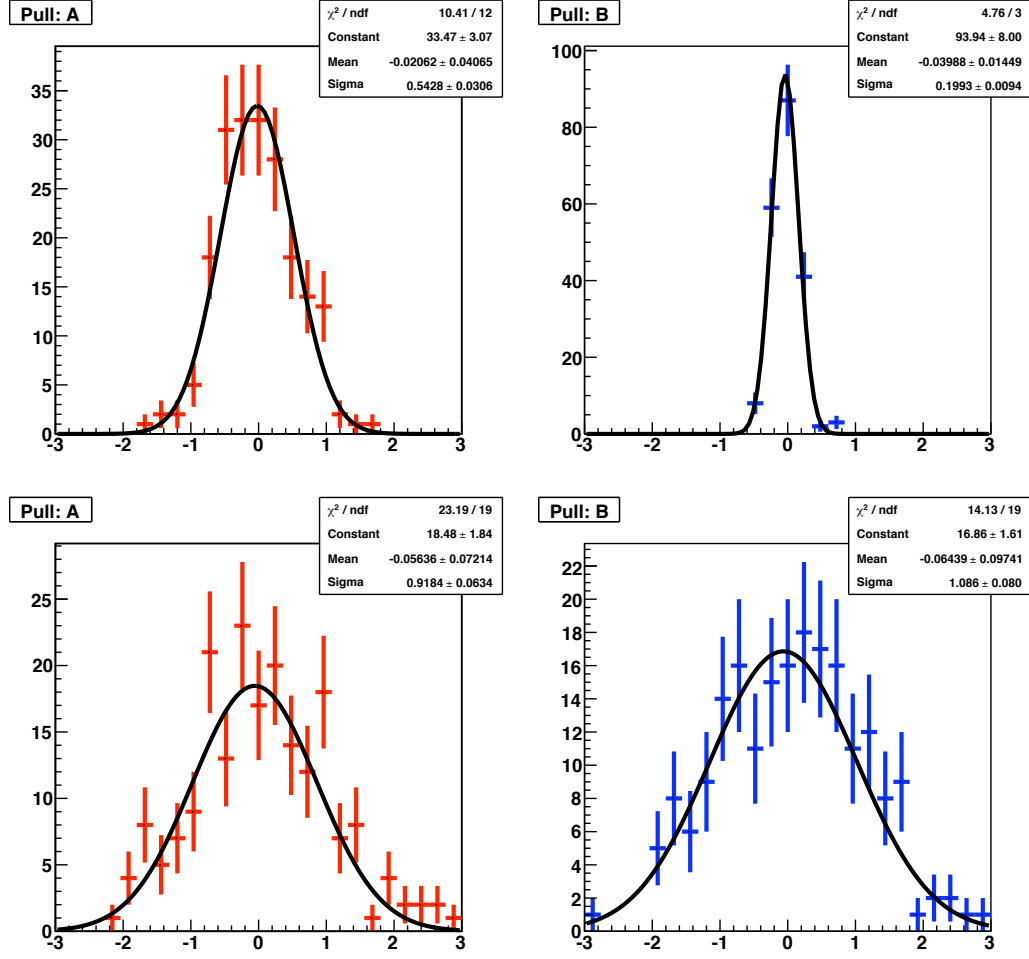
Figure B.4: Pull distributions for the strongly correlated model described in Figure B.1. The top row shows the pull distribution from before, where the external constraint on N_B is kept constant. The bottom row shows the pull distribution when a new constraint is draw for every fit.

289
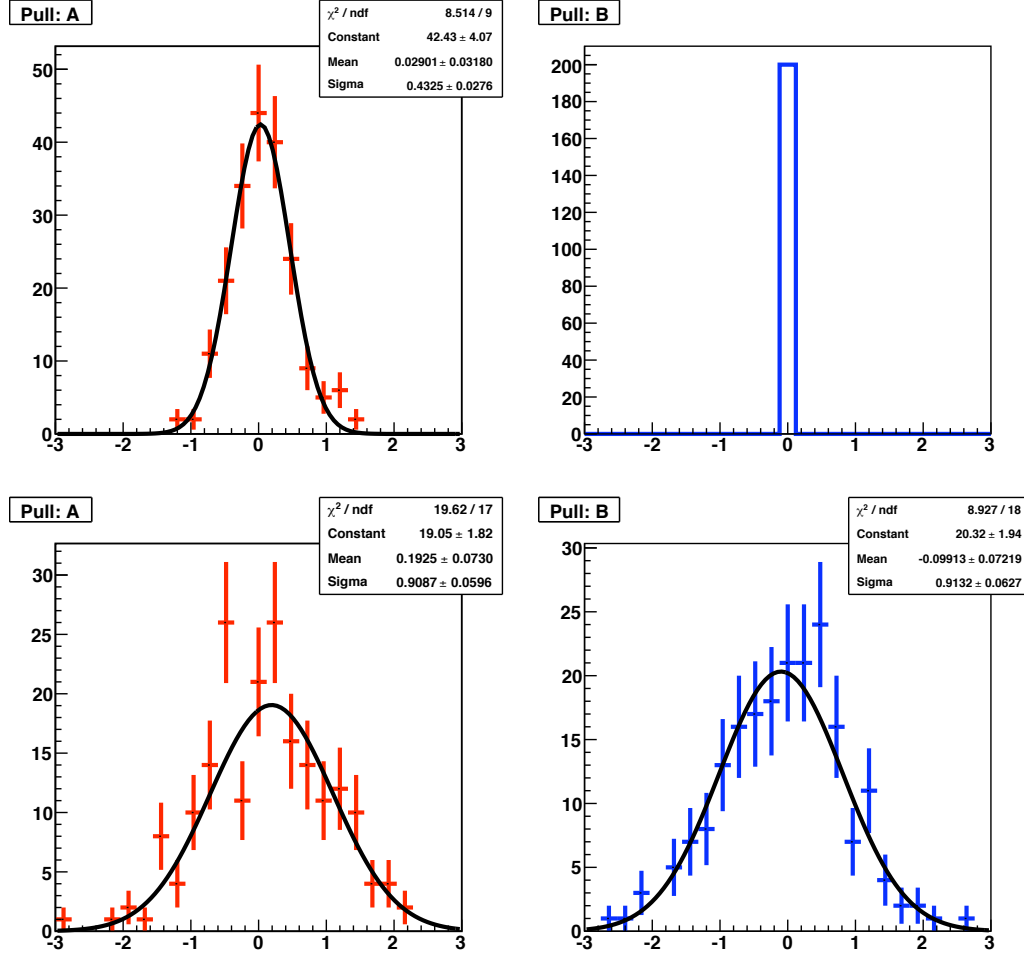
Figure B.5: Pull distributions for the model where A and B are indistinguishable. The top row shows the pull distribution from before, where the external constraint on N_B is kept constant. The bottom row shows the pull distribution when a new constraint is draw for every fit.

Appendix C

nFTK: Energy Estimation for the NCD Phase

FTK, as described in Chapter 5, models direct light, Rayleigh scattering, and reflections off the acrylic vessel and the PMTs. With the introduction of the ^{3}He proportional counters ("neutral current detectors" or "NCDs") into the center of the SNO detector in the third phase of SNO, this model needed to be updated. The outer nickel surfaces of the NCDs are a light grey color, both absorbing and diffusely reflecting photons that strike them.

The original motivation for FTK was to push back on the exponential wall of backgrounds below 5 MeV by improving the resolution of the energy estimator. However this same technique benefits the NCD phase, by compensating for lost photons due to the insertion of opaque NCDs into the central detector volume. Hardware changes were made as well to help mitigate the lost of photons, including a 50V increase in the PMT voltages, and replacement of the electronics crate backplanes with a lower noise design that permitted
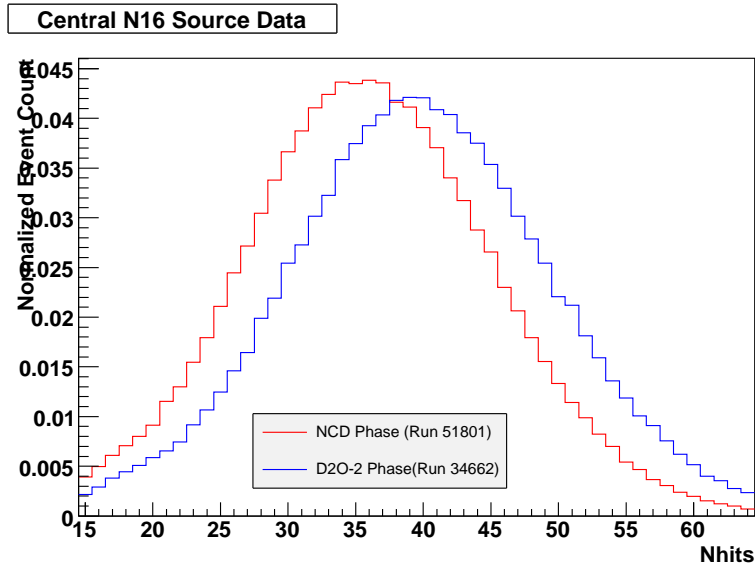
Figure C.1: Comparison of central N16 source runs between from the post-salt phase and NCD phase after hardware modifications to increase collection efficiency. The NCD phase still shows a 10% light loss.

channel thresholds to be reduced. Figure C.1 shows the NHIT spectrum from central N16 runs in both the interim D_2O phase in 2003, and the NCD phase after the hardware upgrades. There is still a 10% loss of photons at the center, which points toward a need for a wider time window in the energy estimators in order to recover lost photon statistics.

FTK's use of the SNOMAN geometry, and reliance on photon tracking makes it fairly straightforward (in principle) to augment the code with new detector components. This appendix describes these changes, as well as the tuning performed for the first reprocessing of the NCD phase data.

C.1 Code Modifications

The FTK processor has several phases of operation:

- Processor initialization: Run at the beginning of the SNOMAN job. Large lookup tables are loaded or generated, including one which is used during fitting to estimate the number of hits from PMT reflections.

- Event setup: Run for each event, prior to fitting. The bulk of the computation is done here as per-PMT hit probabilities are computed at the reconstructed event vertex for the following kinds of light:

  - Direct light

  - Rayleigh scattered light

  - Reflections from the acrylic vessel (both inner and outer surface)

  - PMT reflections

- Maximum likelihood fitting of the number of Cherenkov photons produced by the electron. Here the PMT hit probabilities are multiplied by the energy-dependent angular distribution of Cherenkov photons, and then summed to compute the probability of a single Cherenkov photon generating a hit anywhere in the detector within the FTK time window. The number of Cherenkov photons is used as a proxy for energy in the fit since the conversion of photons produced to hits detected is straightforward with the probabilities computed above.

- Output of the most probable energy given the most likely estimate of the number of Cherenkov photons produced.

Each light component is defined to be composed of photons which have exactly one (or none in the case of direct light) scattering or reflection interaction prior to detection. A photon which reflects off the AV and is then detected by a PMT counts as an "AV reflection hit." However, a photon which reflects off a PMT, then Rayleigh scatters before detection is not modeled here due to its exceedingly low probability. These second order effects, and other such complications, are handled by correction constants which are tuned using Monte Carlo events. Tuning done for the NCD phase will be described in Section C.2.

Since we only model photon processes in FTK to first order, the NCDs only need to be treated as perfectly black objects when we correct the computation of each light component. In addition, there is a new first order process, that of a photon reflecting off of the nickel wall of an NCD and then being detected by a PMT.

When operating on Monte Carlo events, we have perfect knowledge the photon paths and can classify each PMT hit by its "light type" using the above definitions. We can compare this to the estimated number of each type of hit as output by FTK. This technique is our primary tool for verifying the code in fact estimates the light components correctly. Figure C.2 shows the difference between the FTK estimates and the MC-derived number of hits as a function of radius for the salt phase and for the NCD phase prior to the nFTK patches. Deviations from zero indicate what FTK corrections were needed in order to make FTK ready for the NCD phase.
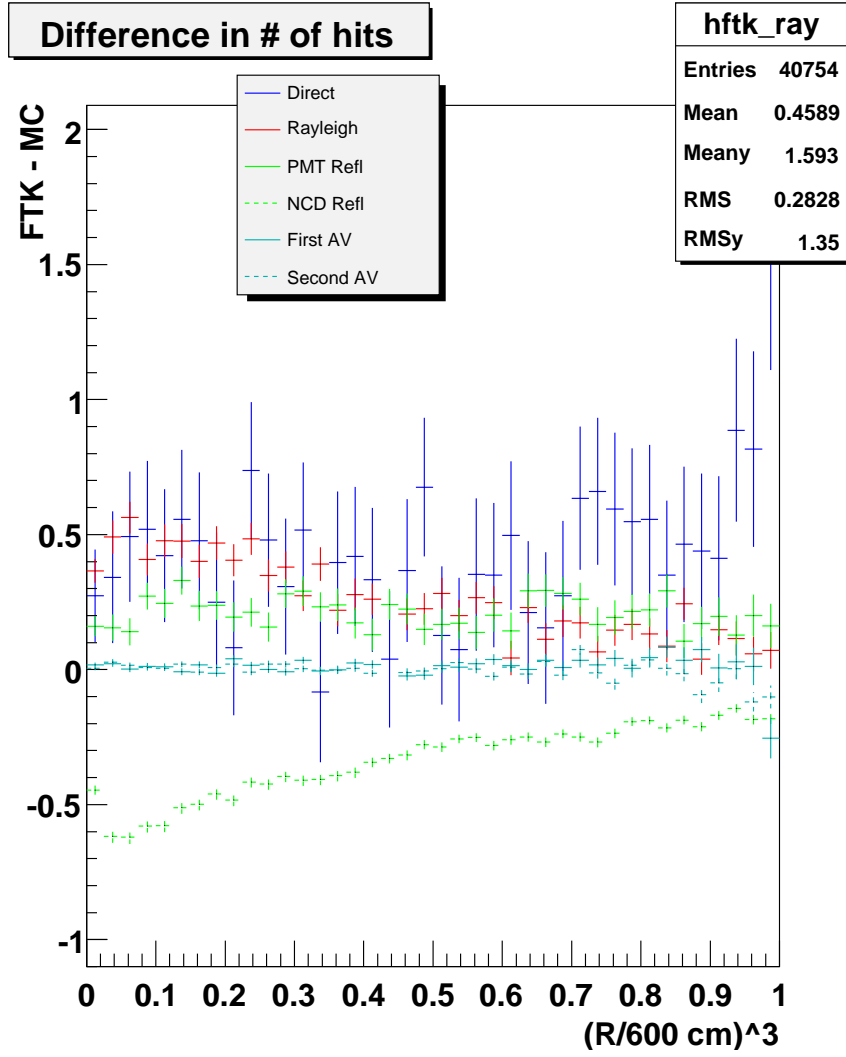
Figure C.2: Average difference between FTK estimate and true number of hits for each light component as a function of R^3, prior to nFTK changes.

C.1.1   Direct Light and PMT Reflections

The direct light calculation traces a test ray to the PMT, iterating up to 5 times to refine the initial direction until the test ray hits the center of the PMT, accounting for the effects of refraction at the AV boundaries. For events in the D_2O, this ray tracing uses the same SNOMAN code which propagates photons in the Monte Carlo, so all components of the detector are accounted for. If the ray intersects any volume which is not D_2O, H2O or acrylic prior to reaching the PSUP, the tracking code reports back to FTK that the photon is blocked. The probability of reaching that PMT by direct light is then recorded as zero. Otherwise, the probability of detection is computed, taking into account the attenuation of the optical media, and the collection efficiency of the PMT, including effects like apparent solid angle with refraction, channel efficiency and angular response of the PMT. The wavelength distribution of Cherenkov light is also integrated over in order to average out wavelength-dependent coefficients.

At the same time the probability of reflection from that PMT and detection at another PMT is computed using the absorption coefficients already obtained, the angle of incidence at the PSUP, and a lookup table generated during FTK initialization.

Because the initial raytracing to the center of the PMT is done with the full SNOMAN geometry, it automatically accounts for NCD shadowing in a binary way. The test ray will intersect the nickel, and the direct light probability for a photon traveling from the event vertex to the PMT in question will be zero. A more accurate estimation of the NCD shadowing would require firing many test rays at the surface of the PMT in order to compute the total

fraction of solid angle which is visible. Effectively, it would be a numerical integral over the surface of the PMT, and would be very computationally expensive.

However, multiple samples are not necessary. Since the apparent size of PMTs is small and the distance between PMTs and NCDs is large, parallax effectively randomizes the apparent NCD positions relative to the PMT centers, as viewed from the event vertex. Thus, just checking the center point is on average a fair estimate of the NCD shadowing fraction. Once you sum over all of the PMTs, the additional variance in the direct light estimate introduced by binary shadowing has been reduced by $1/\sqrt{N}$ and is of little consequence. This argument applies to any other obstruction in the detector (such as the source manipulator) which FTK also includes in its estimate of the direct light.

NCDs are therefore naturally accounted for in the direct light calculation by FTK without code modification. Similarly, the first part of the PMT reflection calculation which tracks the photon from the event to the PSUP also needs no change. The lookup table used to do the second part of the PMT reflection calculation does need to be regenerated, and this is described in Section C.2.

C.1.2 NCD Reflections

The new calculation of hits coming from NCD reflections piggybacks on the direct light calculation. While the direct light code did not need to be modified to account for NCD shadows, an additional test was added to trap rays which were aborted due to intersection with the nickel outer volume of a NCD. At the point of collision with the nickel, a photon bomb is thrown and each photon

is tracked to a PMT, much as is done for the direct light calculation. The direction of bomb photons is drawn from a uniform angular distribution over the hemisphere centered on the surface normal. Diffuse reflection is the only kind of reflection simulated for nickel in SNOMAN, so a uniform photon bomb should produce results equivalent to the Monte Carlo. The probability for NCD reflection is then the product of the transmission probability in all of the optical media both before and after the point of reflection, the probability of reflection from nickel, and the probability of detection at an online PMT. Much as in the direct light calculation, the wavelength distribution of Cherenkov light is integrated over in order to account for the wavelength dependent reflectivity of nickel, as well as transmission coefficients and PMT efficiencies.

The size of the photon bomb was left as a parameter to be controlled in the FTK configuration file. Testing revealed that in fact, there was no advantage to using a bomb size greater than 1! Figure C.3 shows the mean number of NCD reflection hits as function of radius when computed with a bomb size of 1 and a bomb size of 10. At the center, where roughly 15% of the solid angle is blocked by NCDs, we would expect approximately 1400 NCD intersections per event, each accompanied by a "bomb" of size 1. For comparison, the Rayleigh scattering calculation uses a bomb of size 1000 per event to estimate the total number of scattering hits.

A scaling problem remains with the NCD reflection calculation. Figure C.4 shows the comparison as a function of radius of the number of NCD reflection hits predicted by FTK to the number of hits from the Monte Carlo. The shape of the distribution is very close, but the overall scale is wrong. In fact, a scale factor of 2.17 is sufficient to bring the FTK estimate up to the Monte Carlo number, as shown in C.5. The cause of this problem was not
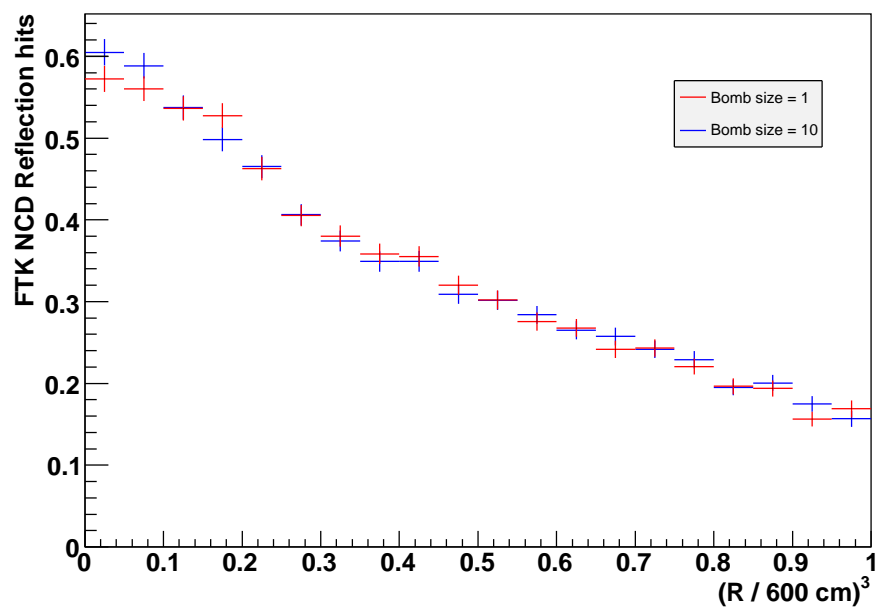
Figure C.3: FTK estimate of NCD reflection hits as a function of R^3 when computed with bomb of size 1 and size 10.

identified before it was necessary to move forward with NCD reprocessing.

C.1.3 Rayleigh Scattering

Rayleigh scattering in the detector model is measured using a bomb of 1000 photons which are forced to scatter inside the PSUP radius. The probability of detection is then

$$P(\text{hit}) = P(\text{hit}|R_{\text{scatter}} < R_{\text{PSUP}}) \times P(R_{\text{scatter}} < R_{\text{PSUP}}).$$

The photon tracking used to compute the first factor in this equation is simplified to only include 4 volumes: D_2O, AV, H2O, and the PSUP boundary.

There are two scenarios in which this calculation will overestimate the amount of Rayleigh scattered hits. First, there is the possibility of intersection with an NCD prior to the scattering point. Second, there is the possibility of collision with an NCD after scattering. The most straightforward way to deal with this is to replace the simplified tracking used to trace photons in this calculation with the full photon tracking method used in the direct light calculation.

Not surprisingly, this is rather slow, so a fast version of the correction was implemented to exchange some accuracy for speed. The attenuation length in the D_2O is on the order of tens of meters, so for events in the D_2O, scattering is most likely to occur well outside the NCD array. Thus, the primary effect of the NCDs is to block photons before they have a chance to scatter. We can just scale the estimated number of scattered hits by the total fraction of solid angle at the event vertex which is unblocked by NCDs. The visible fraction is computed during the direct light calculation by dividing the solid angle of
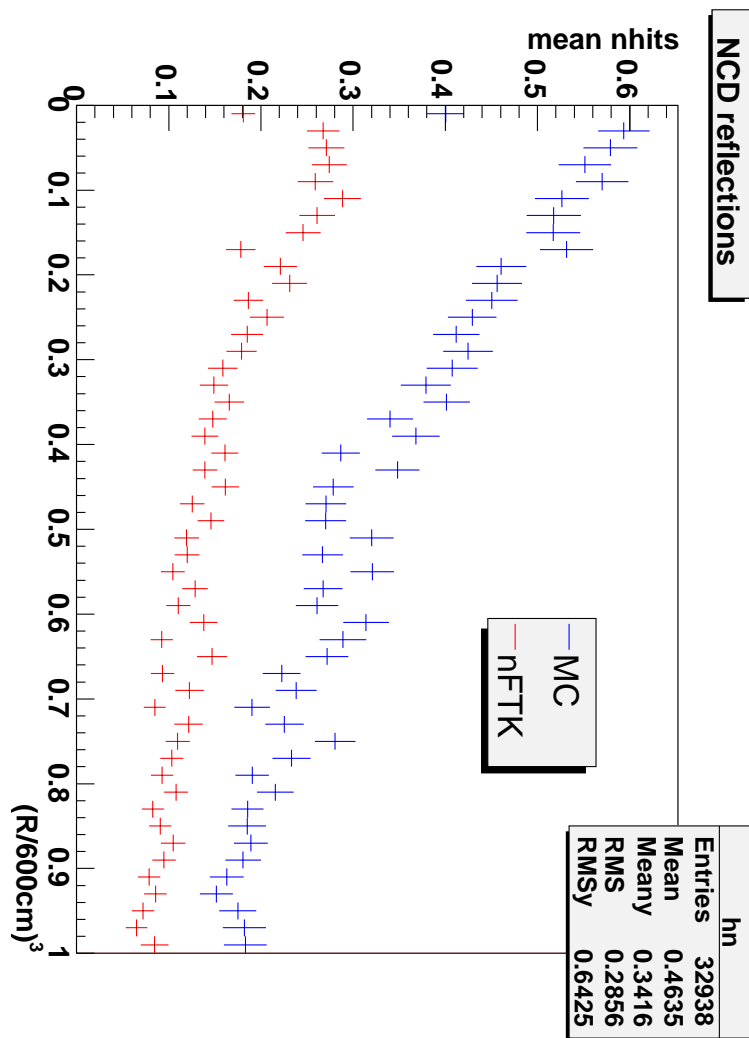
Figure C.4: Number of NCD reflection hits as calculated by FTK compared with Monte Carlo with no scaling.
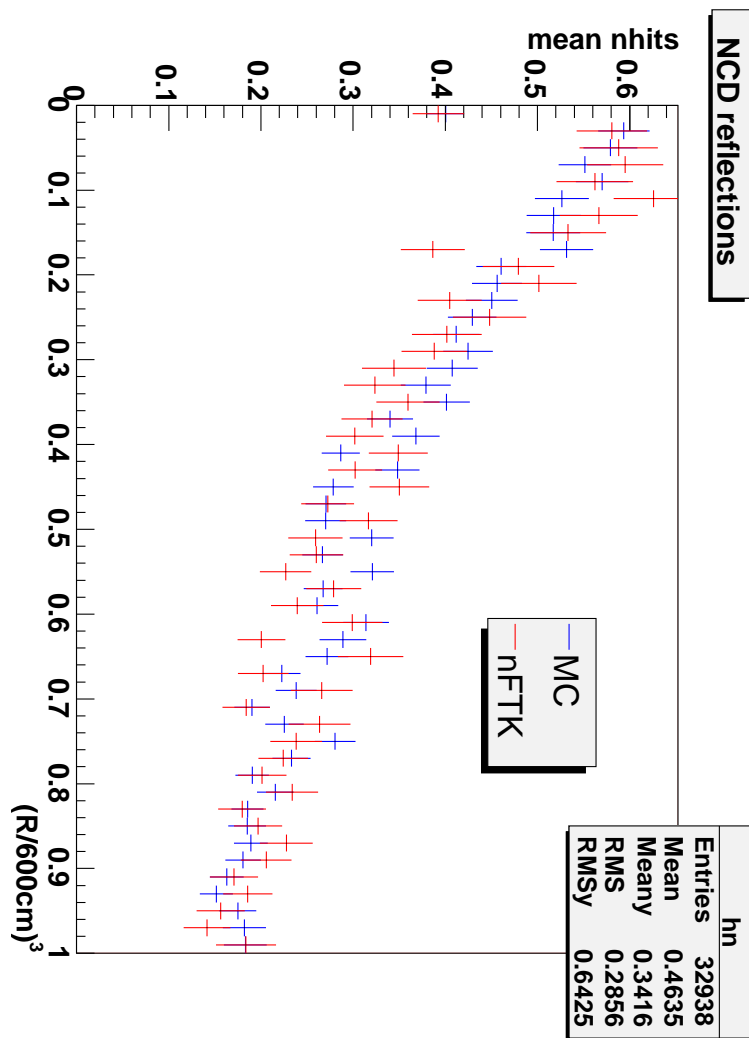
Figure C.5: Number of NCD reflection hits as calculated by FTK compared with Monte Carlo. A scale factor of 2.17 has been applied to the FTK estimate.
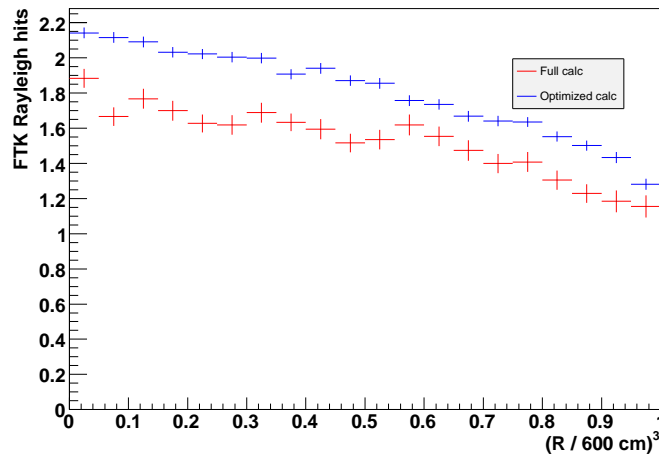
Figure C.6: Comparison of two methods for altering the FTK Rayleigh scattering estimate to include the NCDs. The full version uses the same photon tracking as the Monte Carlo, and the optimized version just scales the old calculation by the fraction of the total solid angle not blocked by NCDs.

all unblocked PMTs by the solid angle of all PMTs. Figure C.6 compares this technique with the slower, full tracking version of the Rayleigh scattering estimate. The difference is about 10-20%, which is acceptable given the overall small contribution of Rayleigh scattering to the total number of hits.

C.1.4   AV Reflections

The AV reflection hits are calculated using a photon bomb of 500 photons. The code uses the simplified tracking model, similar to the Rayleigh scattering, in which there are no NCDs. The AV reflections are a very small part of the total light, so it was decided not to modify this method to correct for the NCD absorption.

C.2   Tuning of FTK Constants

FTK tries to minimize its use of fixed constants in order to allow it to dynamically adapt to the detector state based upon Monte Carlo settings. However, sometimes it is necessary for speed reasons to precompute information and load it at the start of every job. These FTK-specific constants are read from from the DFTK and the TFTK banks.

C.2.1   DFTK Tables

In particular, DFTK banks 1 and 2 are used to speed up the calculation of PMT reflections. Bank 1 holds a table of average reflection probability as a function of wavelength and incidence angle. Bank 2 contains the probability distribution of the outgoing $(\cos\theta', \phi')$ direction as a function of incoming angle. These tables are used in a quick Monte Carlo calculation performed during the initialization phase of FTK. A large photon bomb is used to construct a table which gives the probability of a PMT hit as a function of incoming direction and wavelength. Because this calculation is done at the start of the run, and it uses the SNOMAN geometry code, it needed no code modifications to incorporate the NCDs. The static tables, DFTK 1 and 2, describe quantities local to the PSUP, so they also would not normally need to be regenerated to account for new volumes inserted into the detector.

However, these tables do depend on the angular response of the PMTs, which was modified by G. Orebi Gann to achieve better agreement between the 3D PMT model and laserball measurements. For this reason, DFTK 1 and 2 were regenerated using a custom piece of SNOMAN code provided by Dunford. The code collects the necessary data to generate the tables by running

Monte Carlo of photon bombs on a spherical shell between 820 and 825 cm. The simulation is effectively that of a diffuse, glowing sphere whose surface is just 20 cm away from the PSUP. Incoming and outgoing photon directions and wavelengths are recorded and processed into titles files by an offline FOR-TRAN program, also provided by Dunford. These tables are now stored in ftk_pmt_refl_ncd_20060322.dat and were added to the mcprod directory in the SNOMAN distribution as of release 5.0183. Figure C.7 shows the 2D distributions of reflection probability (DFTK 1) and the PMT angular response used during generation.

C.2.2   TFTK Constants

The TFTK bank holds several fixed constants used to adjust the magnitude of each light component to correct for biases in the FTK estimation methods. These constants are tuned using Monte Carlo events, usually $T = 5\text{MeV}$ uniform and isotropic electrons. Correct values of these constants were found by taking the ratio of FTK and MC-derived hits for several components as a function of R^3 and finding the average value. This will volume weight the correction factor, which is appropriate when we expect to fit events distributed over most of the detector. Constants which were modified from previous phases were:

- Direct light normalization[1]: 0.9852

- PMT reflection normalization: 0.9993

---

[1]This constant is also referred to as the "greydisk/3d PMT normalization" because FTK uses greydisk constants to determine PMT response, but the Monte Carlo uses 3dPMT in the simulation.
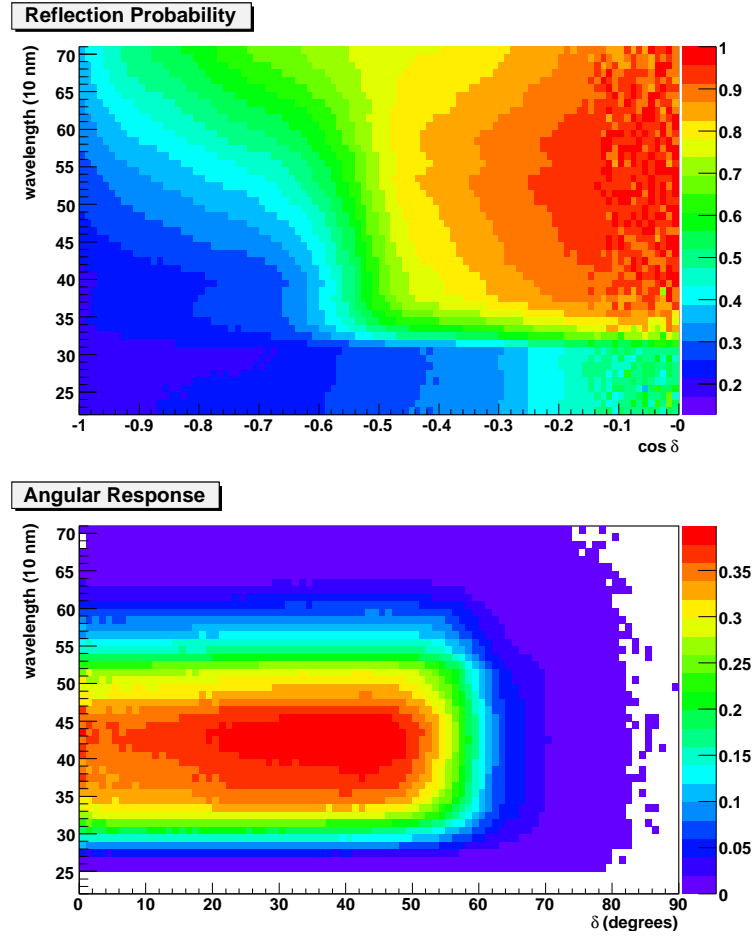
Figure C.7: Top panel shows PMT reflection probability as a function of incidence angle δ, and wavelength, as tabulated and stored in DFTK 1. Bottom panel is the angular response of the PMTs.

There are also the NCD-only FTK constants:

- Enable NCD-phase FTK modifications flag

- Set NCD reflection bomb size to 1.

- NCD reflection normalization: 2.17

Without the NCD flag set, all NCD-related changes to FTK are bypassed to preserve backward compatibility with previous phases. Theoretically, leaving this flag on while processing D_2O or salt phase data should have no effect. Nevertheless, the flag is left off by default and only turned on in the load_ncd_settings.cmd file.

C.3 Performance on Monte Carlo

All of the changes and tuning thus far have been designed to make FTK predict the amount of light in the Monte Carlo. Figure C.8 shows the difference between FTK and Monte Carlo for each light type. Agreement is quite acceptable out to the edge of the fiducial volume, but beyond that point, the deviations become quite pronounced. Figure C.9 shows the reconstructed energy of events as a function of radius.

Figure C.10 shows the reconstructed energy as a function of generated energy for isotropic and uniform electrons. The turn-up at low energies is caused by a minimum energy restriction imposed during fitting. FTK will not allow the event to have energy less than 2.0 MeV (kinetic). This clamps one end of the distribution, pushing the mean reconstructed energy for a given generated energy upwards.
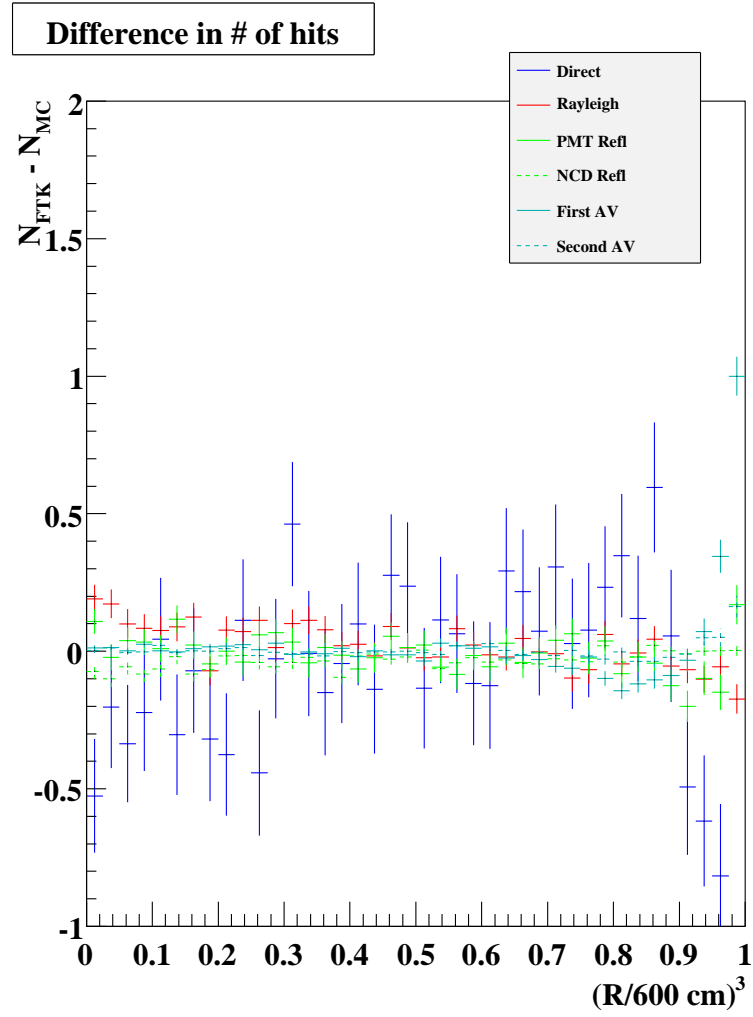
Figure C.8: Average difference in hits between nFTK and the Monte Carlo for $T = 5$ MeV electrons as a function of R^3. Fiducial volume cut is at $R^3 = 0.77$.
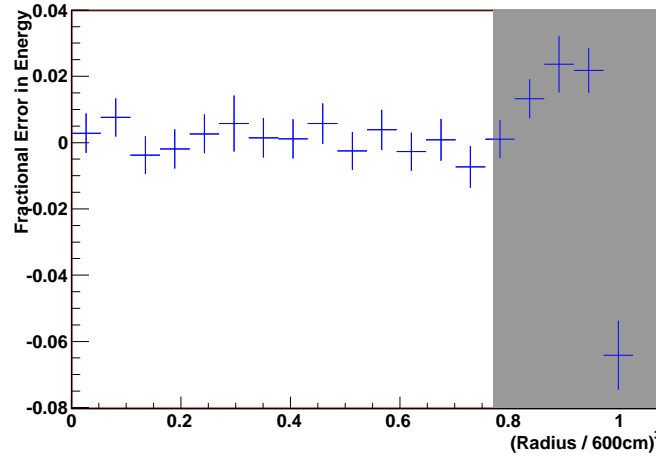
Figure C.9: FTK reconstructed energy as a function of R^3 for isotropic $T = 5$ MeV electrons. Shaded region is outside the fiducial volume.
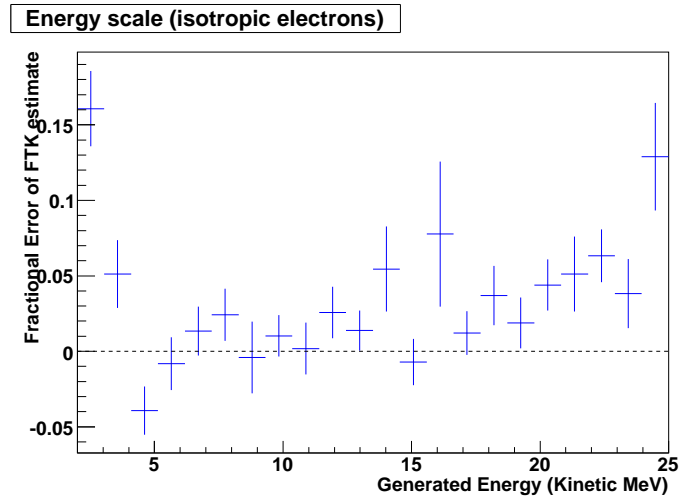
Figure C.10: FTK reconstructed energy as a function of generated energy for isotropic electrons.

At higher energies, the multiphoton effect takes over. FTK applies a correction to account for the probability of two photons hitting the same PMT. However, the probability of 3 or more photons in the same PMT is not included, which causes a visible upward bias beyond 12 MeV.

C.4   Global Tuning of Collection Efficiency

Once FTK has been made consistent with the Monte Carlo, the global collection efficiency must then be tuned in order to make the Monte Carlo consistent with the data. The number does not directly come from any particular physical quantity, but rather is the product of all other unaccounted for scale factors.

The collection efficiency is derived using central N16 calibration runs, under the assumption that the center of the detector is the place best modeled by the Monte Carlo. In order to speed processing, a small sample of runs spread throughout the NCD phase were selected, with the goal of having one central run from each of the approximately monthly N16 scans.

The procedure was then to process the ^{16}N runs with FTK set to "fast mode." In this mode, FTK only records the number of hits within its -25 to 150 ns time window, a quantity called N_{eff}. The equivalent quantities for RSP are N_{cor} and N_{chv}, which are the number of hits and Cherenkov photons, respectively, corrected for attenuation to be equivalent to an event in the center of the detector. The distributions of N_{eff} for data and Monte Carlo are fit with a Gaussian $\pm 1.66\sigma$ around the mean[2]. The ratio of data/MC was plotted, and fit with a constant. Under the assumption that N_{eff} is a linear function of the collection efficiency, the initial guess for the efficiency is scaled by the

---

[2]This interval was selected for consistency with previous RSP-based tunings.

best-fit ratio. Then the ^{16}N Monte Carlo is generated again with the new efficiency, and the process repeated in case non-linear effects caused our new estimate to miss the true value. Finally, the verification pass is run, in which the central runs along with low-rate and off-center runs are processed. FTK is configured to run the full fit and output the energy estimate. The resulting data is checked for unexpected biases as a function of position or rate.

In practice, this procedure is complicated by the presence of two energy estimators in SNOMAN, RSP and FTK. If the Monte Carlo perfectly replicated the time distribution of hits in the data, then tuning the collection efficiency with RSP or FTK would result in the same efficiency factor. However, we know that there are small discrepancies in the late light, which affect RSP and FTK differently. In order to deal with a mismatch, it was decided to split the difference so that FTK and RSP would deviate from the data by approximately the same amount.

Compensating for drift in the detector poses another complication. It is possible to allow the global collection efficiency to be a polynomial function of Julian date in SNOMAN, and thereby model drift in the detector response. Since the drift correction multiplies the global collection efficiency, we first fixed this constant, then fit a linear function to the ratio of data/MC for N16 runs between December 2004 and March 2006. The best-fit lines using both RSP and FTK information were averaged to give the unified drift correction.

Figures C.11 and C.12 show the final result of this tuning, for FTK N_{eff} and RSP N_{cor}, respectively. The Monte Carlo efficiency is 0.5% too high for RSP and 0.5% too low for FTK. In addition, a small residual drift in RSP is left, with a maximum magnitude of 0.2%. Figures C.13 and C.14 show the same N16 runs, but now in terms of reconstructed energy.
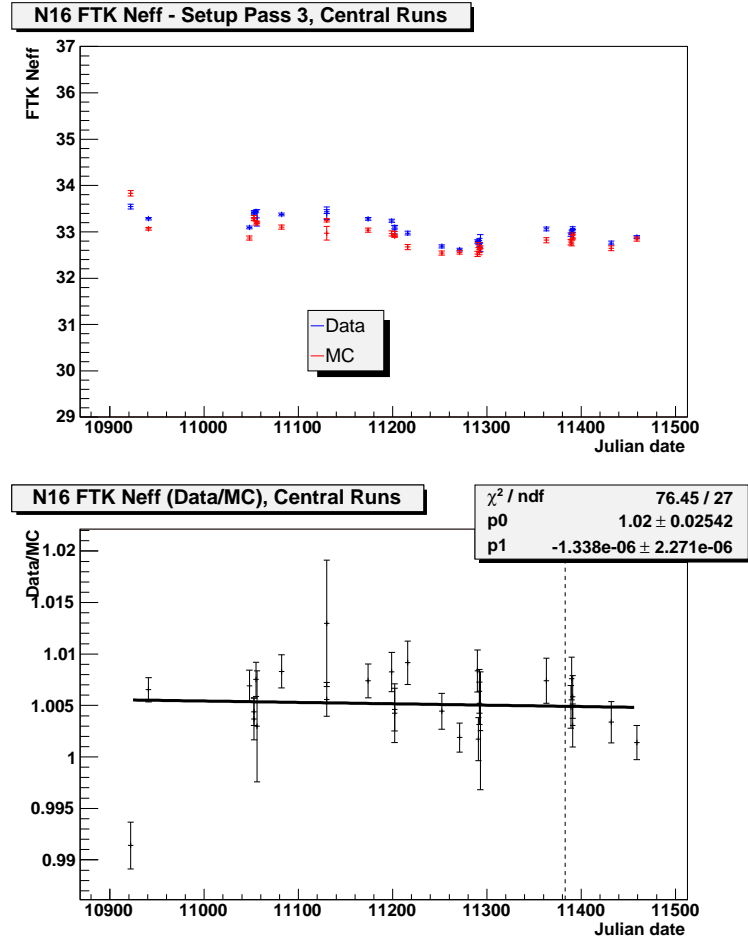
Figure C.11: Comparison of FTK N_{eff} for data and Monte Carlo of central N16 runs. Dashed line shows March 2006, the original cutoff for data to be used in first analysis.
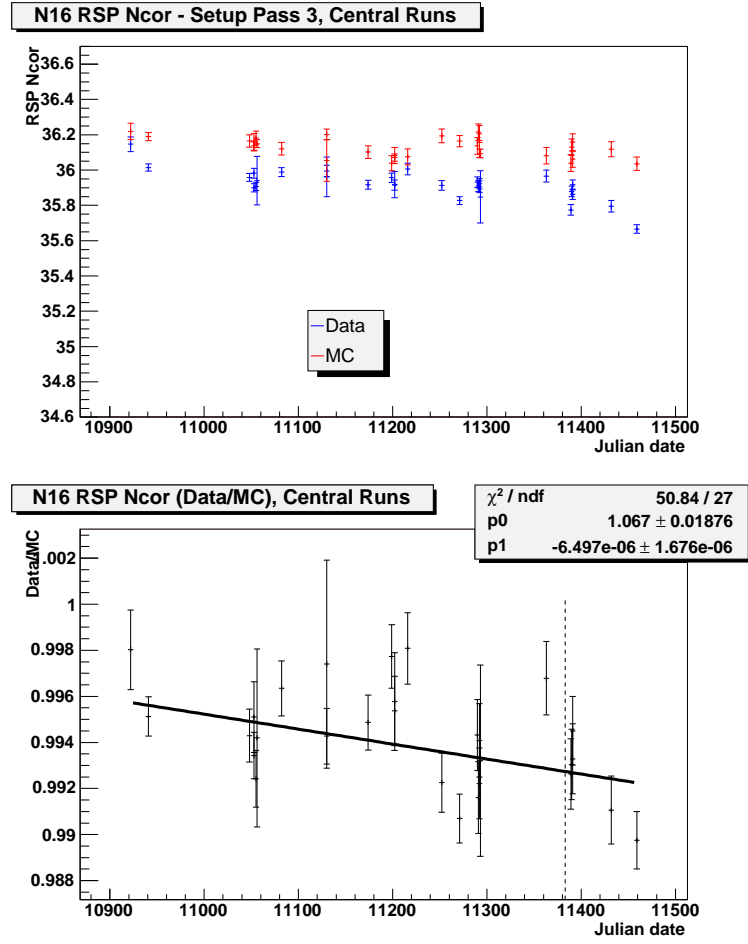
Figure C.12: Comparison of RSP N_{cor} for data and Monte Carlo of central N16 runs. Dashed line shows March 2006, the original cutoff for data to be used in first analysis.
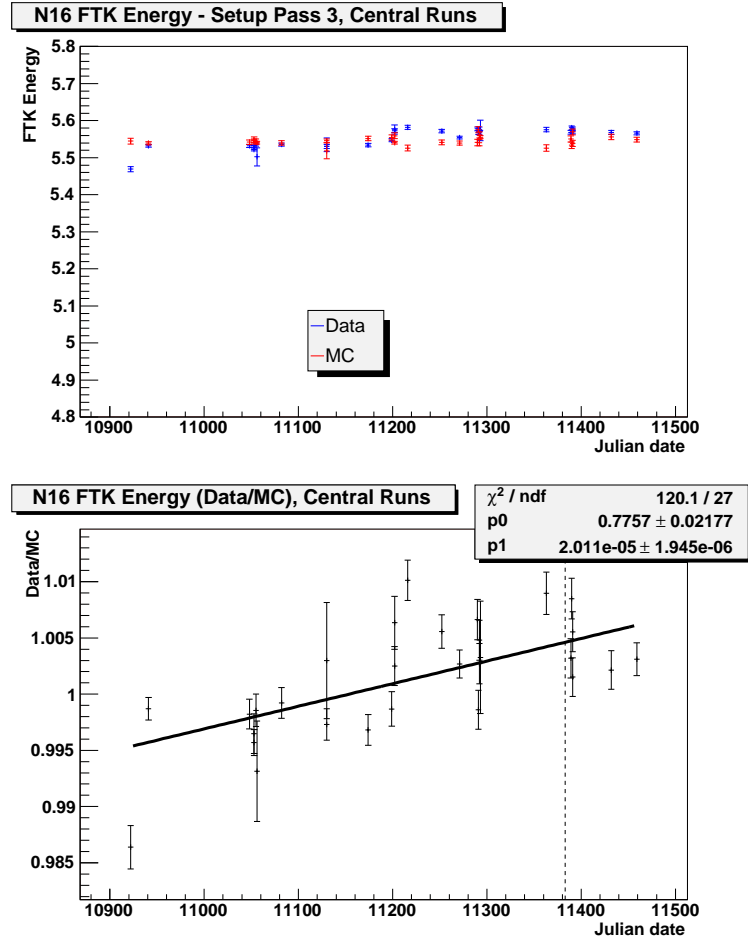
Figure C.13: Comparison of FTK Energy for data and Monte Carlo of central N16 runs. Dashed line shows March 2006, the original cutoff for data to be used in first analysis.
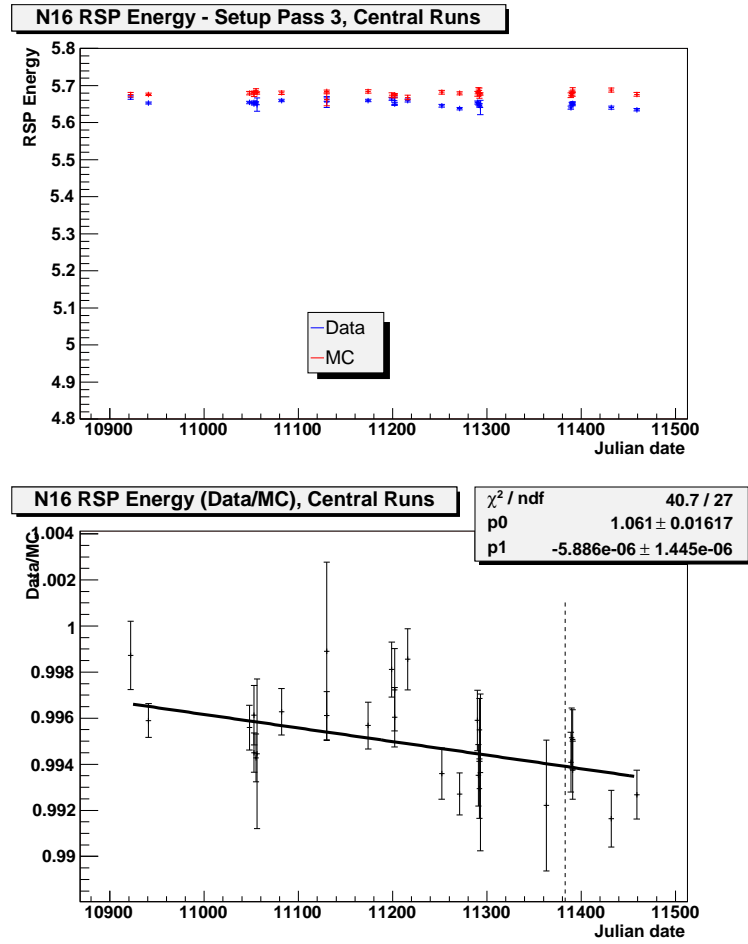
Figure C.14: Comparison of RSP Energy for data and Monte Carlo of central N16 runs. Dashed line shows March 2006, the original cutoff for data to be used in first analysis.

C.4.1   Z-bias

The verification pass of the global efficiency tuning included a small selection of runs at various Z positions from the November 2005 N16 scan. Figure C.15 shows the ratio of data to Monte Carlo mean energy as a function of Z position. Both energy fitters show a roughly a 3–5% span in energy scale bottom to top in the detector.

C.5   Conclusion

The update of FTK for the NCD phase described in this appendix is sufficient for a first pass through the NCD phase data, in order to allow backgrounds to be measured and basic systematics to be studied. NCD-phase updates to FTK started appearing in SNOMAN release 5.0183, but the finalized load_ncd_settings.cmd file was not released until 5.0187.
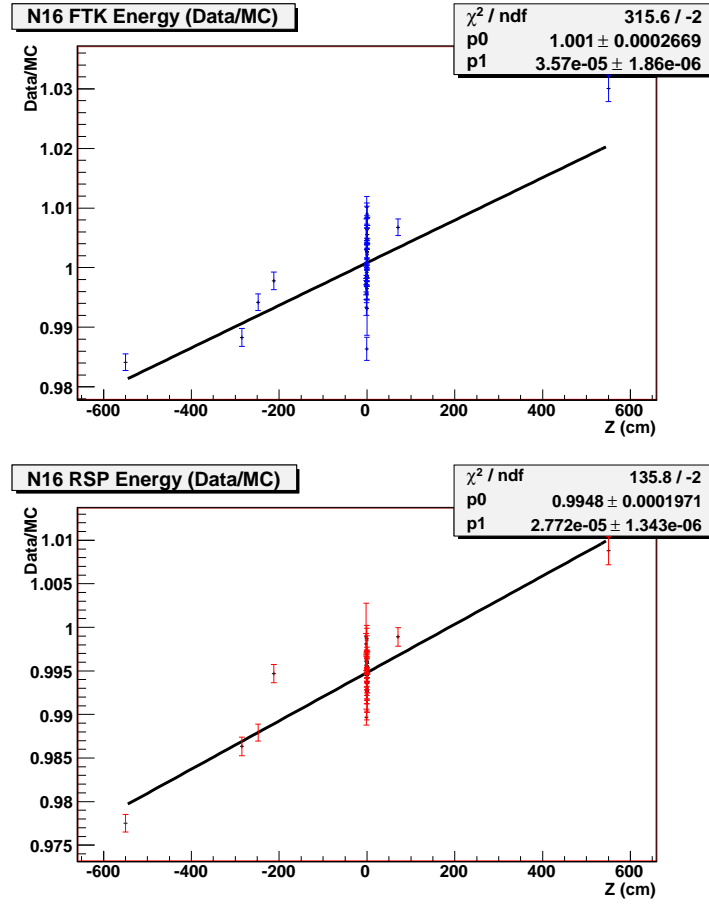
Figure C.15: Ratio of total energy data/Monte Carlo for N16 runs processed during NCD-phase global efficiency tuning. Energy ratios were computed from total energy.

Appendix D

Optical Effects of the NCD Anchors

In Section 5.4, a discrepancy in the energy scale as a function of z of data compared to Monte Carlo was described. The model of the detector in SNOMAN correctly reproduces many details of the optics, but some effects are missing, which show up as a z-bias when we compare data to Monte Carlo, as was shown in Figure 5.3.

It is important to note that interpreting Figure 5.3 has an inherent top-bottom ambiguity. We set the global collection efficiency of the detector using central ^{16}N runs, so data from the center of the detector must match the Monte Carlo, by construction. However, the central runs will average together in equal proportion events whose light went through the top or the bottom of the detector. All of the bias could be caused by a problem just at the top, just at the bottom, or perhaps both places, as long as the sign of the error is reversed between top and bottom. In the case of the bias we have observed,

the real data contains some combination of more light than we expect at the top of the detector and less light at the bottom.

At the top of the vessel, there is the neck, which is a clear culprit for possible optical problems. At the bottom, there are several conjectured causes for loss of light:

- "Batch 3" concentrators—a group of concentrators which might have different optical properties than the rest of the PMT concentrators. Most of the concentrators below the equator are from batch 3.

- Junk at the bottom of the AV or PSUP—this includes the white, flaky "dandruff" observed during NCD deployment, as well as any other debris one could imagine (nuts, washers, screwdrivers, cell phones, etc.) falling into the detector during construction.

- Acrylic NCD anchor points—installed when the SNO acrylic vessel was first constructed, these small blocks were used to hold down the buoyant ^{3}He proportional counters in the third phase.

In this appendix, we investigate the last case: optical problems with the NCD anchors.

D.1 NCD Anchors

On the bottom inner surface of the acrylic vessel there are 96 NCD anchors points. These anchors were fabricated and bonded to the AV during construction, and thus have been present for all of SNO's operating lifetime, including the D_2O and salt phases. These acrylic blocks were to hold down each of the 96

NCD strings that were originally planned for the full NCD array. Ultimately, only 40 anchors were actually needed[1], but all 96 anchors were installed as the decision to reduce the size of the NCD array was made after construction was completed and data taking begun. The positions and dimensions of these anchors were described in detail in [109].

Each anchor is roughly a 3 inch (7.62 cm) diameter, 2.25 inch (5.715 cm) tall cylinder. A notch is cut into the cylinder to accept one of the balls mounted to the bottom of an NCD string during deployment. The upward buoyant force of the NCD pulls the ball against the top of the anchor, locking the entire string assembly in place. Figure D.1 shows a more detailed engineering diagram of an anchor. The anchors are made of commercial *UV-absorbing acrylic*, rather than the UV-transmitting acrylic of the rest of the vessel[110]. This means that the NCD anchors, unlike the acrylic vessel, *absorb* most of the Cherenkov light that strikes them. All of the anchors are located in the bottom hemisphere of the vessel, with $z < -247$ cm.

These anchors are not currently simulated at all[111] in SNOMAN, even in the NCD-phase Monte Carlo with the proportional counters enabled. The NCD string geometry does not include the Vectran braid at the bottom, the anchor balls, or the acrylic anchor points on the AV. If these components, particularly the anchors bonded to the AV, do have a significant optical effect, then their absence would be a source of z-bias in the simulation.

---

[1]Thirty-six were used by the ^{3}He proportional counters, and four were used for nearly identical, but neutron insensitive, ^{4}He proportional counters that provided a pure sample of background events.

#30 DRILL THRU
2 HOLES ON A 2.38 DIA. B.C.

3.00 DIA

1.25 DIA BALL DRILL
OR MILL

B

A   A

B

11°

72°

1.37

R.50

SECTION B - B

.38   3/16" RADIUS BALL END MILL

0.55 SPHERICAL RADIUS

1.25 DIA.
BLIND HOLE

1.25 SPHERICAL
CENTER

.88

125

2.25

SECTION: A - A

NOTE: BREAK ALL SHARP EDGES AND BURRS
VISUAL SURFACE FINISH 125 OR BETTER
MATERIAL 3"DIA. ACRYLIC ROD-SUPPLIED
110 REQUIRED
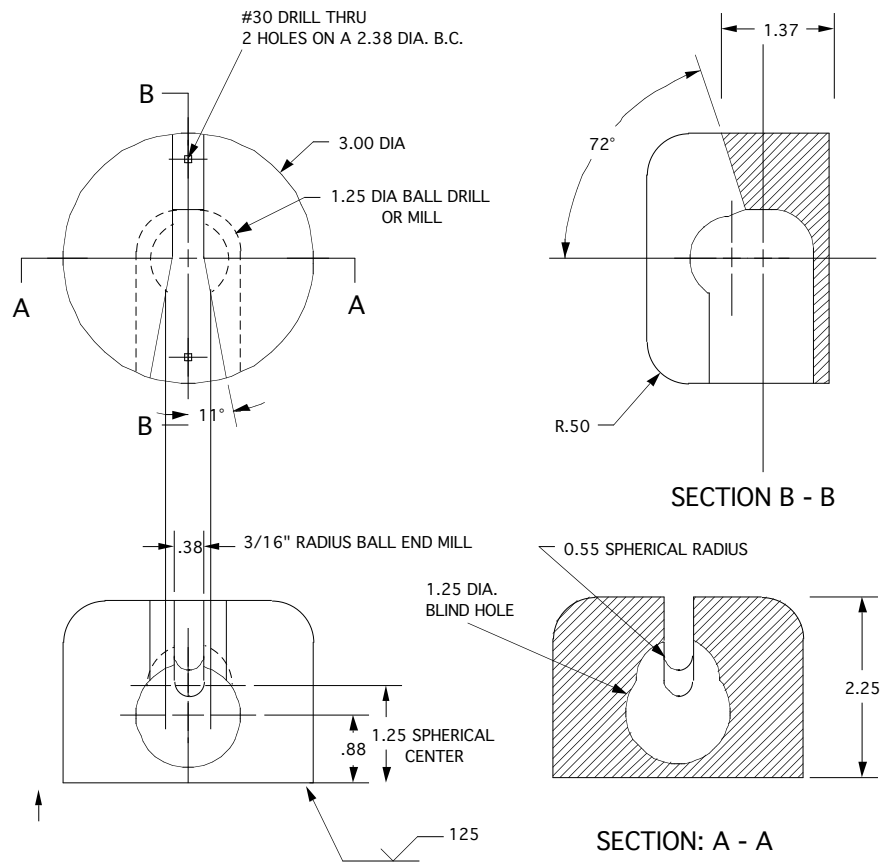
Figure D.1: Engineering diagrams of the NCD anchors, taken from [109].

D.2 Simulation

To determine whether the anchors were a significant effect, the SNOMAN was modified to include a simplified model of them. The anchors were represented as simple cylinders with the same outer dimensions as the real anchors. No attempt was made to model the hollow cavity inside the anchor. The optical response of the acrylic was assumed to be somewhere between the "bad" UV-trasmitting acrylic in the neck (so called because it is not polished) and the same acrylic, but with no light transmittance below 360 nm. Figure D.2 shows the NHIT distribution for 5 MeV events along the z-axis and the ratio of the two anchor simulations to no anchor simulations.

D.3 Conclusion

The optical effects of the NCD anchors are relatively small. At the very bottom of the detector they amount to a 0.5% deficit of photons, depending on the anchor model. While not negligible, anchor shadowing is not the primary component of the 3% deficit of photons seen at the bottom edge of the fiducial volume seen with the ^{16}N source. The calibration source-derived energy correction in Section 5.4 is still required.
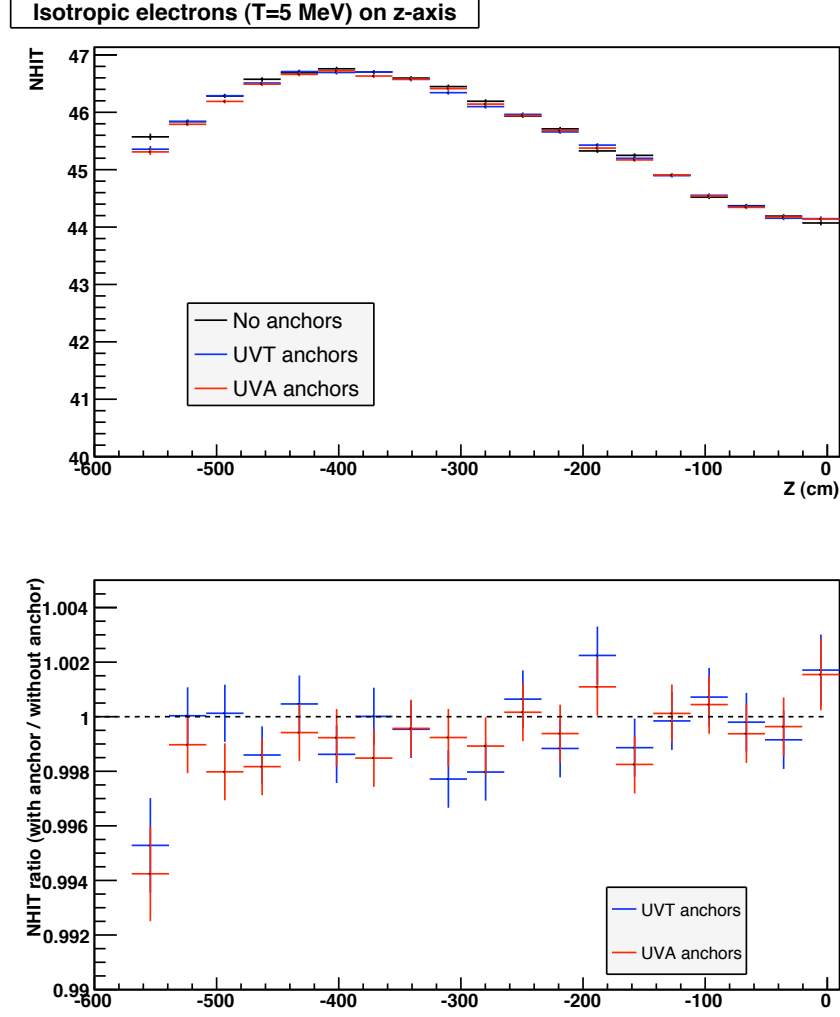
Figure D.2: Average NHIT (top) and ratio to no-anchor simulation (bottom) for T=5 MeV electrons as a function of z in the lower hemisphere. "UVT" anchors are unpolished UV-transmitting acrylic and "UVA" anchors are the same, but opaque for light with wavelengths shorter than 360 nm.

Appendix E

Composition of Monte Carlo Test Data Set

This appendix describes the total amount of Monte Carlo statistics generated, and the composition of the fake data sets used in ensemble testing of signal extraction. Tables E.1 and E.2 show the number of events generated and used per set in the D_2O and salt phases.

Event Type	Rate Generated	# of Events	
		Generated	Avg Test Data Set
^{8}B CC	500 SSM	3109142	2,150.13
^{8}B CC (^{17}O&^{18}O)	500 SSM	22101	15.28
^{8}B ES	1000 SSM	779228	365.89
^{8}B NC	500 SSM	454426	908.85
^{214}Bi in D$_2$O	14 Hz	238410	2,135.50
^{208}Tl in D$_2$O	0.34 Hz	89448	800.20
AV neutrons	0.15 Hz	192188	41.49
^{214}Bi in AV	7 Hz	1613	42.80
^{208}Tl in AV	0.3 Hz	2022	66.60
^{214}Bi in H$_2$O	80 Hz	2435	82.65 (43%)
^{208}Tl in H$_2$O	6 Hz	3363	106.00
^{208}Tl in PMT	700 Hz		2700 (analytic)

Table E.1: Monte Carlo events generated for the D$_2$O phase. Note that 1 SSM equals the BP2000 model.

Event Type	Rate Generated	# of Events	
		Generated	Avg Test Data Set
^{8}B CC	500 SSM	4346313	3,005.70
^{8}B CC (^{17}O&^{18}O)	500 SSM	31276	21.63
^{8}B ES	1000 SSM	1092909	513.17
^{8}B NC	500 SSM	1684711	3,369.42
^{214}Bi in D$_2$O	14 Hz	384942	3,447.00
^{208}Tl in D$_2$O	0.34 Hz	150718	1,137.80
^{24}Na in D$_2$O	0.17 Hz	270821	392.00
AV neutrons	0.15 Hz	694817	150.00
^{214}Bi in AV	7 Hz	2951	81.10
^{208}Tl in AV	0.3 Hz	4292	157.90
^{214}Bi in H$_2$O	80 Hz	4560	155.74 (65%)
^{208}Tl in H$_2$O	6 Hz	6712	145.70
^{208}Tl in PMT	700 Hz		5300 (analytic)

Table E.2: Monte Carlo events generated for the salt phase. Note that 1 SSM equals the BP2000 model.

Bibliography

[1] H. A. Bethe, "Energy Production in Stars," *Phys. Rev.* **55** (Mar, 1939) 434–456.

[2] W. Pauli, "Letter to Physicists at Tubingen," *Translated in Inward Bound, A. Pais.* (1930).

[3] C. L. Cowan, Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, "Detection of the Free Neutrino: A Confirmation," *Science* **124** (July, 1956) 103–104.

[4] **Super-Kamiokande** Collaboration, Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.* **81** (1998) 1562–1567, arXiv:hep-ex/9807003.

[5] J. N. Bahcall, M. C. Gonzalez-Garcia, and C. Pena-Garay, "Does the sun shine by p p or CNO fusion reactions?," *Phys. Rev. Lett.* **90** (2003) 131301, arXiv:astro-ph/0212331.

[6] T. Araki, S. Enomoto, K. Furuno, Y. Gando, K. Ichimura, H. Ikeda, K. Inoue, Y. Kishimoto, M. Koga, Y. Koseki, T. Maeda, T. Mitsui, M. Motoki, K. Nakajima, H. Ogawa, M. Ogawa, K. Owada, J.-S. Ricol,

I. Shimizu, J. Shirai, F. Suekane, A. Suzuki, K. Tada, S. Takeuchi, K. Tamae, Y. Tsuda, H. Watanabe, J. Busenitz, T. Classen, Z. Djurcic, G. Keefer, D. Leonard, A. Piepke, E. Yakushev, B. E. Berger, Y. D. Chan, M. P. Decowski, D. A. Dwyer, S. J. Freedman, B. K. Fujikawa, J. Goldman, F. Gray, K. M. Heeger, L. Hsu, K. T. Lesko, K.-B. Luk, H. Murayama, T. O'Donnell, A. W. P. Poon, H. M. Steiner, L. A. Winslow, C. Mauger, R. D. McKeown, P. Vogel, C. E. Lane, T. Miletic, G. Guillian, J. G. Learned, J. Maricic, S. Matsuno, S. Pakvasa, G. A. Horton-Smith, S. Dazeley, S. Hatakeyama, A. Rojas, R. Svoboda, B. D. Dieterle, J. Detwiler, G. Gratta, K. Ishii, N. Tolich, Y. Uchida, M. Batygov, W. Bugg, Y. Efremenko, Y. Kamyshkov, A. Kozlov, Y. Nakamura, H. J. Karwowski, D. M. Markoff, K. Nakamura, R. M. Rohm, W. Tornow, R. Wendell, M.-J. Chen, Y.-F. Wang, and F. Piquemal, "Experimental investigation of geologically produced antineutrinos with KamLAND," *Nature* **436** (July, 2005) 499–503.

[7] **KAMIOKANDE-II** Collaboration, K. Hirata *et al.*, "Observation of a Neutrino Burst from the Supernova SN 1987a," *Phys. Rev. Lett.* **58** (1987) 1490–1493.

[8] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.

[9] M. C. Gonzalez-Garcia and M. Maltoni, "Phenomenology with Massive Neutrinos," *Phys. Rept.* **460** (2008) 1–129, `arXiv:0704.1800 [hep-ph]`.

[10] **Particle Data Group** Collaboration, W. M. Yao *et al.*, "Review of particle physics," *J. Phys.* **G33** (2006) 1–1232.

[11] **MINOS** Collaboration, D. G. Michael *et al.*, "Observation of muon neutrino disappearance with the MINOS detectors and the NuMI neutrino beam," *Phys. Rev. Lett.* **97** (2006) 191801, `arXiv:hep-ex/0607088`.

[12] L. Wolfenstein, "Neutrino oscillations in matter," *Phys. Rev.* **D17** (1978) 2369.

[13] S. P. Mikheev and A. Y. Smirnov, "Resonant amplification of neutrino oscillations in matter and solar neutrino spectroscopy," *Nuovo Cim.* **C9** (1986) 17–26.

[14] C. E. Rolfs and W. S. Rodney, *Cauldrons in the cosmos: Nuclear astrophysics*. Research supported by NSF, Georgetown University, DFG, et al. Chicago, IL, University of Chicago Press, 1988, 579 p., 1988.

[15] K. Zuber, *Neutrino Physics*. CRC Press, 2004.

[16] H. O'Keeffe, "Low Energy Background in the NCD Phase of the Sudbury Neutrino Observatory,".

[17] J. N. Bahcall, A. M. Serenelli, and S. Basu, "New solar opacities, abundances, helioseismology, and neutrino fluxes," *Astrophys. J.* **621** (2005) L85–L88, `arXiv:astro-ph/0412440`.

[18] J. N. Bahcall and R. K. Ulrich, "Solar models, neutrino experiments, and helioseismology," *Rev. Mod. Phys.* **60** (Apr, 1988) 297–372.

[19] J. N. Bahcall and C. Pena-Garay, "Global analyses as a road map to solar neutrino fluxes and oscillation parameters," *JHEP* **11** (2003) 004, `arXiv:hep-ph/0305159`.

[20] O. Simard, "Private communication." 2008.

[21] B. T. Cleveland *et al.*, "Measurement of the solar electron neutrino flux with the Homestake chlorine detector," *Astrophys. J.* **496** (1998) 505–526.

[22] **GALLEX** Collaboration, P. Anselmann *et al.*, "Solar neutrinos observed by GALLEX at Gran Sasso.," *Phys. Lett.* **B285** (1992) 376–389.

[23] **GNO** Collaboration, M. Altmann *et al.*, "Complete results for five years of GNO solar neutrino observations," *Phys. Lett.* **B616** (2005) 174–190, `arXiv:hep-ex/0504037`.

[24] **SAGE** Collaboration, J. N. Abdurashitov *et al.*, "Measurement of the solar neutrino capture rate in SAGE," *Nucl. Phys. Proc. Suppl.* **118** (2003) 39–46.

[25] J. N. Bahcall, M. H. Pinsonneault, and S. Basu, "Solar models: Current epoch and time dependences, neutrinos, and helioseismological properties," *Astrophys. J.* **555** (2001) 990–1012, `arXiv:astro-ph/0010346`.

[26] Y. Fukuda *et al.*, "The Super-Kamiokande detector," *Nucl. Instrum. Meth.* **A501** (2003) 418–462.

[27] **Super-Kamiokande** Collaboration, J. P. Cravens *et al.*, "Solar neutrino measurements in Super-Kamiokande-II," `arXiv:0803.4312 [hep-ex]`.

[28] **Borexino** Collaboration, C. Arpesella *et al.*, "First real time detection of Be7 solar neutrinos by Borexino," *Phys. Lett.* **B658** (2008) 101–108, `arXiv:0708.2251 [astro-ph]`.

[29] T. B. Collaboration, "New results on solar neutrino fluxes from 192 days of Borexino data," `arXiv:0805.3843 [astro-ph]`.

[30] **KamLAND** Collaboration, K. Eguchi *et al.*, "First results from KamLAND: Evidence for reactor anti- neutrino disappearance," *Phys. Rev. Lett.* **90** (2003) 021802, `arXiv:hep-ex/0212021`.

[31] **KamLAND** Collaboration, S. Abe *et al.*, "Precision Measurement of Neutrino Oscillation Parameters with KamLAND," `arXiv:0801.4589 [hep-ex]`.

[32] **SNO** Collaboration, B. Aharmim *et al.*, "Electron energy spectra, fluxes, and day-night asymmetries of B-8 solar neutrinos from the 391-day salt phase SNO data set," *Phys. Rev.* **C72** (2005) 055502, `arXiv:nucl-ex/0502021`.

[33] P. F. Harrison, D. H. Perkins, and W. G. Scott, "Tri-bimaximal mixing and the neutrino oscillation data," *Phys. Lett.* **B530** (2002) 167, `arXiv:hep-ph/0202074`.

[34] P. F. Harrison and W. G. Scott, "Permutation symmetry, tri-bimaximal neutrino mixing and the S3 group characters," *Phys. Lett.* **B557** (2003) 76, `arXiv:hep-ph/0302025`.

[35] C. I. Low and R. R. Volkas, "Tri-bimaximal mixing, discrete family symmetries, and a conjecture connecting the quark and lepton mixing matrices," *Phys. Rev.* **D68** (2003) 033007, `arXiv:hep-ph/0305243`.

[36] R. N. Mohapatra, M. K. Parida, and G. Rajasekaran, "High scale mixing unification and large neutrino mixing angles," *Phys. Rev.* **D69** (2004) 053007, `arXiv:hep-ph/0301234`.

[37] S. K. Agarwalla, M. K. Parida, R. N. Mohapatra, and G. Rajasekaran, "Neutrino mixings and leptonic CP violation from CKM matrix and Majorana phases," *Phys. Rev.* **D75** (2007) 033007, `arXiv:hep-ph/0611225`.

[38] O. G. Miranda, M. A. Tortola, and J. W. F. Valle, "Are solar neutrino oscillations robust?," *JHEP* **10** (2006) 008, `arXiv:hep-ph/0406280`.

[39] R. Fardon, A. E. Nelson, and N. Weiner, "Dark energy from mass varying neutrinos," *JCAP* **0410** (2004) 005, `arXiv:astro-ph/0309800`.

[40] H. Georgi, "Unparticle Physics," *Phys. Rev. Lett.* **98** (2007) 221601, `arXiv:hep-ph/0703260`.

[41] M. C. Gonzalez-Garcia, P. C. de Holanda, and R. Zukanovich Funchal, "Constraints from Solar and Reactor Neutrinos on Unparticle Long-Range Forces," `arXiv:0803.1180 [hep-ph]`.

[42] A. de Gouvea and C. Pena-Garay, "Probing new physics by comparing solar and KamLAND data," *Phys. Rev.* **D71** (2005) 093002, `arXiv:hep-ph/0406301`.

[43] **SNO** Collaboration, J. Boger *et al.*, "The Sudbury Neutrino Observatory," *Nucl. Instrum. Meth.* **A449** (2000) 172–207, `arXiv:nucl-ex/9910016`.

[44] **SNO** Collaboration, B. Aharmim *et al.*, "Measurement of the nu/e and total B-8 solar neutrino fluxes with the Sudbury Neutrino Observatory phase I data set," *Phys. Rev.* **C75** (2007) 045502, `arXiv:nucl-ex/0610020`.

[45] H. H. Chen, "Direct approach to resolve the solar-neutrino problem," *Phys. Rev. Lett.* **55** (Sep, 1985) 1534–1536.

[46] J. D. Jackson, *Classical Electrodynamics, Third Edition*. Wiley.

[47] J. F. Amsbaugh *et al.*, "An array of low-background ^{3}He proportional counters for the Sudbury Neutrino Observatory," *Nucl. Instrum. Meth.* **A579** (2007) 1054–1080, `arXiv:0705.3665 [nucl-ex]`.

[48] M. Lay, "Creation and Detection of Cerenkov Light in the Sudbury Neutrino Observatory." D.Phil. thesis, University of Oxford, 1994.

[49] **SNO** Collaboration, B. Aharmim *et al.*, "High sensitivity measurement of 224Ra and 226Ra in water with an improved hydrous titanium oxide technique at the Sudbury Neutrino Observatory," `arXiv:0803.4162 [nucl-ex]`.

[50] **SNO** Collaboration, T. C. Andersen *et al.*, "Measurement of radium concentration in water with Mn- coated beads at the Sudbury Neutrino Observatory," *Nucl. Instrum. Meth.* **A501** (2003) 399–417, `arXiv:nucl-ex/0208010`.

[51] M. Dunford, "Measurement of the (8)B solar neutrino energy spectrum at the Sudbury Neutrino Observatory,". Ph.D. thesis, University of Pennsylvania.

[52] J. R. Klein, M. S. Neubauer, R. V. Berg, and F. M. Newcomer, "The SNO Trigger System." SNO technical report SNO-STR-97-036, University of Pennsylvania, 1997.

[53] M. R. Dragowsky *et al.*, "The N-16 calibration source for the Sudbury Neutrino Observatory," *Nucl. Instrum. Meth.* **A481** (2002) 284–296, `arXiv:nucl-ex/0109011`.

[54] R. Ford, "Calibration of SNO for the detection of ^{8}B neutrinos." PhD thesis, Queen's University, 1998.

[55] N. J. Tagg *et al.*, "The Li-8 calibration source for the Sudbury Neutrino Observatory," *Nucl. Instrum. Meth.* **A489** (2002) 178–188, `arXiv:nucl-ex/0202024`.

[56] A. W. P. Poon, M. C. Browne, R. G. H. Robertson, C. E. Waltham, and N. P. Kherani, "Development of a compact 20-MeV gamma-ray source for energy calibration at the Sudbury Neutrino Observatory,". Prepared for International School on Particles and Cosmology (Baksan School 95), Baksan, Russia, 19-27 Apr 1995.

[57] S. B. *et al.*, "First Generation Acrylic Encapsulated U/Th Sources." SNO Technical report SNO-STR-99-023, LANL, LBL, Queen's University, 2000.

[58] R. Lange, "Rn Spike Procedure." Technical Report, Sudbury Neutrino Observatory, 2005.

[59] W. R. Nelson, H. Hirayama, and D. W. O. Rogers, "THE EGS4 CODE SYSTEM,". SLAC-0265.

[60] E. . Briesmeister, Judith F., "MCNP: A General Monte Carlo N-Particle Transport Code,". LA-12625-M.

[61] A. E. G. Ingelman, J. Rathsman, "LEPTO version 6.3 — The Lund Monte Carlo for Deep Inelastic Lepton-Nucleon Scattering." 1995.

[62] P. A. Aarnio *et al.*, "FLUKA: Hadronic benchmarks and applications,". Prepared for International Conference on Monte Carlo Simulation in High-Energy and Nuclear Physics - MC 93, Tallahassee, FL, 22-26 Feb 1993.

[63] C. Zeitnitz and T. A. Gabriel, "The GEANT-CALOR interface and benchmark calculations of ZEUS test calorimeters," *Nuclear Instruments and Methods in Physics Research A* **349** (Sept., 1994) 106–111.

[64] **SNO** Collaboration, B. Aharmim *et al.*, "A search for periodicities in the B-8 solar neutrino flux measured by the Sudbury Neutrino Observatory," *Phys. Rev.* **D72** (2005) 052010, `arXiv:hep-ex/0507079`.

[65] **SNO** Collaboration, B. Aharmim *et al.*, "A search for neutrinos from the solar hep reaction and the diffuse supernova neutrino background with the Sudbury Neutrino Observatory," *Astrophys. J.* **653** (2006) 1545–1551, `arXiv:hep-ex/0607010`.

[66] **SNO** Collaboration, S. N. Ahmed *et al.*, "Constraints on nucleon decay via 'invisible' modes from the Sudbury Neutrino Observatory," *Phys. Rev. Lett.* **92** (2004) 102004, `arXiv:hep-ex/0310030`.

[67] **SNO** Collaboration, B. Aharmim *et al.*, "Electron antineutrino search at the Sudbury Neutrino Observatory," *Phys. Rev.* **D70** (2004) 093014, `arXiv:hep-ex/0407029`.

[68] **SNO** Collaboration, Q. R. Ahmad *et al.*, "Measurement of the charged current interactions produced by B-8 solar neutrinos at the Sudbury Neutrino Observatory," *Phys. Rev. Lett.* **87** (2001) 071301, `arXiv:nucl-ex/0106015`.

[69] **SNO** Collaboration, Q. R. Ahmad *et al.*, "Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory," *Phys. Rev. Lett.* **89** (2002) 011301, `arXiv:nucl-ex/0204008`.

[70] **SNO** Collaboration, Q. R. Ahmad *et al.*, "Measurement of day and night neutrino energy spectra at SNO and constraints on neutrino mixing parameters," *Phys. Rev. Lett.* **89** (2002) 011302, `arXiv:nucl-ex/0204009`.

[71] **SNO** Collaboration, S. N. Ahmed *et al.*, "Measurement of the total active B-8 solar neutrino flux at the Sudbury Neutrino Observatory with enhanced neutral current sensitivity," *Phys. Rev. Lett.* **92** (2004) 181301, `arXiv:nucl-ex/0309004`.

[72] J. Wilson, "The Separation of CC and NC Events in the Sudbury Neutrino Observatory." D.Phil. thesis, University of Oxford, 2004.

[73] M. Neubauer, "Evidence for Electron Neutrino Flavor Change Through Measurement of the 8B Solar Neutrino Flux at the Sudbury Neutrino Observatory." Ph.D. thesis, University of Pennsylvania, 2001.

[74] J. R. Klein, "Tuning the High Level Cuts for LETA." Technical report, University of Texas, 2006.

[75] J. R. Klein and A. Roodman, "Blind analysis in nuclear and particle physics," *Ann. Rev. Nucl. Part. Sci.* **55** (2005) 141–163. `http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.nucl.55.090704.151521`.

[76] M. G. Boulay, "Direct Evidence for Weak Flavour Mixing with the Sudbury Neutrino Observatory." PhD thesis, Queen's University, 2001.

[77] R. P. Brent, *Algorithms for Minimization without Derivatives*. Prentice-Hall, 1973.

[78] R. MacLellan, "^{16}N Source Energy Scale Uncertainty,". Technical Report, Queen's University.

[79] B. Heinzelmann, "Leslie events." Technical Report, University of Pennsylvania, 2008.

[80] C. Tunnell, "Measurement of the External ^{214}Bi and ^{208}Tl Radioactivity at the Sudbury Neutrino Observatory." Senior Thesis, University of Texas at Austin, 2008.

[81] G. D. Orebi Gann, "An Improved Measurement of the ^{8}B Solar Neutrino Energy Spectrum at the Sudbury Neutrino Observatory." PhD thesis, University of Oxford, 2008.

[82] C. Krauss, "Private communication." 2008.

[83] M. P. Wand and M. C. Jones, *Kernel Smoothing*. Chapman Hall/CRC, Boca Raton, FL, USA, 1995.

[84] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley-Interscience, 1992.

[85] D. W. Scott, "On optimal and data-based histograms," *Biometrika* **66** (1979), no. 3, 605–610.
`http://dx.doi.org/10.1093/biomet/66.3.605`.

[86] I. S. Abramson, "On bandwidth variation in kernel estimates-a square root law," *The Annals of Statistics* **10** (Dec, 1982) 1217–1223.

[87] C. K., "Kernel estimation in high-energy physics," *Computer Physics Communications* **136** (15 May 2001) 198–207(10).
`http://arxiv.org/abs/hep-ex/0011057`.

[88] C. Yang, R. Duraiswami, N. A. Gumerov, and L. Davis, "Improved fast gauss transform and efficient kernel density estimation," in *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, p. 464. IEEE Computer Society, Washington, DC, USA, 2003.

[89] "AMD Core Math Library (ACML) Users Guide, Version 4.1.0," tech. rep., Advanced Micro Devices, Inc., Numerical Algorithms Group, Ltd., 2008. `http://developer.amd.com/assets/acml_userguide.pdf`.

[90] D. Goldberg, "What every computer scientist should know about floating-point arithmetic," *ACM Computing Surveys* **23** (1991), no. 1, 5–48. `citeseer.ist.psu.edu/goldberg91what.html`.

[91] J. Wilson, "A Measurement of the 8B Solar Neutrino Energy Spectrum at the Sudbury Neutrino Observatory." D.Phil. thesis, University of Oxford, 2004.

[92] S. Collaboration, "An Independent Measurement of the Total Active 8B Solar Neutrino Flux Using an Array of 3He Proportional Counters at the Sudbury Neutrino Observatory," `arXiv:0806.0989 [nucl-ex]`.

[93] "Python programming language – official website," tech. rep. `http://www.python.org/`.

[94] W. Lavrijsen, "PyROOT: A Python–ROOT Bridge." 2006.

[95] W. Verkerke and D. Kirby, "The roofit toolkit for data modeling," `arXiv:physics/0306116v1`.

[96] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth.* **A389** (1997) 81–86.

[97] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations," *Comput. Phys. Commun.* **10** (1975) 343–367.

[98] G. E. Moore, "Cramming more components onto integrated circuits,"
Electronics **38** (Apr, 1965) 114–117.
`ftp://download.intel.com/museum/Moores_Law/`
`Articles-Press_Releases/Gordon_Moore_1965_Article.pdf`.

[99] "Excerpts from A Conversation with Gordon Moore: Moores Law,"
tech. rep., Intel Corporation, 2005. `ftp:`
`//download.intel.com/museum/Moores_Law/Video-Transcripts/`
`Excepts_A_Conversation_with_Gordon_Moore.pdf`.

[100] N. R. Mahapatra and B. Venkatrao, "The processor-memory
bottleneck: problems and solutions," *Crossroads* (1999) 2.
`http://doi.acm.org/10.1145/357783.331677`.

[101] "Cuda programming guide 1.1," tech. rep., NVIDIA Corporation, Nov,
2007. `http://developer.download.nvidia.com/compute/cuda/1_1/`
`NVIDIA_CUDA_Programming_Guide_1.1.pdf`.

[102] I. American Micro Devices, "Model number comparisons." 2008.

[103] A. Wolfe, "AMD Barcelona Architecture Launch: Native Quad-Core.".

[104] E. Lindholm, J. Nickolls, S. Oberman, and J. Montrym, "NVIDIA
Tesla: A Unified Graphics and Computing Architecture," *Micro, IEEE*
28 (March-April, 2008) 39–55.

[105] R. Duncan, "A survey of parallel computer architectures," *Computer*
23 (Feb 1990), no. 2, 5–16. `http://dx.doi.org/10.1109/2.44900`.

[106] "Cray-1 computer system hardware reference manual," tech. rep., Cray Research, Inc. `http://www.bitsavers.org/pdf/cray/2240004C-1977-Cray1.pdf`.

[107] W. Kahan, "Pracniques: further remarks on reducing truncation errors," *Commun. ACM* **8** (1965), no. 1, 40.

[108] D. H. Bailey, Y. Hida, and K. Jeyabalan, "DSFUN90 Library.".

[109] R. Komar and P. Doe, "Location of NCD Attachment Anchors." Technical Report, SNO-STR-97-045.

[110] P. Doe, "Personal communication." 2006.

[111] J. Loach, "Personal communication." 2006.

Vita

Stanley Reid Seibert attended Mountain Pointe High School in Phoenix, Arizona. In 1998, he enrolled at Arizona State University in Tempe, Arizona. He completed Bachelor of Science degrees in both Physics and Computer Science in May, 2002. In August, 2002, he began graduate studies in Physics at the University of Texas.

Permanent Address: 9608 S. 45th Place

Phoenix, AZ 85044

This dissertation was typeset with LaTeX 2_ε[1] by the author.

---

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay, James A. Bednar, and Ayman El-Khashab.